© by PSP Volume 30– No. 03/2021 pages 2848-2858 Fresenius Environmental Bulletin

GENETIC UNIFORMITY OF A SPECIFIC REGION IN SARS-CoV-2 GENOME AND REPURPOSING OF N-ACETYL-D- GLUCOSAMINE

Omur Baysal[1,*], Ragip Soner Silme[2], Cagatay Karaaslan[3], Alexander N Ignatov[4]

[1]Molecular Microbiology Unit in Department of Molecular Biology and Genetics, Faculty of Science, Mugla Sitki Kocman University, Mentese, Mugla, 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
[4]Federal State Autonomous Educational Institution People's Friendship University of Russia, Miklukho-Maklaya str., Moscow, Russia

ABSTRACT

The causative agent of the viral pneumonia outbreak in the world identified as SARS-CoV-2 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, any useful information is essential for effective control of the disease. Our findings show the existence of unvarying sequence with no mutation in ORF1ab regions from 134 high-quality filtered genome sequences of SARS-CoV-2 downloaded from the GISAID database. We have detected this sequence region by using MAUVE analysis and pairwise alignment using Global Needleman Wunsch algorithm. All these results were also confirmed with the Clustal W analysis. The first 6500 bp of the consensus genome including ORF1ab region is an unvarying sequence in SARS-CoV-2 genome. Unvarying sequence in SARS-CoV-2 genome has been very similar to another spike protein, which belongs to feline infectious peritonitis virus strain UU4 (PDB 6JX7), depending on amino acid sequences encoded, and N-acetyl-D-glucosamine is the ligand of this protein according to the highest TM-score of predicted protein structure analysis. These results have confirmed that N-acetyl-D-glucosamine could play an important effect on pathogenicity 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. The initial reports showed that human-to-human transmission of the virus was not limited [2]. Coronaviruses have error-prone RNA-dependent RNA polymerases, mutations and recombination events occur that is concern with rapidly strengthening and it increases its capacity to cause disease, which enhances also virulence [2]. ACE2, the receptor protein is present in humans in the epithelia of the lung and small intestine [3], and coronavirus binds to this receptor to enter into cell, the virus infects the upper respiratory and gastrointestinal tract of mammals [4]. ORF1ab is a genomic region coding the putative replicase polyprotein. In other coronaviruses, it has been reported that this polyprotein has also protease property encoded by this ORF region involving expression of 10 different 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 expressing for two large polyproteins including pp1a and pp1ab. The expression of pp1ab polyprotein is playing a role for programmed ribosomal frameshifting of signal conducting a bridge between ORF1a and ORF1ab [5]. This frameshifting signal leads to the expression of RNA-

dependent RNA polymerase (RdRP) that is required for the coronavirus replication [6]. The increasing epidemiological and clinical evidence indicates that the SARS-CoV-2 has a stronger transmissibility than SARS-CoV [7]. But the exact mechanism of SARS-CoV-2 is unclear [8]. As a result of a unique mechanism of viral replication, Coronaviruses have a high frequency of recombination [9, 10, 11, 12]. 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 [13, 14].

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 within 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 [15]. 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 SARS-CoV-2.

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 in order to control its replication.

Homology genome blast and genomes information. We retrieved totally 134 complete genome sequences from the GISAID database [16] as of April 19, 2020. Only the complete genomes of high coverage were included in the dataset. The complete genomes of the countries and territories infected by SARS-CoV-2 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 [17]. 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 ClustaI W of MegAlignPro, DNASTAR software [17]. The evolutionary history was inferred using the Neighbor-Joining method in MegAlignPro software. The sequences were analysed and common regions of all genomes detected using MAUVE from pairwise alignment results were obtained with Global Needleman-Wunsch algorithm [18]. 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/) [19] 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 [20].

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 modelling were done using ExPASy proteomics server [19] to study on the protein sequence and further structural details.

Ligand retrieval. The structure of N-acetyl-D-glucosamine (D-GlcNAc) 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 analysed using PRODRG online server (http://prodrg2.dyndns.org/cgi-bin/prodrg.cgi) [21]. Further shape complementarity principle was applied with clustering RMSD 4.0 for docking calculations.

Molecular docking studies. Homology modelling 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 [22, 23, 24, 25]. Analysis on PatchDock yielded results for geometric shape

complementarity score (GSC score) and approximate interface area (AI area). The flexible docking study was carried out using CHIMERA involving AutoDock v 4.0 [26]. 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 results were also confirmed again by Protein dock server SWISSDOCK (http://www.swissdock.ch/docking) within protein receptors and ligand interaction [27]. Later, Pymol software has been used to get insight into their all binding preferences within the active site of these receptors.

>QJA17583.1:1-2042 ORF1a polyprotein [Severe acute respiratory syndrome coronavirus 2] >protein

MSLGAE**N**SVAYS**NN**SIAIPT**N**FTISVTTEILPVSMTKTSVDCTMYICGDSTECS**N**LLLQYGSFCTQL**N**RALTGIAVEQDK

NTQEVFAQVKQIYKTPPIKDFGGF**N**FSQILPDPSKPSKRSFIEDLLF**N**KVTLADAGFIKQYGDCLGDIAARDLICAQKF**N**

GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF**N**GIGVTQ**N**VLYE**N**QKLIA**N**QF**N**SAIGKI

QDSLSSTASALGKLQDVV**N**QN**A**QAL**N**TLVKQLSS**N**FGAISSVL**N**DILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLI

RAAEIRASA**N**LAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEK**N**FTTAPAICHDGKAHFP

REGVFVS**N**GTHWFVTQR**N**FYEPQIITTD**N**TFVSG**N**CDVVIGIV**NN**TVYDPLQPELDSFKEELDKYFK**N**HTSPDVDLGDI

SGI**N**ASVV**N**IQKEIDRL**N**EVAK**N**L**N**ESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC

FIGURE 1

The results of submitted sequence converted to aminoacid sequence using ExPASy proteomics server (https://web.expasy.org/translate) for structural analysis by I-TASSER server. Asparagine regions (N) are marked bold in sequence.

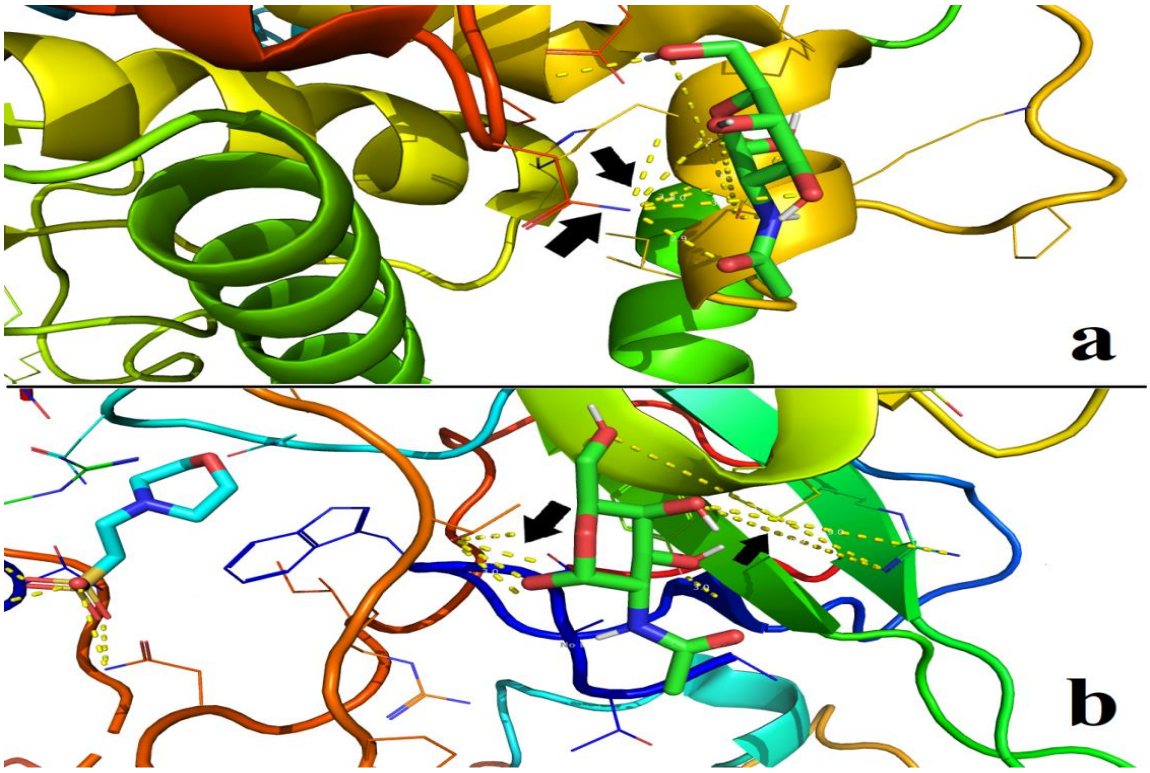

FIGURE 2

Arrow shows binding possibilities (Chimera /Autodock and calculation results) and measurements of D-GlcNAc 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.

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 and Clustal W analysis, respectively (S2a and S2b Files). Genetic diversity among sequences of SARS-CoV-2 genomes clearly indicated that various types of SARS-CoV-2 are present in different locations of the world.

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 [28]. We have detected high mutational changes in the SARS-CoV-2 genomes by pairwise alignment analysis. The isolate of Turkey displayed genetical differentiation compared to other isolates given as 5 gene sequence examples [28].

Homology modelling and protein prediction. The excised uniform regions of sequences subjected to alignment for protein similarity indicated that 6500 bp region including ORF1ab is consists of constantly unvarying sequences [28]. These stabile sequences were selected as a template for further protein structural predictions (Figure 1). The results of submitted sequence converted to aminoacid sequence using ExPASy proteomics server (https://web.expasy.org/translate) 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 I-TASSER analysis and its ligand was determined as N-acetyl-D-glucosamine (D-GlcNAc), a small molecule showing interaction with this stabile sequences (S3, S4 File). Therefore, our further study is relied on testing the same protein-ligand interaction possibilities whether this situation is valid for the spike receptor-binding domain bound with ACE2 (6M0J) and RNA-binding domain of nucleocapsid phosphoprotein (6WKP) of SARS-CoV-2 with our ligand D-GlcNAc.

Protein Docking. For docking analysis of D-GlcNAc with 6M0J and 6WKP, the ligand structure of D-GlcNAc retrieved from PubChem database was analysed (Figure 2a, 2b), 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 D-GlcNAc 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, AutoDock software calculations. Furthermore, binding possibilities of ligand on protein surface have been confirmed with results of SWISDOCK (Figure 3, 4, S6, S7 files).

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 [28]. 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

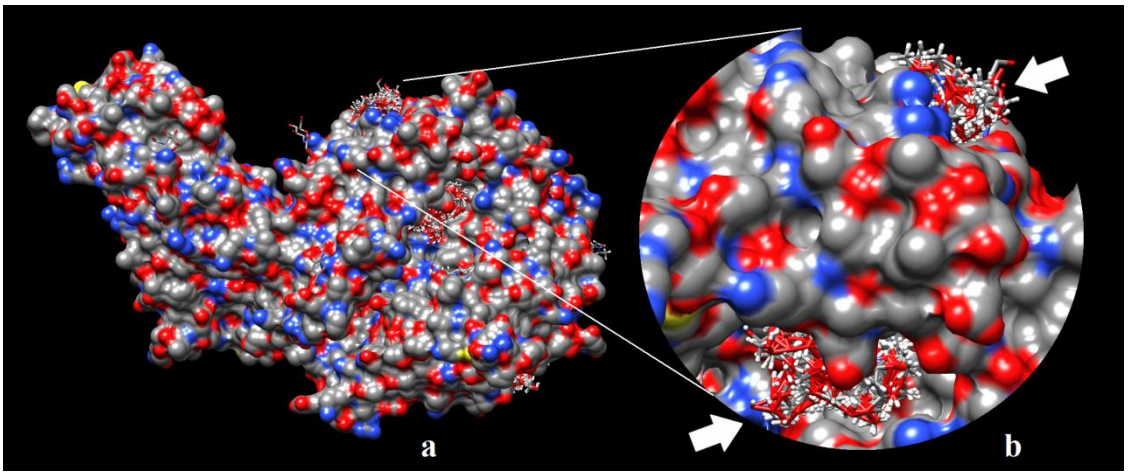

FIGURE 3
a) Protein docking analysis and binding possibilities (SWISSDOCK server results) of protein 6M0J with N-acetyl-D-glucosamine, b) closer view of binding possibilities marked with arrow.

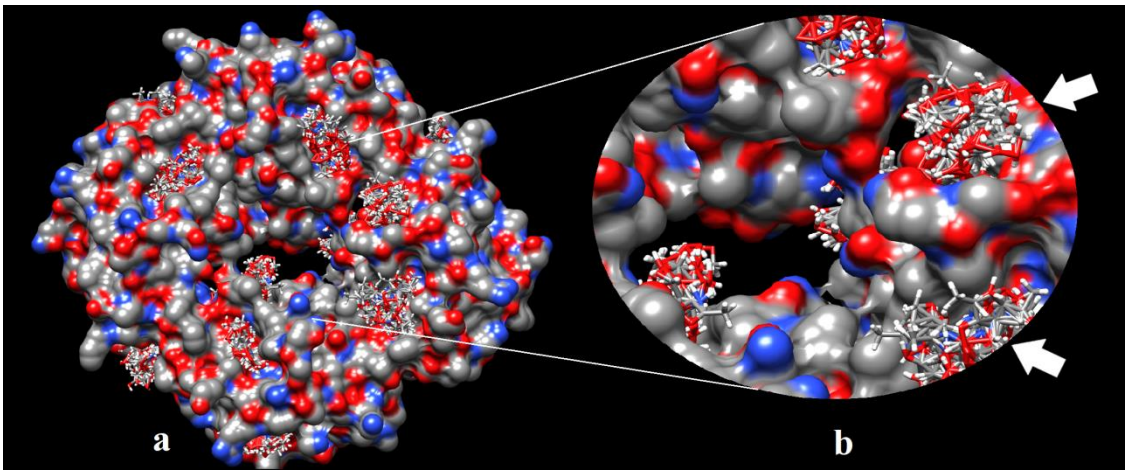

FIGURE 4
a) Protein docking analysis and binding possibilities (SWISDOCK server results) of protein 6WKP with N-acetyl-D-glucosamine, b) closer view of binding possibilities marked with arrows.

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.

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 [29]. 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 non-structural 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 [30]. 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.

We have reported our detected unvarying sequences including ORF1ab to be an important region responsible for the putative replicase polyprotein of proteases secretions [31]. Correspondently, we observed sequence variation with high ratio of genomes (except for 6500 bp fragment) [28] 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 D-GlcNAc against influenza [32]. In another study, gluosamine has been reported to have influence on replication of hepatitis B virus by *in-vitro* and *in-vivo* experiments [33]. Therefore, D-GlcNAc can also be suggested as an antiviral drug for SARS-CoV-2.

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 [34]. N-linked glycans on an immune cell's surface will help for the migration pattern of the cell with specific glycosylations [35]. 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 [35]. Glycans consist of different derivatives of D-GlcNAc and suggests having an important role in the immune system. This could also prevent sniffing of virus from antibody cells.

Moreover, Pant et al. reported asparagine supply is a critical barrier and limiting factor for replication of virus proteins, to development of antiviral drugs [36]. We assume that virus prefers glutamine to glucose for efficient replication, and the viral replication reduces in glutamine-free medium. Asparagine supplementation compensate of glutamine depletion, for viral replication. Asparagine-linked glycosylation is an enzyme-catalysed, co-translational

protein modification which influences either the protein folding process or the stability of the native conjugated glycoprotein form [37]. In our study, we have found D-GlcNAc interacts with proteins encoded by ORF1ab region; we suggest binding of D-GlcNAc to asparagine and inhibition of virus replication as reported by Pant et al. [36]. We have observed by docking analysis that binding of D-GlcNAc 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 D-GlcNAc as ligand molecule (Figure 1). Particularly, the effect of D-GlcNAc has also tested against HIV1 with different concentrations (0.25 mM, 1 mM, 4 mM, and 16 mM) [38]. We can suggest the same mode of action to HIV1 [36, 38]. 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 have reported to seven glycosylation sites playing a role on the S protein, which is critical for DC/L-SIGN-mediated virus attacking to asparagine residues at amino acid positions that are distinct from residues of the ACE2-binding domain [39]. Defections in secretion and infectivity of several flaviviruses concerned with blocking of the N-linked oligosaccharides have confirmed the role of glycosylation [40]. 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 [41]. 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 [42]. Reticulovesicular network of modified endoplasmic reticulum membranes with SARS-coronavirus replicative proteins has also been reported by Knoops et al. [43]. A virus-encoded uridine diphosphate-N-acetylglucosamine pathway associated with N-acetylglucosamine (GlcNAc) is a ubiquitous sugar which represents a fundamental process for virus. Therefore, D-GlcNAc can be a substitute of GlcNAc and could convert all process to support immunity defence mechanism [44]. Our molecular docking analysis on D-GlcNAc, which could mimic GlcNAc, could keep the cell of the SARS-CoV-2's viral integration into ER and Golgi system. Our results also show that D-GlcNAc has an interaction with 6M0J and /or 6WKP (Figures 3, 4).

However, the role of asparagine availability in virus replication remained unclear up to now [37]. The influences of GlcNAc on cell surface signalling proteins alter signal transduction depending on the degree of branching of N-linked glycans [45]. Therefore, this signal transduction could change in immunity system's favour by D-GlcNAc mimicking the same role in signalling instead of GlcNAc. We have found binding possibilities of D-GlcNAc with 6M0J and 6WKP could defect the attachment on human cell and replication mechanism of the virus.

In a previous study, the binding of *Urtica dioica* agglutinin with N-acetylglucosamine-like residues which are present on the glycosylated envelope glycoproteins has been suggested as a molecule preventing by targeting early stages of the replication cycle, adsorption or penetration of virus and attachment to cells [46]. Additionally, Chou et al. has suggested encompassing blocked aminoacids 48-358, which are responsible for ectodomain of the S glycoprotein localisation on the surface, of a clonal cell-line with N-acetyl-glucosamine-terminated carbohydrate structures are important regions in view of neutralizing antibodies for vaccine development and SARS-CoV S protein with its receptor [47]. In the sight of these findings, interrupted interaction between N-acetyl-D-glucosamine and S protein inhibits SARS-CoV-2's replication. We prove that the virus structure interacting with N-acetyl-D-glucosamine is highly conservative, and the forming complex is highly stable. We suggest that instead of using other molecules (like *Urtica dioica* agglutinin or sera) we can apply N-acetyl-D-glucosamine itself.

In a previous study, specifically inhibition of purified reoviruses types 1, 2, and 3 by 250 micromoles or more of N-acetyl-D-glucosamine was reported in their hemagglutination of human erythrocytes. This effect however; could not be obtained by over 20 other sugars tested. N-acetyl-D-glucosamine inhibited reovirus hemaglutination by binding to capsid virus; it did not attach to the erythrocytes. The study indicated that possible reovirus hemagglutination involves union between N-acetyl-D-glucosamine on the surface of the red cell and the glycoprotein of the virus coat [48].

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. Physicochemical properties (Molecular weight, LogP, Hydrogen bond donor and acceptor, polar surface area in 2D, polarizability, Van der Waals surface area in 3D and refractivity of the selected natural compounds) are traits usually evaluated to choose correct chemical as medicine [49]. 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 [50, 51] give scientists an opportunity for rapid detection of efficient molecules target-oriented on SARS-CoV-2 [52].

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 treatment for 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 [53, 54]. That reason our findings support N-acetyl-D-glucosamine is a potential drug. We strongly suggest testing the different concentrations of D-GlcNAc to SARS-CoV-2, considering interaction with proteins involving ORF1ab region which shows constantly unvarying piece of sequence in whole paired genomic data. Our results are likely to increase the underpinning data for drug repurposing in the therapeutic options against SARS-CoV-2 in the future.

ACKNOWLEDGEMENT

The authors wish to thank Dr. Venkatesh Balakkrishnan 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 the world.

Supplementary materials.

S1 file.
https://figshare.com/articles/dataset/S1_File_xls/12326408

S2a file.
https://figshare.com/articles/dataset/S2a_File_pdf/12326402

S2b file.
https://figshare.com/articles/dataset/S2b_File_pdf/12326405

S3 file.
https://figshare.com/articles/dataset/S3_File_pdf/12326417

S4 file.
https://figshare.com/articles/dataset/S4_File_pdf/12326423

S5 file.
https://figshare.com/articles/dataset/S5_File_docx/12326420

S6 file.
https://figshare.com/articles/dataset/S6_File_zip/12326414

S7 file.
https://figshare.com/articles/dataset/S7_File_zip/12326411

REFERENCES

[1] Chan, J.F., To, K.K., Tse, H., Jin, D.Y., and Yuen, K.Y. (2013) Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol. 21, 544-555.

[2] Cui, J., Li, F. and Shi, Z.L. (2019) Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17, 181-192.

[3] Hamming, I., Timens, W., Bulthuis, M.L.C., Lely, A.T., Navis, G.J. and Van Goor, H. (2004) Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus, A first step in understanding SARS pathogenesis. J Pathol. 203, 631-637.

[4] Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., Li, T. and Chen, Q. (2020) High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int. J. Oral. Sci. 12, 1-5.

[5] Plant, E.P., Sims, A.C., Baric, R.S., Dinman, J.D. and Taylor, D.R. (2013) Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. Viruses. 5, 279-294.

[6] Plant, E.P. and Dinman, J.D. (2008) The role of programmed-1 ribosomal frameshifting in coronavirus propagation. Front Biosci. 13, 4873-4881.

[7] Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., Du, B., Li, L.J., Zeng, G., Yuen, K.Y., Chen, R.C., Tang, C.L., Wang, T., Chen, P.Y., Xiang, J., Li, S.Y., Wang, J.L., Liang, Z.J., Peng, Y.X., Wei, L., Liu, Y., Hu, Y.H., Peng, P., Wang, J.M., Liu J.Y., Chen, Z., Li, G., Zheng, Z.J., Qiu, S.Q., Luo, J., Ye, C.J., Zhu, S.Y. and Zhong, N.S. (2020) Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med. 382, 1708-1720.

[8] Ziebuhr, J. (2005) The Coronavirus Replicase. In: Coronavirus Replication and Reverse Genetics. Current Topics in Microbiology and Immunology. (L. Enjuanes), Springer, Berlin, 57-94.

[9] Lai, M.M., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L. and Stohlman, S.A. **(**1985) Recombination between nonsegmented RNA genomes of murine coronaviruses. J Virol. 56, 449-456.

[10] Makino, S., Keck, J.G., Stohlman, S.A. and Lai, M.M. (1986) High-frequency RNA recombination of murine coronaviruses. J. Virol. 57, 729-737.

[11] Kusters, J.G., Jager, E.J., Niesters, H.G and Van der Zeijst, B.A. (1990) Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. Vaccine. 8, 605-608.

[12] Lee, C.W. and Jackwood, M.W. (2000) Evidence of genetic diversity generated by recombination among avian coronavirus IBV. Arch. Virol. 145, 2135-2148.

[13] Domingo, E. and Holland, J. (1997) RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51, 151-178.

[14] Holmes, E.C. and Rambaut, A. (2004) Viral evolution and the emergence of SARS coronavirus. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 1059-1065.

[15] Cherry, J.D. and Krogstad, P. (2004) SARS: the first pandemic of the 21st century. Pediatr. Res. 56, 1-5.

[16] Shu, Y. and McCauley, J. (2017) GISAID: Global initiative on sharing all influenza data – from vision to reality. EuroSurveillance. 22, (13), 30494.

[17] Burland, T.G. (2000) DNASTAR's Lasergene sequence analysis software. Methods Mol. Biol. 71-91.

[18] Needleman, S.B. and Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443-453.

[19] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784-3788.

[20] Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER Suite: Protein structure and function prediction. Nat. Methods. 12, 7-8.

[21] Schüttelkopf, A.W. and Van Aalten, D.M.F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D. 60, 1355-1363.

[22] Duhovny, D., Nussinov, R. and Wolfson, H.J. (2002) Efficient unbound docking of rigid molecules. In: Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics (WABI) Lecture Notes in Computer Science. Rome, Italy. 185-200.

[23] Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R. and Wolfson, H.J. (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. Proteins. 52, 107-112.

[24] Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 33, W363-367.

[25] Mashiach, E., Schneidman-Duhovny, D., Peri, A., Shavit, Y., Nussinov, R. and Wolfson, H.J. (2010) An integrated suite of fast docking algorithms. Proteins. 78, 3197-3204.

[26] Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455-461.

[27] Grosdidier, A., Zoete, V. and Michielin, O. (2011) SwissDock, protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 39 (suppl_2) W270–W277.

[28] Baysal, Ö. (2020) Figshare. Figure. Sequence comparison shows genomic differences of SARS-CoV-2 samples belonging to a) US, b) UK, c)Russia, d) Germany and e) China (reference genome from Wuhan) with Turkey. The analysis using Global Needleman Wunsch DNA fragment algorithm clearly shows genetic stability up to the first 6500 bp with alignment by using MAUVE. https://figshare.com/articles/figure/Fig_1_png/12326345

[29] Tilocca, B., Soggiu, A., Sanguinetti M., Musella, V., Britti D., Bonizzi, L., Urbani, A. and Roncada, P. (2020) Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. Microb. Infect. 22, 188-194.

[30] Zolfaghari Emameh, R., Nosrati, H. and Taheri, R.A. (2020) 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. 22, 8.

[31] Woo, P.C., Huang, Y., Lau, S.K., Tsoi, H.W. and Yuen, K.Y. (2005) In silico analysis of ORF1ab in Coronavirus HKU1 genome reveals a unique putative cleavage site of Coronavirus HKU1 3C-like protease. Microbiol. Immunol. 49, 899-908.

[32] Chen, C.A. and Fang, J.M. (2013) Synthesis of oseltamivir and tamiphosphor from N-acetyl-D-glucosamine. Org Biomol Chem. 11, 7687-7699.

[33] Lin, Y., Wu, C., Wang, X., Liu, S., Zhao, K., Kemper, T., Yu, H., Li, M., Zhang, J., Chen, M., Zhu, Y., Chen, X. and Lu, M. (2020) Glucosamine promotes hepatitis B virus replication through its dual effects in suppressing autophagic degradation and inhibiting MTORC1 signaling. Autophagy. 16, 548-561.

[34] Koch, G. and Kant, A. (1990) Binding of antibodies that strongly neutralise infectious bronchitis virus is dependent on the glycosylation of the viral peplomer protein, in Coronaviruses and their Diseases. In: Advances in Experimental Medicine and Biology. (Cavanagh, D., Brown, T.D.K.), Springer, Boston, MA. 143-150.

[35] Maverakis, E., Kim, K., Shimoda, M., Gershwin, M.E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L.R. and Lebrilla, C.B. (2015) Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. J. Autoimmun. 57, 1-13.

[36] Pant, A., Cao, S. and Yang, Z. (2019) Asparagine is a critical limiting metabolite for Vaccinia virus protein synthesis during glutamine deprivation. J. Virol. 93, e01834-18.

[37] Imperiali, B. and O'Connor, S.E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. Curr. Opin. Chem. Biol. 3, 643-649.

[38] Jochmann, R., Thurau, M., Jung, S., Hofmann, C., Naschberger, E., Kremmer, E., Harrer, T., Miller, M., Schaft, N. and Stürzl, M. (2009) O-Linked N-Acetylglucosaminylation of Sp1 inhibits the Human Immunodeficiency Virus Type 1 promoter. J. Virol. 83, 3704-3718.

[39] Han, D.P., Lohani, M. and Cho, M.W. (2007) Specific Asparagine-Linked glycosylation sites are critical for DC-SIGN- and L-SIGN-Mediated Severe Acute Respiratory Syndrome Coronavirus entry. J. Virol. 81, 12029-12039.

[40] Mondotte, J.A., Lozach, P.Y., Amara, A. and Gamarnik, A.V. (2007) Essential role of Dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation. J. Virol. 81, 7136-7148.

[41] Parodi, A.J. (2000) Protein glucosylation and its role in protein folding. Annu. Rev. Biochem. 69, 69-93.

[42] Van Etten, J.L., Gurnon, J.R., Yanai-Balser, G.M., Dunigan, D.D. and Graves, M.V. (2010) Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. Biochim Biophys. Acta. 1800, 152-159.

[43] Knoops, K., Kikkert, M., Van den Worm, S.H.E, Zevenhoven-Dobbe, J.C., Van der Meer, Y., Koster, A.J., Mommaas, A.M. and Snijder, E.J. (2008) SARS-Coronavirus replication is supported by a reticulovesicular network of modified Endoplasmic Reticulum. PLoS Biol. 6. Article ID: e226.

[44] Piacente, F., Bernardi, C., Marin, M., Blanc, G., Abergel, C. and Tonetti, M.G. (2014) Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology. 24, 51-61.

[45] Konopka, J.B. (2012) N-acetylglucosamine (GlcNAc) functions in cell signalling. Scientifica (Cairo). 2012, 489208.

[46] Kumaki, Y., Wandersee, M.K., Smith, A.J., Zhou, Y., Simmons, G., Nelson, N.M., Bailey, K.W., Vest, Z.G., Li, J.K.K., Chan, P.K.S., Smee, D.F. and Barnard, D.L. (2011) Inhibition of severe acute respiratory syndrome coronavirus replication in a lethal SARS-CoV BALB/c mouse model by stinging nettle lectin, *Urtica dioica* agglutinin. Antivir. Res. 90, 22-32.

[47] Chou, C.F., Shen, S., Tan, Y.J., Fielding, B.C., Tan, T.H.P., Fu, J., Xu, Q., Lim, S.G. and Hong, W. (2005) A novel cell-based binding assay system reconstituting interaction between SARS-CoV S protein and its cellular receptor. J. Virol. Methods. 123, 41-48.

[48] Gelb, L.D. and Lerner, A.M. (1965) Reovirus hemagglutination: inhibition by N-Acetyl-D-Glucosamine. Science. 147, 404-405.

[49] Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug. Deliv. Rev. 23, 3-25.

[50] Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin W. and Cheng, F. (2020a) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell. Discov. 6, 14.

[51] Leelananda, S.P. and Lindert, S. (2016) Computational methods in drug discovery. Beilstein J. Org. Chem. 12, 2694-2718.

[52] Li, G. and De Clercq, E. (2020) Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov. 19, 149-150.

[53] Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen H. and B. Cao. (2020b) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet. 395, P1054-1062.

[54] Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M., Li, X., Xia, J., Chen, N., Xiang, J., Yu, T., Bai, T., Xie, X., Zhang, L., Li, C., Yuan, Y., Chen, H., Li, H., Huang, H., Tu, S., Gong, F., Liu, Y., Wei, Y., Dong, C., Zhou, F., Gu, X., Xu, J., Liu, Z., Zhang, Y., Li, H., Shang, L., Wang, K., Li, K., Zhou, X., Dong, X., Qu, Z., Lu, S., Hu, X., Ruan, S., Luo, S., Wu, J., Peng, L., Cheng, F., Pan, L., Zou, J., Jia, C., Wang, J., Liu, X., Wang, S., Wu, X., Ge, Q., He, J., Zhan, H., Qiu, F., Guo, L., Huang, C., Jaki, T., Hayden, F.G., Horby, P.W., Zhang, D. and Wang, C. (2020) A trial of Lopinavir–Ritonavir in adults hospitalized with severe Covid-19. N. Engl. J. Med. 382, 1787-1799.

Received: **26.10.2020**
Accepted: **26.12.2020**

CORRESPONDING AUTHOR

Omur Baysal
Molecular Microbiology Unit in
Department of Molecular Biology and Genetics,
Faculty of Science,
Mugla Sitki Kocman University,
48121 Mentese-Mugla – Turkey

e-mail: omurbaysal@mu.edu.tr

View publication stats