N-terminal domain (NTD) of SARS-CoV-2 spike-protein structurally resembles MERS-CoV NTD sialoside-binding pocket

Mayanka Awasthi
Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA
https://orcid.org/0000-0002-4395-0589

Sahil Gulati
Gatan Inc., Pleasanton, California, USA https://orcid.org/0000-0002-8255-479X

Debi P. Sarkar
Department of Biochemistry, University of Delhi South Campus, New Delhi, India
https://orcid.org/0000-0002-8886-8415

Swasti Tiwari
Department of Molecular Medicine & Biotechnology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, U.P., India

Suneel Kateriya
School of Biotechnology, Jawaharlal Nehru University, New Delhi, India. https://orcid.org/0000-0001-5428-4297

Peeyush Ranjan (rpeeyush@umd.edu)
Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA. https://orcid.org/0000-0001-6823-2033

Santosh Kumar Verma (santoshkv@sgpgi.ac.in)
Department of Molecular Medicine & Biotechnology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, U.P., India. https://orcid.org/0000-0001-6058-6494

Short Report

Keywords: SARS-CoV-2, N-terminal domain, Motif, Spike protein, MERS-CoV.

Posted Date: June 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37300/v1

Version of Record: A version of this preprint was published on August 19th, 2020. See the published version at https://doi.org/10.3390/v12090909.

Abstract

COVID-19 novel coronavirus disease caused by SARS-CoV-2 causes severe lethal respiratory illness in humans and has recently developed into a worldwide pandemic. The lack of effective treatment strategy and vaccines against SARS-CoV-2 poses a threat to human health. Extremely high infection rate and multi-organ secondary infection within a short time period makes this virus more deadly and challenging for therapeutic interventions. Despite of high sequence similarity and utilization of common host-cell receptor, human angiotensin-converting enzyme-2 (ACE2) for virus entry, SARS-CoV-2 is much infectious than SARS-CoV. Structure-based sequence comparison of the N-terminal domain (NTD) of spike protein of MERS-CoV, SARS-CoV and SARS-CoV-2 illustrate three short stretches of amino acid motifs in SARS-CoV-2 , which appears to be the reminiscent of MERS-CoV sialoside binding pockets. These key differences with SARS-CoV and similarity with MERS-CoV, suggest an evolutionary adaptation of SARS-CoV-2 spike protein reciprocal interaction with host surface sialosides.

Introduction

Multiple coronaviruses are known to cause infection in humans of which β-coronavirus family members namely SARS-CoV, MERS-CoV and recently SARS-CoV-2 outbreak is a serious threat to the public health and has infected over 7 million people including over 400 K deaths worldwide with latency period of 3-14 days [1]. Recently, the world health organization (WHO) officially declared COVID-19 (caused by SARS-CoV-2) a global pandemic. SARS-CoV-2 is a positive strand RNA virus and like SARS and MERS-CoV, attacks the lower respiratory system, causing acute respiratory distress in lungs [2]. Recent reports suggest that SARS-CoV-2 also target the multiple organ system like the heart, liver, kidney, gastrointestinal system and central nervous system [3-5]. The rapid infection rate and spread of SARS-CoV-2 in multiple organs could be understood well by studying the virus-host cell interaction and fusion. Spike glycoprotein of both SARS-CoV and SARS-CoV-2 interacts with human ACE2 receptor of host cell, and mediates the adhesion of virus to the host cell. After series of cleavages and conformational changes in the spike protein, SARS-CoV-2 fuses with target cellular membrane [6, 7]. X-ray crystallography and cryo-electron microscopy studies along with *in-vivo* experiments confirm that SARS-CoV-2 entry in host cells is also achieved by the spike-ACE2 interaction [8]. However, recent study suggest limited ACE2 expression in the human respiratory system [9]. Thus, the role of ACE2 as the sole receptor for SARS-CoV-2 infection needs to be further revisited.

Apart from the ACE2 and other surface protein receptor, many coronaviruses namely MERS-CoV, HCoV-OC43, and HCoV-HKU1 utilizes an alternate receptor to infect host cells, where the NTD domain of spike protein binds with host sialic

acid receptors. To understand this high infectivity and possibility of SARS-CoV-2 spike protein interaction with alternate receptor, we analyzed the primary sequence, secondary and tertiary structure of the NTD of SARS-CoV-2. Interestingly, we found the potential sialic acid binding pocket at the NTD of SARS-CoV-2 that appears to be the reminiscent of MERS-CoV sialic acid binding pocket. Further analysis of the spike

protein of SARS-CoV-2 illustrates three short amino acid motifs in the NTD region. These three motifs are the part of the same reminiscent sialic acid binding pocket similar to the one in MERS-CoV [10]. Recent studies report the presence of such motifs independently, but yet the potential significance of these motifs cooperatively could not be established [11, 12]. We used *in-silico* docking algorithms and molecular dynamic simulations to test the binding potential of this SARS-CoV2 NTD pocket with different sialosides. Remarkably, the dynamic nature of the β14-β15 loop possessing one of the motifs potentiates the binding of SARS-CoV-2 NTD with diverse and larger sialosides. Our result suggests that SARS-CoV-2 NTD is similar to the MERS- CoV and accommodate sialosides at neutral pH. Thus, our observation potentiates further study involving functional validation of reciprocal interaction of SARS-CoV-2 during virus entry, tissue tropism and in identifying therapeutic targets preventing such interaction.

Methods

Phylogenetic tree and sequence alignment

Phylogenetic analyses of Spike proteins of human coronaviruses were performed computationally by Neighbour – joining (NJ) method on MEGA-X with thousand bootstrap replicates [13]. The same was also verified by maximum likelihood ML method on MEGA-X and topology was viewed by MEGA-X as well as tree view and NJ plot [13]. Multiple sequence alignment of the NTD domains was performed with Clustal W program [14].

Structure preparation

The cryo-EM structures of SARS-CoV-2 (PDB ID: 6VXX) [8] and MERS-CoV (PDB ID: 6Q04) [10] spike glycoproteins were used as the starting point for all studies [8]. The full-length model of SARS-CoV-2 spike glycoprotein (YP_009724390.1) was strongly biased on the crystal structure of SARS-CoV-2, while the extended loops were first modeled using MODELLER, version 9v24 using the model-loop procedure [15]. The 100 resulting models were ranked using Discrete Optimized Protein Energy (DOPE) [16] statistical potentials. The best scoring model was protonated, and energy minimized by using Amber99Sb force field [17] and an RMS gradient of 0.1 kcal mol^{-1} A^{-2}. The minimized model was explicitly solvated by a spherical water box with cell borders placed at least 6 Å away from any protein atom using TIP3P as water model. To neutralize the total charge, Na^{+}/Cl^{-} counterions were added to a final salt concentration of 0.150. The solvated system was energy minimized by tethering all atoms by a harmonic potential to its starting coordinates with a 0.5 Å deviation. The MD simulations were conducted with Nosé-Poincaré-Anderson (NPA) algorithm and canonical ensemble (NVT). In the initialization step, the simulation was performed for 100 ps at 300 K, followed by a 100 ps equilibration step gradually increasing the temperature from 300 K to 310.15 K. Finally, the production step was carried out for 5000 ps using a 1 fs time step and applying harmonic positional constraints on protein by a force constant of 1 kcal mol^{-1} A^{-2} and non- rigid water molecules. During this step, the temperature was maintained at 310.15 K by a Langevin thermostat, the pressure at 1 atm by a Berendsen barostat and a heavy atom tether standard deviation of 0.5 Å deviation

Molecular docking and simulations

The molecular docking calculations were performed using the Lamarckian Genetic Algorithm implemented in Autodock 4.2.6 [18]. Initial ligand files were generated from SMILES strings using the Grade Web Server full-length model of SARS-CoV-2 glycoprotein obtained after NPA simulations and the ligands were prepared using the Autodock Tools program [18]. A local search was performed using Autosite [19] to identify possible binding sites on the NTD of SARS-CoV-2 glycoprotein. A grid box of 80 x 80 x 80 Å was defined centered on the selected binding pocket to allow all sialosides to rotate freely. A minimum ten docked poses were generated for each sialoside derivative and the best pose was selected based on the highest binding affinity. All sialoside-SARS-CoV-2 glycoprotein complexes were explicitly solvated by a spherical water box with cell borders placed at least 10 Å away from any protein or ligand atom. Na^+/Cl^- counterions were added to a final salt concentration of 0.150 M to neutralize the total charge of the system. The solvated protein-ligand complexes were then protonated, and energy minimized by using Amber12 [20] EHT force field and an RMS gradient of 0.1 kcal mol^{-1} A^{-2}. Molecular dynamics simulations were carried out at 310.15 K, 1 atm pressure and a heavy atom tether standard deviation of 0.5 Å. All systems were equilibrated for 50 ps followed by a 500 ps production step with a 2 fs time step and applying a harmonic positional constraints of 1 kcal mol^{-1} A^{-2} and non-rigid water molecules.

Results And Discussion

SARS-CoV-2 spike-protein NTD retain extended loop region structurally analogous to MERS-CoV

Receptor binding with host cell is the initial step in virus infection, tissue tropism and cell spread. Coronaviruses utilize complex pattern of receptor recognition to infect diverse host cells. COVID-19 caused by SARS-CoV-2 has been a cause of increase global burden with high co-morbidity and mortality [21]. Limited understanding of the diverse range of tissues targeted by the virus and its potential receptors, there is an immediate need to understand SARS-CoV-2 entry mechanism and pathogenesis to develop effective therapeutics. It is recently identified that trimeric transmembrane spike glycoprotein of SARS-CoV-2 and SARS-CoV that share a high degree of similarity within the RBD, binds to the common human ACE-2 receptor, yet SARS-CoV-2 is highly infectious than SARS-CoV [22]. To investigate the difference in these highly similar spike proteins, we analyzed the phyletic relatedness of the spike proteins from coronaviruses that are known to infect humans. As expected, spike proteins of SARS-CoV and SARS-CoV-2 are highly similar and both groups together in the same clade (Fig. 1a). The closest spike protein to SARS clade is of the MERS- CoV, which suggests that MERS-CoV shares a higher similarity to SARS clade than other coronaviruses. Additionally, the spike protein of HCoV-229E and HCoV-NL63 forms a separate clade (Fig. 1a). Even though, the spike proteins of HCoV-OC43 and MERS-CoV are distantly related, they both bind to host sialic acids as an alternate host-receptor during infection. To date, nothing is known about the interaction of SARS clade with host sialic acid receptors. Despite of the 76% homology between spike-proteins, SARS-CoV-2 is more infectious than SARS-CoV, which suggests a possible structural or mechanistic difference [7]. One stark difference between their spike

proteins is the presence of a furin-like cleavage site on SARS-CoV-2 spike protein [23]. SARS-CoV-2 has 12 extra nucleotides upstream to the single Arg↓ cleavage site forming PRRAR ↓SV sequence, which is a canonical furin-like cleavage site [23]. The presence of this furin-like cleavage site in SARS-CoV-2 is predicted as a possible reason for its efficient spread as compared to the other beta coronaviruses [23]. Alternatively, by comparative sequence analysis of the NTD of spike protein, we identified three extended region in SARS-CoV-2 and MERS-CoV (Fig 1b), but not in SARS-CoV (Fig 1b). To identify if these three regions forms a part of domain or a functional module of NTD sequence, we modeled the full-length SARS-CoV- 2 spike glycoprotein strongly biasing on the cryo-EM structure of SARS-CoV-2 spike protein (Fig 1c). The cryo-EM structures of SARS-CoV-2 spike protein display a well ordered β-strand rich NTD, RBD and the core helical domain [8]. Owing to their flexibility, all β-β loops, except for β14-β15 displays almost no cryo-EM density even after *B*-factor sharpening. The missing β-β loops were modeled *ab-initio* and the model with the best DOPE score was further energy minimized and used for computational analyses as discussed in the methods section. We compared this modeled structure and found a major difference between SARS-CoV and SARS-CoV-2 with respect to the loop lengths. SARS- CoV-2 has larger, β4-β5, β9-β10 and β14-β15 loops in comparison to SARS-CoV (Fig 1b and c), however these loop lengths were comparable to MERS-CoV. The β14-β15 loop is particularly interesting owing to its length and flexibility due to the presence of interspersed glycine residues and a flanking poly-alanine region (Fig 1b). This putative function of β14-β15 loop is reminiscent to the β6-β7 loop (Thr129-Thr136) of MERS-CoV [10]. MERS-CoV β6-β7 loop has a similar long arm loop that forms critical electrostatic anchor points to host sialoside receptor engagement and stability [10].

SARS-CoV-2 NTD motifs share loop region and predicted to bind sialosides.

To test the capacity of the SARS-CoV-2 spike protein to engage host-cell sialic acid receptors, we selectively docked Neu5Ac, 2,3-SLN, 2,6-SLN, Neu5Gc and sLeX (Fig. 2) on to the SARS-CoV-2 NTD. The selected sialosides represents large family of more than 500 human sialoside and have been previously shown to bind with the S1A domain of MERS-CoV [10]. A recent study by Milanetti *et. al*, also predicted a sialoside-binding pocket in the NTD of SARS-CoV-2 by surface iso-electron density mapping, further highlights the importance of NTD interaction with sialic acid [24]. The amino acid residues Leu18-Gln23, His66- Thr78 of β4-β5 loop, and Gly252-Ser254 of β14-β15 loop forms the sialic acid- binding site in SARS-CoV-2 spike protein (Fig. 2 and 3). While, the β4-β5 loop is involved in the engagement with all sialosides, the β14-β15 loop is specific to larger sialic acids such as sLex (Figs 2 and 3) which possibly suggest to have preferential interaction with α2,3-Linked sialosides as in MERS-CoV [25, 26]. The predicted interacting sites of the tested sialosides are mapped in Fig. 3. The presence of key electrostatic and hydrophobic interactions with each of these sialosides suggests possibility of a physiological interaction with the NTD domain of SARS-CoV-2. Molecular dynamics simulation of SARS-CoV-2 NTD-sialoside complexes highlights the flexibility of β14-β15 loop and its induced ability to accommodate larger sialosides (Fig 2f). The superimposition of all produced SARS-CoV-2 NTD-sialoside complexes show an outward movement of the β14- β15 loop, allowing the sialic acid-binding site to accommodate larger sialosides such as sLex (Fig 2f). On the other hand, the spike protein of SARS-CoV features a shorter 9 amino acid β14-β15 loop (Fig. 1b), which offers reduced degrees of freedom, with a decreased

capacity to engage host sialosides. In addition, a single-turn alpha helix (Thr20-Leu24) formed key interactions with all sialosides tested. Interestingly, the NTD of MERS-CoV also displays a single-turn helix (Gln37-Phe40), which is important for sialoside binding [25]. However, both the SARS-CoV and HCoV-OC43 spike proteins lack this element [7, 10, 13, 18, 26]. Taken together, these findings suggest that SARS-CoV-2 spike protein might have independently evolved to recognize sialosides using its NTD. This acquired ability of SARS-CoV-2 to accommodate and engage diverse sialosides might be a reason to ascribe potential role of sialosides as an alternative reciprocal host cell receptor supporting SARS-CoV-2 pathogenesis with broad tissue tropism. Such differential distribution of sialic acid in the respiratory tract and other organs along with limited ACE2 expression in human airway epithelia [9] explains the differential SARS-CoV-2 infectivity, transmission and tropism [27]. In connotation to this, the recent preprint report suggests that the spike glycoprotein also recognizes different Siglecs (Sialic acid–binding Ig-like lectins) and C-type lectins which indicates that the spike protein interact in ACE2-independent infection pathways with the immune cells [28]. In addition, the human ABO blood group and COVID-19 susceptibility [29, 30] may relate to modulation of sialosides distribution pattern on target membrane, possibly regulating SARS-CoV-2 transmission and tropism [31] despite of high affinity with ACE2. The comprehensive *in-silico* structural analysis reported in this study; provide a basis for further research to explore the functional role of the reciprocal interaction of SARS-CoV-2 with host sialic acid during virus entry and spread.

Declarations

Acknowledgement

We are grateful to Dr. Kartika Padhan, NIAID, NIH, USA for scientific comments on this manuscript. This study is not funded by any financial support.

Declaration of competing interest

The authors declare no conflicts of interest.

References

1. Jiang, S. Rayner, M.H. Luo, Does SARS-CoV-2 has a longer incubation period than SARS and MERS?, J Med Virol, 92 (2020) 476-478.
2. Zumla, J.F. Chan, E.I. Azhar, D.S. Hui, K.Y. Yuen, Coronaviruses - drug discovery and therapeutic options, Nat Rev Drug Discov, 15 (2016) 327-347.
3. Shi, M. Qin, B. Shen, Y. Cai, T. Liu, F. Yang, W. Gong, X. Liu, J. Liang, Q. Zhao, H. Huang, B. Yang, C. Huang, Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China, JAMA Cardiol, (2020).
4. Su, M. Yang, C. Wan, L.X. Yi, F. Tang, H.Y. Zhu, F. Yi, H.C. Yang, A.B. Fogo, X. Nie, C. Zhang, Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China, Kidney Int,

(2020).
5. Mao, H. Jin, M. Wang, Y. Hu, S. Chen, Q. He, J. Chang, C. Hong, Y. Zhou,
6. Wang, X. Miao, Y. Li, B. Hu, Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China, JAMA Neurol, (2020).
7. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Muller, C. Drosten, S. Pohlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, Cell, 181 (2020) 271-280 e278.
8. Othman, Z. Bouslama, J.T. Brandenburg, J. da Rocha, Y. Hamdi, K. Ghedira, N. Srairi-Abid, S. Hazelhurst, Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: Similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism, Biochem Biophys Res Commun, (2020).
9. C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, Cell, 181 (2020) 281-292 e286.
10. Hikmet, L. Méar, M. Uhlén, C. Lindskog, The protein expression profile of ACE2 in human tissues, bioRxiv preprint (2020).
11. J. Park, A.C. Walls, Z. Wang, M.M. Sauer, W. Li, M.A. Tortorici, B.J. Bosch, F. DiMaio, D. Veesler, Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors, Nat Struct Mol Biol, 26 (2019) 1151-1157.
12. Behloul, S. Baha, R. Shi, J. Meng, Role of the GTNGTKR motif in the N- terminal receptor-binding domain of the SARS-CoV-2 spike protein, Virus Res, (2020) 198058.
13. B., Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as the binding of host sialic acid glycans, Computers in Biology and Medicine, Volume 122 (2020).
14. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Mol Biol Evol, 35 (2018) 1547-1549.
15. A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics, 23 (2007) 2947-2948.
16. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using Modeller, Curr Protoc Bioinformatics, Chapter 5 (2006) Unit-5 6.
17. Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci, 15 (2006) 2507-2524.
18. W. Ponder, D.A. Case, Force fields for protein simulations, Adv Protein Chem, 66 (2003) 27-85.
19. M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J Comput Chem, 30 (2009) 2785-2791.

20. A. Ravindranath, M.F. Sanner, AutoSite: an automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms, Bioinformatics, 32 (2016) 3142-3149.
21. A. Case, T.E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J Comput Chem, 26 (2005) 1668-1688.
22. Roussel, A. Giraud-Gatineau, M.T. Jimeno, J.M. Rolain, C. Zandotti, P. Colson, D. Raoult, SARS-CoV-2: fear versus data, Int J Antimicrob Agents, 55 (2020) 105947.
23. Zheng, SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat, Int J Biol Sci, 16 (2020) 1678-1685.
24. Coutard, C. Valle, X. de Lamballerie, B. Canard, N.G. Seidah, E. Decroly, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade, Antiviral Res, 176 (2020) 104742.
25. Milanetti, M. Miotto, L. Di Rienzo, M. Monti, G. Gosti, G. Ruocco, In-Silico evidence for two receptors based strategy of SARS-CoV-2, BioRxiv [Preprint], (2020).
26. Li, R.J.G. Hulswit, I. Widjaja, V.S. Raj, R. McBride, W. Peng, W. Widagdo, M.A. Tortorici, B. van Dieren, Y. Lang, J.W.M. van Lent, J.C. Paulson, C.A.M. de Haan, R.J. de Groot, F.J.M. van Kuppeveld, B.L. Haagmans, B.J. Bosch, Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein, Proc Natl Acad Sci U S A, 114 (2017) E8508-E8517.
27. A. Tortorici, A.C. Walls, Y. Lang, C. Wang, Z. Li, D. Koerhuis, G.J. Boons, B.J. Bosch, F.A. Rey, R.J. de Groot, D. Veesler, Structural basis for human coronavirus attachment to sialic acid receptors, Nat Struct Mol Biol, 26 (2019) 481-489.
28. Schwegmann-Wessels, G. Herrler, Sialic acids as receptor determinants for coronaviruses, Glycoconj J, 23 (2006) 51-58.
29. C. F., B.S. C.M., R. E., L. R.J.E., M. A., S. A., L.F. Di, G.-R. D., V.-B. Y., V.-V., K. Y.V., Novel ACE2-Independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota, BioRxiv [Preprint], (2020). [29] Z. J., Y. Y., H. H., L. D., G. D., L. X., Z. Z., L. L., L. T., L. Y., H. Y., S. B., W. M., Y. G., W. X, Z. L, Z. X., X. M., W. P.G, Relationship between the ABO Blood Group and the COVID-19 Susceptibility, BioRxiv [Preprint], (2020).
30. Breiman, N. Ruven-Clouet, J. Le Pendu, Harnessing the natural anti- glycan immune response to limit the transmission of enveloped viruses such as SARS-CoV-2, PLoS Pathog, 16 (2020) e1008556.
31. Cohen, N. Hurtado-Ziola, A. Varki, ABO blood group glycans modulate sialic acid recognition on erythrocytes, Blood, 114 (2009) 3668-3676.

Figures

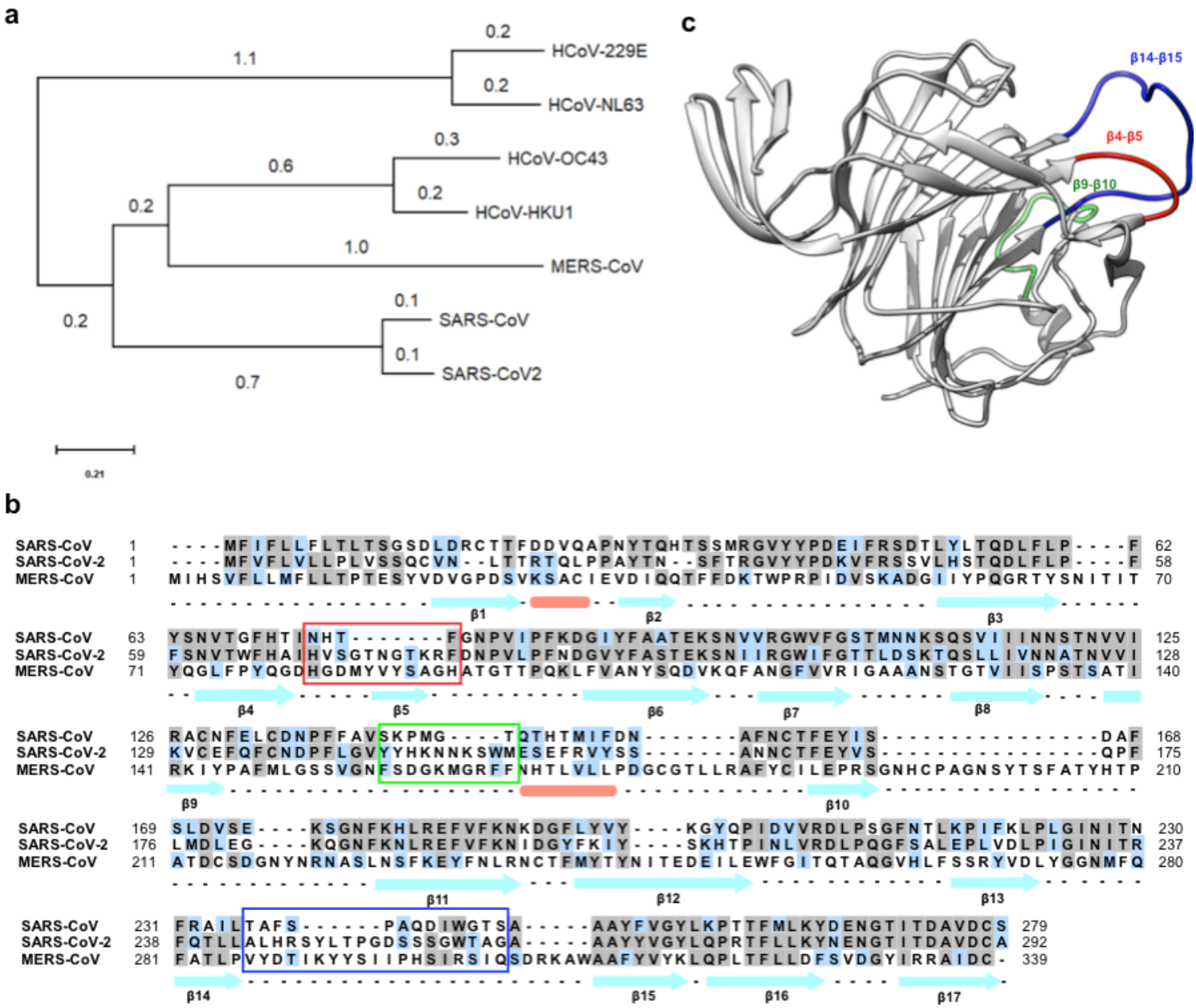

Figure 1

(a) Evolutionary analysis of the spike protein of human-infecting coronaviruses. Phylogenetic tree was drawn by using the Maximum Likelihood method and scale with branch lengths measured in the number of substitutions per site. (b) Sequence alignment of NTD of SARS-CoV, SARS-CoV-2 and MERS-CoV illustrates the divergent region of otherwise highly similar protein sequence. The beta strands are highlighted and the box depicts longer length of β loop of SARS-CoV-2 and MERS-CoV. (c) Tertiary structure of NTD of spike protein of SARS-CoV-2. The loops regions are highlighted in blue, which is adjacent to the predicted sialic acid binding pocket depicted by red and green.

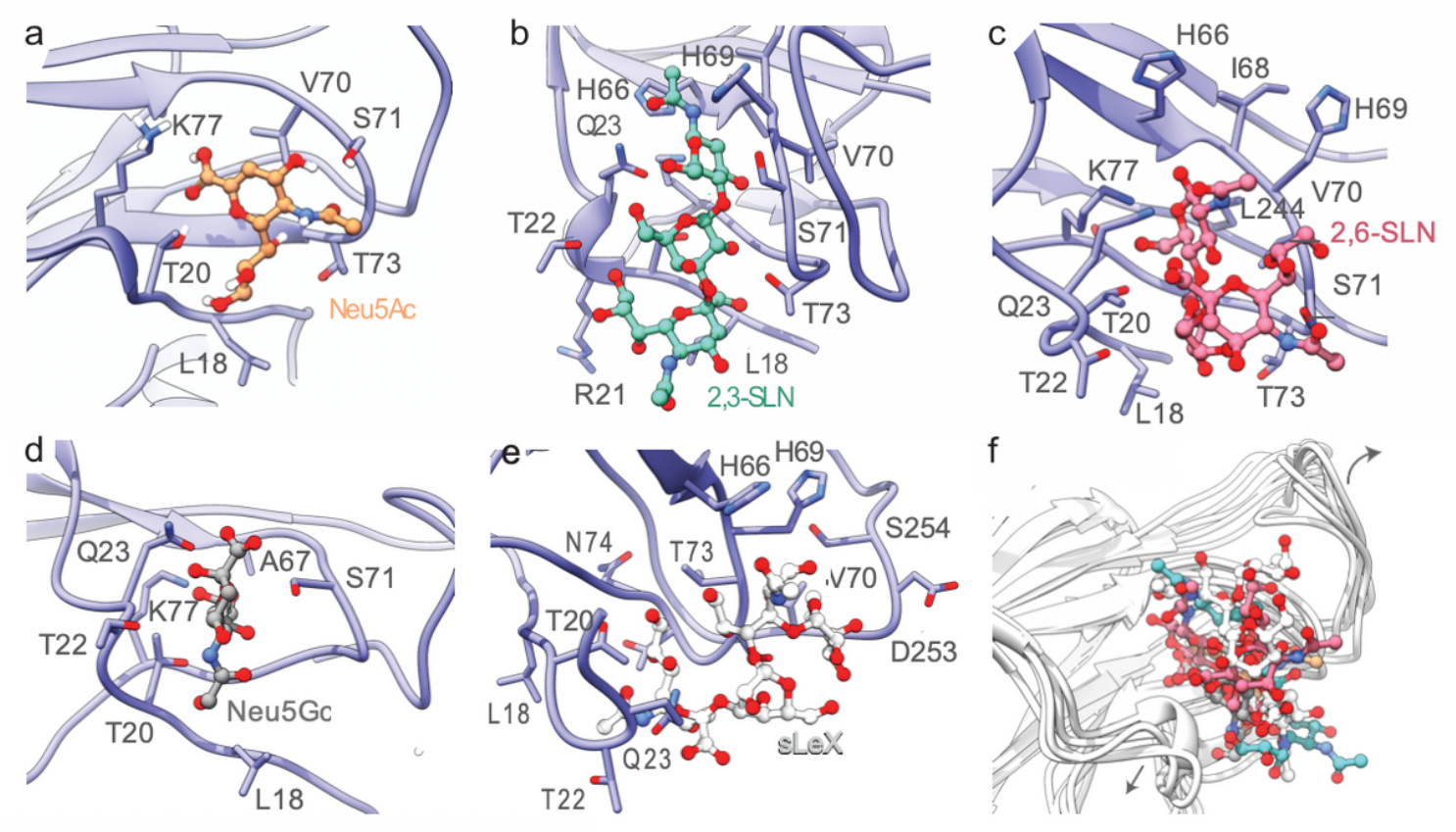

Figure 2

Binding between the docked sialoside derivatives and the NTD of SARS-CoV-2 glycoprotein. The key amino acid side chains surrounding (a) 2- Deoxy-2,3-dehydro-n-acetyl-neuraminic acid, (b) 3-Sialyl-N-acetyllactosamine, (c) 6-Sialyl-N-acetyllactosamine, (d) N-glycoloyl-beta-neuraminic acid, and (e) Sialyl-Lewis X are shown (purple sticks). (e) The flexibility of β4-β5 loop (Leu244-Gly261) enables binding of a wide variety of sialosides. The NTD of SARS-CoV-2 glycoprotein is shown as purple ribbons and sialoside derivatives are shown as ball and stick models

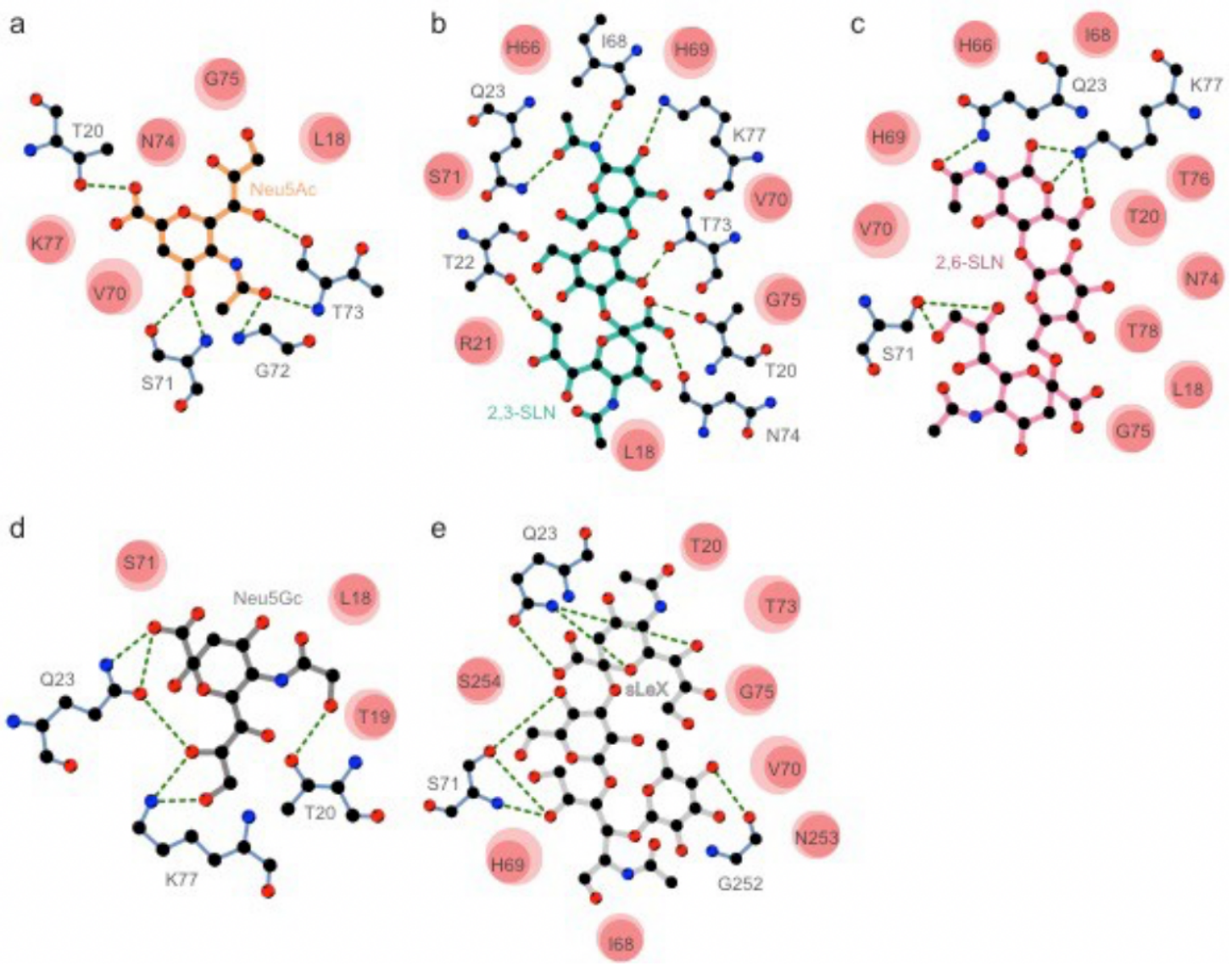

Figure 3

Protein-ligand interactions between the NTD of SARS-CoV-2 glycoprotein and (a) 2-Deoxy-2,3-dehydro-n-acetyl-neuraminic acid, (b) 3-Sialyl- N-acetyllactosamine, (c) 6-Sialyl-N-acetyllactosamine, (d) N-glycoloyl-beta- neuraminic acid, and (e) Sialyl-Lewis X. Dashed green lines show electrostatic interactions formed between the SARS-CoV-2 glycoprotein amino acid residues and the ligand. Hydrophobic contacts are shown as filled circles, where the orientation and size of the opaque ellipsoid marks the directionality and strength of hydrophobic interactions. In all panels, nitrogen, carbon and oxygen atoms are colored blue, black and red, respectively.