

Use of Isoelectric Point for Fast Identification of Anti-SARS CoV-2 Coronavirus Proteins

Kanad Mallik*

Abstract

A novel approach has been suggested to use isoelectric points of viral and human proteins to quickly identify proteins that are effective in not allowing virus particles to attach to human receptor cells by virtue of their electrical charge. The method has been applied to SARS CoV-2 to suggest potentially important human proteins that can be suitable for making anti-viral drugs.

Keywords: Isoelectric point, anti-viral, SARS CoV-2

Introduction

The on-going SARS CoV-2 Coronavirus pandemic is unprecedented in a century, but it had its precursors in the forms of SARS and MERS outbreaks since 2003 and there are possibilities that similar zoonotic transmissions will increase in the coming years¹. Such a scenario needs international preparedness in many areas including effective viral treatments. Reliance on vaccines alone is not adequate because of their narrow applicability, and long times required for their development and deployment. Broad-spectrum anti-viral drugs working on general principles to fight viruses will be effective to complement vaccines and save lives.

For any viral infection it is important for the treatment to minimise or prevent attachment of viral particles (virions) to human cells within the body. Remembering that a virus is essentially a nanoparticle of sizes up to a few hundred nanometres and is electrically charged²⁻⁶, its attachment to any surface is dominantly controlled by electrostatic forces which depend on the charge of the virus. As its surface-to-volume ratio is inversely proportional to the radial dimension of the virus, the charge on/near the viral surface becomes a dominant factor. Further it depends on the pH of the ambient medium (generally, blood) and the charge of the receptors in the body¹. Charges of both the virus and the receptors are, in turn, determined by their constituent proteins. This suggests that it is possible to 'entice' virions to get attached to preferred receptors other than their usual prey human cells depending on their charges, engineered by the constituent proteins of the receptors. In this communication it has been found that it is feasible to quickly assess the 'affinity' of SARS CoV-2 virions to specific cells based on their isoelectric point (pI) determined by proteins they are made of.

Methods

Isoelectric point⁷⁻¹² is a well-known measure of charge of proteins with reference to the pH of their environment. Proteins having $pI > pH$ are positively charged, and $pI < pH$ are negatively charged. This principle has been the basis of established bio-chemical techniques like electrophoresis and isoelectric focusing. In this work values of pI of Coronaviral proteins and the human ACE2 receptors have been calculated using the IPC tool⁸ which employs standard experimental values of acid dissociation constants (pKa) of constituent amino acids and FASTA sequences of the proteins.

Results

SARS CoV-2 Coronavirus has an average diameter of 125 nm¹³ and consists of central RNA enclosed in the nucleocapsid surrounded by the envelope and membrane and spikes form the outermost layer of the virus, diagrammatically shown¹⁴ in Fig. 1. UniProt¹⁵ IDs of proteins found in different parts of the

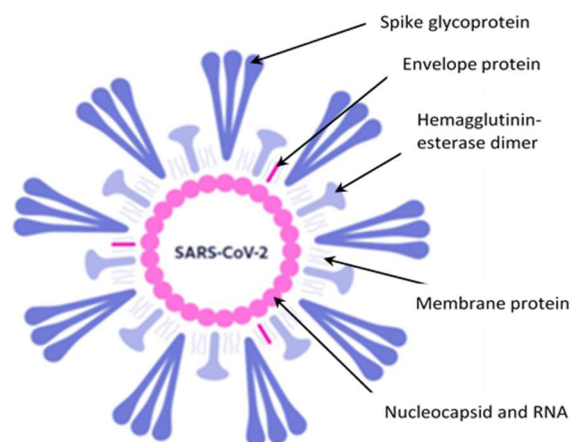


Figure 1 - A diagram showing the radial protein structure of SARS CoV-2 virus. Courtesy: Drug Bank¹⁴ white paper, "COVID-19: Finding the Right Fit".

structure along with their calculated values of pI and charge are given in Fig. 2. In a set of work, the protein of the S230 antigen-binding natural antibody was investigated before (PDB ID 6NB8)¹⁷ and after (PDB ID 2DD8)¹⁸ binding with the virus. Table 1 lists the proteins and their values of pI, calculated according to the IPC Protein formalism⁹ from the respective FASTA sequence.

*Proxisense Ltd, 148 Sixth Street, Didcot, Oxfordshire OX11 0TR, United Kingdom. E-mail:kanad.mallik@gmail.com, ORCID ID: 0000-0003-2568-9556

Table 1. Comparison of calculated isoelectric points (pI) according to the IPC protein formalism⁹ of various proteins with UniProt (*) and PDB (+) IDs from their FASTA sequences in different conditions.

| Protein ID in database | Protein | Stage | Isoelectric Point |
|------------------------|-----------------------------------------------------|-------------------------------|-------------------|
| Q202F5 (*) | Coronavirus Spike protein | Before virus-ACE2 cell fusion | 5.41 |
| J9TDZ0 (*) | Coronavirus Spike protein | | 5.321 |
| Q9BYF1 (*) | Human receptor Angiotensin-converting enzyme (ACE2) | After virus-ACE2 cell fusion | 5.219 |
| 6M18, 6M1D (+) | Human ACE2 receptor | | 5.283 |
| 2AJF (+) | Coronavirus spike protein | | 7.132 |
| 2DD8 (+) | Coronavirus Spike glycoprotein | Before attachment to antigen | 7.059 |
| 2AJF (+) | Human ACE2-related Carboxypeptidase | | 4.902 |
| 2DD8 (+) | Human ACE2 IGG heavy chain | | 7.044 |
| 2DD8 (+) | Human ACE2 IGG light chain | | 5.703 |
| 6W7Y (+) | Human antibody CR3022 heavy chain | | 7.79 |
| 6W7Y (+) | Human antibody CR3022 light chain | Before attachment to antigen | 5.906 |
| 6NB8 (+) | Human S230 antigen-binding fragment heavy chain | | 8.083 |
| 6NB8 (+) | Human S230 antigen-binding fragment light chain | | 6.046 |

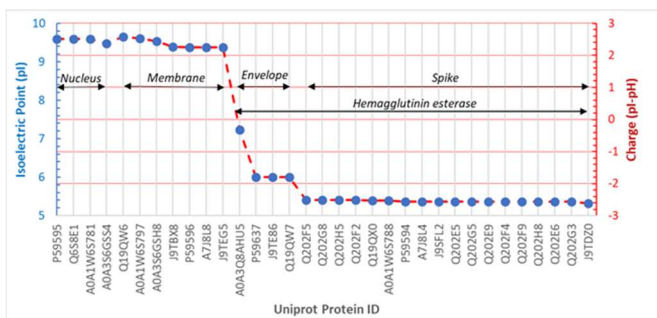


Figure 2 - A plot showing isoelectric points of proteins (UniProt IDs in x-axis) from SARS CoV-2 virus nucleus, membrane, envelope, and spikes as indicated by the double-headed arrows. Corresponding FASTA sequences have been used to calculate values of pI. Hemagglutinin esterase proteins (ID-s not shown in the plot for brevity) have pI-s in the range of 7.51-4.85.

In a recent work, Gordon and co-workers¹⁹ have reported detailed mapping of 332 high-confidence interactions between SARS Cov-2 viral proteins and human proteins to identify 66 drugs that can potentially be re-purposed for Coronavirus treatments. It is further instructive to look at the Biogrid²⁰ interaction data of the publication from electrical charge

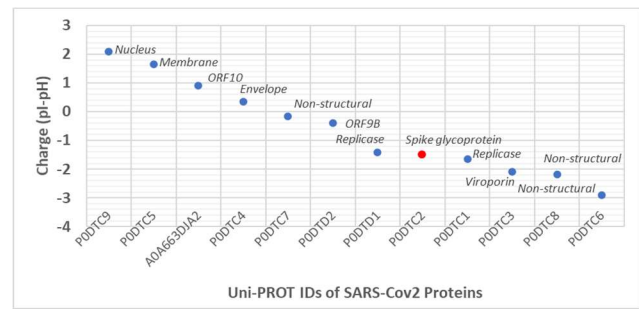


Figure 3 - Calculated charge on principal SARS CoV-2 viral proteins with UniProt IDs in normal human blood pH of 7.4 from isoelectric points and FASTA sequences.

considerations. For this, pI values for all the proteins reported¹⁹ in these interactions were calculated using the IPC tool using respective FASTA sequences. Subsequently, charge for each protein was calculated according to $Q = (pI - pH)$ with $pH = 7.4$. Figure 3 shows the resultant charges of these different Coronavirus proteins. Among them, PODTC2 is the viral spike glycoprotein that bears negative charge and is the key in docking interactions with human ACE2 receptor cells as structurally the spikes form the outermost layer of the virus. Figure 4 shows the difference, $\Delta pI = pI (\text{host protein}) - pI (\text{viral protein})$, plotted for the proteins which have the five highest values of ΔpI among the human proteins reported in the Gordon et. al.'s work. Corresponding interaction scores also are plotted, which show that interactions between these pairs of proteins are high.

Discussion

If we follow the radial structure of the SARS CoV-2 virion, a gradual fall in the pI values is observed in Fig. 2 from 9.59 in the nucleocapsid to 5.32 in the spike proteins. This implies that at the normal human blood pH of 7.4 (average of the

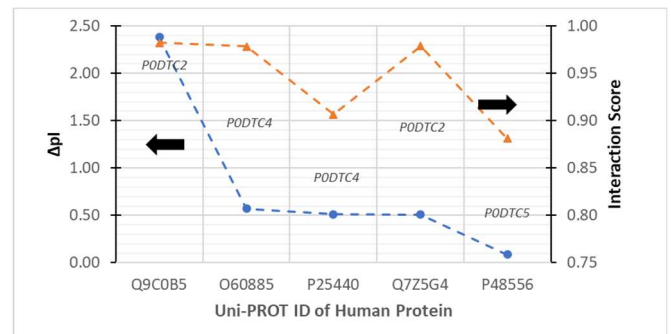


Figure 4 - Excess charge of human proteins featured in Gordon et. al.'s work in interactions between the human protein and principal SARS CoV-2 viral protein pairs calculated from isoelectric points and FASTA sequences are plotted on the left-hand vertical axis. Corresponding interaction scores reported are plotted in the right-hand vertical axis.

range, 7.35 - 7.45) a virion has a positive core surrounded by a negatively charged envelope and spikes. As the spikes are instrumental in the docking mechanism to human ACE2 receptors (PDB IDs 6M18, 6M1D), pI-s of the spike proteins (and so, their charge) play a major role in the mechanism. A spike protein was studied following fusion with the ACE2 receptor (PDB ID 2AJF)²¹. It is observed from Table 1 that pI values for the Coronavirus spike proteins before fusion with human ACE2 receptors range from 5.41 - 5.32. As a result, free coronavirus particles are heavily negatively charged in the

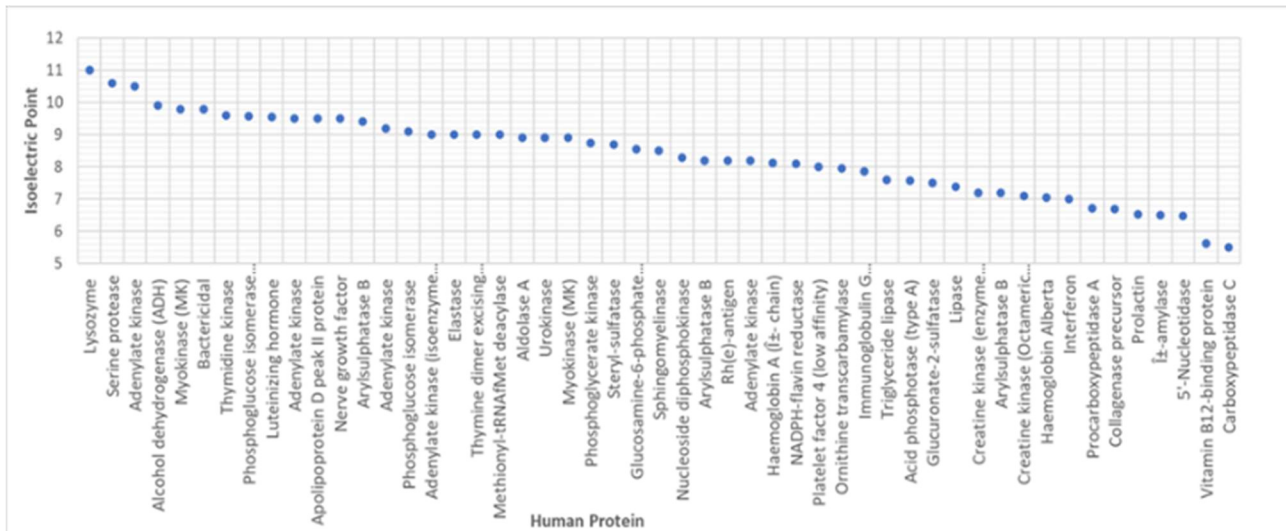


Figure 5 - Experimental values of isoelectric point for human proteins in the 5.5-11 range from the PIP-DB¹¹. These have potential to be utilised for making anti-viral drugs to engage the SARS CoV-2 spike protein away from its natural ACE2 receptor in the human body.

blood stream. This is consistent with reports that a large variety of viruses are acidic in nature²². On the other hand, the pI-s of the viral spike proteins increase to a range of 7.04 - 7.13 following their fusion with the human ACE2 receptors indicating that they bear smaller negative charges on them. The values of pI of the ACE2 cells change from a range of 5.28 - 8.08 before the virion attachment to a range of 4.92 - 7.04 following the attachment. Hence the negative charge on the ACE2 receptors increases and/or the positive charge decreases. If we combine this observation with changes in charge on the viral spike proteins, it turns out that the attachment process of the virions to the receptors leads to a transfer of negative charges from the viral spike protein to that of the receptor, which is analogous to oxidation of the virus spike protein by the ACE2 receptor protein. From this dynamics of charge, it may be hypothesised that attachment of the Coronavirus to human cells in the body can be prevented or minimised if, compared to the ACE2 receptor protein, a stronger oxidising species which has pI values in ranges higher than 5.28 – 8.08 is made available in the blood stream. It is notable that proteins (PDB ID 6W7Y) in the CR3022 human antibody for SARS CoV-2 have pI values 7.79 and 5.906 which are higher than those of the ACE2 protein and antibodies can attract virion spikes better than ACE2 proteins. This observation supports the above hypothesis.

Figure 4 shows a plot of excess charge (ΔpI) of five human proteins over the Coronaviral spike, envelope, and membrane proteins they were reacted with in Gordon et. al.'s work. These proteins have the highest values of ΔpI among the 332 interactions they studied. The corresponding values of their interaction scores also are shown. As negative charges get transferred from the spike protein to the host cells in the docking process, host proteins that are more capable of accepting such charges would be preferred for attachment. In other words, higher the pI of the host cell protein compared to that of the spike protein, larger would be the affinity of the spike protein to attach to it. Hence from the electrostatics point of view, Palmitoyltransferase (UniProt ID Q9C0B5, pI = 8.29) and Golgin (UniProt ID Q7Z5G4, pI = 6.42) are the best

candidates among the reported set of proteins to engage the SARS CoV-2 spike protein (UniProt ID P0DTC2). Bromodomain-containing proteins (UniProt ID O60885, pI = 8.33 and UniProt ID P25440, pI = 8.27) are similarly effective for engaging the Coronavirus envelope protein (UniProt ID P0DTC4), and 26S Proteasome (UniProt ID P48556, pI = 9.14) for the viral membrane protein (UniProt ID P0DTC5). A combination of drugs to deliver the proteins suitable for engaging the viral spike, envelope, and membrane proteins in preference to the human ACE2 receptor cell protein may be a quick good way forward to combat the SARS CoV-2 virus.

Figure 5 shows experimental values¹¹ of pI of human proteins in the range of 5.5 – 11 which have the potential of attracting the Coronavirus in preference to the human ACE2 receptor cells as from the charge transfer considerations, larger the pI more efficient the protein is for this purpose. It is better, therefore, to select proteins from the higher end of this list to make anti-viral drugs for SARS CoV-2. From the pI consideration, application of human Interferon- β (PDB ID 1AU1) which is currently being trialled for Coronavirus treatment has a pI of 7 (calculated value is 8.22) and lies in the lower end of the list is not expected to be effective.

Conclusions

A novel analysis based on isoelectric points of SARS CoV-2 and various human proteins shows that it is possible to quickly identify proteins that are potentially efficient as anti-viral drugs that would engage the virus saving human ACE2 cells from infection. This method of prevention of viral infection using control of virion electrical charge could be another general way to tackle viral infections.

Conflicts of interest

There are no conflicts to declare.

References

- 1 D. Schoeman and B. C. Fielding, Coronavirus envelope protein: current knowledge, *Virology*, 2019, **16**, 69.
- 2 S. Karlin and V. Brendel, Charge configurations in viral proteins, *Proc. Natl. Sci.*, 1988, **85**, 9396–9400.
- 3 F. Patolsky, G. Zheng, O. Hayden, M. Lakadamyali, X. Zhuang and C. M. Lieber, Electrical detection of single viruses, *Proc National Acad Sci*, 2004, **101**, 14017–14022.
- 4 F. Caridi, S. López-Argüello, A. Rodríguez-Huete, E. Torres, M. J. Bustos, R. Cañas-Arranz, M. A. Martín-Acebes, M. G. Mateu and F. Sobrino, Negatively charged amino acids at the foot-and-mouth disease virus capsid reduce the virion-destabilizing effect of viral RNA at acidic pH, *Sci Rep-uk*, 2020, **10**, 1657.
- 5 M. Hernando-Pérez, A. X. Cartagena-Rivera, A. L. Božič, P. J. P. Carrillo, C. S. Martín, M. G. Mateu, A. Raman, R. Podgornik and P. J. de Pablo, Quantitative nanoscale electrostatics of viruses., *Nanoscale*, 2015, **7**, 17289–98.
- 6 M. V. Arkhipenko, N. A. Nikitin, O. A. Baranov, E. A. Evtushenko, J. G. Atabekov and O. V. Karpova, Surface Charge Mapping on Virions and Virus-Like Particles of Helical Plant Viruses, *Acta Naturae*, 2019, **11**, 73–78.
- 7 C. P. Gerba, Advances in Applied Microbiology, *Adv Appl Microbiol*, 1984, **30**, 133–168.
- 8 L. P. Kozlowski, IPC – Isoelectric Point Calculator, *Biol Direct*, 2016, **11**, 55.
- 9 L. P. Kozlowski, Proteome-pI: proteome isoelectric point database, *Nucleic Acids Res*, 2017, **45**, D1112–D1116.
- 10 B. Michen and T. Graule, Isoelectric points of viruses, *J Appl Microbiol*, 2010, **109**, 388–397.
- 11 E. Bunkute, C. Cummins, F. J. Crofts, G. Bunce, I. T. Nabney and D. R. Flower, PIP-DB: The Protein Isoelectric Point database, *Bioinformatics*, 2015, **31**, 295–296.
- 12 B. D. Halligan, ProMoST: A tool for calculating the pI and molecular mass of phosphorylated and modified proteins on 2 dimensional gels, *Methods Mol. Biol.*, 2009, **527**, 283.
- 13 B. W. Neuman, B. D. Adair, C. Yoshioka, J. D. Quispe, G. Orca, P. Kuhn, R. A. Milligan, M. Yeager and M. J. Buchmeier, Supramolecular Architecture of Severe Acute Respiratory Syndrome Coronavirus Revealed by Electron Cryomicroscopy, *J Virol*, 2006, **80**, 7918–7928.
- 14 D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018., *Nucleic Acids Res*, 2017, **46**, D1074–D1082.
- 15 A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, E. Alpi, B. Bely, M. Bingley, R. Britto, B. Bursteinas, G. Busiello, H. Bye-A-Jee, A. D. Silva, M. D. Giorgi, T. Dogan, L. G. Castro, P. Garmiri, G. Georghiou, D. Gonzales, L. Gonzales, E. Hatton-Ellis, A. Ignatchenko, R. Ishtiaq, P. Jokinen, V. Joshi, D. Jyothi, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, M. Menchi, A. Nightingale, J. Onwubiko, B. Palka, K. Pichler, S. Pundir, G. Qi, S. Raj, A. Renaux, M. R. Lopez, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, P. Vasudev, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. Bridge, I. Xenarios, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.-C. Blatter, J. Bolleman, E. Boutet, L. Breuza, C. Casals-Casas, E. de Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, U. Hinz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, G. Keller, A. Kerhornou, V. Lara, P. Lemerrier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. B. Neto, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, C. Sigrist, K. Sonesson, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. Cowart, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L.-S. Yeh and J. Zhang, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res*, 2018, **47**, D506–D515.
- 16 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, 2000, **28**, 235–242.
- 17 A. C. Walls, X. Xiong, Y.-J. Park, M. A. Tortorici, J. Snijder, J. Quispe, E. Cameroni, R. Gopal, M. Dai, A. Lanzavecchia, M. Zambon, F. A. Rey, D. Corti and D. Veesler, Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion, *Cell*, 2019, **176**, 1026-1039.e15.
- 18 P. Prabaharan, J. Gan, Y. Feng, Z. Zhu, V. Choudhry, X. Xiao, X. Ji and D. S. Dimitrov, Structure of Severe Acute Respiratory Syndrome Coronavirus Receptor-binding Domain Complexed with Neutralizing Antibody, *J Biol Chem*, 2006, **281**, 15829–15836.
- 19 D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O’Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Huettenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. M. Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorbha, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataraman, J. Liboy-Lugo, Y. Lin, X.-P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O’Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu, H.-Y. Wang, R. Trenker, J. M. Young, D. A. Caverro, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d’Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet and N. J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing., *Nature*, 2020, 1–13.
- 20 C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, BioGRID: a general repository for interaction datasets, *Nucleic Acids Res*, 2006, **34**, D535–D539.
- 21 F. Li, W. Li, M. Farzan and S. C. Harrison, Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor, *Science*, 2005, **309**, 1864–1868.
- 22 C. D. Lytle and L. B. Routson, Minimized Virus Binding for Tests of Barrier Materials, *Applied and Environmental Microbiology*, 1995, **61**, 643–649.