The role of host cell glycans on virus infectivity: The SARS-CoV-2 case

Silvia Acosta-Gutiérrez^{1,2,3,*‡}, **Joseph Buckley**^{1,2,3,‡} and **Giuseppe Battaglia**^{1,2,3,4,5,*}

¹Department of Chemistry

²Institute for the Physics of Living Systems

³Institute of Structural and Molecular Biology, University College London, London, UK.

⁴Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute for Science and Technology (BIST), Barcelona, Spain.

⁵Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

***Corresponding authors:** Dr Silvia Acosta-Gutiérrez s.gutierrez@ucl.ac.uk and Prof Giuseppe Battaglia: g.battaglia@ucl.ac.uk and gbattaglia@ibecbarcelona.eu

‡These authors have contributed equally

Long and complex chains of sugars, called glycans, often coat both the cell and protein surface. Glycans both modulate specific interactions and protect cells. On the cell surface, these sugars form a cushion known as the glycocalyx. Here, we show that Heparan Sulfate (HS) chains - part of the glycocalyx - and other glycans - expressed on the surface of both host and virus proteins - have a critical role in modulating both attractive and repulsive potentials during viral infection. We analyse the SARS-CoV-2 virus, modelling its spike proteins binding to HS chains and two key entry receptors, ACE2 and TMPRSS2. We include the volume exclusion effect imposed on the HS chains during virus insertion into glycocalyx and the steric repulsion caused by changes in the conformation of the ACE2 glycans involved in binding to the spike. We then combine all these interactions, showing that the interplay of all these components is critical to the behaviour of the virus. We show that the virus tropism depends on the combinatorial expression of both HS chains and receptors. Finally, we demonstrate that when both HS chains and entry receptors express at high density, steric effects dominate the interaction, preventing infection.

Keywords: multivalent polymers, super-selectivity, virus, nanoparticles, multiplexing

Introduction

The COVID-19 pandemic, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has caused over 124 million infections and over 3 million deaths worldwide (World Health Organisation, WHO). SARS-CoV-2 belongs to the betacoronavirus genus of the coronaviridae family, the same as the Middle East Respiratory Syndrome coronavirus (MERS-CoV), and the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), two well established pathogens which emerged in the middle east in 2012 and the south of China in 2003, respectively [1]. These outbreaks have demonstrated that coronaviruses have a high zoonotic potential, and therefore, understanding their mechanism of infection is necessary to improve our therapeutical approach and lessen the burden on our healthcare systems[2].

The virus life cycle involves a game of multivalency to enable binding to and infection of the host cell (**Figure 1**) [3]. Viruses recognise and bind to multiple cell surface receptors, either via protein/protein or protein/glycan interaction. Different viruses target different receptors – receptors which may have diverse sequences, structures, and cellular functions[4]. Viruses prefer the membrane proteins involved in cell adhesion [5]. Nature tightly regulates the single viral protein/receptor interaction or affinity and, moreover, their combination into multivalent and multiplexed association profiles which lead to sharp response to gradients of receptors density and give rise to the so-called super-selective binding [6].

Most human cells, particularly those that comprise biological barriers such as endothelial and epithelial cells, have a membrane coated by a dense and complex mixture of long polysaccharides, known as glycosaminoglycan (GAG) chains. These chains bind to specific proteins forming compounds called proteoglycans. This chemical cushion, known as the glycocalyx, has several functions, including protecting cells from chemical injury and hindering the entry of pathogens. Several viruses have evolved to exploit these glycans, using them as an attachment factor to facilitate infection [7, 8]. SARS-CoV-2 is an example of this, binding to Heparan Sulfate (HS), one of the most abundant components of the glycocalyx [9, 10, 11]. Additionally, it has been shown that Heparin – a Heparan Sulfate analogue and widely used anticoagulant, antithrombotic, and blood-thinning drug – inhibits infection in-vitro [9, 12, 11]. Several clinical reports have also suggested the therapeutic effect of heparin in both moderate and severe COVID-19, and there are currently over 90 current clinical trials ongoing investigating this [13].

As shown in **Figure 1**, the SARS-CoV-2 virus spike glycoprotein interacts with the Heparan Sulfate (HS) chains expressed by proteoglycans[14]. There are two families of proteoglycans, Syndecans (from 1 to 4) and Glypicans (from 1 to 6); in **Figure 1**, we show the structure of the most common Syndecan-1 and -4 and Glypican-1. The affinity between the virus spike and the HS chains is dependent on the HS sequence[11]. Once the virus reaches the cell membrane, the specific binding of the receptor-binding-domain (RBD) of the spike glycoprotein (S) to the cellular-entry receptors (R) takes place (**Figure 1**). Different entry receptors have been identified for SARS-CoV-2, including angiotensin-converting enzyme 2 or ACE2 (HCoV-NL63, SARS-CoV and SARS-CoV-2) [15], TMPRSS2, [15], SRB1 [16], neuropilin-1 [17]. The expression and tissue distribution of entry receptors influence viral tropism and pathogenicity, but the cell glycocalyx composition also plays a key role. In this paper, we present an analytical and predictive model for viral host tropism, based on the components of the cell-virus interaction during viral attachment.

Figure 1: **SARS-Cov 2 entry into the host cell.** (a) Representation of the SARS-CoV-2 virion landing on the membrane of a host cell. SARS-CoV-2 spike proteins are shown as a surface with all-atom glycosylation. Protomers are coloured according to the RDB conformation: up (purple, PDBid 6VXX) or down (cyan, PDBid 6VSB). On the host membrane, all-atom structures of different components are represented as surface: Syndecan 1 and 4, Glypican, TMPRSS2 and the ACE2-BT0 complex (PDBid 6M1D). (b) Insertion into the glycocalyx: the insertion of one spike protein (in the closed state) into the glycocalyx. (c) Host receptor (ACE2) recognition and binding: the ACE2-BT0 complex structure bound to one spike protein in the open conformation. The binding region between one ACE2 monomer and the RBD domain of one protomer of the spike protein is highlighted as an inset. (d) Glycan conformational volume: an all-atom representation of a high-mannose glycan and its conformational volume.

A simple model for Viral Tropism

The specificity of infection for certain viruses to hosts and host tissues, viral-host tropism, can be described in terms of the number of viruses bound to a given combination of receptors, or phenotype. The binding to the cell surface can be estimated using an Langmuir isotherm. We define two surface coverages, the glycocalyx surface coverage, θ_G , which represents the fraction of the glycocalyx occupied by virions, and the receptor surface coverage, θ_R , which represents the fraction of the cell-surface occupied by viruses,[18, 6, 19] equal to

$$\theta_G = \frac{\rho v_B Q_G}{1 + \rho v_B Q_G} \quad \text{and} \quad \theta_R = \frac{\rho_G v_B Q_R}{1 + \rho_G v_B Q_R}, \quad (1)$$

where ρ is the bulk viral titre (the number of virus copies per unit volume or viral load), ρ_G is the viral titre in the glycocalyx, v_B is the binding volume, and Q_R and Q_G are the grand canonical partition functions for the virus in the receptors and the glycocalyx surface, respectively. The surface coverage, θ , effectively normalises the virus binding: $\theta = 0$ indicates no interaction and $\theta = 1$ indicates full saturation of the available binding sites on the surface considered. Such an approach allows us to assess the nonlinear nature of viral binding and how it changes from zero to saturation for very small changes of the cell's receptor/HS composition and the spike affinities towards them, viral size, spike number and initial viral load. The grand canonical partition functions, Q_G and Q_R , can be broken down into the product of the component partition functions arising from all the possible binding and steric potentials acting on the virus. We can divide this into HS binding ($q_{HS}^{binding}$), HS insertion ($q_{HS}^{insertion}$), and receptor binding ($q_R^{binding}$) components respectively, as

$$Q_G = q_{HS}^{binding} q_{HS}^{insertion} \quad \text{and} \quad Q_R = q_R^{binding} q_{HS}^{insertion} \quad (2)$$

where

$$q_R^{binding} = \sum_{j=1}^{N_R} \Omega_R(j) e^{-j\beta(\epsilon_R + u_R)}; \quad q_{HS}^{binding} = \sum_{j=1}^{N_{HS}} \Omega_{HS}(j) e^{-j\beta\epsilon_{HS}} \quad \text{and} \quad q_{HS}^{insertion} = e^{-\beta u_{HS}} \quad (3)$$

Here, N_R is the number of linkers, $\Omega_R(j)$ the number of possible combinations between the R linkers and the virus proteins, ϵ_R the binding energy of the single bond between the virus protein and the R linker, u_R the R steric potential arising from compression of the receptor glycans ($u_R=0$ for TMPRSS2). N_{HS} is number of HS chains per binding μm^2 , $\Omega_{HS}(j)$ the number of possible combinations between the HS chains and the virus proteins, ϵ_{HS} the binding energy of the single bond between the virus protein and the HS chain, u_{HS} , the steric potential arising from insertion of the virus into the glycocalyx, β is the inverse thermodynamic temperature. Equations (1-3) allow us to estimate the virus tropism as a function of the individual interactions, each of which can be derived separately and analysed down to the molecular level. Further derivation is provided in the Supporting Information.

Figure 2: The insertion step into the glycocalyx. (a) All-atom representation of the glypican 1 proteoglycan together with an inset of an HS highly sulphated, IdoA(2S)-GlcNS(6S), monomer. (b) Heparan Sulfate (HS)-spike protein binding: binding interaction (kT) behaviour with increasing number of monomers per HS chain and HS chains density per μm^2 . (c) Heparan Sulfate (HS)-spike protein insertive repulsion: steric repulsion evolution with increasing number of monomers per HS chain and HS chains density per μm^2 .

As shown in **Figure 1**, during the insertion into the glycocalyx, coronaviruses bind to Heparan Sulfate. Each Heparan Sulfate chain, consists of a variably sulfated repeating disaccharide unit (**Figure 2a**). Only a fraction of these units bind to the virion spike proteins, normally 8 to 10 monomers or binding motifs [20]. Therefore, the total binding energy between a single HS-chain and the virus, ϵ_{HS} , (Supporting Information) is dependent on the number of binding motifs, N_L . The more binding motifs the higher the attractive energy (**Figure 2b**). This energy also increases with the number of HS chains, proportional to the density of proteoglycans

(Figure 2b). The repulsive energy generated by the insertion of the virus into the HS, q_{HS}^{steric} , (Supporting Information) increases dramatically with the number of monomers per chain and the number of chains (proteoglycan density) (Figure 2c).

Once a virus has inserted into the glycocalyx, it can bind to the ACE2 receptor via the RBD of the spike glycoprotein (Figure 1). Upon binding, glycans at the binding site are compressed, inducing a repulsive component to this interaction (Figure 1 (lower right panel), Figure 3a,b, Supporting Information). The composition of these glycans can vary widely between species, with human glycans being particularly large and complex[21]. Both host-cell proteins and viral proteins are glycosylated, and this post-translational glycosylation process is known to play a major role in infection virulence and interaction with the host immune system [22]. In SARS-CoV-2, the spike protein typically contains 22 N-glycans [23, 24] and 2 O-glycans [23]. The spike protein glycans shield the virus from the immune system[25], but also seems to regulate the opening and closing of the RBD domain[26].

Figure 3: The effect of glycosylation in the recognition and binding of the spike protein to the host cell entry receptor. . (a, b) ACE2-RBD complex dissociation: multiple snapshots from an out-of-equilibrium simulation of the dissociation of the ACE2-RBD complex are superimposed for the glycans and RBD domain. The RBD-domain is coloured according to the distance to the binding site (red bound, pink (a) or blue (b) unbound). Glycans are represented as lines and coloured radially in shades of pink. In both panels an schematic representation of the glycans is depicted as well as an all-atom representation of the glycan trapped in the binding site. (c,d) ACE2-RBD KD variation with glycan morphology: ACE2-RBD dissociation constant plotted as a function of the number of branches and coloured according to number of monomers per branch in c and vice versa in d. (e,f) ACE2 Glycans steric potential variation with glycan morphology: the steric potential generated by glycan compression in kT units is plotted as a function of the number of branches and coloured according to number of monomers per branch in e and vice versa in f.

In Figure 3 we also analyse the influence of the trapped glycans on the dissociation constant (K_D) of the ACE2-spike protein complex. Both the number of branches and branch length (number of monomers) decreases the K_D of the complex. Our analysis explains why the same spike proteins interacting with the same receptors are reported by several studies in human

cells with different dissociation constants; proteins expressed in HEK293T-gave a $K_D = 44.8$ nM [27], FreeStyle293F $K_D = 34.6$ nM [28], while proteins expressed in insect Sf9 cells, reported $K_D = 4$ nM [29]. Recently, Allen et al. [30] attempted to understand the influence of ACE2 glycosylation on the binding of the RBD domain of the SARS-CoV-2 spike protein. The authors chemically modified various glycan positions involved in the binding and measured the binding affinity using surface plasmon resonance (SPR). The predominant type of glycan observed across all five ACE2 glycosylated sites were fucosylated with two or three branches complex-type glycans which lack galactose and sialic acid. They found that ACE2 treated with sialidase has a higher affinity for SARS-CoV-2 compared to WT ACE2, K_D 30% lower. The presence or absence of fucose or the presence of a single N-acetylglucosamine residue did not impact ACE2 binding to a significant extent. But when all N-linked glycans on ACE2 were converted to oligomannose-type glycans (predominantly Man9GlcNAc2) the K_D increases by 50%, corresponding to a decrease in binding. These differences, though subtle, are in perfect agreement with our model. In **Figure 3c,d** we observe that the increasing glycan complexity (branching and number of monomers per branch) increases the ACE2-RBD K_D , due to the steric potential generated by the glycan compression in the binding site. The difference between an insect glycan (**Figure 3c**) and a more complex glycan (human-like) with 3 to 4 branches is nearly 20 nM (~ 17.7 kT, for one glycan trapped in the binding region). According to our model, differences in K_D can be as high as 50 nM (~ 16.8 kT) between a single-branched short glycan and multiple-branched long glycan, like the ones found in the human lungs (**Figure 3b**) [31]. Single chemical modification does not seem to have a major role in the complex binding affinity, as glycans have a steric role rather than a binding role. Therefore, the steric repulsion aids the dissociation of the complex - the shorter the glycan the higher the binding energy and the lower the K_D . Additionally, we also performed out-of-equilibrium simulations of the RBD-ACE2 complex dissociation, using three different types of glycans: insect-like, high-mannose and complex. Our simulations reveal how the force needed to break the complexes decreases when the glycan complexity increases (**Figure S1**).

Viral Tropism in a nutshell

The attractive/repulsive interactions described above defines the viral tropism, via the grand canonical partition function, $Q_{HS/R}$. In our model, different cell phenotypes are characterised by the expression of proteoglycans (proteoglycan density, μm^2) and the entry receptors (receptor density, μm^2). Likewise, a virus can be described by the binding affinity of its spike protein to HS and the entry receptors. In **Figure 4a**, we see how the HS density and the virus spike-HS K_D affects θ for cells with a specific proteoglycan density, where above a certain proteoglycan density the virus-cell interaction is switched off - this is known as range selectivity [18]. This change from bound to unbound occurs within a very narrow range of HS densities. We can use the experimental data collected for SARS-CoV-2, SARS-CoV and MERS-CoV binding with Heparin [10], to show that SARS-CoV-2 has a higher affinity toward HS: this allows SARS-CoV-2 to bind to higher HS densities, compared to the other coronavirus (**Figure 4a**). By analysing the proteoglycan expression on the different tissues and cells found on the human upper and lower airways, **Figure 4b**, we found that the upper airways have a proteoglycan density which favours the attachment of SARS-CoV-2, but not SARS-CoV or MERS-CoV.

The ideal location for attachment of MERS-CoV and SARS-CoV is in regions with lower proteoglycans densities, such as the lower airways, while SARS-CoV-2 can attach and replicate in the epithelium of the upper airways. This mean SARS-CoV-2 can infect the upper respiratory tract and can navigate the airways more easily, as reported by virological assessment of SARS-CoV-2, SARS-CoV and MERS-CoV patients [32, 33]. Moreover, we can observe in **Figure 4c** that the affinity of the virus for HS determines the initial viral titre require for successful attachment. At low HS densities, **Figure 4d**, very low initial titres result in full surface coverage for SARS-CoV-2 (**Figure 4d**, left) and MERS-CoV (**Figure 4d**, middle), while SARS-CoV needs at least 103 copies/ml in bulk to have 50% surface coverage. At medium HS densities, **Figure 4e**, SARS-CoV needs at least 109 copies/ml to achieve 50% surface coverage (**Figure 4e**, left) and MERS-CoV requires 105 copies/ml (**Figure 4e**, middle). At high HS densities, **Figure 4f**, MERS-CoV and SARS-CoV, cannot bind under any condition, while SARS-CoV-2 binds for initial viral titres above

Figure 4: **Viral tropism: cell proteoglycans density and initial viral titre effect.** (a) Cell Surface coverage by viral particles variation with HS density and HS-spike binding affinity: Contour plot of the cell surface coverage by viral particle as a function of the HS density and the spike protein-HS binding affinity. Surface coverage equal to 1 implies that the initial viral titre successfully attaches to the cell surface. SARS-CoV-2, SARS-CoV and MERS-CoV spike proteins HS binding affinities are displayed. (b) Proteoglycans gene expression across cell types for the upper and lower human airways: Each gene is normalised to its maximum value across all tissues. The tissues correspond to the indicated positions along the airway from nasal to distal lung. Dataset obtained from Durant et al. Initial viral titre threshold for binding. (c) Cell surface coverage by viral particles variation with HS density and initial viral titre: Contour plot of the cell surface coverage by viral particle as a function of the HS density and the initial viral titre for SARS-CoV-2 (left), MERS-CoV (middle) and SARS-CoV (right). Cell surface coverage by viral particles variation with the initial viral titre for (d) ~ 440 HS chains per μm^2 ; for SARS-CoV-2 (left), MERS-CoV (middle) and SARS-CoV (right).

152 10^3 copies/ml. This is in agreement with respiratory samples extracted from SARS-CoV-2 pa-

153 tients with mild and severe COVID-19 symptoms revealed viral titres in between 10^3 and 10^9
154 copies/ml [12].

155 Several receptors have been found responsible for SARS-CoV-2 entry and while our anal-
156 ysis can be expanded to all of them, we focus on the most widely discussed elements in the
157 SARS-CoV-2 infection, the receptors ACE2 and TMPRSS2[34]. In **Figure 5a** we observe how
158 successful viral attachment occurs when the receptors act in a synergistic manner. Cells for
159 which the number of ACE2 receptors account for less 40 % of the total receptor density per
160 μm^2 , like the olfactory sensory neurons (OSN) [35, 36], have low cell surface coverage for any
161 percentage of TMPRSS2, therefore, low viral titre on the surface. While for cell expressing high
162 densities of ACE2, >60%, even moderate values of TMPRSS2 are enough to obtain high cell sur-
163 face coverage values, as is the case for many cells from the human airways [37, 36], including
164 olfactory epithelial sustentacular cells. This suggests that SARS-CoV-2 infection of non-neuronal
165 cell types leads to olfactory dysfunction (anosmia) in patients with COVID-19.

Figure 5: **Viral Tropism: Entry receptors binding synergy.** (a) Contour plots of the cell surface coverage by viral particle as a function of the ACE2 and TMPRSS2 cell densities and (b) affinity towards the viral spike protein. (c) Viral morphology: radius and spike-proteins available for binding influence on viral the number of viral particles attached to the cell surface: Contour plot of the virus titre on the cell surface (Log) as a function of the virion radius and the number of spikes proteins in open conformation available for binding. (d) Viral morphology: radius and number of protein spikes. Probability distributions of the different SARS-CoV-2 viral radius and number of protein spikes. Data was replotted from Extended Data Fig. 1 in Ke, Z. et al. different sample conditions are reported on the original article [38].

166 Mutations of the RBD domain of the SARS-CoV-2 spike protein are worrying due to the
167 higher transmissibility of these variants, especially the Brazilian (P1., K417T, E484K, N501Y),
168 South African (B.1.351, K417N, E484K, N501Y) and British variants (B.1.1.7, N501Y) [39, 40,
169 41, 42, 43, 44]. This is related to the ability of these mutations to hide the virus from antibod-
170 ies [39, 42, 44] and cause reinfection [41]. The N501Y mutation shared by all three variants

also increases the binding affinity of the RBD-domain towards the ACE2, decreasing its K_D [40]. However, for an optimum ratio of receptors (ACE2, TMPRSS2) densities, our model does not predict major changes in cell surface coverage with decreasing ACE2-RBD-domain K_D (**Figure 5b**). For high ACE2 K_D values (low binding affinity) on the other hand, changes in the affinity towards TMPRSS2 might lead to a decrease in viral titre on the cell surface according to our model (**Figure 5b**).

Finally, we found that both the virion morphology and the number of available RBD-domains in open conformation available for binding to the ACE2 affects virus titre on the cell surface. Smaller virions ~ 40 nm radius are able to attach to the host cell with fewer spike proteins available for binding to the entry receptors, while a bigger virus with same number of spikes available for binding would be less successful in this case and the virus titre on the cell surface will be lower (**Figure 5c**). Cryogenic electron microscopy on SARS-CoV-2 virions populations have revealed an average radius of ~ 45 nm with a standard deviation of ~ 10 nm, with individual virions containing 24 ± 9 spike proteins not uniformly distributed and highly flexible [45, 38]. There is a very nonlinear response with the morphology of the virus, as shown in Figure 5c, and a virus with a radius larger than 45 nm is less effective at attaching to cell. According to our model, for larger virions to be as effective as viruses with a radius of 45 nm, they require more spike proteins on their surface. By fitting the measurements provided by Ke et al. [38] (**Figure 5d**) we observe that for all three preparations there is a clear distribution of size, but the number of spikes per virion distributions are very broad, indicating that the size of the virus is the critical parameter for viral attachment.

Conclusions

Overall, our model can explain the viral tropism of SARS-CoV-2 in terms of interactions with the various host-cell components. We find that viruses are extremely selective, not only in terms of entry receptors distribution on the host cell but also on the spike protein affinity towards the key elements of the cell phenotype. Additionally, we find that Heparan Sulfate plays a major role in viral infection, as the repulsion potential can dramatically decrease the number of viral particles that can interact with the entry receptors - this increases the viral load threshold required for infection. We have also shown that the role of the glycans on the entry receptors is not enthalpic but entropic, decreasing the binding affinity of the RBD-ACE2 complex at increasing morphological complexity. Finally, we have shown how the size of the viral particle is a critical parameter with a nonlinear response, and therefore, induces a selection of the optimal viral size for infection.

Acknowledgements

Author contributions

J.B and G.B perform the mathematical derivation of the model. J.B. coded the model. S.A.G performed the all-atom simulations and the generation, analysis and representation of the synthetic data presented in the manuscript. All three authors wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Methods

Out-of-equilibrium molecular dynamics . The ACE2-RDB SARS-CoV2 complex was extracted from the cryo-em pdb structure 6M17 [46]. The protein complex presented no mutations or structural gaps. Two different systems (different glycan types) were generated by adding glycans using the Glycam web server Glycoprotein Builder tool [47], the insect-like system (DManpa1-3[DManpa1-6]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH) and high mannose (DManpa1-6[DManpa1-3]DManpb1-4DGlcpNAcb1-4DGlcpNAcb1-OH). The two systems were

218 protonated at pH 7.4 and solvated (150mM NaCl) using a TIP3P[48] water model and Charmm36m[49]
 219 protein, glycans and ions force field. The final simulated systems consisted of 240K atoms, re-
 220 sulting in a cubic box simulation cell of 12nm x 12nm x 17nm. All simulations were performed
 221 with Gromacs v2019.2 [50] with a 2fs integration step. We performed 2000 steps of minimiza-
 222 tion (steepest descent algorithm), followed by 5000 steps of isotropic NPT at 300K, using the
 223 velocity-rescaling thermostat[51] and Parrinello-Rahman barostat[52]. After 50000 steps of NVT
 224 the system was fully equilibrated and ready for production, which was performed in the NVT
 225 ensemble for 100 ns. A hundred steered molecular dynamics simulations were performed for
 226 each system using PLUMED2.6[53, 54].

227 **Synthetic data generation** . In Table 1 we specify the model parameters fixed to generate
 228 the data for each of the figures displayed in this article.

Viral radius, nm	Figure 2	Figure 3	Figure 4a	Figure 4c	Figure 4d	Figure 4e	Figure 5a	Figure 5b	Figure 5c
Viral titre, copies/ml	50	50	50	50	50	50	50	50	50
Number of Spikes	3.01E+07	6.02E+07	3.01E+07	3.01E+07	3.01E+07	3.01E+07	6.02E+07	6.02E+07	6.02E+07
Spike length, m	15	15	15	15	15	15	15	15	15
HS-Spike binding energy, kT	1.45E-08	1.45E-08	1.45E-08	1.45E-08	1.45E-08	1.45E-08	1.45E-08	1.45E-08	1.45E-08
HS fraction of binding motif	-16			-25,-18,-16	-25,-18,-16	-25,-18,-16	-23	-23	-23
HS chain length, m	0.1		0.1	0.1	0.1	0.1	0.1	0.1	0.1
HS number of chains, per			1.3E-07	1.3E-07	1.3E-07	1.3E-07			1.5E-07
HS Kuhn Length, m	9E-09		9E-09	4.40E-07	5.0E-07	6.0E-07			100
ACE2 mean binding distance		2.7E-09		9E-09	9E-09	9E-09	2.7E-09	2.7E-09	9E-09
ACE2 glycan branch length							1.25E-09	1.25E-09	1E-09
ACE2 glycan number of branches		2					2	2	3
ACE2 number of trapped glycans							2	2	2
ACE2-Spike KD		2.06E-09					2.06E-09		2.06E-09
Number of ACE2 per		56					3	56	45
Number of ACE2 binding sites		3						3	3
Number of TMPRSS2 per		100						100	100
TMPRSS2-Spike KD		2.06E-09					2.06E-09		1.52E-08
Number of TMPRSS2 binding sites		1					1	1	1
Cell size, m2	1E-14	1E-14	1E-14				1E-14	1E-14	1E-14

Table 1. Model fixed parameters used to generate the synthetic data.

References

- [1] P. V'Kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, "Coronavirus biology and replication: implications for sars-cov-2," *Nat Rev Microbiol*, vol. 19, no. 3, pp. 155-170, 2021.
- [2] C. Salata, A. Calistri, C. Parolin, and G. Palu, "Coronaviruses: a paradigm of new emerging zoonotic diseases," *Pathog Dis*, vol. 77, no. 9, 2019.
- [3] D. S. Dimitrov, "Cell biology of virus entry," *Cell*, vol. 101, 2000.
- [4] E. D. Eric Baranowski, Carmen M. Ruiz-Jarabo, "Evolution of cell recognition by viruses," *SCIENCE*, vol. 292, 2001.
- [5] D. S. Dimitrov, "Virus entry: molecular mechanisms and biomedical applications," *Nat Rev Microbiol*, vol. 2, no. 2, pp. 109-22, 2004.
- [6] F. J. Martinez-Veracoechea and D. Frenkel, "Designing super selectivity in multivalent nanoparticle binding," *Proc Natl Acad Sci U S A*, vol. 108, no. 27, pp. 10963-8, 2011.
- [7] V. Cagno, E. D. Tseligka, S. T. Jones, and C. Tapparel, "Heparan sulfate proteoglycans and viral attachment: True receptors or adaptation bias?," *Viruses*, vol. 11, no. 7, 2019.
- [8] M. Koehler, M. Delguste, C. Sieben, L. Gillet, and D. Alsteens, "Initial step of virus entry: Virion binding to cell-surface glycans," *Annu Rev Virol*, vol. 7, no. 1, pp. 143-165, 2020.
- [9] T. M. Clausen, D. R. Sandoval, C. B. Spliid, J. Pihl, H. R. Perrett, C. D. Painter, A. Narayanan, S. A. Majowicz, E. M. Kwong, R. N. McVicar, B. E. Thacker, C. A. Glass, Z. Yang, J. L. Torres, G. J. Golden, P. L. Bartels, R. N. Porell, A. F. Garretson, L. Laubach, J. Feldman, X. Yin, Y. Pu, B. M. Hauser, T. M. Caradonna, B. P. Kellman, C. Martino, P. Gordts, S. K. Chanda, A. G. Schmidt, K. Godula, S. L. Leibel, J. Jose, K. D. Corbett, A. B. Ward, A. F. Carlin, and J. D. Esko, "Sars-cov-2 infection depends on cellular heparan sulfate and ace2," *Cell*, vol. 183, no. 4, pp. 1043-1057 e15, 2020.
- [10] S. Y. Kim, W. Jin, A. Sood, D. W. Montgomery, O. C. Grant, M. M. Fuster, L. Fu, J. S. Dordick, R. J. Woods, F. Zhang, and R. J. Linhardt, "Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (sars-cov-2) spike glycoprotein binding interactions," *Antiviral Res*, vol. 181, p. 104873, 2020.
- [11] L. Liu, P. Chopra, X. Li, M. A. Wolfert, S. M. Tompkins, and G. J. Boons, "Sars-cov-2 spike protein binds heparan sulfate in a length- and sequence-dependent manner," *bioRxiv*, 2020.
- [12] S. Zheng, J. Fan, F. Yu, B. Feng, B. Lou, Q. Zou, G. Xie, S. Lin, R. Wang, X. Yang, W. Chen, Q. Wang, D. Zhang, Y. Liu, R. Gong, Z. Ma, S. Lu, Y. Xiao, Y. Gu, J. Zhang, H. Yao, K. Xu, X. Lu, G. Wei, J. Zhou, Q. Fang, H. Cai, Y. Qiu, J. Sheng, Y. Chen, and T. Liang, "Viral load dynamics and disease severity in patients infected with sars-cov-2 in zhejiang province, china, january-march 2020: retrospective cohort study," *BMJ*, vol. 369, p. m1443, 2020.
- [13] National Health Library. "https://clinicaltrials.gov/ct2/results?cond=Covid19&term=heparin&cntry=&state=&city=&dist=". Accessed:May 2021.
- [14] A. Milewska, P. Nowak, K. Owczarek, A. Szczepanski, M. Zarebski, A. Hoang, K. Berniak, J. Wojarski, S. Zeglen, Z. Baster, Z. Rajfur, and K. Pyrc, "Entry of human coronavirus nl63 into the cell," *Journal of Virology*, vol. 92, no. 3, 2018.
- [15] M. 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, and S. Pohlmann, "Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor," *Cell*, vol. 181, no. 2, pp. 271-280 e8, 2020.

- 272 [16] C. Wei, L. Wan, Q. Yan, X. Wang, J. Zhang, X. Yang, Y. Zhang, C. Fan, D. Li, Y. Deng, J. Sun,
273 J. Gong, X. Yang, Y. Wang, X. Wang, J. Li, H. Yang, H. Li, Z. Zhang, R. Wang, P. Du, Y. Zong,
274 F. Yin, W. Zhang, N. Wang, Y. Peng, H. Lin, J. Feng, C. Qin, W. Chen, Q. Gao, R. Zhang,
275 Y. Cao, and H. Zhong, "Hdl-scavenger receptor b type 1 facilitates sars-cov-2 entry," *Nat*
276 *Metab*, vol. 2, no. 12, pp. 1391-1400, 2020.
- 277 [17] L. Cantuti-Castelvetri, R. Ojha, L. D. Pedro, M. Djannatian, J. Franz, S. Kuivanen, F. van der
278 Meer, K. Kallio, T. Berk Kaya, M. Anastasina, T. Smura, L. Levanov, L. Szivovicsa, A. Tobi,
279 H. Kallio-Kokko, P. Österlund, M. Joensuu, F. A. Meunier, S. J. Butcher, M. S. Winkler,
280 B. Mollenhauer, A. Helenius, O. Gokce, T. Teesalu, J. Hepojoki, O. Vapalahti, C. Stadel-
281 mann, G. Balistreri, and M. Simons, "Neuropilin-1 facilitates sars-cov-2 cell entry and infec-
282 tivity," *Science*, vol. 310, pp. 856-860, 2020.
- 283 [18] M. Liu, A. Apriceno, M. Sipin, E. Scarpa, L. Rodriguez-Arco, A. Poma, G. Marchello,
284 G. Battaglia, and S. Angioletti-Uberti, "Combinatorial entropy behaviour leads to range
285 selective binding in ligand-receptor interactions," *Nat Commun*, vol. 11, no. 1, p. 4836,
286 2020.
- 287 [19] G. B. Xiaohe Tian, Stefano Angioletti-Uberti, "On the design of precision nanomedicines,"
288 *Science Advances*, vol. 6, 2020.
- 289 [20] Z. Shriver, I. Capila, G. Venkataraman, and R. Sasisekharan, "Heparin and heparan sulfate:
290 analyzing structure and microheterogeneity," *Handb Exp Pharmacol*, no. 207, pp. 159-76,
291 2012.
- 292 [21] G. Lauc, M. Pezer, I. Rudan, and H. Campbell, "Mechanisms of disease: The human n-
293 glycome," *Biochim Biophys Acta*, vol. 1860, no. 8, pp. 1574-82, 2016.
- 294 [22] D. J. Vigerust and V. L. Shepherd, "Virus glycosylation: role in virulence and immune inter-
295 actions," *Trends Microbiol*, vol. 15, no. 5, pp. 211-8, 2007.
- 296 [23] A. Shajahan, N. T. Supekar, A. S. Gleinich, and P. Azadi, "Deducing the n- and o-
297 glycosylation profile of the spike protein of novel coronavirus sars-cov-2," *Glycobiology*,
298 vol. 30, no. 12, pp. 981-988, 2020.
- 299 [24] D. W. J. S. M. M. C. Yasunori Watanabe, Joel D. Allen, "Site-specific glycan analysis of the
300 sars-cov-2 spike," *Science*, vol. 369, 2020.
- 301 [25] O. C. Grant, D. Montgomery, K. Ito, and R. J. Woods, "Analysis of the sars-cov-2 spike pro-
302 tein glycan shield reveals implications for immune recognition," *Sci Rep*, vol. 10, no. 1,
303 p. 14991, 2020.
- 304 [26] L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fog-
305 arty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, and R. E. Amaro, "Beyond shielding:
306 The roles of glycans in the sars-cov-2 spike protein," *ACS Cent Sci*, vol. 6, no. 10, pp. 1722-
307 1734, 2020.
- 308 [27] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, and F. Li, "Structural
309 basis of receptor recognition by sars-cov-2," *Nature*, vol. 581, no. 7807, pp. 221-224, 2020.
- 310 [28] K. S. C. J. A. G. C.-L. H. O. A. B. S. G. J. S. M. Daniel Wrapp, Nianshuang Wang, "Cryo-em
311 structure of the 2019-ncov spike in the prefusion conformation," *Science*, vol. 367, p. 1260-
312 1263, 2020.
- 313 [29] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and
314 X. Wang, "Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2
315 receptor," *Nature*, vol. 581, no. 7807, pp. 215-220, 2020.
- 316 [30] J. D. Allen, Y. Watanabe, H. Chawla, M. L. Newby, and M. Crispin, "Subtle influence of ace2
317 glycan processing on sars-cov-2 recognition," *J Mol Biol*, vol. 433, no. 4, p. 166762, 2021.

- [31] N. Jia, L. Byrd-Leotis, Y. Matsumoto, C. Gao, A. N. Wein, J. L. Lobby, J. E. Kohlmeier, D. A. Steinhauer, and R. D. Cummings, "The human lung glycome reveals novel glycan ligands for influenza a virus," *Sci Rep*, vol. 10, no. 1, p. 5320, 2020.
- [32] K. P. Y. Hui, M.-C. Cheung, R. A. P. M. Perera, K.-C. Ng, C. H. T. Bui, J. C. W. Ho, M. M. T. Ng, D. I. T. Kuok, K. C. Shih, S.-W. Tsao, L. L. M. Poon, M. Peiris, J. M. Nicholls, and M. C. W. Chan, "Tropism, replication competence, and innate immune responses of the coronavirus sars-cov-2 in human respiratory tract and conjunctiva: an analysis in ex-vivo and in-vitro cultures," *The Lancet Respiratory Medicine*, vol. 8, no. 7, pp. 687-695, 2020.
- [33] R. Wolfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Muller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brunink, J. Schneider, R. Ehmann, K. Zwirgmaier, C. Drosten, and C. Wendtner, "Virological assessment of hospitalized patients with covid-2019," *Nature*, vol. 581, no. 7809, pp. 465-469, 2020.
- [34] M. Cevik, K. Kuppalli, J. Kindrachuk, and M. Peiris, "Virology, transmission, and pathogenesis of sars-cov-2," *BMJ*, vol. 371, p. m3862, 2020.
- [35] C. W. M. L. David H. Brann, Tatsuya Tsukahara, R. C. I. C. M.-H.-J. C. Koen Van den Berge, Boying Gong, K. S. H. R. d. B.-Y. G. C. Russell B. Fletcher, Diya Das, E. P. J. M. R. A. H. Davide Risso, Sandrine Dudoit, B. J. G. M. S. G. Hiroaki Matsunami, Darren W. Logan, and S. R. D. John Ngai, "Non-neuronal expression of sars-cov-2 entry genes in the olfactory system suggests mechanisms underlying covid-19-associated anosmia," *Science Advances*, vol. 6, 2020.
- [36] M. A. Durante, S. Kurtenbach, Z. B. Sargi, J. W. Harbour, R. Choi, S. Kurtenbach, G. M. Goss, H. Matsunami, and B. J. Goldstein, "Single-cell analysis of olfactory neurogenesis and differentiation in adult humans," *Nat Neurosci*, vol. 23, no. 3, pp. 323-326, 2020.
- [37] M. Deprez, L. E. Zaragosi, M. Truchi, C. Becavin, S. Ruiz Garcia, M. J. Arguel, M. Plaisant, V. Magnone, K. Lebrigand, S. Abelanet, F. Brau, A. Paquet, D. Pe'er, C. H. Marquette, S. Leroy, and P. Barbry, "A single-cell atlas of the human healthy airways," *Am J Respir Crit Care Med*, vol. 202, no. 12, pp. 1636-1645, 2020.
- [38] Z. Ke, J. Oton, K. Qu, M. Cortese, V. Zila, L. McKeane, T. Nakane, J. Zivanov, C. J. Neufeldt, B. Cerikan, J. M. Lu, J. Peukes, X. Xiong, H. G. Krausslich, S. H. W. Scheres, R. Bartsch-Schlager, and J. A. G. Briggs, "Structures and distributions of sars-cov-2 spike proteins on intact virions," *Nature*, vol. 588, no. 7838, pp. 498-502, 2020.
- [39] A. J. Greaney, A. N. Loes, K. H. D. Crawford, T. N. Starr, K. D. Malone, H. Y. Chu, and J. D. Bloom, "Comprehensive mapping of mutations in the sars-cov-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies," *Cell Host Microbe*, vol. 29, no. 3, pp. 463-476 e6, 2021.
- [40] D. N. e. a. Horby P, Huntley C, "Nervtag note on b.1.1.7 severitypdf icon," *SAGE meeting report. January 21*, 2021.
- [41] 2021. <https://virological.org/>.
- [42] Y. Weisblum, F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J. C. Lorenzi, F. Muecksch, M. Rutkowska, H. H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C. D. Hillyer, M. Caskey, D. F. Robbiani, C. M. Rice, M. C. Nussenzweig, T. Hatziioannou, and P. D. Bieniasz, "Escape from neutralizing antibodies by sars-cov-2 spike protein variants," *Elife*, vol. 9, 2020.
- [43] K. Wu, G. Peng, M. Wilken, R. J. Geraghty, and F. Li, "Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus," *J Biol Chem*, vol. 287, no. 12, pp. 8904-11, 2012.
- [44] X. Xie, J. Zou, C. R. Fontes-Garfias, H. Xia, K. A. Swanson, M. Cutler, D. Cooper, V. D. Menachery, S. Weaver, P. R. Dormitzer, and P. Y. Shi, "Neutralization of n501y mutant sars-cov-2 by bnt162b2 vaccine-elicited sera," *bioRxiv*, 2021.

- 367 [45] B. Turonová, M. Sikora³, C. Schürmann, W. J. H. Hagen, S. Welsch, , F. E. C. Blanc, S. von
368 Bülow, M. Gecht, K. Bagol, C. Hörner, , G. van Zandbergen, J. Landry, N. T. D. de Azevedo,
369 , S. Mosalaganti, A. Schwarz, R. Covino, M. D. Mühlebach, , G. Hummer, J. K. Locker, and
370 M. Beck, "In situ structural analysis of sars-cov-2 spike reveals flexibility mediated by three
371 hinges," *Science*, vol. 370, pp. 203-208, 2020.
- 372 [46] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, "Structural basis for the recognition of
373 sars-cov-2 by full-length human ace2," *Science*, vol. 367, pp. 1444-1448, 2020.
- 374 [47] Glycoprotein Builder. Woods Group. (2005-2021) GLYCAM Web. Complex Carbohydrate
375 Research Center, University of Georgia, Athens, GA. (<http://glycam.org>).
- 376 [48] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison
377 of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*,
378 vol. 79, no. 2, pp. 926-935, 1983.
- 379 [49] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and
380 J. MacKerell, A. D., "Charmm36m: an improved force field for folded and intrinsically dis-
381 ordered proteins," *Nat Methods*, vol. 14, no. 1, pp. 71-73, 2017.
- 382 [50] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs:
383 High performance molecular simulations through multi-level parallelism from laptops to
384 supercomputers," *SoftwareX*, vol. 1-2, pp. 19-25, 2015.
- 385 [51] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling,"
386 *J Chem Phys*, vol. 126, no. 1, p. 014101, 2007.
- 387 [52] M. P. Rahman and A., "Polimorphic transitions in single crystals: A new molecular dynamics
388 method," *Journal of Applied Physics*, vol. 52, p. 7182, 1981.
- 389 [53] T. P. consortium, "Promoting transparency and reproducibility in enhanced molecular sim-
390 ulations," *Nat Methods*, vol. 16, no. 8, pp. 667-669, 2019.
- 391 [54] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "Plumed 2: New
392 feathers for an old bird," *Computer Physics Communications*, vol. 185, no. 2, pp. 604-613,
393 2014.

Supporting information

1 Theoretical Model

As discussed in the main text, binding between coronaviruses and the host cells proceeds in a two step process. First, the virus binds to Heperan Sulfate, the long-chain sugars that coat the cell surface. After binding to the Heperan Sulfate, the virus then can form bonds with the receptors on the surface, particularly ACE2 and TMPRSS2.

We will define the glycocalyx surface coverage, θ_G , as the probability that a virus will be bound to a HS chain at a given site, and the cell-surface coverage, θ_R , as the probability that a virus will be bound to a receptor at a given site. In equation 1 in the main text, we give the general form of the surface coverage, derived from the Langmiur isotherm, which we will specialise to deal with each case, giving

$$\theta_G = \frac{\rho v_B Q_G}{1 + \rho v_B Q_G} \quad \text{and} \quad \theta_R = \frac{\rho_G v_B Q_R}{1 + \rho_G v_B Q_R}, \quad (4)$$

where the binding volume, v_B , is determined from the viral radius, R , and the spike length, d , as $v_B = \pi((R+d)^3 - \frac{\pi}{3}R^3)$ and ρ is the viral bulk concentration, and ρ_G is the viral concentration in the glycocalyx. The cell-surface coverage can be linked to the glycocalyx surface coverage as $\rho_G = \frac{\theta_G}{v_B}$, however, in order to limit the complexity for our calculation, we treat ρ_G as a parameter, effectively decoupling the system.

The partition functions, Q_G and Q_R , contains information about the possible states of the system when the virus is bound to either the surface or the HS chains respectively. We can calculate the partition function taking the product of the individual contributions to the respective partition function.

For Q_G , we have contributions from the HS binding, $q_{HS}^{binding}$ as well as a contribution from the steric repulsion arising from insertion into the chain, $q_{insertion}^{HS}$.

$$Q_G = q_{binding}^{HS} q_{insertion}^{HS}. \quad (5)$$

Likewise, for Q_R we have contributions from the receptor binding $q_{binding}^R$, as well as the steric repulsion generated by the surrounding HS $q_{insertion}^{HS}$.

$$Q_R = q_{binding}^R q_{insertion}^{HS*}. \quad (6)$$

It should be noted that $q_{insertion}^{HS}$ and $q_{insertion}^{HS*}$ are slightly different, due to differences in the geometry of binding in the different cases. The individual partition functions will be discussed in more detail in the following subsections.

S1 HS Binding

Each HS chain contains multiple sites to which the spike can bind. The number of these binding sites, N_L is some fraction of the number of units, N_p . The number of units can be calculated from the HS chain length, d_{HS} and HS Kuhn length, b_{HS} as $N_p = (d_{HS}b_{HS})^2$.

Only a single bond can form between a given HS chain and the virus. This is likely due to the mechanical properties of the HS chain, suggesting the entropic cost of forming an additional bond is too high. We can, therefore, state the partition function for a single HS chain binding to the virus in terms of the number of spikes available for binding, N_S , and the HS-spike binding energy in solution ϵ_{HS} :

$$q_{binding}^{chain} = N_L N_S \exp\left(-\frac{\epsilon_{HS}}{k_B T}\right). \quad (7)$$

Then, we can calculate the total partition function for HS-spike binding from the binding energy of a given chain, $\epsilon_{chain} = -\ln(q_{HS}^{binding})$, and the number of chains, as [?]

$$q_{binding}^{HS} = \sum_{i=0}^{N_{chains}} \binom{N_S}{i} \binom{N_{chains}}{i} i! \exp\left(\frac{-i\epsilon_{chain}}{k_B T}\right). \quad (8)$$

S1 HS Repulsion

As multiple HS chains will bind to a single proteoglycan, we expect the density of the HS chains to be non-uniformly distributed, with areas of locally high density around each proteoglycan. This means when the virus is bound only to HS chains, there will be effectively a high local density then when the virus is bound to the receptors.

We, therefore, have to consider two cases for the steric repulsion. First, when the virus is bound only to HS chains, and the second when the virus is bound to the receptors. For the receptor binding case, we will take the global HS density as ρ_{HS} , while for the HS binding case we will take the higher local density.

We can then calculate the energy of insertion in terms of the density of chains, the inserted volume, $V_\mu(z)$, and the insertion parameter, δ . This gives

$$f_{insertion} = k_B T V_\mu(z) \rho_{HS} (1 - \delta^2)^{\frac{9}{4}}. \quad (9)$$

From this expression, we can calculate the partition function as $q_{insertion} = \exp\left(-\frac{f_{insertion}}{k_B T}\right)$.

The inserted volume can be calculated geometrically by considering the distance from the surface, z , as

$$V_\mu(z) = \begin{cases} \frac{1}{3}\pi z^2(3R - z), & \text{for } z < 2R \\ \frac{4}{3}\pi R^3, & \text{for } z \geq 2R. \end{cases} \quad (10)$$

Insertion Parameter

The insertion parameter quantifies how far into the brush the virus has penetrated, and is given as the ratio between the brush length and the distance between the virus and surface. It can be expressed in terms of the HS chain length, d_{HS} , the distance between the virus and the cell surface, z , and the receptor tether length, d_r , as

$$\delta_p = \frac{z}{d_r + d_{HS}}. \quad (11)$$

In the presence of receptor-spike binding, the cell-virus distance will be equal to the receptor tether length, d_r . However, in the absence of receptor-spike binding, situation is more complex.

We will number the binding sites of a HS chain, from $n = 1, \dots, N_L$, where $n = 1$ is the site farthest from the cell surface. We can then define the distance between two sites, n and $n+1$, as d_n . It is not possible to identify each distance between sites, so we will assume that the distances can be approximated by the average distance, an assumption that holds for long chain lengths with roughly equal distribution of sites. We can then express the inter-site distance as

$$d_n = \bar{d}_n = \frac{b\sqrt{N_P}}{(N_L + 1)}. \quad (12)$$

The virus-cell distance can then be expressed as a function of the closest binding site to the surface, n_{max} ;

$$z = \begin{cases} d_r & \text{for receptor-spike binding} \\ d_r + d_{HS} - n_{max}d & \text{otherwise.} \end{cases} \quad (13)$$

Mean Field Approximation

As we cannot know the distance of each virus from the cell surface, we will apply a mean field approximation, treating the repulsion as the repulsion at the average distance.

To do this, we must find the probability that a given site, n , is the closest bound site to the surface on a given chain. This can be expressed in terms of the probability that no bonds are formed on any HS chain at a given site, $p_\kappa = (1 - \kappa)^{N_{HS}}$, where κ is the probability a given site is bound (which can be calculated from the binding energy of a given site as described in equation (19))

$$p^L(n) = (1 - p_\kappa)p_\kappa^{N_L - n}. \quad (14)$$

468 The expectation value of the cell-virus distance is then given by

$$\langle d_c \rangle = d_r + \sum_{n=1}^{N_L} p_L(n) (d_{HS} - nd_n). \quad (15)$$

469 S1 Receptor Interaction

470 As there are multiple possible receptors that the spike can bind to, we need to take a different
471 but equivalent approach for calculating the receptor-spike partition function. First, we define
472 the partition function as

$$q_{receptor} = \exp \left(-\frac{f_{receptor}}{k_B T} \right) - 1. \quad (16)$$

473 Then, we can define the energy of binding in terms of the probability that a given site is
474 unbound, p_i .

$$\frac{f_{receptor}}{k_B T} = \sum_i \ln p_i + \frac{1}{2} (1 - p_i), \quad (17)$$

475 The probability of a linker being unbound can be calculated by closure (i.e. all probabilities
476 must sum to 1). Defining p_{ij} as the probability linker i is bound to linker j , we have

$$p_i + \sum_j p_{ij} = 1. \quad (18)$$

477 Finally, we can calculate p_{ij} can be defined as the product of the probability that both i and j
478 are unbound and the probability that a bond will form, which is given by the Boltzmann factor
479 - the negative exponential of the binding energy. This gives the expression

$$p_{ij} = p_i p_j \exp \left(\frac{f_{R-Spike}}{k_B T} \right). \quad (19)$$

480 Where $f_{R-Spike}$ is the Receptor-Spike binding energy.

481 Receptor-Spike Binding Energy

482 The binding energy between the spike and a given receptor has both an attractive component,
483 arising from the binding interaction, and a repulsive component, arising from the steric repul-
484 sion caused by trapped glycans, as discussed in the main body.

485 As the spike may have multiple binding sites for a given receptor, we can calculate the
486 receptor-spike binding energy in terms of the number of binding sites, N_{Site} , the binding en-
487 ergy for a single site, ϵ_R , and the repulsive contribution u_R , as

$$f_{R-Spike} = N_{Site} (\epsilon_R + u_R). \quad (20)$$

488 For TMPRSS2 we have only a single binding site, and we do not see the presence of gly-
489 cans upon binding. Therefore, the TMPRSS2-Spike binding energy simplifies to $f_{R-Spike}^{TMPRSS2} =$
490 $\epsilon_R^{TMPRSS2}$. ACE2, meanwhile, has three binding sites, and we expect to see the presence of
491 glycans upon binding. Therefore we need to calculate the repulsive contribution form.

492 We start by considering the Kuhn length of the glycans, b_G ; the unit size at which each unit
493 can freely move, and orient in any direction. All details of chemical arrangements, bond rota-
494 tion constraints, monomer-monomer interaction and monomer interaction with its local envi-
495 ronment, are described by this parameter. A larger Kuhn length corresponds to a more rigid,
496 and therefore less coiled, chain.

497 The mean squared end-to-end distance of a coiled polymer is then given by

$$\bar{r}_0^2 = N_G b_G^2, \quad (21)$$

498 while the length of a fully extended chain is given by $r_{max} = N_G b_G$.

499 The probability that the end-to-end distance of a glycan is a given value, r , can be deter-
500 mined from a Gaussian distribution,

$$p(r, N_G) = \left(\frac{3}{2\pi N_G b_G^2} \right)^{\frac{3}{2}} \exp \left(-\frac{3r^2}{2N_G b_G^2} \right). \quad (22)$$

501 Each branch can explore a hemispherical volume of radius r . Therefore, integrating equation
502 (22) over this volume gives the partition function for a given branch,

$$q_j(r) = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^r \left(\frac{3}{2\pi N_G b_G^2} \right)^{\frac{3}{2}} \exp \left(-\frac{3r^2}{2N_G b_G^2} \right) d\phi d\theta dr = \frac{3\pi}{4N_G b_G^2} \operatorname{erf} \left(\sqrt{\frac{3}{2N_G b_G^2}} r \right). \quad (23)$$

503 Hence the energy loss associated with the compression of a glycan with B branches upon
504 binding is determined by the change in r from the average chain length $\bar{r}_0$ to the binding dis-
505 tance $\bar{r}_B$,

$$U_{glycan} = \ln \left[\frac{\sum_{j=1}^B q_j(\bar{r}_0)}{\sum_{j=1}^B q_j(\bar{r}_B)} \right] = \ln \left[\sum_{j=1}^B \frac{\operatorname{erf} \left(\sqrt{\frac{3}{2}} \right)}{\operatorname{erf} \left(\sqrt{\frac{3}{2N_j b_G^2}} \bar{r}_B \right)} \right]. \quad (24)$$

506 The total free energy of repulsion can now be found by summing over each glycan, $u_R =$
507 $\sum_{N_{Glycan}} U_{glycans}$, where N_{Glycan} is the number of glycans. Assuming that each glycan contains
508 a similar amount of units, we can approximate

$$u_R = N_{Glycan} U_{glycan} = N_{Glycan} \ln \left[\sum_{j=1}^B \frac{\operatorname{erf} \left(\sqrt{\frac{3}{2}} \right)}{\operatorname{erf} \left(\sqrt{\frac{3}{2N_j b_G^2}} \bar{r}_B \right)} \right]. \quad (25)$$