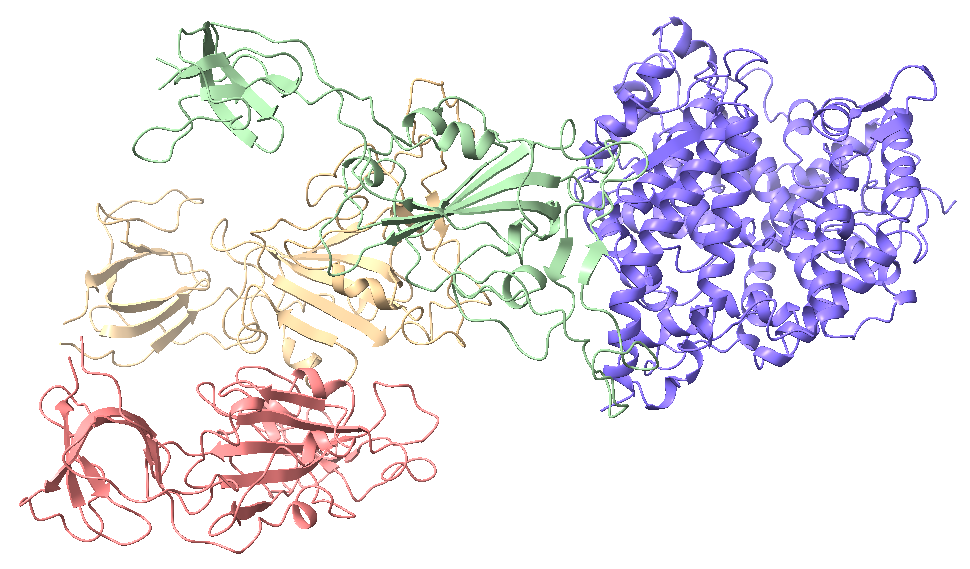
**Engineering a recombinant protein pair for interaction assay using FRET, with human ACE2 and SARS-CoV-2 spike glycoprotein:**

SARS-Cov-2 represents a stark threat to global public health. Entry of this coronavirus into host cells is mediated by a homotrimeric spike glycoprotein present within the viral envelope, which interacts with hACE2 receptor [1]. This S-protein comprises two subunits: S1 and S2. Once the RBD from one S1 component of the spike trimer binds hACE2, the S-protein adopts a more open conformation that allows rearrangement of the S2 helix such that SARS-CoV-2 can fuse with the plasma membrane of human endothelial cells [1]. The interaction between hACE2 and one RBD of the S1 subunit of SARS-Cov-2 S-protein is therefore essential for viral entry into human cells. Consequently, here we investigate molecular contacts between these two proteins, as computational prediction of residues implicated in this interaction may aid design of therapeutic small-molecule inhibitors that prevent S1 binding hACE2.

**Question 1:**

To identify which regions of each peptide are responsible for this protein-protein interaction, we employ SwissModel to predict the theorised formation of RBD S1-hACE2 complexes. Each component of the S1 trimer contains a RBD between residues 319 and 591, therefore we model the interactions of this domain with residues 19 to 615 of hACE2 [2]. The first eighteen hACE2 amino acids are not implicated in interfacial interactions between the two proteins; these residues are therefore expendable in lieu of the atom number constraints of SwissModel software [2]. Our model, displayed in Figure 1, is considered an accurate three-dimensional representation of an *in vivo* protein-protein interaction, evidenced by a strong GMQE score of 0.86 and 100% sequence identity with our target RBD and hACE2 sequences.

Figure 1 demonstrates that S1 adopts predominantly β-sheet based topology, with only one subunit of the trimer interfacing with hACE2. Previous studies have confirmed that whilst each S1 component of the trimer contains a receptor-binding domain, two of the three RBDs do not engage with the hACE2 receptor [3]. However, through an extensive network of hydrogen bonding, the remaining RBD, adopting a protruding ‘up-conformation’ in contrast from the closed conformation of unengaged RBDs, predominantly interfaces with three hACE2 domains: two N-terminal helices, H1 and H2, and a short β-hairpin known as EE3 [3].

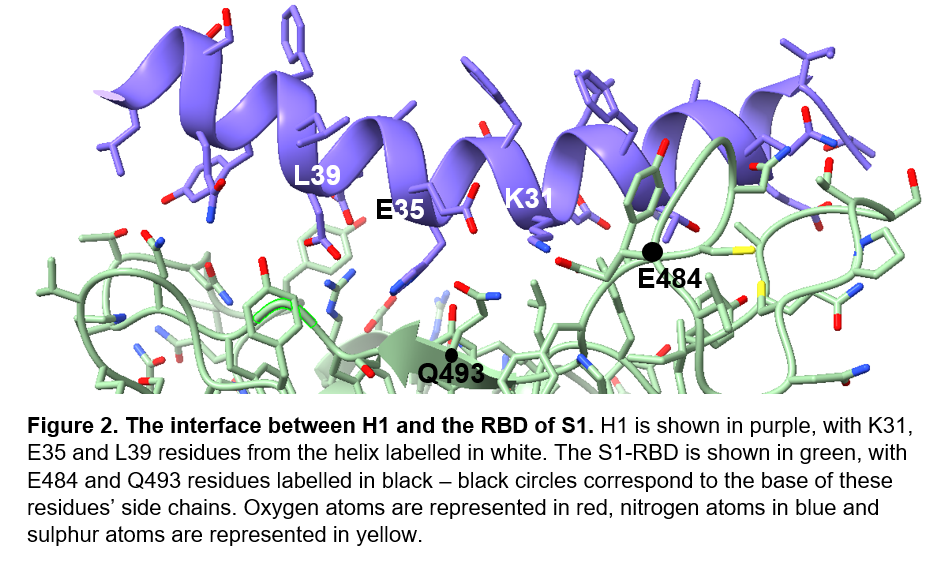


**Figure 1. The three-dimensional configuration of the hACE2-S1 complex.** hACE2 is shown in blue, whilst the three components of the S1 trimer are shown in red, yellow and green. Only one of these three S1 chains interfaces with hACE2, as circled in black.

Here, we demonstrate the H1 N-terminal helix, located between residues 19 and 52, is the leading region of hACE2 interaction with the protruding RBD of S1, stabilised by additional hydrogen bonds with an antiparallel β-sheet-rich EE3 domain. One face of H1 is more prominently implicated in binding the RBD, as displayed in Figure 2, with multiple charged contact residues located in positions corresponding to canonical α-helical periodicity (i.e., i, i+4, i+7…). One contact residue located within this H1 face is K31, which facilitates RBD-H1 interaction through salt bridge formation between the slight Nζ negative charge of K31 and the slight Oε1 positive charge of the E484 residue from S1 [4]. This 484-residue of the S1-RBD represents a site of significant interest to the research community, as convergent evolution from multiple SARS-CoV-2 lineages has favoured an E484K mutation at this position [5]. Introduction of lysine at residue-484 provokes conformational changes in the RBD that produce more favourable electrostatic forces of attraction between hACE2 and S1. E484K also enables formation of additional hydrogen bonds that increase binding affinity between the two proteins, and is therefore deemed a variant-of-concern in the Covid-19 pandemic as this mutation enhances the ability of SARS-CoV-2 to bind and ultimately infect human cells [5]. This pair of K31 and E/K484 residues is therefore highly relevant to the protein-protein interface.

**K31**

**L39**



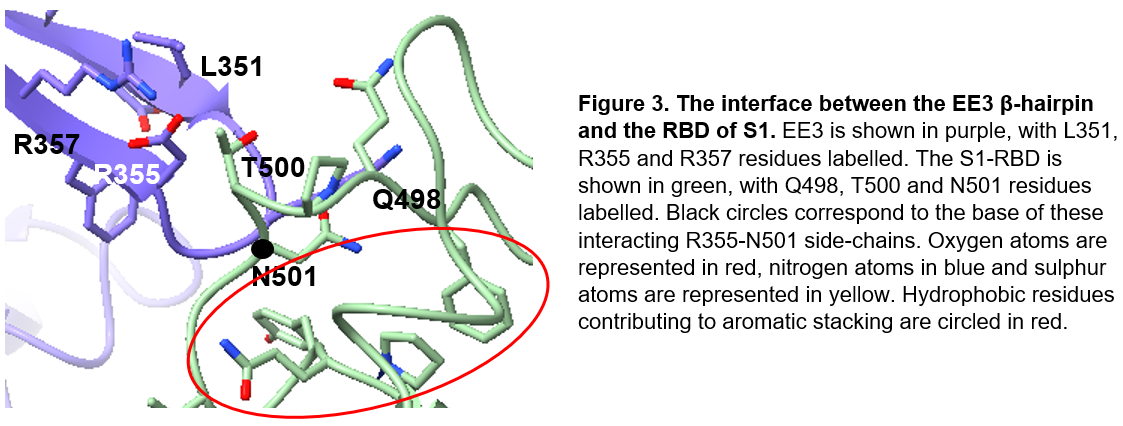


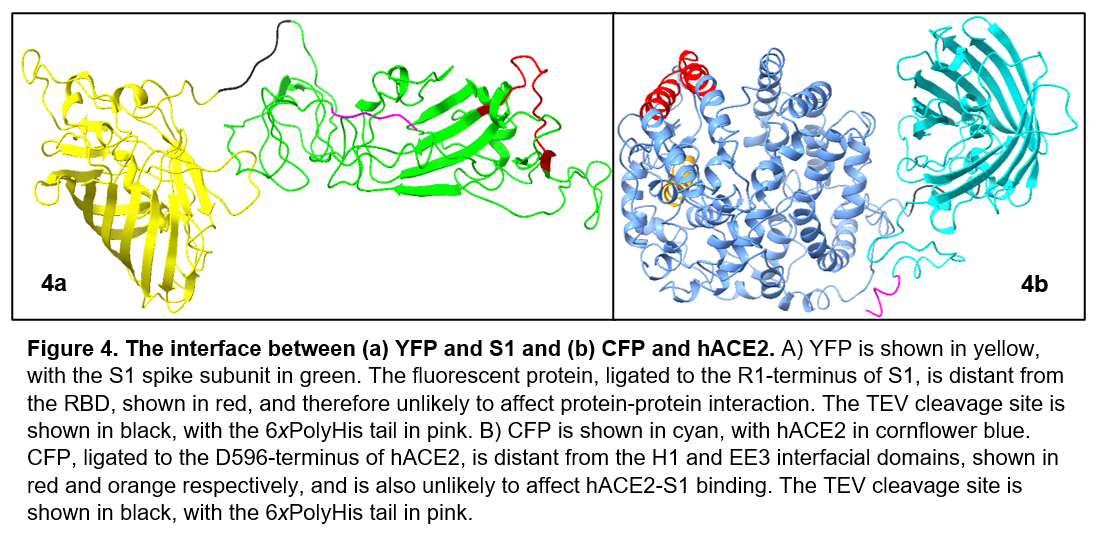
Figure 3 demonstrates that the EE3 domain of hACE2, located between residues 346 and 360, is requisite for stabilising the H1-RBD interface. This short β-hairpin interacts with a highly conserved S1-RBD binding loop between residues 498-508, which contains multiple hydrophobic aromatic residues that form cation-π interactions with the charged residue sidechains of EE3 in a process termed aromatic stacking [6]. Formation of helix-stabilising electrostatic forces of attraction between R355 of hACE2 and N501 of the S1-RBD is one example of a pair of residues involved at the interface between these two binding domains, as displayed in Figure 3. An N501Y substitutive mutant, present in the genome of several variant SARS-Cov-2 lineages including B.1.1.7 and B.1.351, has been shown to strengthen binding affinity between the S1-RBD and hACE2 [7]. The presence of another tyrosine residue in addition to Y505 and Y508 further contributes to aromatic stacking implicated in interfacial EE3-RBD interactions, to such an extent that N501Y confers an evolutionary advantage for SARS-CoV-2 through enhanced binding affinity to hACE2, demonstrating the importance of this pair of residues to our protein-protein interaction.

We can employ site-specific fluorescence probes as FRET pairs to quantify distances between residues implicated at the hACE2-S1 interface. Demonstrating specific amino acids on each protein are spatially close will further validate our modelling results. Here, we plan to selectively label cysteine-488 from the S1-RBD with coumarin maleimide, which fluoresces in the presence of cysteine [8]. We use cysteine-associated fluorophores as our FRET-pair donor due to the relative scarcity of cysteine residues amongst amino acids; this reduces the likelihood our maleimide label will react with other cysteines in the protein and cause false positive fluorescence from our FRET acceptor. Cysteine-488 is one of only two cysteine residues at the hACE2-S1 interface – to ensure nearby cysteine-480 does not also induce fluorescence of the maleimide label, we conduct point mutagenesis of this residue into alanine, the least destabilising substitute with comparable hydrophobicity.

There are no cysteine residues from hACE2 at the protein-protein interface, therefore we insert an acridonylalanine fluorophore into position-31 of the receptor amino acid sequence as our acceptor molecule of the FRET pair. Cysteine-488 interacts with a cluster of hACE2 residues present on one face of H1, including K31 and adjacent amino acids, hence this domain within hACE2 is an attractive region to investigate. Insertion of acridonylalanine-31 may disrupt the hACE2-S1 interface, however the periodic helical nature of H1-RBD interaction suggests multiple residues are involved - one point mutation is therefore unlikely to significantly alter protein-protein interactions. The acridonylalanine acceptor will fluoresce using light energy transferred from the cysteine-488 donor, provided distance between our FRET-paired residues is within the detectable range of initial cysteine-488 fluorescence. Quantification of the relative rate of emission from this acceptor residue will enable calculation of the distance between the two residues – close spatial proximity will further support this residue pair is implicated in hACE2-RBD interaction.

**Question 2:**

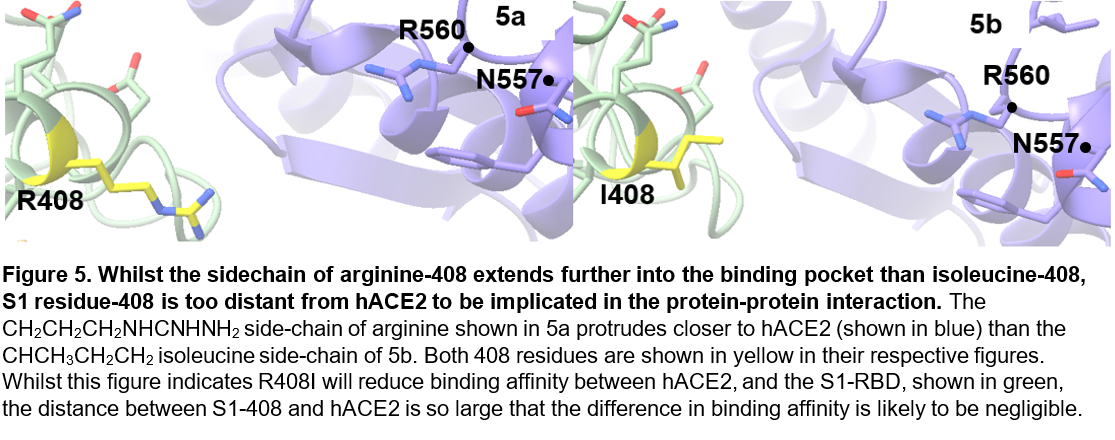
To confirm that a protein-protein interaction exists between hACE2 and S1-RBDs in the first place, we must also ligate fluorescent proteins to termini from each protein that do not interfere with docking. If FRET can be detected between the two fluorophores, it will confirm that the two proteins are in close spatial proximity to one another, supporting the existence of an hACE2-S1 interaction. We employ CFP and YFP as our FRET pair; in the presence of YFP, the rate of fluorescent CFP photobleaching decreases, therefore we can quantify a decreased rate of fluorescence decay from CFP to demonstrate FRET is occurring between the two fluorescent proteins, and hence recombinant hACE2 and S1 are spatially close [9]. We ligate these CFP and YFP fluorophores to the D596-terminus of hACE2 and the R1-terminus of one S1 chain respectively; these termini are significant distances from our modelled protein-protein interface, and therefore introducing fluorescent proteins at these positions should not affect hACE2-S1 interaction. Future experiments using ClusPro software could subsequently determine whether fluorophore-tagged hACE2 and S1 indeed interact, and whether our FRET-pair labels interfere with this protein-protein interaction.

Figure 4 demonstrates the overall topology of both hACE2 and S1 fused with their respective FRET labels, modelled using Phyre2 and visualised in ChimeraX. Figure 4a demonstrates the barrel-shaped CFP ligated to the D596-terminus is located well away from the main H1, H2 and EE3 domains of hACE2, circled in red. Figure 4b shows that YFP bound at the R1-terminus is also distant from interfacial regions of the S1-RBD, circled in blue. With 98% of residues modelled at above 90% confidence by Phyre2, we can therefore reliably confirm that neither fluorescent label will disrupt hACE2-S1 interaction.

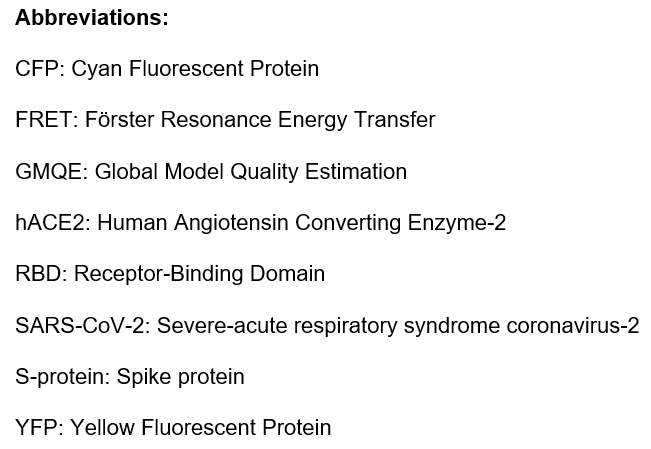
Our homology models from Figure 4 also contain a poly-6*x*His tag at the C-terminus of both hACE2 and the S1-RBD. When investigating the hACE2-RBD interface in the laboratory, a polyhistidine label is employed to improve purity of protein complexes; it is therefore important to model the structural effects of this tag here as a control to show the label is not implicated in the hACE2-S1 interaction [10]. Both models also contain a TEV cleavage site, primary sequence ENLYFQS, inserted between concatenated target protein and fluorophore sequences. Laboratory techniques often incorporate this cleavage site into recombinant protein complexes, as it enables both components of each recombinant protein pair to be separated and recycled for future use – this allows the structural interface between hACE2 and S1 to be investigated both with and without ligated FRET-pairs [11]. It is therefore essential to include this cleavage site in our models as a control to demonstrate TEV sites are not involved at the hACE2-S1 interface.

**Question 3:**

In Question 1, we introduced E484K and N501Y mutants from SARS-Cov-2 S-protein. Both these point mutations line the interface between the S1-RBD and hACE2, expediting viral entry into human cells by enhancing binding affinity between the two proteins. Whilst both these SARS-CoV-2 variants strengthen the hACE2-S1 interaction, they each increase protein-protein affinity through a different mechanism: E484K strengthens electrostatic forces of attraction between the two proteins, whilst aromatic stacking cation-π interactions enhance binding affinity between the S1-RBD and the EE3 domain of hACE2. Another characterised mutation in the S1-RBD, S477G, increases S1-hACE2 affinity by altering the local conformation of a flexible set of residues in the RBD, providing a more compatible structural arrangement for binding hACE2 [12]. Variable residues within the S1-RBD therefore manifest changes in hACE2-S1 affinity through a range of diverse mechanisms.

To determine additional SARS-CoV-2 variants from earlier in the evolutionary history of the virus, we examine the phylogeny of the S-protein to identify further variable regions of the S1-RBD. Here, we employ UNIPROT to align SARS-CoV-2 sequences entries from their database, and identify variable residues between positions 319-591 (the S1-RBD). One UNIPROT entry contains isoleucine at position-408 instead of canonical arginine. Isoleucine is non-polar, in contrast from the basic arginine, therefore R408I reduces electrostatic forces of attraction between S1 and hACE2. However, Figure 5 demonstrates that whilst the long arginine side-chain extends further into the hACE2-S1 binding interface than the shorter isoleucine R-group, the distance between hACE2 and S1 is consistently too large for significant protein-protein interaction.

Another variable residue is V483A; like R408I, the alternative residue is non-polar with a smaller side-chain than the canonical amino acid, and correspondingly reduces hACE2-S1 affinity. However, any V483A-induced change in S1-hACE2 binding affinity is likely to be negligible, as the distance between position-483 of S1 and hACE2 is again too large for significant protein-protein interaction. Whilst V483 is therefore not involved at the S1-hACE2 interface, identifying residues implicated in this protein-protein interaction is critical to designing Covid-19 therapies. For example, artificial peptide mimics could be synthesised with similar RBD-binding interfaces to hACE2, inhibiting entry of SARS-CoV-2 into human cells and therefore reducing patient symptoms following Covid-19 infection.



**References:**

1. Wang, L., & Xiang, Y. (2020). Spike-Glycoprotein-Mediated Entry of SARS Coronaviruses. *Viruses,* **12(11):** 1289.
2. Huang, Y., Yang, C., Xu, Xf. *et al*. (2020). Structural and functional properties of SARS-CoV-2 spike-protein: potential antivirus drug development for COVID-19. *Acta-Pharmacol Sin,* **41:**1141–1149.
3. Song, W., Gui, M., Wang, X., & Xiang, Y. (2018). Cryo-EM structure of the SARS-coronavirus spike-glycoprotein in complex with its host-cell receptor ACE2. *PLoS pathogens,* **14(8):** e1007236.
4. Karoyan, P., Vieillard, V., Gómez-Morales, L. *et al*. (2021). Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection. *Commun Biol,* **4**:197.
5. Nelson, G., Buzko, O., Spilman, P. *et al.* (2021). Dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *bioRxiv,* **13:**426558.
6. Veeramachaneni, G. K., Thunuguntla, V., Bobbillapati, J., & Bondili, J. S. (2020). Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor. *Journal of biomolecular structure & dynamics,* **1–11.**
7. Collier, D.A., De Marco, A.,Tortorici, M.A. *et al*. (2021). SARS-CoV-2 B.1.1.7 sensitivity to mRNA vaccine-elicited, convalescent and monoclonal antibodies. *medRxiv.*
8. Ferrie, J. J., Ieda, N., Haney, C. M., Walters, C. R., Sungwienwong, I., Yoon, J., & Petersson, E. J. (2017). Multicolor protein FRET with tryptophan, selective coumarin-cysteine labeling, and genetic acridonylalanine encoding. *Chemical Communications,* **53(80):** 11072-11075.
9. Tramier, M., Zahid, M., Mevel, J. C., Masse, M. J., & Coppey‐Moisan, M. (2006). Sensitivity of CFP/YFP and GFP/mCherry pairs to donor photobleaching on FRET determination by fluorescence lifetime imaging microscopy in living cells. *Microscopy research and technique,* **69(11):** 933-939.
10. Guebre-Xabier, M., Patel, N., Tian, J. H., Zhou, B., Maciejewski, S., Lam, K., ... & Smith, G. (2020). NVX-CoV2373 vaccine protects cynomolgus macaque upper and lower airways against SARS-CoV-2 challenge. *Vaccine,* **38(50):** 7892-7896.
11. Laffeber, C., de Koning, K., Kanaar, R., & Lebbink, J. (2021). Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants. *bioRxiv.*
12. Singh A, Steinkellner G, Köchl K, Gruber K, Gruber CC. (2021). Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. *Scientific Reports,* **11(1):**4320.