

***In Silico* Analysis of DC-SIGN Binding to SARS-CoV-2 Variants of Concern**

Lian Snow

Abstract:

Since its appearance in 2019, multiple variants of SARS-CoV-2 have emerged. The World Health Organization (WHO) classifies variants based on their levels of potential harm. Variants of Concern (VOCs) is a classification that indicates potential for increased transmissibility, increased virulence, and reduced effectiveness to preventative measures or treatments. The VOCs investigated in this study were Alpha, Beta, and Gamma. Published research suggests that SARS-CoV-2 and its variants bind with angiotensin-converting enzyme 2 (ACE2). SARS-CoV-2 may also bind to the dendritic cell-specific intercellular adhesion molecule grabbing non-integrin (DC-SIGN). This dendritic cell receptor, after recognizing and binding to a variety of pathogens, including coronaviruses, enables binding between dendritic cells and T-lymphocytes.

This study worked to extend the knowledge of the COVID-19 disease process through conducting *in silico* analyses of the initial binding interactions of DC-SIGN and the SARS-CoV-2 spike proteins of Alpha, Beta, and Gamma variants, at increasing temperatures. This was intended to mimic a fever in humans. The collected data included: dissociation constants, Gibbs free energy, and interfacial contact points of DC-SIGN bound to the Alpha, Beta, and Gamma spike protein variants. The results suggest that the initial binding interactions were strongest between DC-SIGN and wild-type SARS-CoV-2, followed by the variants Beta, Gamma, and lastly Alpha. The Gibbs free energy values for Beta, Gamma, and Alpha, respectively, were -14.1, -13.1, 10.9, and -10.1 kcal/mol. Following the same order, the dissociation constants of their interactions at 25°C were 0.049, 0.26, 11, and 42 nM. When modeled at 37°C, the values were 0.12, 0.6, 22, and 81 nM. At 40°C the values were 0.15, 0.74, 26, and 95 nM. These data indicate that as temperature increases, the dissociation constant increases between DC-SIGN and the Alpha, Beta, and Gamma variants. This suggests that fevers decrease the binding affinity between DC-SIGN and VOCs. The number of interfacial residues involved in the binding between DC-SIGN and the variants is consistent with this data. This study provides further evidence that there are additional host-virus interactions beyond ACE2 that should be studied to provide insights on the function of SARS-CoV-2 and the development of treatments.

Introduction:

Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2) is a strain of coronavirus that was discovered in Wuhan China, in December 2019, and has since spread globally (1). The virus can pose a serious health risk to the public by accessing and attacking the lungs through oral and nasal passages; infection can lead to death if the immune system is too weak to fight it off (2). Other symptoms include cardiovascular issues, such as cardiac arrest, myocarditis, and cardiomyopathy (2). As of August 24, 2021, nearly 213 million have become infected and nearly 4.45 million have succumbed to the virus (3). Variants of SARS-CoV-2 have evolved as well, which cause varying harm to the public; the Center for Disease Control and Prevention has classified each variant into categories to organize the level of danger that the variants could impose on the public (4). The categories include Variants of Interest (VOI), Variants of Concern

(VOC), and Variants of High Consequence (VOHC) (4).

Research suggests that the VOIs have significant point mutations from Wild Type SARS-CoV-2 (W.T.), but do not appear to be dangerous to the public. Additionally, they have varying transmissibility. VOCs have significant point mutations from W.T., appear to cause more hospitalizations and deaths, and have increased transmissibility. VOHCs have differing point mutations from W.T., have varying transmissibility, and

pose as a significant health risk to patients (4). No SARS-CoV-2 variants have been classified under the VOHC category (4). The World Health Organization has also assigned a Greek letter to each variant in the order in which they've been discovered to assist non-scientific audiences in discussing them and to limit geographical stigmatization (5).

Despite differing point mutations between SARS-CoV-2 variants, all have similar tertiary structures, a 3-D protein shape created by side chain interactions. The SARS-CoV-2 virions are composed of four structural proteins, one of which is the spike glycoprotein, located on the lipid membrane matrix (6, 7). This is cleaved by host proteases into the subdomains S1, which binds to host receptors, and S2, which fuses with the host receptors. S1 can be branched further to form an N-terminal domain (NTD) and a C-terminal domain (CTD) that both act as receptor binding domains (RBDs) to bind with host-cell receptors (8, 6, 9). It is currently unclear whether host-cell receptors bind to the viral NTD or the CTD (8, 10, 11). Once the RBD binds to a host-cell receptor, the virion successfully infects the body by replicating virus particles through cellular hijacking (9).

The variants used in this study were selected from the VOC category: B.1.1.7 (α), B.1.351 (β), P.1 (γ), and B.1.617.2 (δ). The α variant was discovered in the United Kingdom in December 2020 and is currently one of the most common variants in the U.S (4). Transmissibility is approximately 50% higher than W.T., and point mutations on the spike glycoprotein include 69del, 70del, 144del, E484K, S494P, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H, and K1191N (4). The β variant was discovered in South Africa in December 2020 and is approximately 50% more transmissible than W.T.. The spike glycoprotein point mutations include D80A, D215G, 241del, 242del, 243del, K417N, E484K, N501Y, D614G, and A701V (4). The γ variant was discovered in Brazil in November 2020, is 95% more transmissible than W.T., and the spike glycoprotein mutations include L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, and T1027I (4). The δ variant was discovered in India in December 2020 and replaced the α variant as the most common SARS-CoV-2 variant in the U.S. in July (4). It is 60% more transmissible than W.T. and spike glycoprotein point mutations include T19R, V70F, T95I, G142D, E156-, F157-, R158G, A222V, W258L, K417N, L452R, T478K, D614G, P681R, and D950N (4).

Most of the research regarding SARS-CoV-2 interactions with host cell receptors in the human body has been dedicated to the Angiotensin Converting Enzyme 2 (ACE2) receptor. However, research suggests that the Dendritic Cell-Specific Intercellular Adhesion Molecule – 3 – Grabbing Non-Integrin (DC-SIGN, or CD209) also acts as a host-cell entry point for SARS-CoV-2 virions, and potentially with higher dissociation constant and Gibbs free energy

values (12, 13, 14). As the SARS-CoV-2 variants have differing spike glycoprotein mutations, the dissociation constants and Gibbs free energy values may differ accordingly.

DC-SIGN is a type II transmembrane protein, C-type lectin receptor found on dendritic cells that mediate cell-cell interactions between DC and T-Lymphocyte cells (15, 16). C-type lectins are a calcium (Ca^{2+}) dependent, glycan-binding receptor in the human body, responsible for identifying and binding to carbohydrates to enact immune responses in the body (17). DC-SIGN binds with pathogens, specifically ligands bearing mannose and related sugars, using a carbohydrate recognition domain composed of a tetramer and an α helical neck peptide, located in the C terminal of the protein (17, 18). However, DC-SIGN may bind to a variety of viruses as well as bacteria (18, 19). It is currently unknown why, however scientists theorize that DC-SIGN can adapt and bind with mammalian-type glycan structures as DC-SIGN's CRD consists of additional oligosaccharides, and research suggests that they can arrange themselves into microdomains to adapt with larger receptors on viruses (19).

Research suggests that if DC-SIGN acts as a successful host-cell entry point for SARS-CoV-2 variants, DC-SIGN may be unable to efficiently identify and bind with pathogens in the body, potentially increasing the risk for co-infection (20). This suggests the importance of considering and researching the binding, and how it may differ between variants, between DC-SIGN and SARS-CoV-2 variants alongside ACE2.

Materials and Methods

Locating Protein Models

Models of DC-SIGN and the α , β , and γ variants were found on the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), a global database that provides free three-dimensional protein and nucleic acid models to the scientific community (21, 22). The IDs of the models used in this study are 1K9I (2.50 Å) for DC-SIGN, 7LWT (3.19 Å) for the α variant, 7LYN (3.32 Å) for the β variant, and 7NXC (3.14 Å) for the γ variant (23, 24, 25, 26).

No experimentally solvable structure for the δ variant was located; a structure was attempted to be created with its primary structure using the server Robetta (27). However, the server did not produce an optimal structure. This resulted in a limited amount of supporting data gathered between δ and DC-SIGN. Troubleshooting suggests that the input of the primary structure was too large. Additional research will be conducted to reattempt the creation of an experimentally solvable structure of the δ variant.

For W.T. docked to DC-SIGN, the same data was used from previously conducted research, predicted with a model we created using the RCSB PDB IDs 6M0J (2.45 Å) and 1K9I (2.50 Å), for W.T. and DC-SIGN, respectively (28, 23).

Docking Protein Models

The W.T. was docked to DC-SIGN in a previous study by isolating from RCSB PDB model 6M0J (2.45 Å) and docking it to the DC-SIGN RCSB PDB model 1K9I (2.50 Å). The

server High-Ambiguity Driven Docking 2.4 (HADDOCK 2.4), created by Bonvin Labs, was used (29). The W.T. and VOCs were all docked at the residues 445-455, 485-502. For DC-SIGN, the residue sequence 347-349, 357-360, and 364-367 was used. After submitting the docking request, HADDOCK 2.4 provided multiple models docked at differing angles; post visual analysis, the most accurate model was chosen to gather data.

Data Gathering

Using a server produced by Clustal, Clustal Omega, the primary structures of the VOCs and W.T. were compared through a multiple sequence alignment (30). These structures were provided by UniProt (31).

The Gibbs free energy values (measured in kcal/mol), dissociation constants (measured in K_d (nM)), and interfacial data was gathered with the Protein Binding Energy Predicting (PRODIGY) (32, 33).

For the dissociation constants, each model was predicted at three different temperatures to determine their differing values when the body enacts the immune system response with a fever. The temperatures used were 25 °C, the default setting for PRODIGY and the temperature that most thermodynamic calculations are conducted at, 37 °C, the average temperature of the human body, and 40 °C, the temperature of a high-grade fever (34, 35).

Results

The figures, graph, and tables below display the data of the multiple sequence alignment, Gibbs free energy values, dissociation constants, and interfacial data.

For the multiple sequence alignment, there are approximately 31 residues (indicating the point mutations) that are not asterisks between W.T., α , β , and γ .

Multiple Sequence Alignment of W.T. and VOCs

α	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS	60
Γ	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS	60
δ	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS	60
W.T.	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS	60
β	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS	60

α	NVTWFHAI--SGTNGTKRFDPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIV	118
Γ	NVTWFHAIHVS GTNGTKRFDPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIV	120
δ	NVTWFHAIHVS GTNGTKRFDPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIV	120
W.T.	NVTWFHAIHVS GTNGTKRFDPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIV	120
β	NVTWFHAIHVS GTNGTKRFDPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIV	120

α	NNATHNVVIKVC EFQFCNDPFLGVY-HKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLE	177
Γ	NNATHNVVIKVC EFQFCNYPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLE	180
δ	NNATHNVVIKVC EFQFCNDPFLGVYYHKNNKSWMES--GVYSSANNCTFEYVSQPF LMDLE	178
W.T.	NNATHNVVIKVC EFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLE	180
β	NNATHNVVIKVC EFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLE	180

α	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT	237
Γ	GKQGNFKNLSEFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT	240
δ	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT	238
W.T.	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT	240
β	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRGLPQGFSALEPLVDLPIGINITRFQT	240
α	LLALHRSYLT PGDSSSGWTAGAAAYVGYLQPTFLFKYNEGTITDAVDCALDPLSETK	297
Γ	LLALHRSYLT PGDSSSGWTAGAAAYVGYLQPTFLFKYNEGTITDAVDCALDPLSETK	300
δ	LLALHRSYLT PGDSSSGWTAGAAAYVGYLQPTFLFKYNEGTITDAVDCALDPLSETK	298
W.T.	LLALHRSYLT PGDSSSGWTAGAAAYVGYLQPTFLFKYNEGTITDAVDCALDPLSETK	300
β	--LHRSYLT PGDSSSGWTAGAAAYVGYLQPTFLFKYNEGTITDAVDCALDPLSETK	297

α	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISN	357
Γ	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISN	360
δ	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISN	358
W.T.	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISN	360
β	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISN	357

α	CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTINVYADSFVIRGDEVQRQIAPGQTGKIAD	417
Γ	CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTINVYADSFVIRGDEVQRQIAPGQTGTIAD	420
δ	CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTINVYADSFVIRGDEVQRQIAPGQTGKIAD	418
W.T.	CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTINVYADSFVIRGDEVQRQIAPGQTGKIAD	420
β	CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTINVYADSFVIRGDEVQRQIAPGQTGTIAD	417

α	YNYKLPDDFTGCVIAWNSNLD SKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPC	477
Γ	YNYKLPDDFTGCVIAWNSNLD SKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPC	480
δ	YNYKLPDDFTGCVIAWNSNLD SKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSKPC	478
W.T.	YNYKLPDDFTGCVIAWNSNLD SKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPC	480
β	YNYKLPDDFTGCVIAWNSNLD SKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPC	477

α	YNYKLPDDFTGCVIAHNSINLDSKVGGMYNYLYRLFRKSNLKPFERDSTEIYQAGSTPC	47
Γ	YNYKLPDDFTGCVIAHNSINLDSKVGGMYNYLYRLFRKSNLKPFERDSTEIYQAGSTPC	48
δ	YNYKLPDDFTGCVIAHNSINLDSKVGGMYNYLYRLFRKSNLKPFERDSTEIYQAGSKPC	47
W.T.	YNYKLPDDFTGCVIAHNSINLDSKVGGMYNYLYRLFRKSNLKPFERDSTEIYQAGSTPC	48
β	YNYKLPDDFTGCVIAHNSINLDSKVGGMYNYLYRLFRKSNLKPFERDSTEIYQAGSTPC	47

α	NGVEGFNCYFPLQSYGFQPTYGVGYPYRWWLSFELLHAPATVCGPKKSTNLVKNKCVN	53
Γ	NGVKGFNCYFPLQSYGFQPTYGVGYPYRWWLSFELLHAPATVCGPKKSTNLVKNKCVN	54
δ	NGVEGFNCYFPLQSYGFQPTNGVGYPYRWWLSFELLHAPATVCGPKKSTNLVKNKCVN	53
W.T.	NGVEGFNCYFPLQSYGFQPTNGVGYPYRWWLSFELLHAPATVCGPKKSTNLVKNKCVN	54
β	NGVEGFNCYFPLQSYGFQPTYGVGYPYRWWLSFELLHAPATVCGPKKSTNLVKNKCVN	53

α	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP	597
Γ	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP	600
δ	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP	598
W.T.	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP	600
β	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP	597

α	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSMVFQTRAGCLIGAETHVNSY	657
Γ	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSMVFQTRAGCLIGAETHVNSY	660
δ	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSMVFQTRAGCLIGAETHVNSY	658
W.T.	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSMVFQTRAGCLIGAETHVNSY	660
β	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSMVFQTRAGCLIGAETHVNSY	657

α	ECDIPIGAGICASYQTQTNSRRARSVASQSIIAYTMSLGAENSVAYSMNSIAIPNFTI	717
Γ	ECDIPIGAGICASYQTQTNSRRARSVASQSIIAYTMSLGAENSVAYSMNSIAIPNFTI	720
δ	ECDIPIGAGICASYQTQTNSRRARSVASQSIIAYTMSLGAENSVAYSMNSIAIPNFTI	718
W.T.	ECDIPIGAGICASYQTQTNSRRARSVASQSIIAYTMSLGAENSVAYSMNSIAIPNFTI	720
β	ECDIPIGAGICASYQTQTNSRRARSVASQSIIAYTMSLGAENSVAYSMNSIAIPNFTI	717

α	SVTTEILPVSMKTSVDCTMYICGDSIECSNLLQYGSFCTQLNRALTGIAVEQDKNTQE	777
Γ	SVTTEILPVSMKTSVDCTMYICGDSIECSNLLQYGSFCTQLNRALTGIAVEQDKNTQE	780
δ	SVTTEILPVSMKTSVDCTMYICGDSIECSNLLQYGSFCTQLNRALTGIAVEQDKNTQE	778
W.T.	SVTTEILPVSMKTSVDCTMYICGDSIECSNLLQYGSFCTQLNRALTGIAVEQDKNTQE	780
β	SVTTEILPVSMKTSVDCTMYICGDSIECSNLLQYGSFCTQLNRALTGIAVEQDKNTQE	777

α	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDG	837
Γ	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDG	840
δ	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDG	838
W.T.	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDG	840
β	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDG	837

α	LGDIAARDLICAQKFNGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM	897
Γ	LGDIAARDLICAQKFNGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM	900
δ	LGDIAARDLICAQKFNGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM	898
W.T.	LGDIAARDLICAQKFNGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM	900
β	LGDIAARDLICAQKFNGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM	897

α	QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQHAQALN	957
Γ	QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQHAQALN	960
δ	QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQHAQALN	958
W.T.	QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQHAQALN	960
β	QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQHAQALN	957

α	TLVKQLSSNFGAISSVLNDILARLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA	1017
Γ	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA	1020
δ	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA	1018
W.T.	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA	1020
β	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA	1017

α	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSPHGVVFLHVTYVPAQEKNTTAPA	1077
Γ	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSPHGVVFLHVTYVPAQEKNTTAPA	1080
δ	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSPHGVVFLHVTYVPAQEKNTTAPA	1078
W.T.	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSPHGVVFLHVTYVPAQEKNTTAPA	1080
β	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSPHGVVFLHVTYVPAQEKNTTAPA	1077

α	ICHGKAHFPRGCVFVSNQTHMFVTQRNFYEPQIITTHNTFVSGNCDVIGIVNNTVYDP	1137
Γ	ICHGKAHFPRGCVFVSNQTHMFVTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDP	1140
δ	ICHGKAHFPRGCVFVSNQTHMFVTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDP	1138
W.T.	ICHGKAHFPRGCVFVSNQTHMFVTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDP	1140
β	ICHGKAHFPRGCVFVSNQTHMFVTQRNFYEPQIITDNTFVSGNCDVIGIVNNTVYDP	1137

α	LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLSLIDL	1197
Γ	LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLSLIDL	1200
δ	LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLSLIDL	1198
W.T.	LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLSLIDL	1200
β	LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLSLIDL	1197

α	QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCMTSCCCLKGCCSCGSCCKFDEDD	1257
Γ	QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCMTSCCCLKGCCSCGSCCKFDEDD	1260
δ	QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCMTSCCCLKGCCSCGSCCKFDEDD	1258
W.T.	QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCMTSCCCLKGCCSCGSCCKFDEDD	1260
β	QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCMTSCCCLKGCCSCGSCCKFDEDD	1257

α	SEPVLKGVKLHYT	1270
Γ	SEPVLKGVKLHYT	1273
δ	SEPVLKGVKLHYT	1271
W.T.	SEPVLKGVKLHYT	1273
β	SEPVLKGVKLHYT	1270

Figure 1. Multiple Sequence Alignment of W.T. and the variants α, β, γ, and δ, conducted through Clustal Omega. Each row contains one letter amino acid abbreviations for each variant (indicated by the names in the left column), each color representing a different amino acid group: red represents small and hydrophobic; green represents hydroxyl, sulfhydryl, amine, and G; blue represents acidic; magenta represents basic-H, gray represents unusual amino acids. The gray lettered columns on the right state the residue lines of the variants. The symbols beneath the amino acids indicate similarities between the variants: asterisks indicate identical amino acids, colons indicate highly similar amino acids, and periods indicate low similarity of amino acids.

As for the Gibbs free energy values of the variants when bound to DC-SIGN, W.T. has a value of -14.1 kcal/mol, α a value of -10.1 kcal/mol, β a value of -13.1 kcal/mol, and γ a value of -10.9 kcal/mol.

Gibbs Free Energy

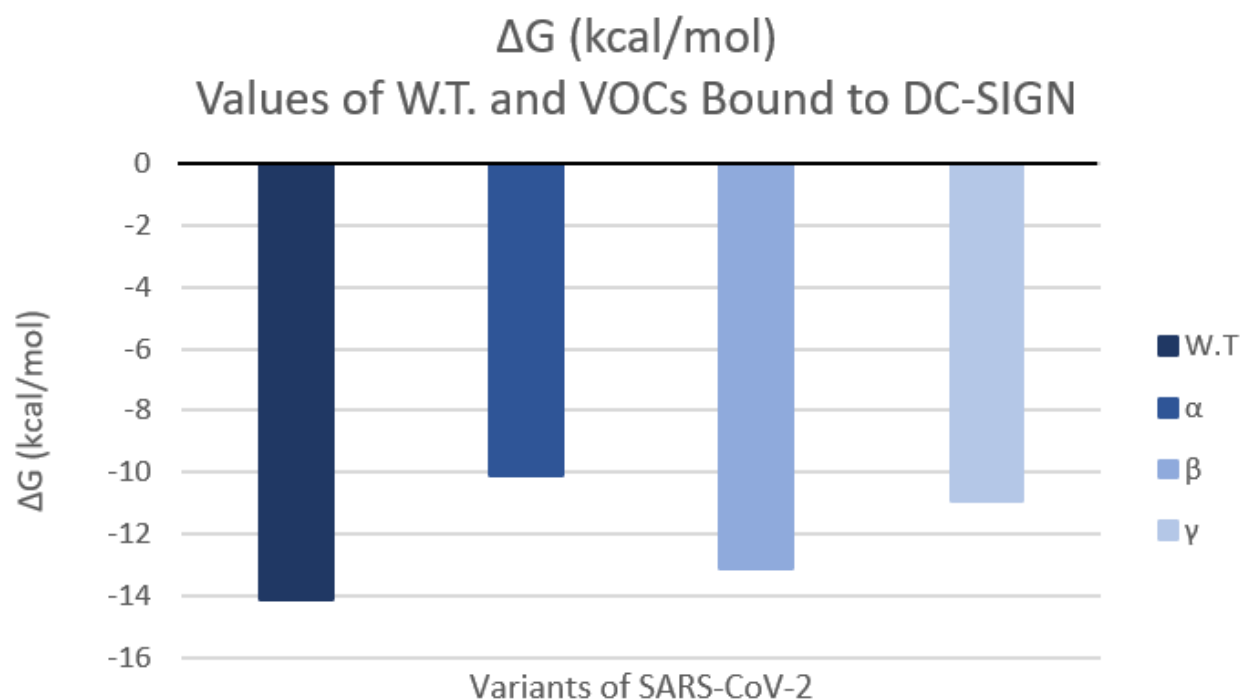


Figure 2. Gibbs Free Energy values (ΔG , in kilocalories/mole and indicated in the y axis) of W.T. and the variants α , β , γ , and δ , predicted by the server PRODIGY. The x axis contains the variants, each colored bar representing a variant, indicated by the legend to the right of the bars: dark blue represents W.T. bound to DC-SIGN, blue represents α bound to DC-SIGN, light blue represents β bound to DC-SIGN, and the lightest blue represents δ bound to DC-SIGN.

The dissociation constants of W.T. α , β , and γ at 25°C are 0.049, 42, 0.26, and 11 K_d (nM), respectively; at 37°C they are 0.12, 81, 0.6, and 22 K_d (nM), respectively; and at 40°C they are 0.15, 95, 0.74, and 26, respectively K_d (nM).

Dissociation Constant

Variants	Kd (nM) @ 25°C	Kd (nM) @ 37°C	Kd (nM) @ 40°C
W.T.	0.049	0.12	0.15
α	42	81	95
β	0.26	0.6	0.74
γ	11	22	26

Table 1. Predicted dissociation constants (K_d) of W.T. and the variants α , β , γ , and δ bound to DC-SIGN (contained in the black column) at 25°C (contained in the yellow column), 37°C (contained in the orange column), and 40°C (contained in the red column). This data was predicted with the server PRODIGY.

The W.T. variant and DC-SIGN have an accumulated interfacial residue count of 92 and an accumulated non-interacting surface percentage of 32.26. B.1.1.7. and DC-SIGN have an accumulated interactive residue count of 28 and an accumulated non-interacting surface percentage of 56.15. B.1.351 and DC-SIGN have an accumulated interfacial residue count of 85 and an accumulated non-interacting surface percentage of 58.26%. Finally, P.1 and DC-SIGN have an accumulated interfacial residue count of 62 and an accumulated non-interactive surface percentage of 56.28.

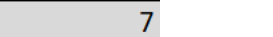
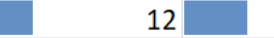
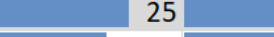
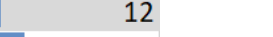
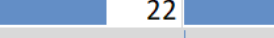
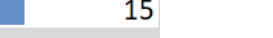
Interfacial Data				
Variants	W.T.	B.1.1.7.	B.1.351	P.1
ICs charged-charged	 4	 5	 6	 7
ICs charged-polar	 9	 8	 12	 8
ICs charged-apolar	 22	 19	 16	 17
ICs polar-polar	 10	 5	 4	 3
ICs polar-apolar	 32	 10	 25	 12
ICs apolar-apolar	 15	 18	 22	 15
NIS charged	19.09%	19.26%	19.83%	19.03%
NIS apolar	13.17%	36.89%	38.43%	37.25%

Table 2. Interfacial data of the binding between W.T. (second column), α (third column), β (fourth column), and γ (fifth column) bound to DC-SIGN. The blue column on the far left contains the interfacial contact residues (ICs) and the non-interacting surface (NIS) percentage with charged, polar, and apolar charges. This data was gathered through PRODIGY.

Discussion

Indicated by *Figure 1*, the multiple sequence alignment of the VOCs and W.T. predictably suggests that there are minimal differences between the variants' primary structure. *Figure 1* indicates that there are approximately 31 residues not indicated by asterisks, most likely the point mutations of the VOCs. These residue differences could be quite significant indicated by their differing transmissibility rates of approximately 50% for α and β , and 95% for γ) and the differing Gibbs free energy values and dissociation constants gathered in this study (4).

Displayed by *Figure 2*, the Gibbs free energy values suggest that W.T. binds strongest to DC-SIGN at -14.1 kcal/mol, followed by β (-13.1 kcal/mol), γ (-10.9 kcal/mol), and α (-10.1 kcal/mol). Regarding the dissociation constant, *Table 1* indicates that the variant that binds strongest to DC-SIGN is W.T., β , γ , and α , consistent with the Gibbs free energy values. The interfacial data supports the predicted values as the most to least interfacial data between W.T., the VOCs and DC-SIGN are W.T., β , γ , and α . In that corresponding order, the accumulated interfacial residue count of W.T. and the variants bound to DC-SIGN is 92, 85, 62, and 28. The non-interacting surface percentage of the variants and DC-SIGN differ from this order though, as despite W.T. having the least amount at 32.26%, α has the second least at 56.15%, then γ at

56.28%, and lastly β at 58.26%. A potential explanation for this inconsistent order is that β and γ have larger virions and stronger interfacial residues and bonds at their RBD than α .

Concluding this study, the data gathered provides more evidence of the interactions between W.T. and the VOCs of SARS-CoV-2. The W.T. may bind strongest with DC-SIGN, followed by the β , γ , and α variants. However, this does not indicate that the variants do not pose a threat to the public as their transmissibility rates are higher than W.T. (4). Additionally, the δ variant has one of the strongest transmissibility rates of the VOCs, indicating that it could bind strongly with DC-SIGN. Further experimentation may be conducted to investigate (4).

The Gibbs free energy and dissociation constant predictions indicate that there could be a strong affinity between DC-SIGN and the VOCs, the order from strongest to weakest being W.T., β , γ , and α . This infers that the DC-SIGN receptor should be considered as a primary host-cell entry point for the virus, and it should be studied as such alongside the Angiotensin Converting Enzyme 2 receptor. Further research could assist the development of treatments against SARS-CoV-2 variants, as well as inhibit binding between SARS-CoV-2 and DC-SIGN which could increase co-infection.

Acknowledgement

This author would like to thank the New Hampshire Academy of Science for providing mentorship and support throughout this research, specifically Dr. Alyson Michael, Dr. Kelly Salmon, and Dr. Peter Faletra.

Additionally, thanks to the Hypertherm HOPE Foundation, the Science Education Partnership Award, and the Couch Family Foundation for funding the New Hampshire Academy of Science.

References

1. Johns Hopkins University of Medicine. (n.d.). *COVID-19 Dashboard*. Retrieved from <https://coronavirus.jhu.edu/>
2. Wu, L.; O'Kane, A.; Peng, H.; Bi, Y.; Motriuk-Smith, D.; Ren, J. (n.d.). *SARS-CoV-2 and cardiovascular complications: from molecular mechanisms to pharmaceutical management*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0006295220303506>
3. Zheng J. (n.d.). *SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat*. Retrieved from <https://doi.org/10.7150/ijbs.45053>
4. Center for Disease and Control Prevention. (n.d.) *SARS-CoV-2 Variant Classifications and Definitions*. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>

5. World Health Organization. (n.d.) *Tracking SARS-CoV-2 Variants*. Retrieved from <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>
6. Belouzard, S.; Millet, J.; Licitra, B.; Whittaker, G. (2012, June 20). *Mechanisms of coronavirus cell entry mediated by the viral spike protein*. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/22816037/>
7. Perrotta, F.; Matera, M.; Cazzola, M.; Bianco, A. (n.d.) *Severe respiratory SARS-CoV-2 infection: does ACE2 receptor matter?* Retrieved from <https://www.sciencedirect.com/science/article/pii/S0954611120301360>
8. Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J. (2020, April 9). *Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2*. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/32275855/>
9. Adhikari, P.; Li, N.; Shin, M.; Steinmetz, N.; Twarock, R.; Podgornik, R.; Ching, W. (2020, August 5). *Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding*. Retrieved from <https://pubs.rsc.org/en/content/articlehtml/2020/cp/d0cp03145c>
10. Cheorl-Ho, K. (2020, June 26). *SARS-CoV-2 evolutionary adaptation toward host entry and recognition of receptor O-acetyl sialylation in Virus–Host interaction*. Retrieved from [doi:http://dx.doi.org.library.capella.edu/10.3390/ijms21124549](http://dx.doi.org.library.capella.edu/10.3390/ijms21124549)
11. Wec, A.; Wrapp, D.; Herbert, A.; Maurer, D.; Haslwanter, D.; Sakharkar, M.; Jangra, R.; Dieterle, M.; Lilov, A.; Huang, D.; Tse, L.; Johnson, N.; Hsieh, C.; Wang, N.; Nett, J.; Champney, E.; Burnina, I.; Brown, M.; Lin, S.; Sinclair, M.; Johnson, C.; Pudi, S.; Bortz III, R.; Wirchnianski, A.; Lauder milch, E.; Florez, C.; Fels, J.; Cecilia M. O'BrienGraham, B.; Nemazee, D.; Burton, D.; Baric, R.; Voss, J.; Chandran, K.; Dye, J.; McLellan, J.; Walker, L. (2020, August 7). *Broad neutralization of SARS-related viruses by human monoclonal antibodies*. Retrieved from <https://science.sciencemag.org/content/369/6504/731.full>
12. Chiodo, F.; Bruijns, S.; Rodriguez, E.; Li, R.; Molinaro, A.; Silipo, A.; Lorenzo, F.; Garcia-Rivera, D.; Valdes-Balbin, Y.; Verez-Bencomo, V.; van Kooyk, Y. (2020, May 14). *Novel ACE2-independent carbohydrate-binding of SARS-CoV-2 spike protein to host lectins and lung microbiota*. Retrieved from <https://www.biorxiv.org/content/10.1101/2020.05.13.092478v1>

13. Snow, L. (n.d.). *Computational Analysis of Spike Glycoprotein Bound To ACE2 and DC-SIGN*. Retrieved from <https://aaas.confex.com/aaas/2021/meetingapp.cgi/Paper/28959>
14. Gao, C.; Zeng, J.; Jia, N.; Stavenhagen, K.; Matsumoto, Y.; Zhang, H.; Li, J.; Hume, A.; Mühlberger, E.; van Die, I.; Kwan, J.; Tantisira, K.; Emili, A.; Cummings, R. Cummings. (2020, July 30). *SARS-CoV-2 spike protein interacts with multiple innate immune receptors*. Retrieved from <https://www.biorxiv.org/content/10.1101/2020.07.29.227462v1.full>
15. Snyder, G.; Ford, J.; Torabi-Parizi, P.; Arthos, J.; Schuck, P.; Colonna, M.; Sun, P. (2020, December 22). *Characterization of DC-SIGN/R interaction with human immunodeficiency virus type 1 gp120 and ICAM molecules favors the receptor's role as an antigen-capturing rather than an adhesion receptor*. Retrieved from <https://doi.org/10.1128/JVI.79.8.4589-4598.2005>
16. Geijtenbeek, T.; Kooyk, Y. (2008, October 9). *Pathogens target DC-SIGN to influence their fate DC-SIGN functions as a pathogen receptor with broad specificity*. Retrieved from <https://onlinelibrary.wiley.com/doi/full/10.1034/j.1600-0463.2003.11107803.x>
17. Cummings, R.; McEver, RP. (n.d.). *C. Type Lectins*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1943/#_NBK1943_pubdet_
18. Cambi, A.; Coopman, M.; Fidgor, C. (2005, February 28). *How C-type lectins detect pathogens*. Retrieved from <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1462-5822.2005.00506.x>
19. Mitchel, D.; Fadden, A.; Drickamer, K. (n.d.). *A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR*. Retrieved from <https://www.jbc.org/content/276/31/28939.short>
20. Chiodo, F.; Bruijns, S.; Rodriguez, E.; Li, R.; Molinaro, A.; Silipo, A.; Lorenzo, F.; Garcia-Rivera, D.; Valdes-Balbin, Y.; Verez-Bencomo, V.; van Kooyk, Y. (2020, May 14). *Novel ACE2-independent carbohydrate-binding of SARS-CoV-2 spike protein to host lectins and lung microbiota*. Retrieved from <https://www.biorxiv.org/content/10.1101/2020.05.13.092478v1>
21. Berman, J.; Westbrook, Z.; Feng, G.; Gilliland, T.; Bhat, H.; Weissig, I.; Shindyalov, P.; Bourne. (n.d.). *Nucleic Acids Research*. Retrieved from <https://www.rcsb.org/>
22. Van Zundert, G.; Rodrigues, M.; Trellet, C.; Schmitz, P.; Kastitis, E.; Karaca, A.; Melquiond, M.; van Dijk, S.; de Vries, A. (n.d.). *Bonvin*. Retrieved from

<https://www.bonvinlab.org/>

23. Ellgaard, L; Bettendorff, P; Braun, D; Herrmann, T; Fiorito, F; Jelesarov, I; Güntert, P; Helenius, A; Wüthrich, K. (2002, September 27). *NMR Structures of 36 and 73-residue Fragments of the Calreticulin P-domain*. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0022283602008124?via%3Dihub#!>
24. Gobeil, S; Janowska, K; McDowell, S; Mansouri, K; Parks, R; Stalls, V; Kopp, M; Manne, K; Li, D; Wiehe, K; Saunders, K; Edwards, R; Korber, B; Haynes, B; Henderson, R; Acharya, P. (2021, August 6). *Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity*. Retrieved from <https://science.sciencemag.org/content/373/6555/eabi6226>
25. Gobeil, S; Janowska, K; McDowell, S; Mansouri, K; Parks, R; Stalls, V; Kopp, M; Manne, K; Saunders, K; Edwards, R; Haynes, B; Henderson, R; Acharya, P. (2021, March 11). *Effect of Natural Mutations of SARS-CoV-2 on Spike Structure, Conformation and Antigenicity*. Retrieved from <https://www.biorxiv.org/content/10.1101/2021.03.11.435037v2>
26. Dejnirattisai, W., Zhou, D., Supasa, P., Liu, C., Mentzer, A. J., Ginn, H. M., Zhao, Y., Duyvesteyn, H., Tuekprakhon, A., Nutalai, R., Wang, B., López-Camacho, C., Slon-Campos, J., Walter, T. S., Skelly, D., Costa Clemens, S. A., Naveca, F. G., Nascimento, V., Nascimento, F., Fernandes da Costa, C., ... Sreaton, G. (2021, May 27). *Antibody evasion by the P.1 strain of SARS-CoV-2*. Retrieved from <https://doi.org/10.1016/j.cell.2021.03.055>
27. Robetta. (n.d.). *Multiple Sequence Alignment*. Retrieved from <https://robetta.bakerlab.org/>
28. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. (2020, March 30). *Structural basis of receptor recognition by SARS-CoV-2*. Retrieved from <https://doi.org/10.1038/s41586-020-2179-y>
29. Van Zundert, G.; Rodrigues, M.; Trellet, C.; Schmitz, P.; Kastitis, E.; Karaca, A.; Melquiond, M.; van Dijk, S.; de Vries, A. (n.d.). *The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes*. Retrieved from <https://bianca.science.uu.nl/haddock2.4>
30. Clustal Omega. (n.d.). *Multiple sequence alignment*. Retrieved from <https://www.ebi.ac.uk/Tools/msa/clustalo/>
31. Bateman, A. (2008, November 5). *The UniProt Consortium UniProt: a*

worldwide hub of protein knowledge. Retrieved from
<https://academic.oup.com/nar/article/47/D1/D506/5160987>

32. Kastitis, P.; Moal, I.; Hwang, H.; Weng, Z.; Bates, P.; Bonvin, A.; Janin, J. (2011, January 6). *A structure-based benchmark for protein-protein binding affinity*. Retrieved from <https://doi.org/10.1002/pro.580>

33. Xue, L.; Rodrigues, J.; Kastitis, P.; Bonvin, A.; Vangone, A. (2015, July 20). *PRODIGY: a web-server for predicting the binding affinity in protein-protein complexes*. Retrieved from <https://elifesciences.org/articles/07454>

34. Alvarenga, A.; Teixeira, C.; Ruano, M.; Pereira, W. (2009, September 2). *Influence of temperature variations on the entropy and correlation of the Grey-level co-occurrence amtrak from B-Mode images*. Retrieved from
<https://www.sciencedirect.com/science/article/abs/pii/S0041624X09001036>

35. Calvo-Romero, María.; Lima-Rodríguez, Esther.; Pérez-Miranda, M.; Bureo-Dacal, P. (n.d.). *Low-grade and high-grade fever at presentation of acute pulmonary embolism*. Retrieved from
<https://journals.lww.com/bloodcoagulation/pages/articleviewer.aspx?year=2004&issue=06000&article=00007&type=abstract>