Chapter 4

EXPLORING NOVEL INTERACTIONS BETWEEN CHONDROITIN SULFATE AND THE EPHB3 RECEPTOR

Adam R. Griffith,[a,b,1] Claude J. Rogers,[b,c,1] Ravinder Abrol,[a,b] William A. Goddard III[a,ab,*]
Greg Miller,[b,c] and Linda C. Hsieh-Wilson,[b,c]
[a]Materials and Process Simulation Center,
California Institute of Technology, Pasadena, CA 91125
[b]Division of Chemistry and Chemical Engineering
California Institute of Technology, Pasadena, CA 91125,
[c]Howard Hughes Medical Institute,
[1] These two authors contributed equally
[*] To whom correspondence should be addressed; email wag@wag.caltech.edu

Introduction

Recent work in the Hsieh-Wilson group has identified chondroitin sulfate-E (CS-E) binds to the EphB3 receptor with high specificity and physiologically relevant affinity and is responsible for the direction of retinal neuron growth.[1] However, CS-A, which is less sulfated than CS-E, does not bind to EphB3, nor does it direct neuronal growth. Furthermore, while EphB2 shares ~60% sequence identity (extracellular region) with EphB3, experiments show that it does not bind any glycosaminoglycans (GAGs). We have previously applied computational methods (GAG-Dock) to similar GAG-protein systems with great success, both for systems with known crystal structures and for identifying novel interactions with the protein tyrosine phosphatase σ (RPTPs) and Nogo receptors (NgR).[2] We believe that predictions of the interaction between CS-E and EphB3 will be useful in studying and understanding the role of this interaction in neuron growth.

Using the GAG-Dock method, we explain the differential binding of CS ligands to the EphB3 and EphB2 receptors. Our results identify the previously unknown binding site for

CS-E on the EphB3 receptor and suggest experiments that can be used to validate our predictions.

GAG-Dock Overview

GAG-Dock[2] is a docking method based on the DarwinDock[3-11] and GenDock methodology that has been accommodated to work with large, highly charged, surface-binding ligands characteristic of GAGs. Because the binding sites for proteins that bind GAGs are typically not known, it is necessary to sample the entire surface of the protein. The surface of the protein is broken into regions, which are then evaluated using a "coarse" level of docking, which generates 10,000 ligand poses for each region. Based on the ranking of these regions by energy, a subset is docked to using a "fine" level of docking. The "fine" docking is carried out to a completeness threshold of 5%; however, due to the computational difficulty of these systems, a limit of 50,000 ligand poses is placed on the completeness.

Modifications to GAG-Dock

GAG-Dock is used almost identically to the way that it was used in our work on RPTPs and NgR. The key difference has to do with the way in which regions of the protein were sampled. The extracellular domains of EphB3 and EphB2 are very large and the location of the CS binding site was not previously known. It was therefore necessary to sample the majority of the protein surface. As in our previous work, spheres were generated that cover the entire surface of the protein. These spheres were divided into overlapping boxes/regions, however at a smaller size: 15Å/side (instead of 20Å) with 3Å overlap

(instead of 5Å). This was done to reduce the computational cost of working with an octasaccharide ligand. However, this resulted in an excessive number of boxes to test. Knowing that the CS binding site must be positively charged to match the negative charge of the ligand, we used electrostatics to eliminate most of the sphere regions. Specifically, we calculated the electrostatic potential for the proteins (Fig. 4-1A) using the Adaptive Poisson-Boltzmann Solver (APBS[12-14]) method and mapped the potential onto the spheres. Regions were ordered based on the number of positively charged spheres, and the 25% with the largest number of positive spheres were kept for docking. This resulted in 45 regions (238-1108 positive spheres) for EphB3 and 47 regions (180-1211 positive spheres) for EphB2. (Fig. S4-11)

All other parts of the GAG-Dock procedure were the same. Because CS ligands of sufficient size had already been prepared for our prior work, we used the same CS-A, CS-D, and CS-E octasaccharides.

"Coarse" docking was applied using CS-A, CS-D, and CS-E to the 45 EphB3 regions and 47 EphB2 regions. The top 13 EphB3 regions for CS-E binding were reexamined using "fine" docking for CS-A, CS-D, and CS-E.

EphB2 and EphB3 Models

Because no crystal structures of the full EphB3 or EphB2 extracellular regions exist, it was necessary to use homology modeling to generate the protein structures. EphB3 (PDB: 3P1I[15]) and EphB2 (PDB: 2QBX[16]) ephrin ligand binding domain crystal structures were

used with a crystal structure of the EphA2 ectodomain (PDB: 2X11[17]) to generate the homology models for EphB3 and EphB2 using SWISS-MODEL[18-21].

The human EphB2 model was constructed for the sequence corresponding to protein residues 20-529 by using the 2.3 Å resolution structure for human EphB2 (PDB: 2QBX) for protein residues 20-194 and combining it with a homology structure for residues 195-529 based on a lower resolution (4.3 Å) human Ephrin type-A receptor 2 (EphA2) structure (PDB: 2X11). This required aligning 2QBX structure to the full 2X11 homology structure and extracting residues 195-529 to attach to 2QBX structure. This was followed by minimizing hinge residues 192-197 using the DREIDING[22] force field in MPSIM[23] while keeping all other residues fixed and then minimizing all the residues.

The human EphB3 model was constructed for the sequence corresponding to protein residues 39-544 by using the 2.1 Å resolution structure for human EphB3 (PDB: 3P1I) for protein residues 39-209 and combining it with a homology structure for residues 210-544 based on human Ephrin type-A receptor 2 (EphA2) structure (PDB: 2X11). This required aligning 3P1I structure to the full 2X11 homology structure and extracting residues 210-544 to attach to 3P1I structure. This was followed by minimizing hinge residues 207-212 using the DREIDING force field in MPSim, while keeping all other residues fixed and then minimizing all the residues. A schematic of the domains present in our EphB2 and EphB3 models is shown in Fig. 4-1A.

Results

We observed significant differences in the amount and placement of positive charge on the electrostatic potential surfaces of the EphB2 and EphB3 models (Fig. 4-1B, S4-11). Since CS-E is a highly sulfated GAG, we expected this to provide a structural basis for the selectivity of CS-E toward EphB3. This was verified by the GAG-Dock predictions for the CS-E octa-saccharide bound to each of the two proteins. We found from coarse docking that CS-E bound to EphB3 (–345 kcal/mol) more strongly than to EphB2 (–119 kcal/mol). We also docked two other CS octa-saccharides, CS-A and CS-D, to EphB3 and EphB2. The binding energies from coarse binding for these ligands also indicated better binding to EphB3 than to EphB2 (Fig. 4-2).

Comparisons of the binding energies from fine docking of the three CS octasaccharides (Fig. 4-3) showed that CS-E bound strongly to EphB3 (–381 kcal/mol) while CS-A did not (–280 kcal/mol). This is in agreement with experimental results for CS-E and CS-A binding to EphB3 found by the Hsieh-Wilson group. In our calculations CS-D (–374 kcal/mol) bound comparably to CS-E; however, there are no experimental results for CS-D binding as it is difficult to obtain pure molecules of CS-D for ligand binding experiments.

Overall GAG-Dock predicts binding sites and energies that correspond well with the known experimental data for CS binding to EphB2 and EphB3. The predicted CS-E binding region on EphB3 contains eight arginines (R309, R344, R363, R391, R408, R420, R440, and R478) as well as two lysines (K378, K434). However, no single binding pose can access more than six of these attractive positive residues. Furthermore, distinct binding

motifs were apparent in the docking output. Therefore, for CS-E bound to EphB3, we identified five different binding motifs (Modes 1-5), all in the area of the first fibronectin III domain (Fig 4-1E, S4-10). Modes 1 and 2 (Fig 4-1C, 4-1D) are predicted to have comparable binding energy (-377.5 kcal/mol and -381 kcal/mol, respectively) and each is found 10 times in the best 25 poses making them the most likely candidates for the actual ligand binding site. Detailed images for Mode 1 and Mode 2 are shown in Fig. 4-1D/E. Mode 3 has an energy of -380.5 kcal/mol, making it comparable in energy to Modes 1 and 2, but is only represented by 3 poses. Modes 4 and 5 each have one pose, with energies of -351.9 kcal/mol, and -318.7 kcal/mol, respectively. Given the presence of multiple competing binding sites and the inability of any one pose to interact with all of the charged residues in the region, it is possible that more than one position of the ligand is biologically relevant. Furthermore, it is possible that the less represented binding modes might leave available charged sites that could allow dimerization of two EphB3 proteins, which is a possible mechanism for activation.

Per-residue nonbond energies for each of the five binding modes is shown in Table S4-1. Table S4-2 focuses on the arginine and lysine residues in the binding sites and clearly shows that, while each of these charge residues contributes to the binding energy, the pattern of interactions with these residues differs between the binding modes.

While the best pose from Mode 1 is ~3 kcal/mol worse in energy than the best pose from Mode 2, the poses from Mode 1 are very consistent in their placement (Fig S4-14) and make very good contact with the protein (Fig S4-13). A detailed image of the best Mode 1

pose is shown in Fig S4-12. The poses for Mode 2 are less consistent in their placement (Fig S4-15B). The best pose for Mode 2 (detailed, Fig S4-15A) shows that while the mode generally fits to the protein well (Fig S4-15C), the middle part of the ligand loses contact with the protein (Fig S4-15D). Similar analysis for Modes 3-5 are shown in Fig S4-16 – S4-20. Pharmacophores for all five binding modes are shown in Fig 4-4 – 4-8.

Suggested Post-Prediction Validations

To provide a means for experimentally validating our novel CS-E/EphB3 binding site, we propose several targeted mutations of key residues involved in CS-E binding. The most significant contributions for ligand binding come from eight arginines (R309, R344, R363, R391, R408, R420, R440, R478) and two lysines (K378, K434) in the binding site, as expected for a highly negatively charged ligand. (Table S4-2) We suggest that mutation of these residues to glutamine (or asparagine) should dramatically reduce the binding while minimizing the risk of large structural changes that more severe mutations (e.g., to alanine) could cause. Our methodology in determining suggested mutations is described in the supplemental information.

The differences in the orientations of our five predicted binding modes suggests that specific residues may play a larger role in binding, leaving others to play a lesser role. Since Modes 1 and 2 represented 80% of the top 25 poses, we will focus our results on mutations for these two modes. While all ten positively charged residues contributed to the overall binding energies, the strongest five contributions for Mode 1 were R440 (-174.5 kcal/mol), R363 (-137.7 kcal/mol), R309 (-128.0 kcal/mol), K434 (-125.7 kcal/mol), and

R344 (-120.9 kcal/mol). The strongest six contributions for Mode 2 were R440 (-167.1 kcal/mol), K434 (-142.9 kcal/mol), K378 (-130.6 kcal/mol), R363 (-120.6 kcal/mol), R420 (-115.5 kcal/mol), and R309 (-98.4 kcal/mol). R440 was the strongest contributor for both Mode 1 and Mode 2, suggesting that it should be the first target for specific mutations. R309 and R344 both contributed more strongly to Mode 1, and K378 and R420 both contributed more strongly to Mode 2. Mutation of these residues may be able to provide experimental evidence for which Mode is best. Since R391, R408, and R478 did not contribute strongly to either Mode 1 or Mode 2 mutations of these residues could provide experimental information on whether Modes 3-5 are relevant. Contributions for all residues are presented in Table S4-1.

We carried out *in silico* mutations of the key Mode 1 residues to glutamine, which led to the following changes to the binding energy (positive indicates weaker interactions): R440Q +165.9 kcal/mol, R363Q +131.0 kcal/mol, R309Q +122.7 kcal/mol, R344Q +120.6 kcal/mol, K434Q +114.5 kcal/mol. For Mode 2 the changes to binding energy were: R440Q +160.7 kcal/mol, K434Q +133.8 kcal/mol, K378Q +111.2 kcal/mol, R363Q +100.4 kcal/mol, R420Q +95.2 kcal/mol, R309Q +94.0 kcal/mol.

We recommend that numerous simultaneous mutations be done for tests of our predictions. The reason is that because a large number of charged residues contribute to the binding, mutation to a single residue may be insufficient to significantly alter binding. Moreover since other positive residues are available in the same regions, the ligand might find new interactions in the absence of just one or two key residues.

A more rigorous validation of our predicted binding modes would be to perform mutations that unambiguously increase binding affinity. Consequently we identified mutations of several residues that GAG-Dock suggests should increase binding affinity. We selected these mutations to allow additional contacts with the charged and polar groups on CS-E. Again we considered mutations to glutamine, since the mutated structures may be more likely to fold to the proper structure, than say mutations to alanine. Eight individual mutations for Mode 1 predicted to make new contacts with the ligand are (negative indicates stronger binding): T448Q (-18.77 kcal/mol), V339Q (-13.65 kcal/mol), I446Q (-12.48 kcal/mol), A442Q (-11.97 kcal/mol), N445Q (-11.42 kcal/mol), T319Q (-11.20 kcal/mol), N323Q (-4.23 kcal/mol), and N322Q (-0.78 kcal/mol). Seven individual mutations for mode 2 predicted to make new contacts are: E424Q (-117.08 kcal/mol), V339Q (-16.14 kcal/mol), T422Q (-16.14 kcal/mol), T338Q (-14.13 kcal/mol), N445Q (-7.44 kcal/mol), N323Q (-2.97 kcal/mol), and S341Q (-0.92 kcal/mol). These single residue mutations are summarized in Table S4-3 for Mode 1 and Table S4-5 for Mode 2. (Modes 3-5 in Tables S4-7, -8, -10)

Based on their individual predicted contributions to binding, we suggest the following set of 7 mutations for the first experiments to test Mode 1: T319Q, N322Q, V339Q, A442Q, A443N, I446Q, and T448Q. We predicted that this set of mutations for Mode 1 improves binding energy by 66.03 kcal/mol, or 16.5% better than binding to the wild-type. The predicted binding site for this set of mutations for Mode 1 is shown in Fig 4-9A. Energies for all sets of mutations tested for Mode 1 are shown in Table S4-4.

The presence of E424 in the neighborhood of Mode 2 is puzzling, since it has a repulsive interaction with the ligand. Mutating E424 to glutamine resulted in a significant increase in binding energy, but might also modify the binding site. Therefore, we propose two sets of mutations for Mode 2. The first set is: N323Q, T338N, V339Q, S341Q, T422Q, and N445Q. This improved binding energy by 46.64 kcal/mol or 11.4% better than the wild-type. The second set for mode 2 adds the E424Q mutant, resulting in an improvement of 163.84 kcal/mol or 40%. The predicted binding site for the non-E424Q set for Mode 2 is shown in Fig 4-9B. Energies for all sets of mutations tested for Mode 2 are shown in Table S4-6.

We applied this same procedure also to Modes 3-5, with the results reported in Tables S4-7, S4-9, and S4-11. Detailed images and pharmacophores of the predicted mutant binding sites for the selected sets of mutations are shown in Figures S4-21 to S4-28.

Since we found five competitive binding modes for CS-E/EphB3, it may be that CS-E binding recognizes a binding region or ensemble of binding sites rather than a specific binding site that is typical for binding of small molecules. We selected CS-E octasaccharide as a representative of the natural, extended polysaccharide. The experimental system may well be more complicated with interactions beyond a single octasaccharide binding mode. Indeed none of our five predicted binding modes interacts with all 10 positively charged residues within the binding region. We suggest that these additional charged residues may serve two purposes. First, the extra, non-shared residues could allow for a single polysaccharide to bind to two proteins using one mode for the first

protein and a different mode for the second protein, possibly allowing for dimerization and activation of the proteins. Second, the presence of extra positive residues could allow for the ligand to migrate within the binding region without losing adhesion to the protein. To test this second possibility we suggest mutations expected to increase binding affinity. A single mutation from arginine or lysine to glutamine or asparagine might not change the binding as much as we predict, because the CS-E might move its preferred binding region slightly to account for the reduced arginines. This suggests that validation be done with multiple simultaneous mutations. Of course, mutating multiple residues simultaneously may increase the likelihood of misfolding, rendering the study useless. For a single beneficial mutation, such misfolding is less likely, although the change in binding affinity may be less dramatic.

Conclusions

Studying the CS-E/EphB3 system computationally was a difficult challenge: a large, highly negatively charged ligand, and a protein with a completely unknown binding site. Furthermore, the related CS-A ligand was shown not to bind experimentally, and neither CS-A nor CS-E bound to the similar EphB2 protein. Our goal was to identify structural explanations for these differences. In both cases we successfully identified the cause to be related to the charges on the ligand and/or the protein. EphB2 lacks the positively-charged region of EphB3 and thus cannot bind the negatively-charged CS ligands. Similarly, the reduced negative charge of CS-A relative to CS-E means that it does not bind with sufficient strength to EphB3. The *pattern* of sulfation does not appear to be a significant factor, as CS-D binds comparably to CS-E. This is likely due to flexibility of the sulfate

groups on the ligand and the arginine and lysine sidechains on the protein. Specific patterns are not needed for such a general interaction.

We have further used our predicted structural information to suggest mutation experiments that would validate one or more of our binding modes for CS-E. As mentioned previously, we consider mutations from arginine/lysine to alanine to lack subtlety. Loss of binding from such mutations could be due to larger structural changes than simple binding site modification. Instead, we have suggested sets of mutations to *improve* binding, which would validate our binding modes with much less ambiguity. We encourage our experimental colleagues to attempt these sets of mutations:

- T319Q, N322Q, V339Q, A442Q, A443N, I446Q, and T448Q
- N323Q, T338N, V339Q, S341Q, T422Q, and N445Q (optionally E424Q)

The first set should increase binding affinity for CS-E if our predicted Mode 1 is the correct binding pose, and the second set should increase binding affinity for Mode 2.

This project highlights the role that computation can have in studying complicated biological systems, and in complementing and directing experiment. The specificity of the binding site predictions suggests clear follow-up experiments to further understanding of the role of CS-E in EphB3 activation, which, hopefully, will suggest new directions for computation.

Acknowledgements

ARG, RA, and WAG were supported from NIH (R01-NS071112, R01-NS073115, and R01-AI040567) and other funds donated to the Materials and Process Simulation Center. The computers used in this research were funded by grants (to WAG) from DURIP (Defense University Research Instrument Program) and from NSF (equipment part of the NSF-MRSEC-CSEM). CJR, GM, and LCH-W were supported by the National Institutes of Health (grant R01 GM084724 to L.C.H-W).

References

1. Rogers, C. J., Miller, G., Hsieh-Wilson, L. C. (to be published)
2. Griffith, A. R., Rogers, C. J., Miller, G., Abrol, R., Hsieh-Wilson, L. C., Goddard, W. A., III. (2016, to be published). Predicting glycosaminoglycan-surface protein interactions: Implications for studying axonal growth.
3. Scott, C. E., Ahn, K. H., Graf, S. T., William A Goddard, I., Kendall, D. A., & Abrol, R. (2016). Computational Prediction and Biochemical Analyses of New Inverse Agonists for the CB1 Receptor. *Journal of Chemical Information and Modeling*, *56*(1), 201–212. http://doi.org/10.1021/acs.jcim.5b00581
4. Li, Q., Kim, S.-K., Goddard, W. A., III, Chen, G., & Tan, H. (2015). Predicted Structures for Kappa Opioid G-Protein Coupled Receptor Bound to Selective Agonists. *Journal of Chemical Information and Modeling*, *55*(3), 614–627. http://doi.org/10.1021/ci500523z
5. Abrol, R., Trzaskowski, B., Goddard, W. A., III, Nesterov, A., Olave, I., & Irons, C. (2014). Ligand- and mutation-induced conformational selection in the CCR5 chemokine G protein-coupled receptor. *Proceedings of the National Academy of Sciences*, *111*(36), 13040–13045. http://doi.org/10.1073/pnas.1413216111
6. Bray, J. K., Abrol, R., Goddard, W. A., III, Trzaskowski, B., & Scott, C. E. (2014). SuperBiHelix method for predicting the pleiotropic ensemble of G-protein–coupled receptor conformations. *Proceedings of the National Academy of Sciences*, *111*(1), E72–E78. http://doi.org/10.1073/pnas.1321233111
7. Kim, S.-K., & Goddard, W. A. (2014). Predicted 3D structures of olfactory receptors with details of odorant binding to OR1G1. *Journal of Computer-Aided Molecular Design*, *28*(12), 1175–1190. http://doi.org/10.1007/s10822-014-9793-4
8. Kim, S.-K., Goddard, W. A., Yi, K. Y., Lee, B. H., Lim, C. J., & Trzaskowski, B. (2014). Predicted Ligands for the Human Urotensin-II G Protein-Coupled Receptor

with Some Experimental Validation. *ChemMedChem*, *9*(8), 1732–1743. http://doi.org/10.1002/cmdc.201402087
9. Tan, J., Abrol, R., Trzaskowski, B., & William A Goddard, I. (2012). 3D Structure Prediction of TAS2R38 Bitter Receptors Bound to Agonists Phenylthiocarbamide (PTC) and 6-n-Propylthiouracil (PROP). *Journal of Chemical Information and Modeling*, *52*(7), 1875–1885. http://doi.org/10.1021/ci300133a
10. Kim, S.-K., Fristrup, P., Abrol, R., & William A Goddard, I. (2011a). Structure-Based Prediction of Subtype Selectivity of Histamine H3 Receptor Selective Antagonists in Clinical Trials. *Journal of Chemical Information and Modeling*, *51*(12), 3262–3274. http://doi.org/10.1021/ci200435b
11. Kim, S.-K., Riley, L., Abrol, R., Jacobson, K. A., & Goddard, W. A. (2011b). Predicted structures of agonist and antagonist bound complexes of adenosine A3 receptor. *Proteins: Structure, Function, and Bioinformatics*, *79*(6), 1878–1897. http://doi.org/10.1002/prot.23012
12. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, *98*(18), 10037–10041. http://doi.org/10.1073/pnas.181342398
13. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, *35*(suppl 2), W522–W525. http://doi.org/10.1093/nar/gkm276
14. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*, *32*(suppl 2), W665–W667. http://doi.org/10.1093/nar/gkh381
15. Ligand binding domain of human ephrin type-B receptor 3. (2011) http://dx.doi.org/10.2210/pdb3p1i/pdb
16. Chrencik, J. E., Brooun, A., Recht, M. I., Nicola, G., Davis, L. K., Abagyan, R., et al. (2007). Three-dimensional Structure of the EphB2 Receptor in Complex with an Antagonistic Peptide Reveals a Novel Mode of Inhibition. *Journal of Biological Chemistry*, *282*(50), 36505–36513. http://doi.org/10.1074/jbc.M706340200
17. Seiradake, E., Harlos, K., Sutton, G., Aricescu, A. R., & Jones, E. Y. (2010). An extracellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nature Structural & Molecular Biology*, *17*(4), 398–402. http://doi.org/10.1038/nsmb.1782
18. Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, *22*(2), 195–201. http://doi.org/10.1093/bioinformatics/bti770
19. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(W1), gku340–W258. http://doi.org/10.1093/nar/gku340
20. Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical

perspective. *Electrophoresis*, *30*(S1), S162–S173. http://doi.org/10.1002/elps.200900140

21. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, *37*(suppl 1), D387–D392. http://doi.org/10.1093/nar/gkn750
22. Mayo, S. L., Olafson, B. D., & Goddard, W. A. (2002). DREIDING: a generic force field for molecular simulations. *The Journal of Physical Chemistry*, *94*(26), 8897–8909. http://doi.org/10.1021/j100389a010
23. Lim, K. T., Brunett, S., Iotov, M., & McClurg, R. B. (1997). Molecular dynamics for very large systems on massively parallel computers: the MPSim program. *Journal of Computational Chemistry*, *18*(4), 501-521
24. Kam, V. W. T., & William A Goddard, I. (2008). Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements. *Journal of Chemical Theory and Computation*, *4*(12), 2160–2169. http://doi.org/10.1021/ct800196k

Figures & Tables

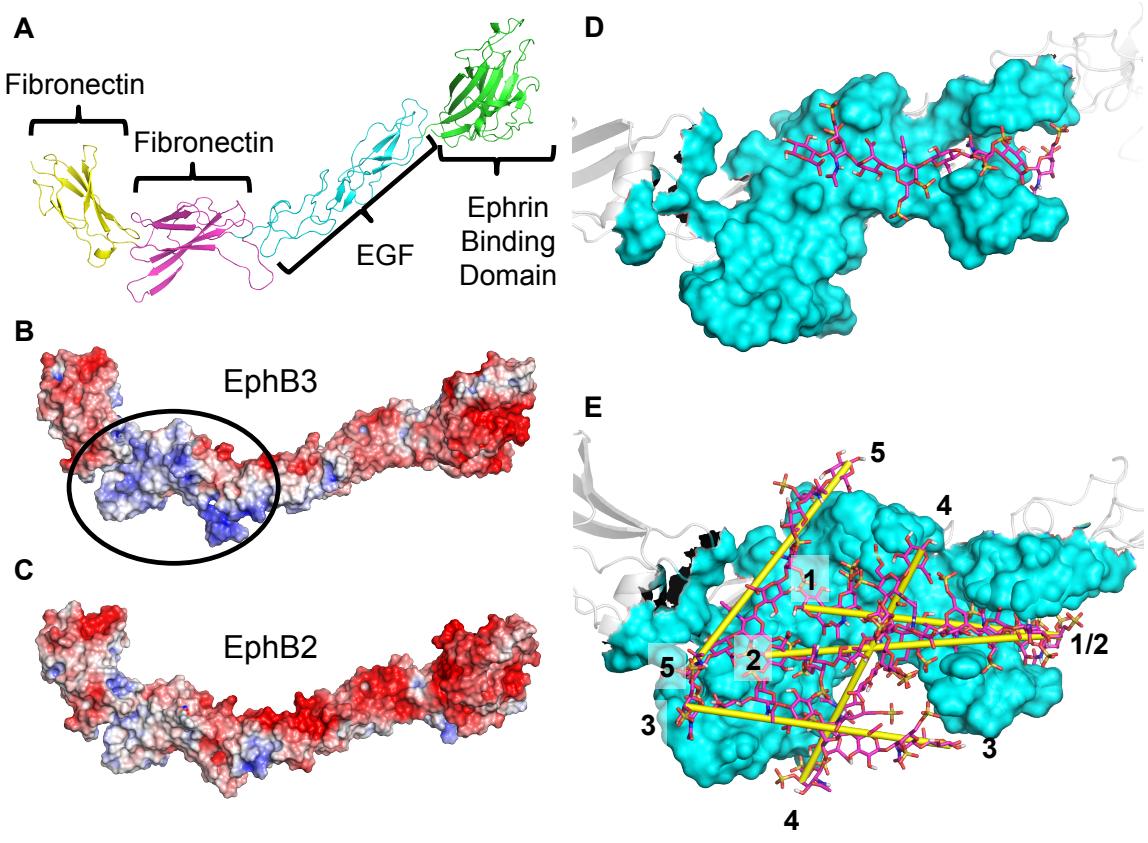

Figure 4-1 – (A) Model of EphB3. (B-C) Electrostatics mapped onto the surfaces of EphB3 and EphB2. Circled region denotes binding region for top five EphB3/CS-E binding modes (cyan region in D-E). (D) Predicted best EphB3/CS-E binding mode. (E) Overlay of predicted Top five EphB3/CS-E binding modes. The general orientation of binding modes shown in yellow.

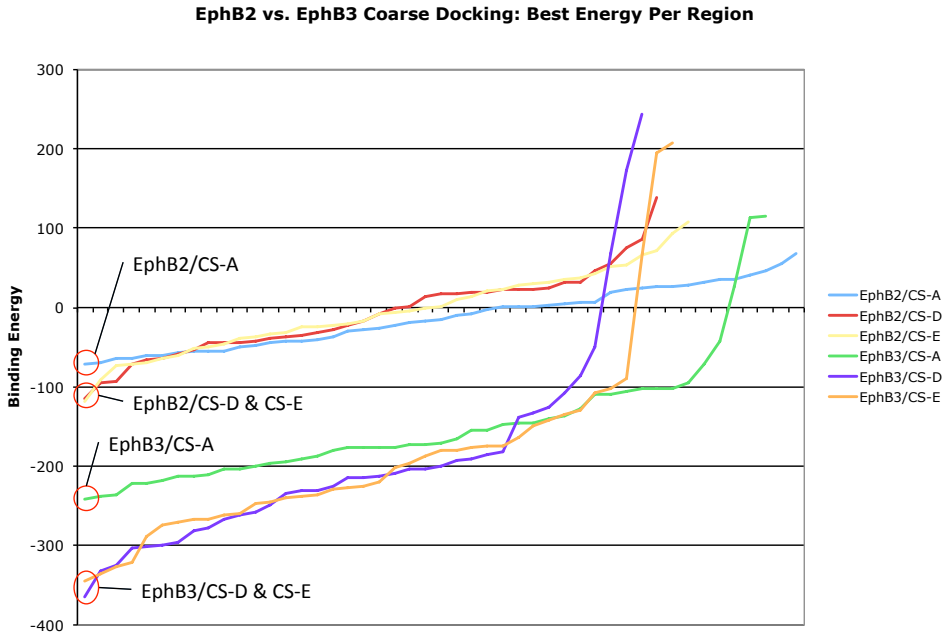

Figure 4-2 –Plot of the energy of the best pose in each region after coarse docking for CS-A, CS-D, and CS-E docked to EphB2 and EphB3. It is clear from the chart that the binding energies are much worse for EphB2 than EphB3. Additionally, CS-A has a much worse binding energy to EphB3 than CS-D and CS-E.

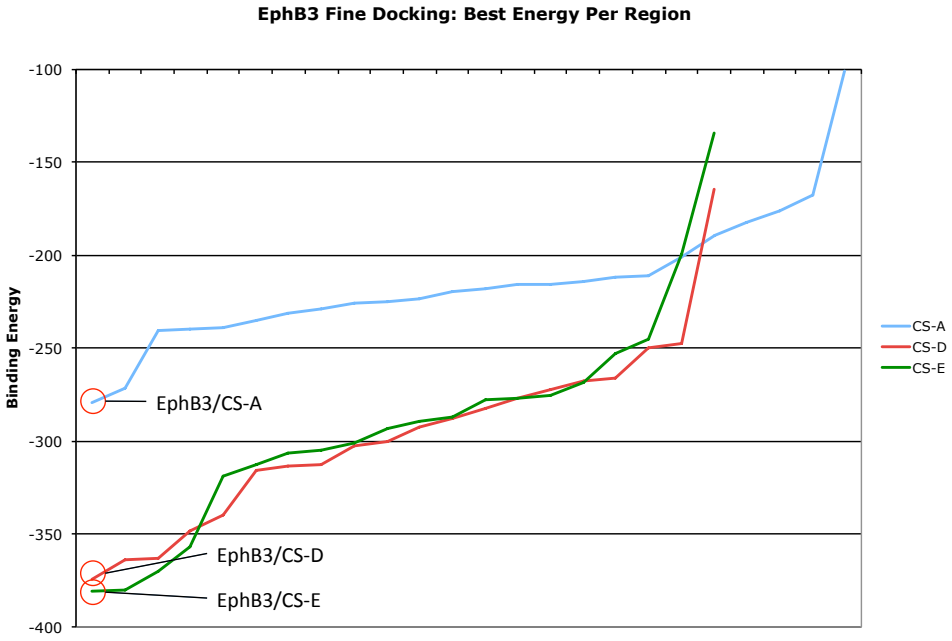

Figure 4-3 – Plot of the energy of the best pose in each region after fine docking for CS-A, CS-D, and CS-E docked to EphB3. After fine-level docking, CS-E binds slightly better than CS-D, and both bind significantly better than CS-A.

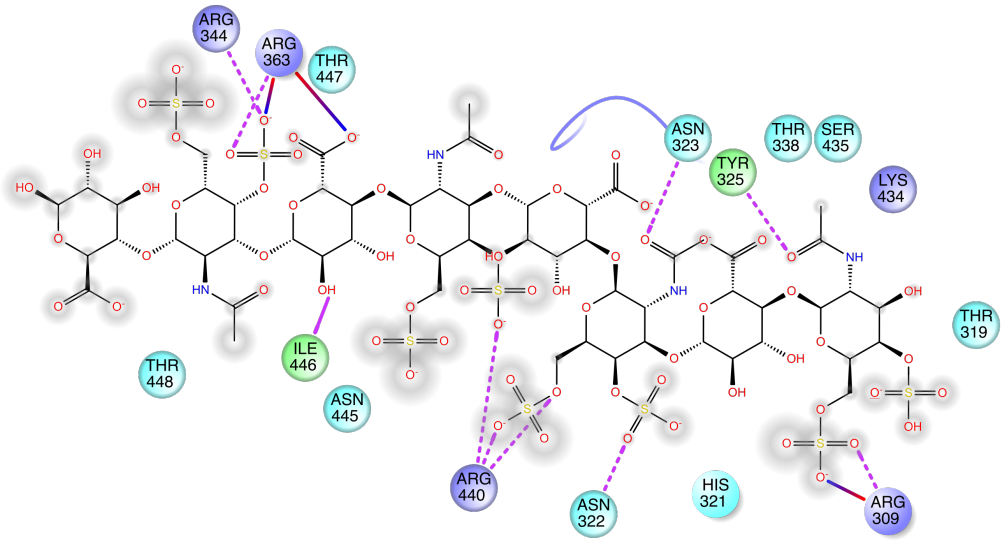

Figure 4-4 – Pharmacophore for best pose in EphB3/CS-E mode 1.

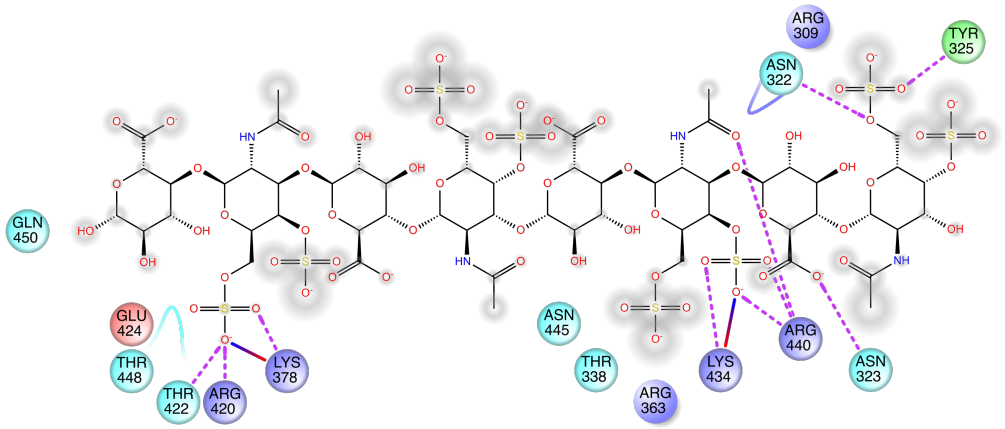

Figure 4-5 – Pharmacophore for best pose in EphB3/CS-E mode 2.

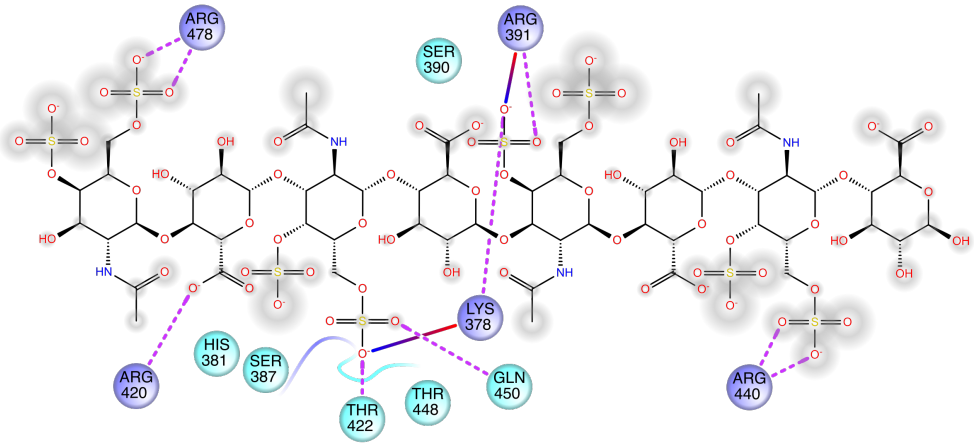

Figure 4-6 – Pharmacophore for best pose in EphB3/CS-E mode 3.

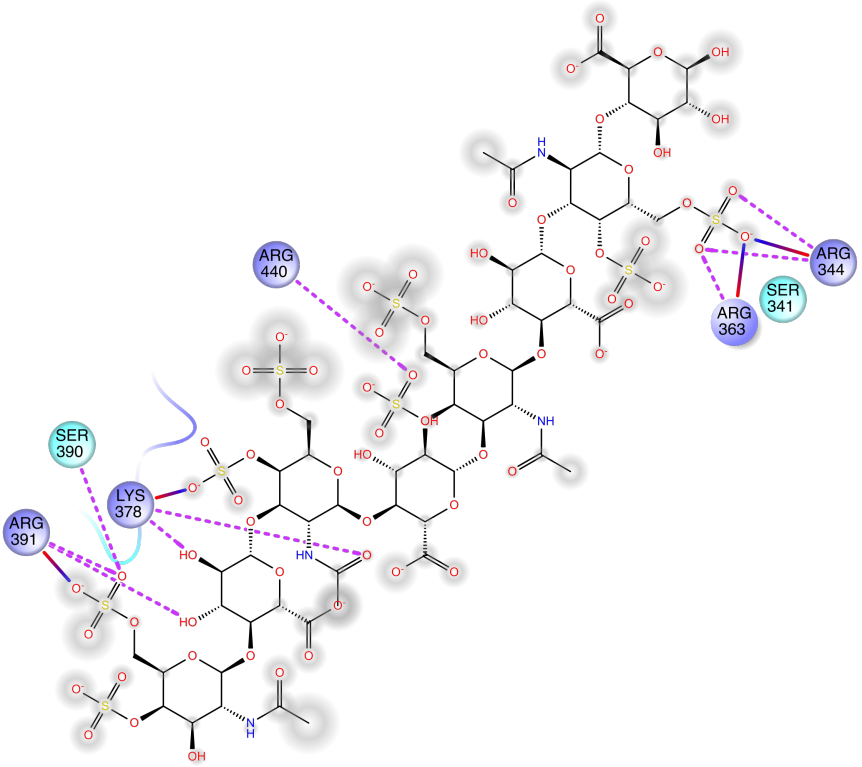

Figure 4-7 – Pharmacophore for best pose in EphB3/CS-E mode 4.

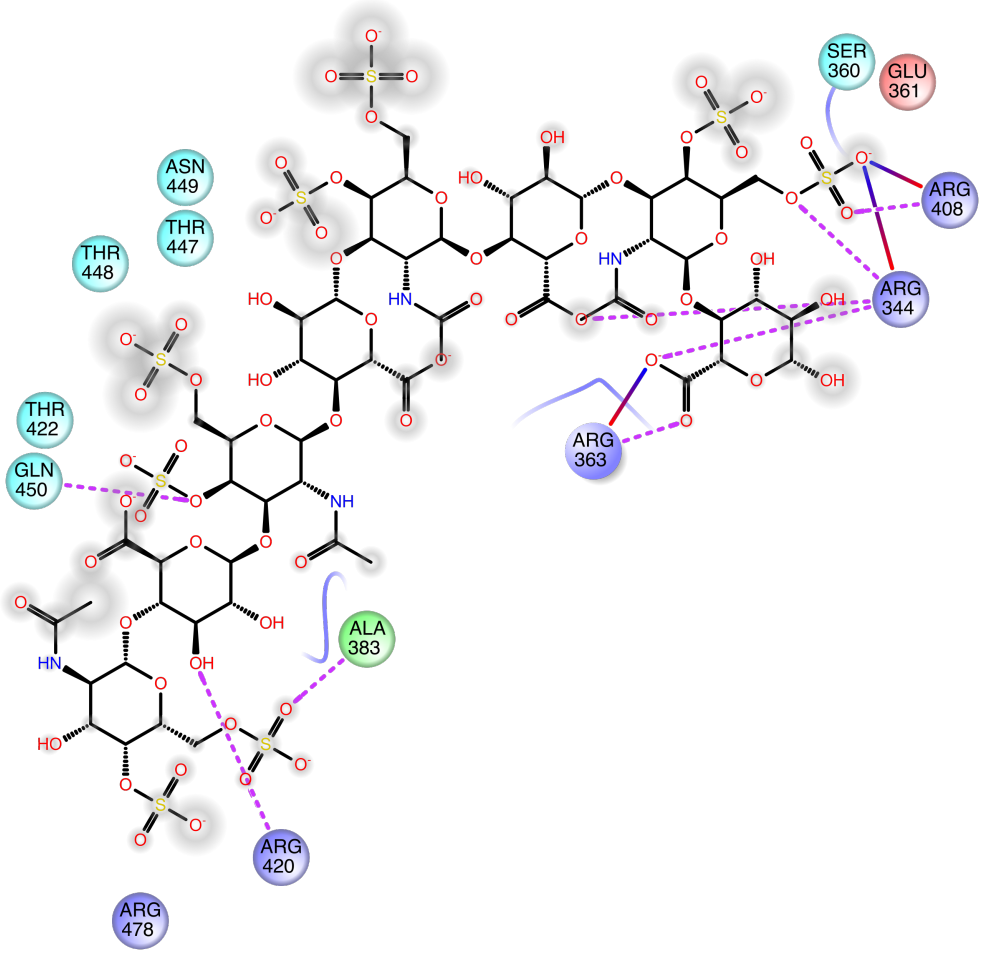

Figure 4-8 – Pharmacophore for best pose in EphB3/CS-E mode 5.

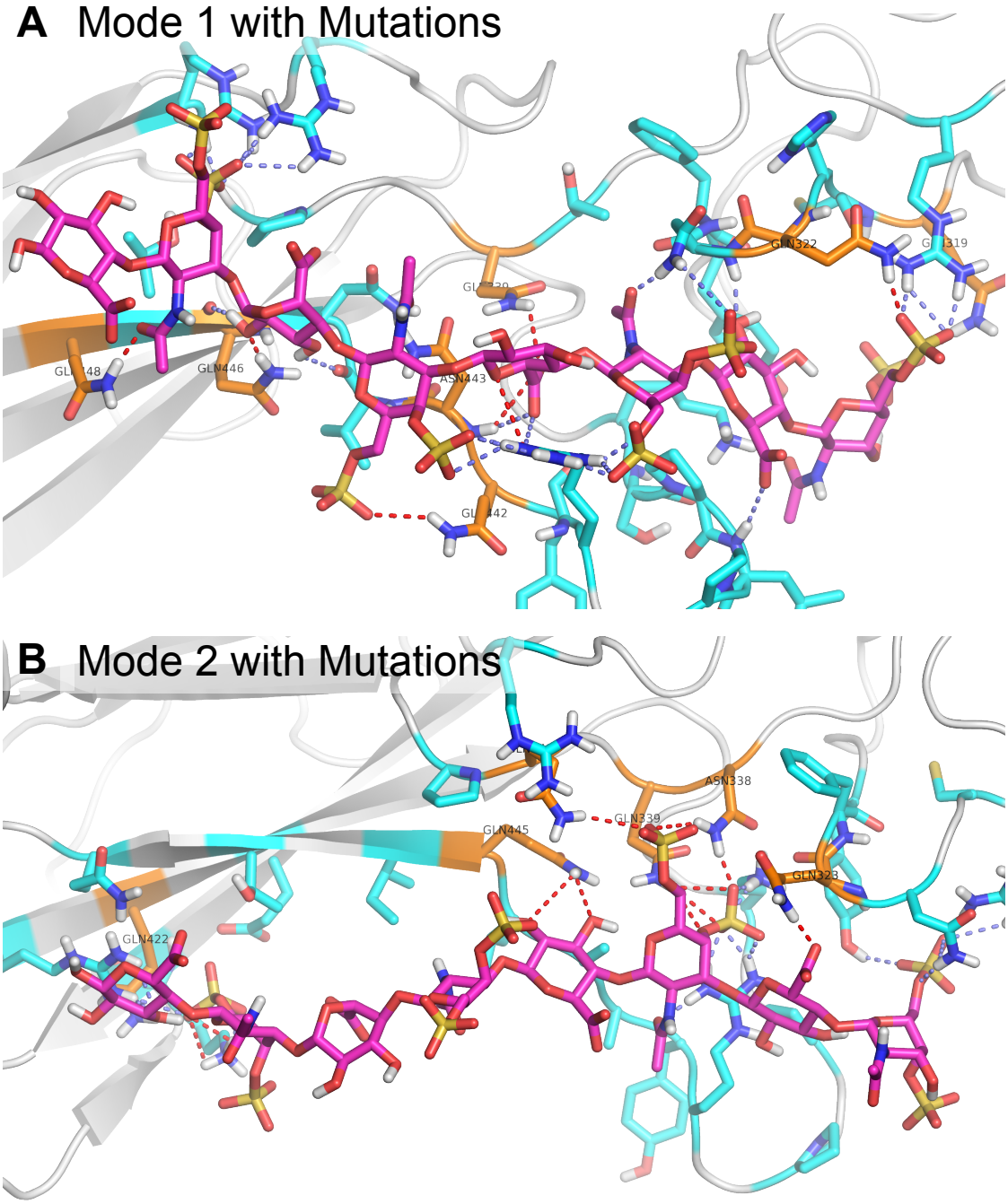

Figure 4-9 – Mutations to Gln predicted to increase EphB3/CS-E binding. Mutated residues are colored orange. Red hydrogen bond markers denote new hydrogen bonds with the ligand due to mutations and blue markers denote hydrogen bonds to the ligand that are common to both mutant and wild type. (A) Mutations for binding mode 1: T448Q, V339Q, I446Q, A442Q, T319Q, A443N, N322Q. Binding energy improved by 66.0 kcal/mol or 16.5% over wild type. (B) Mutations for binding mode 2: V339Q, T422Q, T338N, N445Q, N323Q, S341Q. Binding energy improved by 46.6 kcal/mol or 11.4% over wild type.

Supplemental Information

Mutation Methodology

In order to identify mutations that could validate our predicted CS-E/EphB3 binding modes we performed *in silico* mutations. Each residue – excluding proline and glycine – in the 5Å binding site was individually mutated to glutamine using SCREAM[24]. Simultaneously, the rest of the sidechains were also optimized to allow them to accommodate the mutated sidechain's position. The binding site and ligand were then minimized for 50 steps of conjugate gradient minimization using DREIDING[22] in MPSim[23]. At the end of this procedure mutations were identified that increased the binding energy of the ligand, summarized in Tables S4-3, S4-5, S4-7, S4-8, and S4-10. Based on these single mutants, sets of combined mutants that should increase binding were identified and tested. Again, SCREAM was used to perform the mutations as well as optimize the remaining sidechains in the binding site, followed by 50 steps of minimization. In some cases two mutant sidechains would clash, resulting in non-optimal interactions with the ligand. Thus additional sets that omitted some mutations were tested. Additionally, glutamine proved to be too large to make a good interaction with the ligand in some cases, thus asparagine was tried instead. In the end, one set of mutants was identified for each mode that maximized ligand binding and resulted in each mutated residue making a new hydrogen bond with the ligand. An additional set was generated each for Mode 2 and Mode 5. These modes have nearby glutamic acids (E424 and E361, respectively). We are wary of mutating these residues because they may have a special role in the structure or function of the EphB3 receptor or binding site. However, sets of mutations were generated for Mode 2 and Mode

5 that included the respective E424Q and E361Q mutations. The binding energies for the sets of mutations are summarized in Tables S4-4, S4-6, S4-7, S4-9, and S4-11.

Supplemental Figures & Tables

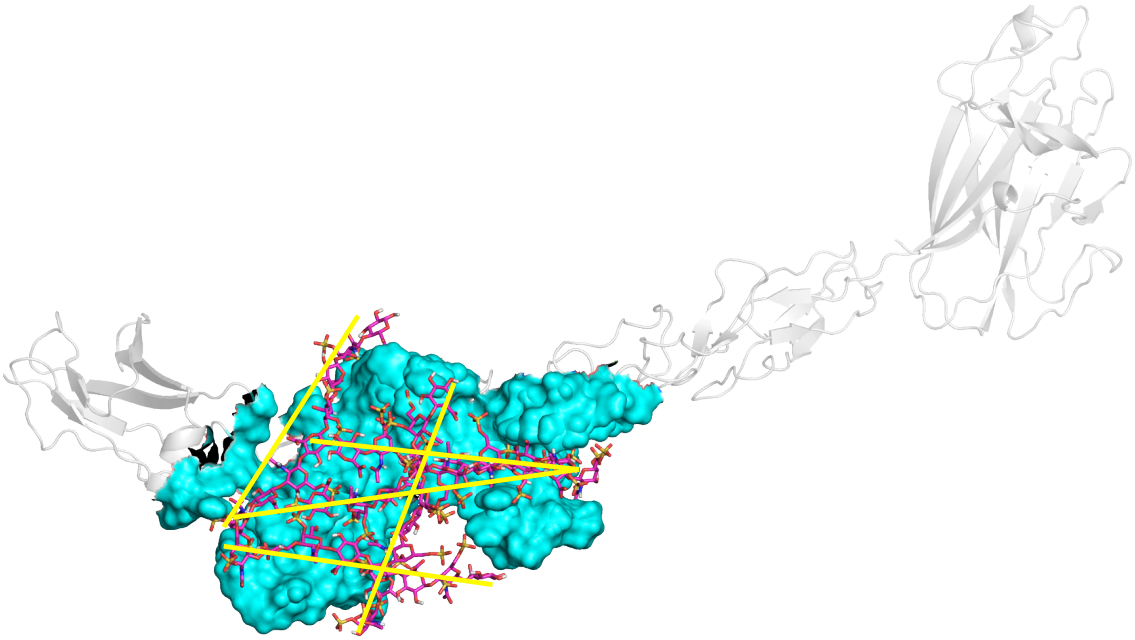

Figure S4-10 – Schematic showing placement of CS-E binding modes bound to EphB3.

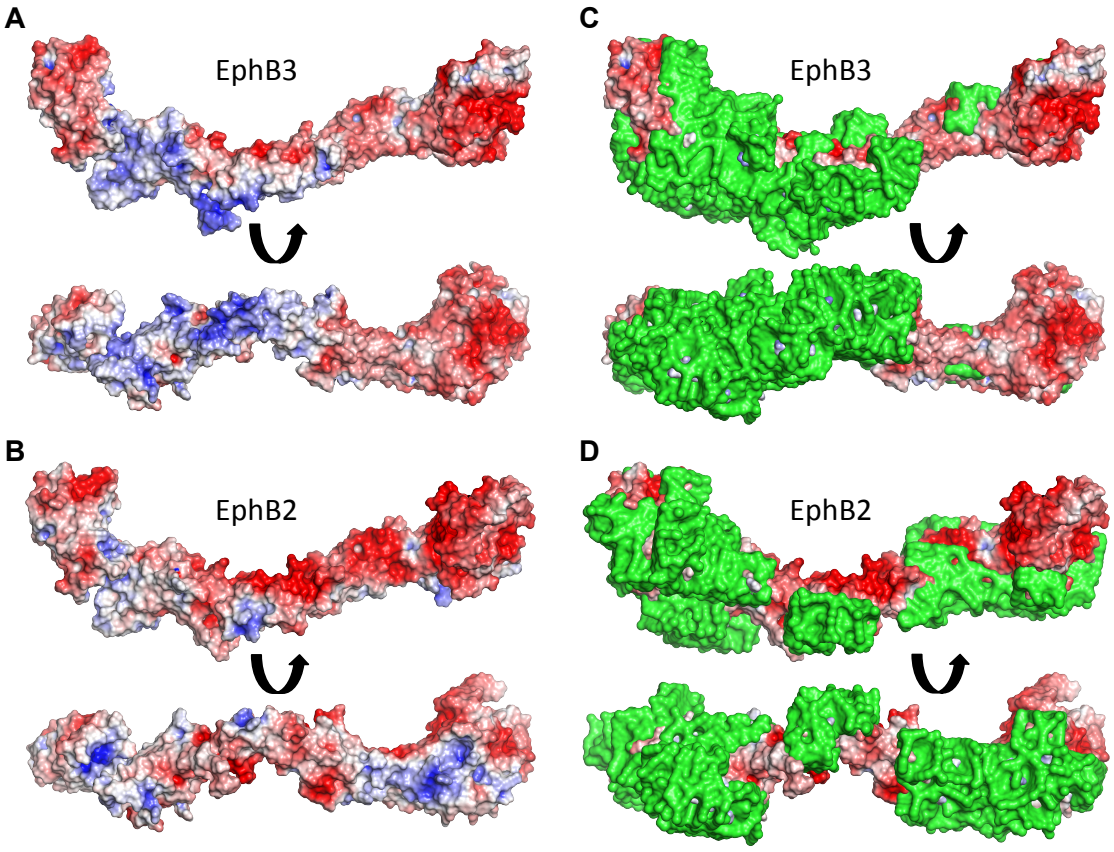

Figure S4-11 – Electrostatic surfaces of (A) EphB3 and (B) EphB2. Sphere regions used for coarse docking are shown in green for (C) EphB3 and (D) EphB2. Note that the regions sampled cover the positively charged regions of the proteins.

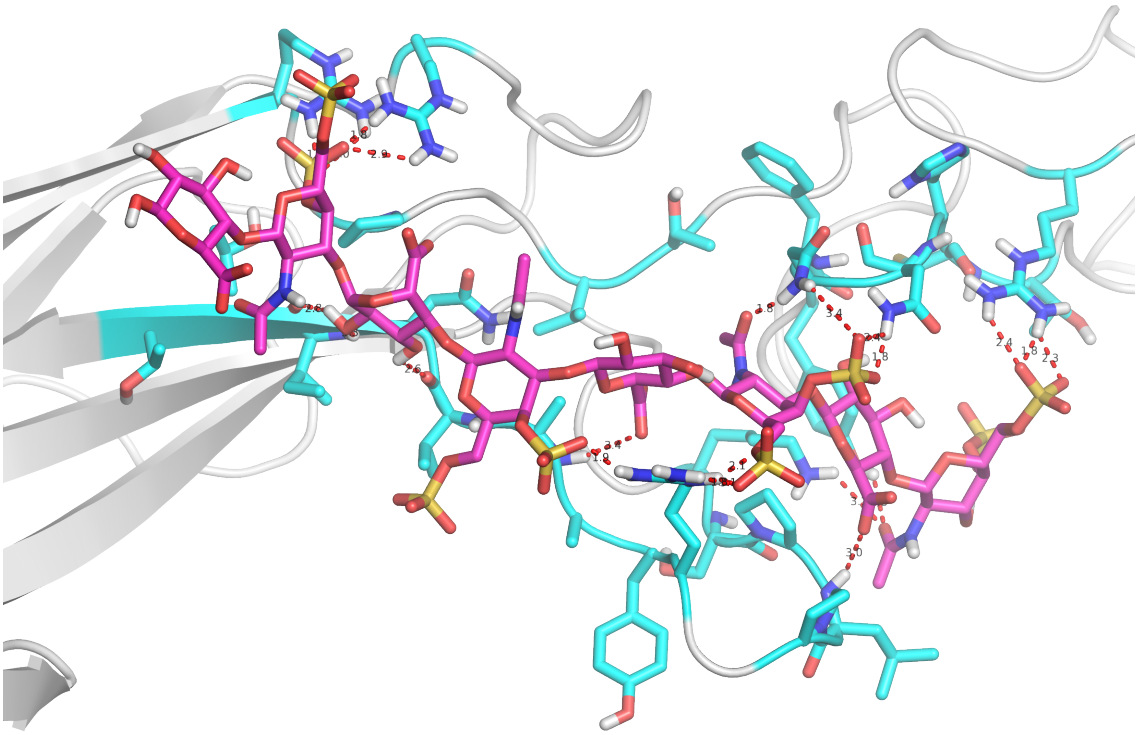

Figure S4-12 – Detailed view of the best CS-E/EphB3 Mode 1 binding pose.

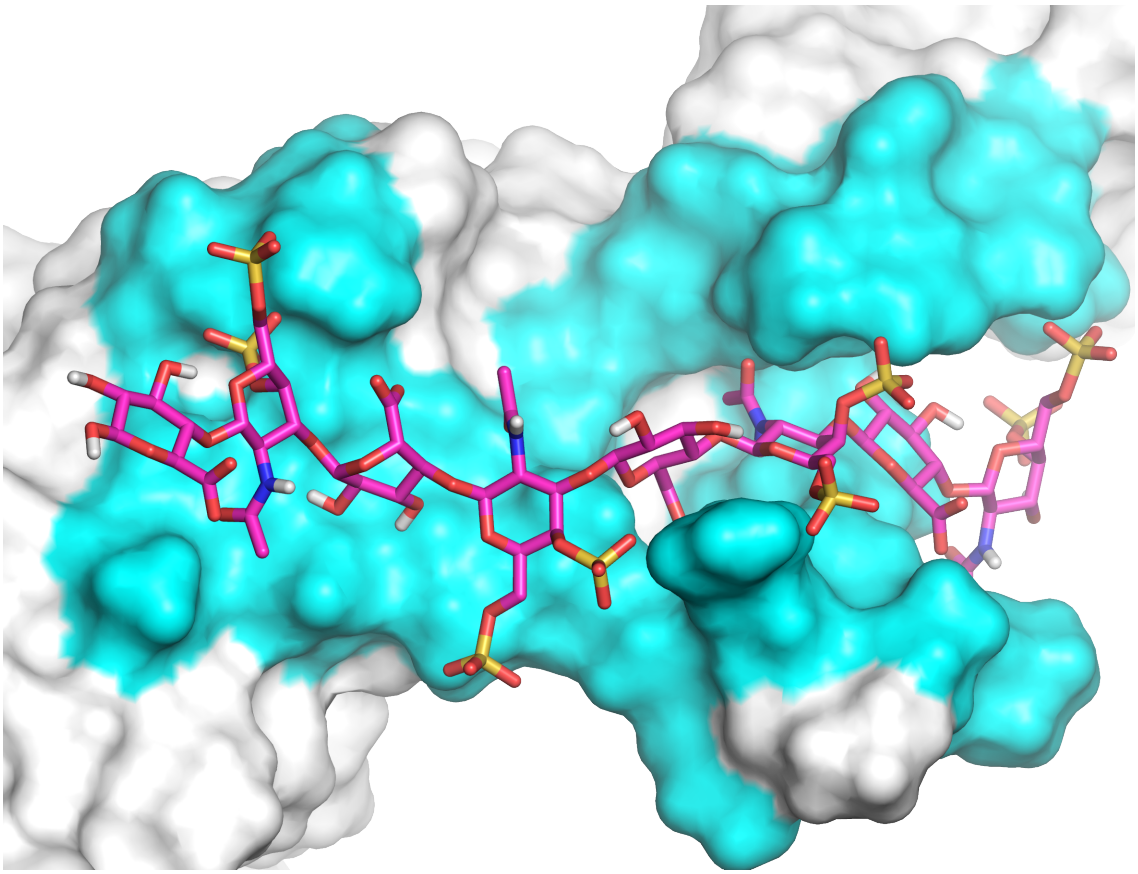

Figure S4-13 - The best CS-E/EphB3 Mode 1 binding pose, with the VDW surface of the protein shown to illustrate how well the ligand fits to the protein.

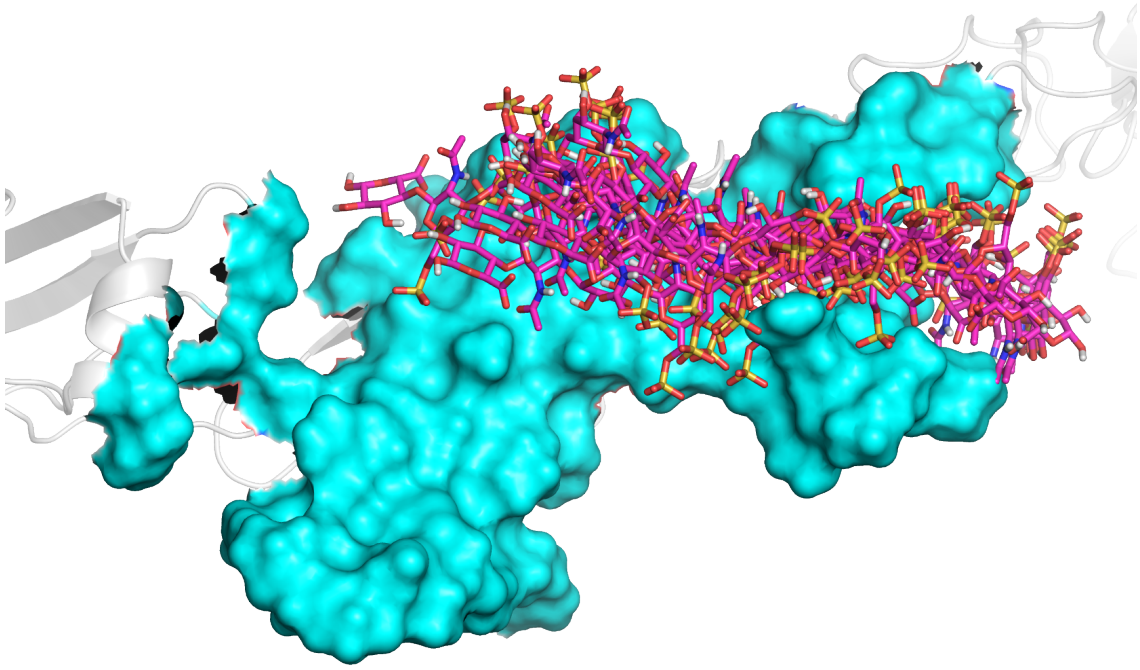

Figure S4-14 – The placement of all CS-E/EphB3 Mode 1 poses. The top pose in this mode is the #3 pose overall (-377.46 kcal/mol), but this mode shows the most consistency in placement.

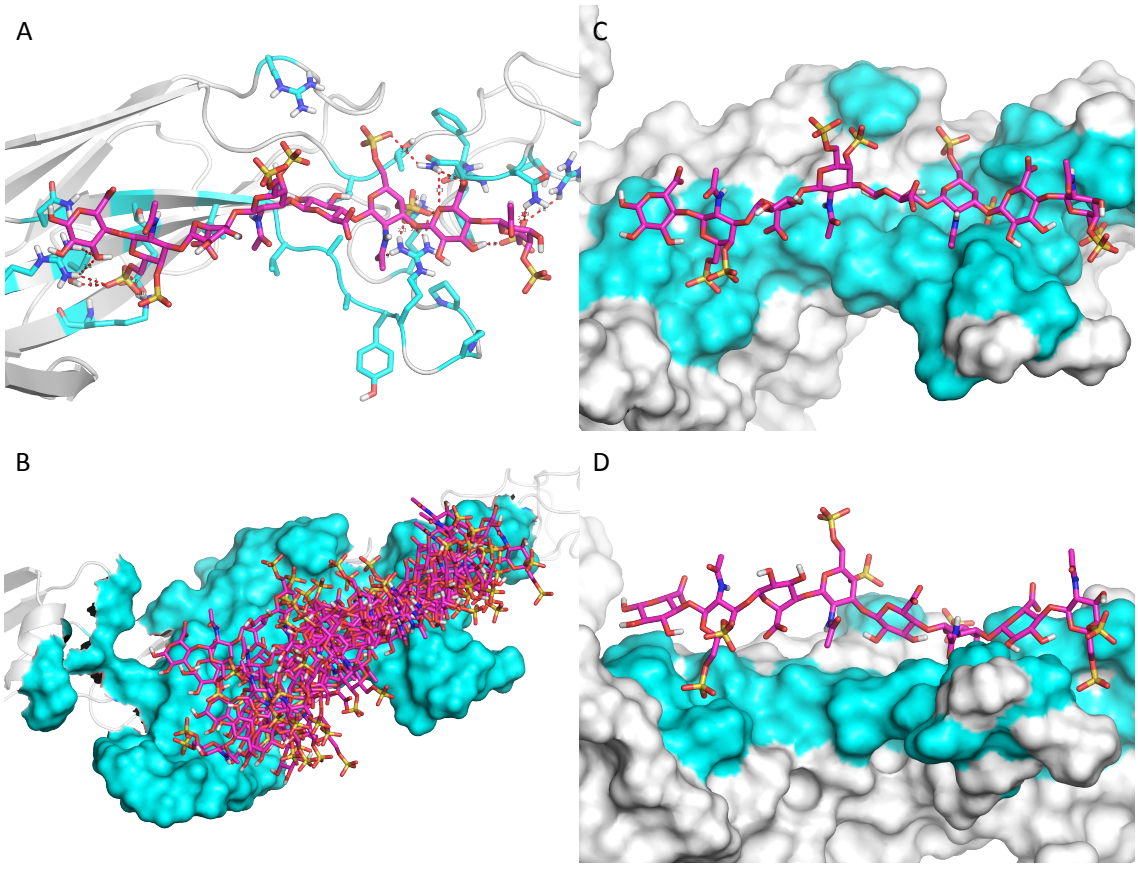

Figure S4-15 – (A) Detailed view of CS-E docked to EphB3 in top pose from binding mode 2. (B) Placement of all CS-E poses docked to EphB3 in binding mode 2. The top pose in this mode is the #1 pose overall (-380.80 kcal/mol), but shows less consistency in pose placement than Mode 1. (C) Top view of the best Mode 2 pose appears to fit closely to the protein surface, but the rotated view (D) shows that the middle section of the octasaccharide is separated from the surface.

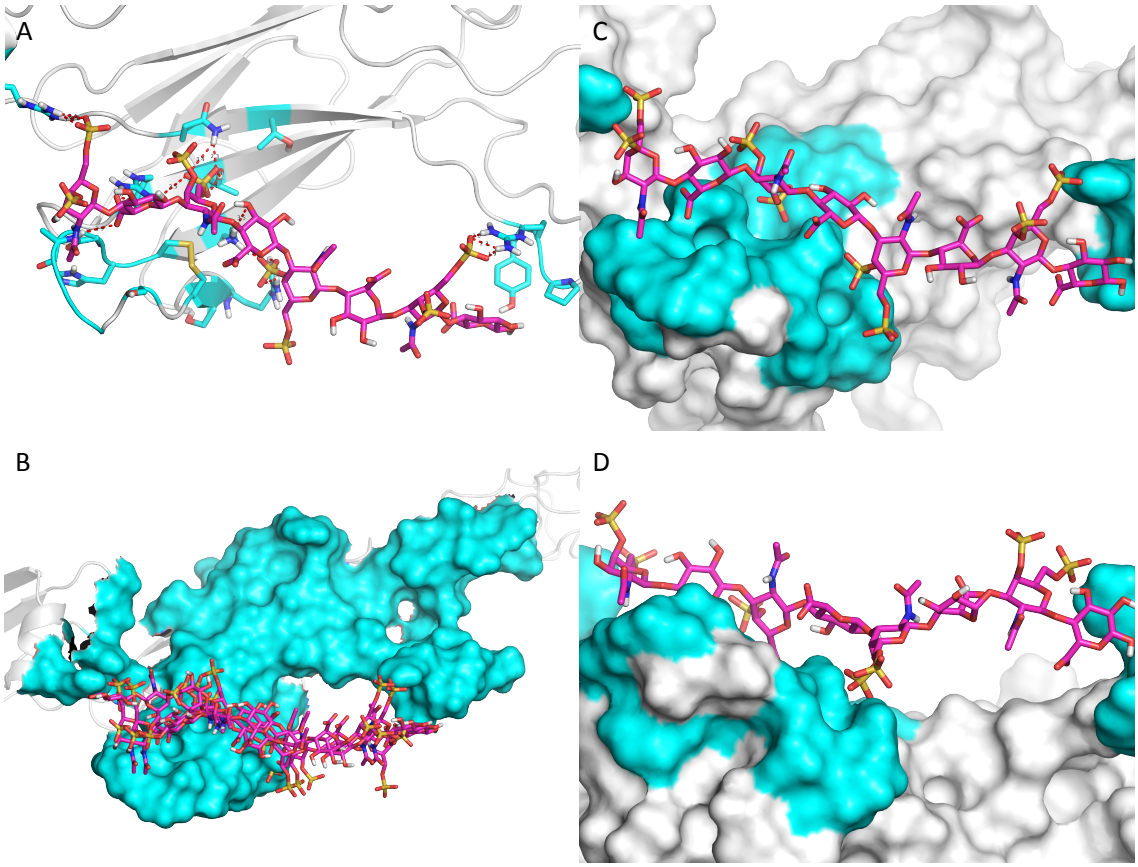

Figure S4-16 – (A) Detailed view of CS-E docked to EphB3 in top pose from binding mode 3. (B) Placement of all CS-E poses docked to EphB3 in binding mode 3. The top pose in this mode is the #2 pose overall (-380.53). This mode shows less contact with the surface of the protein.. (C) Top view of the best Mode 3 pose appears to fit closely to the protein surface, but the rotated view (D) shows that the much of the octasaccharide is separated from the surface.

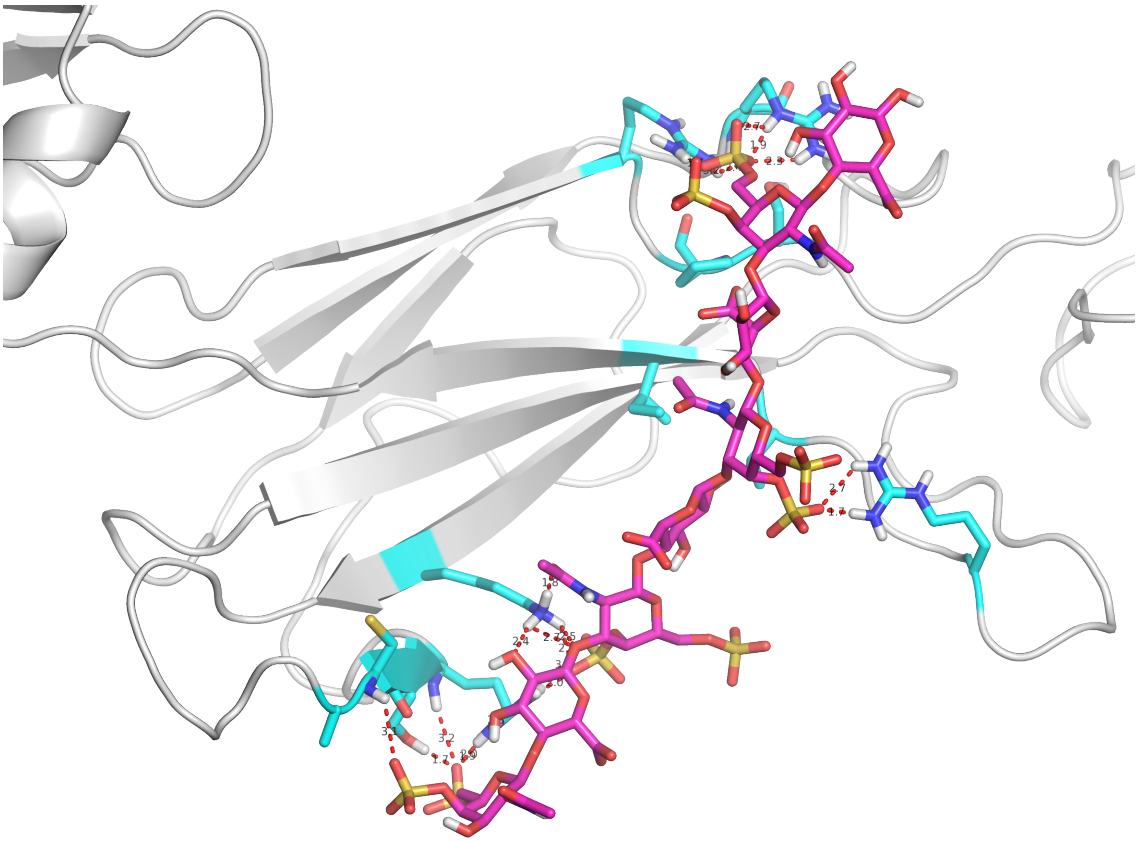

Figure S4-17 – Placement of only CS-E pose docked to EphB3 in binding mode 4. This mode contains only one pose in the top 25 poses. This pose is #6 overall (-351.91 kcal/mol).

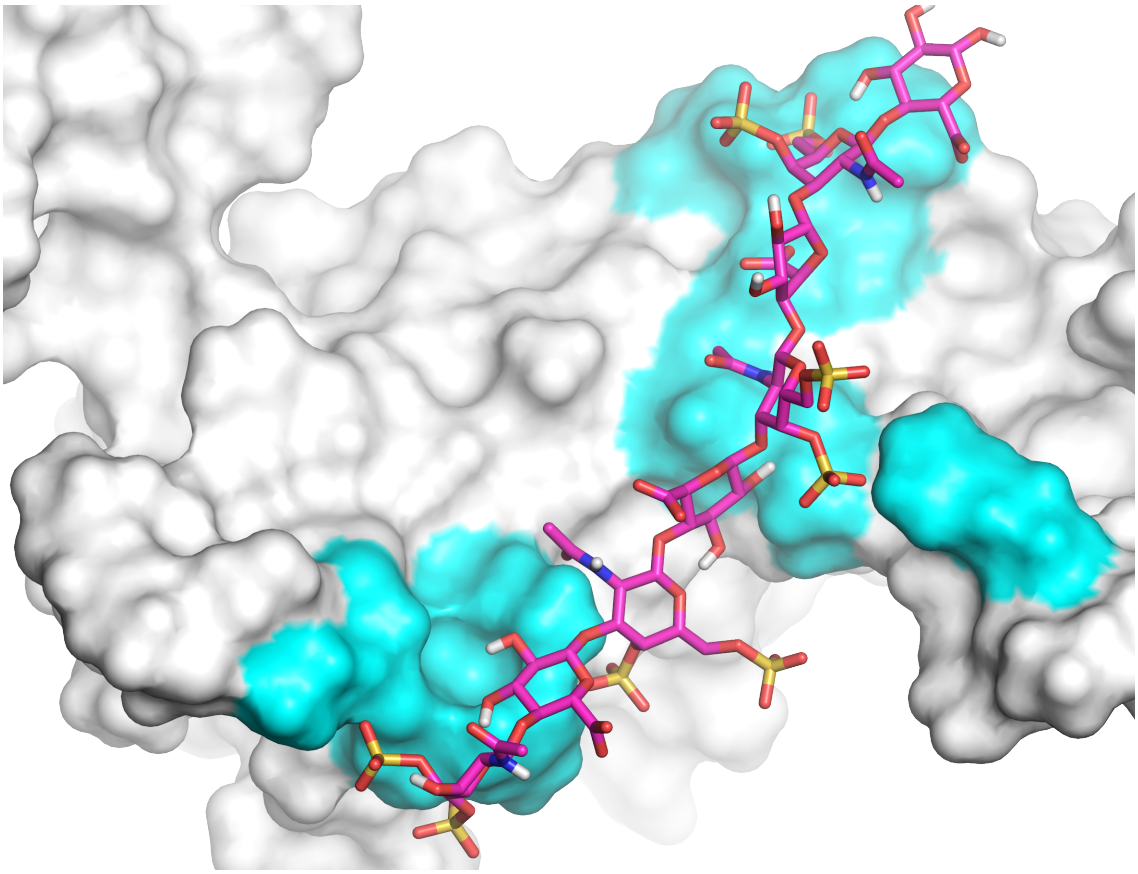

Figure S4-18 – Placement of the only CS-E pose docked to EphB3 in binding mode 4, with the protein surface shown.

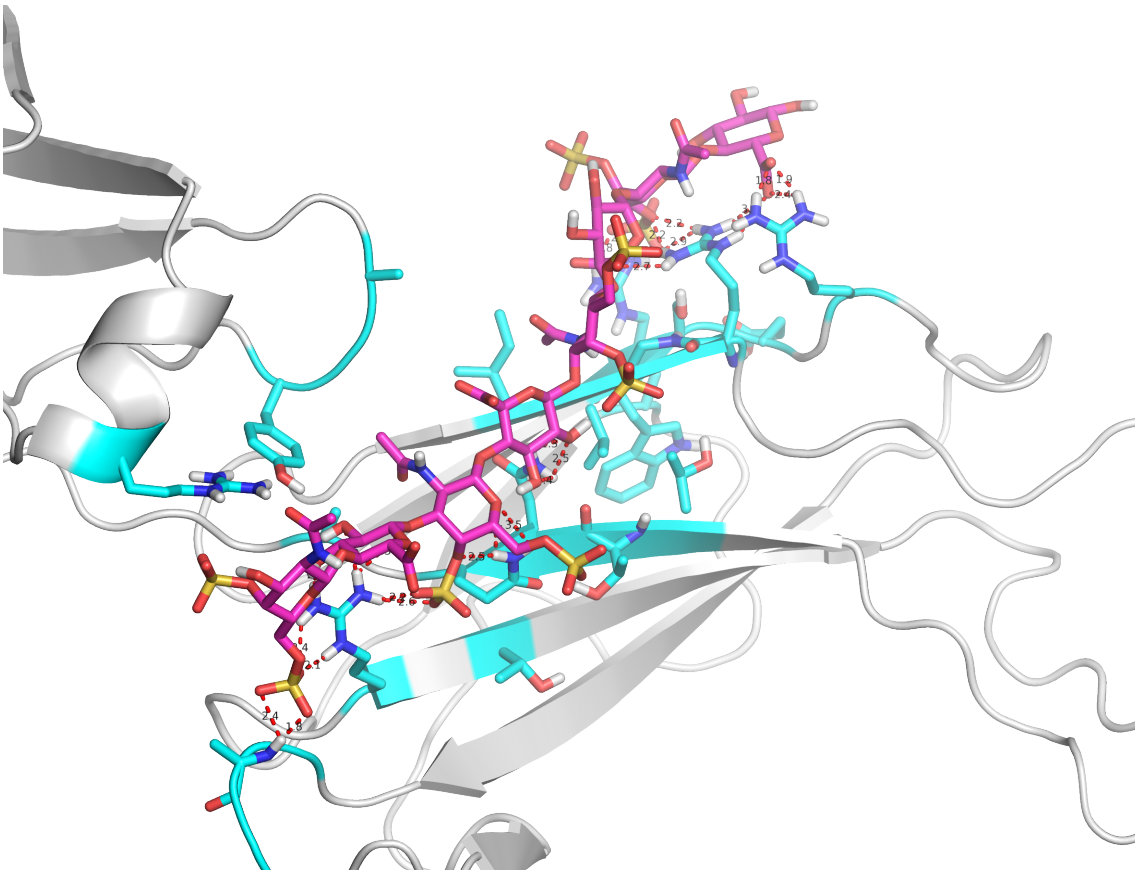

Figure S4-19 – Placement of the only CS-E pose docked to EphB3 in binding mode 5. This mode contains only one pose in the top 25 poses. This pose is #22 overall (-318.65 kcal/mol).

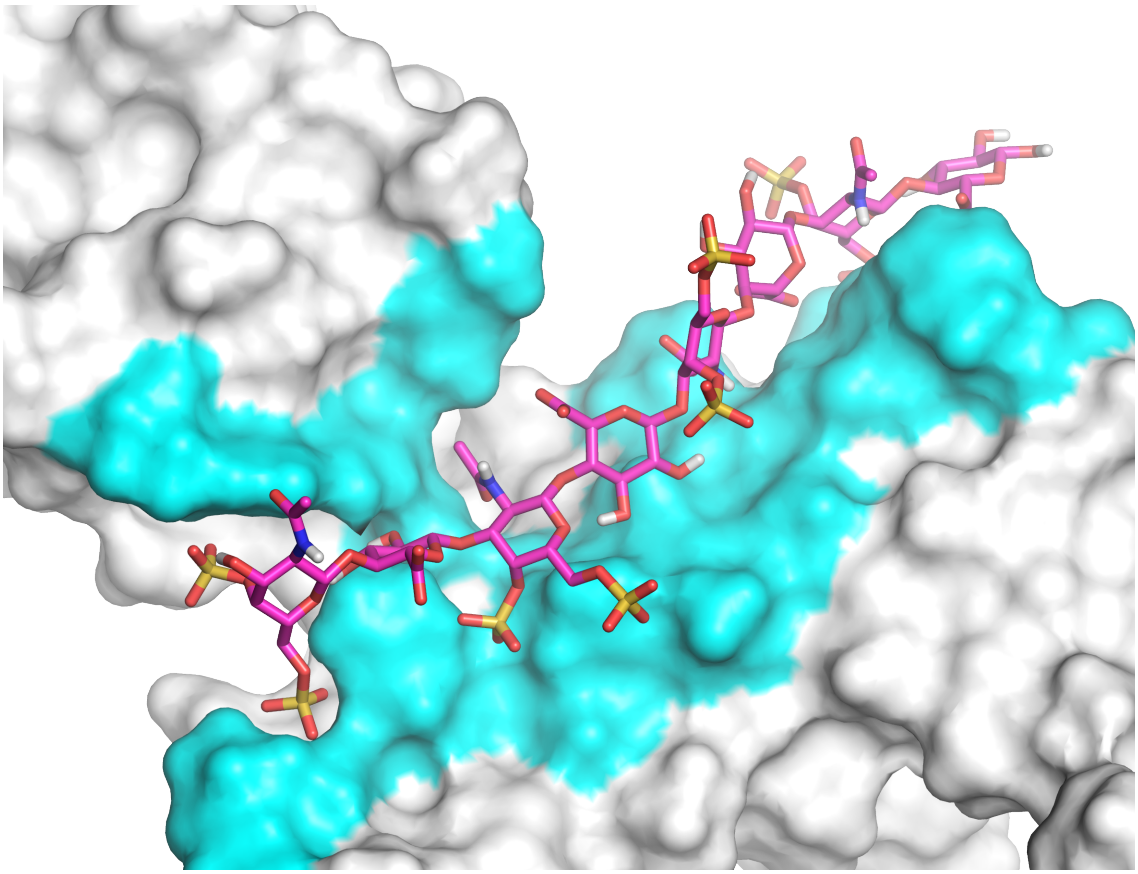

Figure S4-20 – Placement of the only CS-E pose docked to EphB3 in binding mode 5, with the protein surface shown.

Table S4-1 – Nonbond interactions by residue for the top pose in each of the five binding modes. As expected, favorable interactions are dominated by arginines and lysines (green), unfavorable interactions are dominated by glutamic acids (red). Ordered by nonbond energy for Mode 1.

Res	Num	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
ARG	440	-174.5	-167.1	-128.4	-137.4	-57.1
ARG	363	-137.7	-120.6	-57.8	-127.1	-138.0
ARG	309	-128.0	-98.4	-40.8	-45.4	-38.8
LYS	434	-125.7	-142.9	-48.7	-61.4	-54.9
ARG	344	-120.9	-71.5	-50.8	-115.5	-143.2
LYS	378	-85.9	-130.6	-167.5	-175.6	-92.9
ARG	420	-59.6	-115.5	-149.5	-72.9	-162.5
ARG	391	-57.7	-69.6	-139.1	-158.2	-56.5
ARG	408	-56.3	-55.9	-46.9	-55.7	-124.8
ARG	478	-40.6	-47.1	-120.7	-48.4	-121.1
ASN	323	-17.0	-6.2	1.8	2.3	0.5
ASN	322	-10.5	-14.5	-1.5	-1.4	0.2
ALA	443	-9.2	-8.0	-1.6	-3.8	-0.7
THR	319	-7.0	-2.9	-0.6	-0.4	0.2
TYR	325	-6.4	-5.0	0.1	0.3	0.8
THR	338	-6.4	-10.4	-0.5	0.1	0.1
SER	341	-6.3	-5.0	-0.2	-4.0	0.6
PRO	342	-3.8	-2.5	-2.0	-2.2	0.6
ASN	449	-3.3	-3.8	-4.2	0.4	-8.0
ALA	442	-3.1	-2.9	0.6	0.9	0.4
THR	422	-2.7	-9.5	-10.6	-2.3	-5.5
TRP	359	-2.2	-1.7	-0.8	-1.8	-0.7
SER	435	-2.1	-2.3	-0.7	-0.1	-0.4
THR	447	-1.7	-0.1	-0.3	0.6	-6.3
PRO	438	-1.6	-1.3	0.0	0.8	0.4
LEU	437	-1.6	-2.1	1.4	1.0	0.2
ASN	445	-1.5	-2.4	-1.0	-2.3	0.2
ALA	529	-1.0	-0.8	-0.1	-0.8	-2.8
HIS	321	-0.9	-1.1	0.8	-0.3	1.1

Res	Num	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
THR	448	-0.9	-5.6	-2.0	-1.4	-0.3
GLY	530	-0.9	-1.4	-2.3	-1.0	-4.1
GLY	382	-0.7	-1.3	-7.2	-0.2	-9.7
ALA	383	-0.7	-1.2	-6.3	-0.8	-6.9
CYX	389	-0.7	-1.8	-2.2	-0.3	-1.0
ALA	452	-0.2	-0.5	-1.0	-0.5	-2.2
GLN	450	-0.1	-6.8	-16.2	-2.8	-10.1
ALA	388	-0.1	0.1	-1.1	-6.0	2.3
SER	390	0.0	1.1	0.6	-8.3	1.9
VAL	339	0.0	1.4	-0.9	0.4	0.2
GLY	345	0.1	-0.1	-0.6	0.2	-5.4
ILE	347	0.1	-0.1	-0.6	0.2	-3.4
CYX	380	0.2	0.2	-2.3	-0.3	-0.4
SER	360	0.2	0.3	0.6	0.9	-7.1
TYR	531	0.3	0.2	-1.8	0.1	-5.7
PRO	439	0.7	0.8	1.1	2.7	0.8
GLY	385	0.7	1.1	1.8	2.5	0.5
GLY	384	0.8	1.3	2.9	1.3	2.4
CYX	320	0.8	-0.1	-0.2	-0.7	-0.6
VAL	346	0.8	0.8	0.7	0.5	1.4
HSE	381	1.0	1.9	-1.8	-0.8	-3.2
VAL	444	1.5	2.8	0.3	0.5	0.5
TYR	441	2.0	3.1	-5.0	-4.6	-0.1
SER	387	2.0	4.1	0.4	0.9	-0.7
ILE	446	2.2	1.3	-0.5	0.7	0.6
PRO	436	2.9	-5.5	1.2	0.6	0.0
PHE	324	3.2	2.4	0.2	-0.1	-0.7
GLU	361	65.7	64.1	48.1	62.9	98.6
GLU	424	79.6	100.8	103.3	108.3	92.6

Table S4-2 – Comparison of the arginine and lysine interactions with the five binding modes. Note that the best interactions differ for each binding mode. Ordered by nonbond energy for Mode 1. For instance, R344 contributes much more strongly to mode 1 than to mode 2. Similarly, R420 contributes much more strongly to mode 2 than to mode 1. Mutations targeting mode-specific residues could be used to identify which mode is correct.

Res	Num	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
ARG	440	-174.5	-167.1	-128.4	-137.4	-57.1
ARG	363	-137.7	-120.6	-57.8	-127.1	-138.0
ARG	309	-128.0	-98.4	-40.8	-45.4	-38.8
LYS	434	-125.7	-142.9	-48.7	-61.4	-54.9
ARG	344	-120.9	-71.5	-50.8	-115.5	-143.2
LYS	378	-85.9	-130.6	-167.5	-175.6	-92.9
ARG	420	-59.6	-115.5	-149.5	-72.9	-162.5
ARG	391	-57.7	-69.6	-139.1	-158.2	-56.5
ARG	408	-56.3	-55.9	-46.9	-55.7	-124.8
ARG	478	-40.6	-47.1	-120.7	-48.4	-121.1

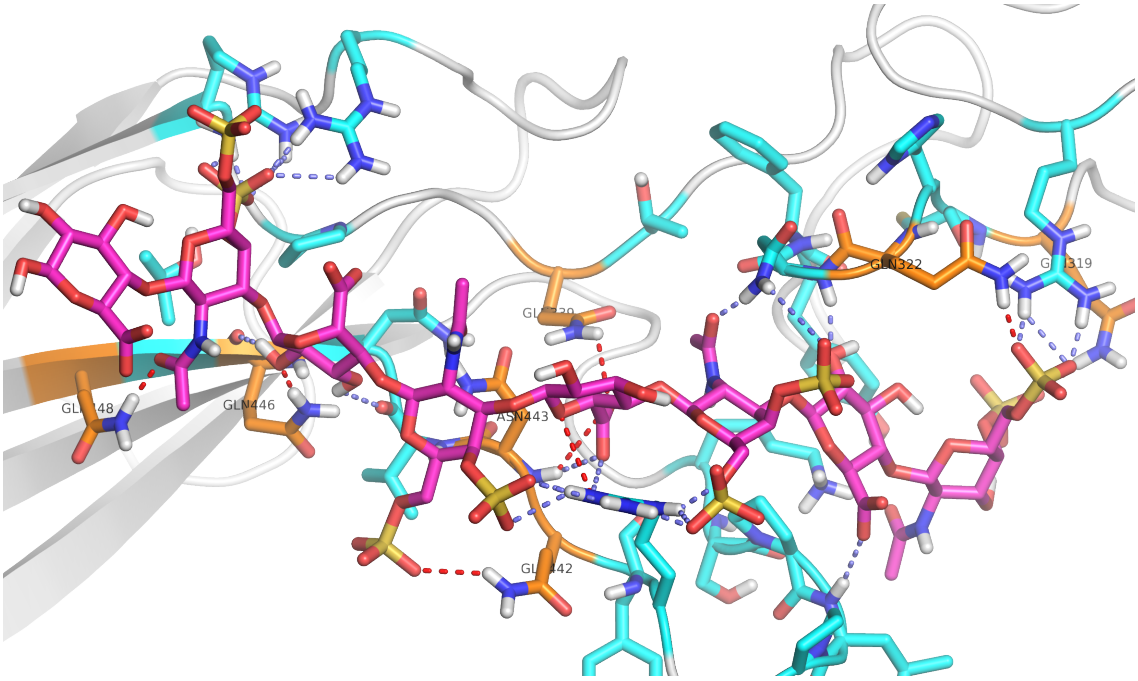

Figure S4-21 – Structure with proposed mutations for Mode 1: T319Q, N322Q, V339Q, A442Q, A443N, I446Q, and T448Q. We predict that this set of mutations for Mode 1 improves binding energy by 66.03 kcal/mol, or 16.5% better than binding to the wild-type.

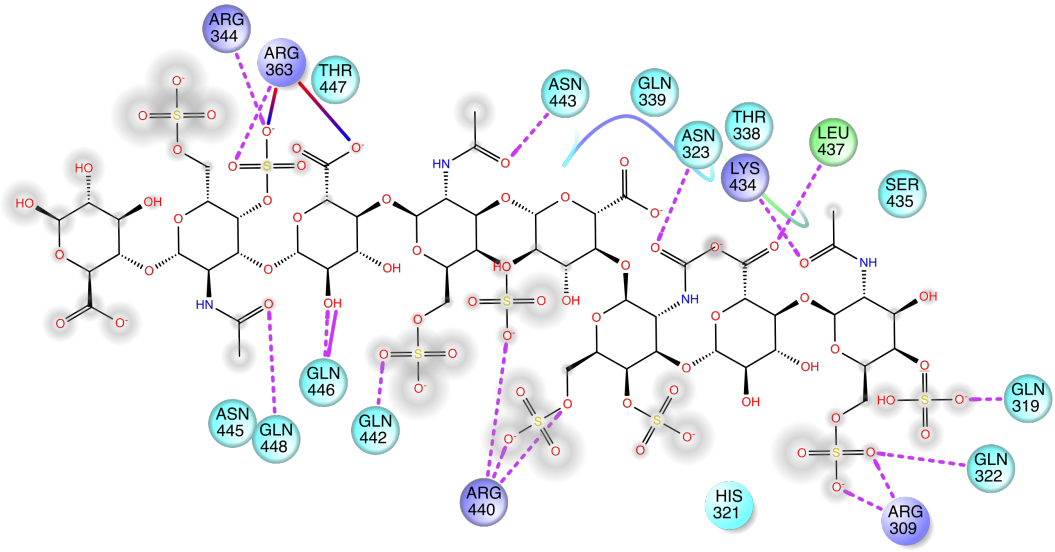

Figure S4-22 – Pharmacophore with proposed mutations for Mode 1: T319Q, N322Q, V339Q, A442Q, A443N, I446Q, and T448Q. We predict that this set of mutations for Mode 1 improves binding energy by 66.03 kcal/mol, or 16.5% better than binding to the wild-type.

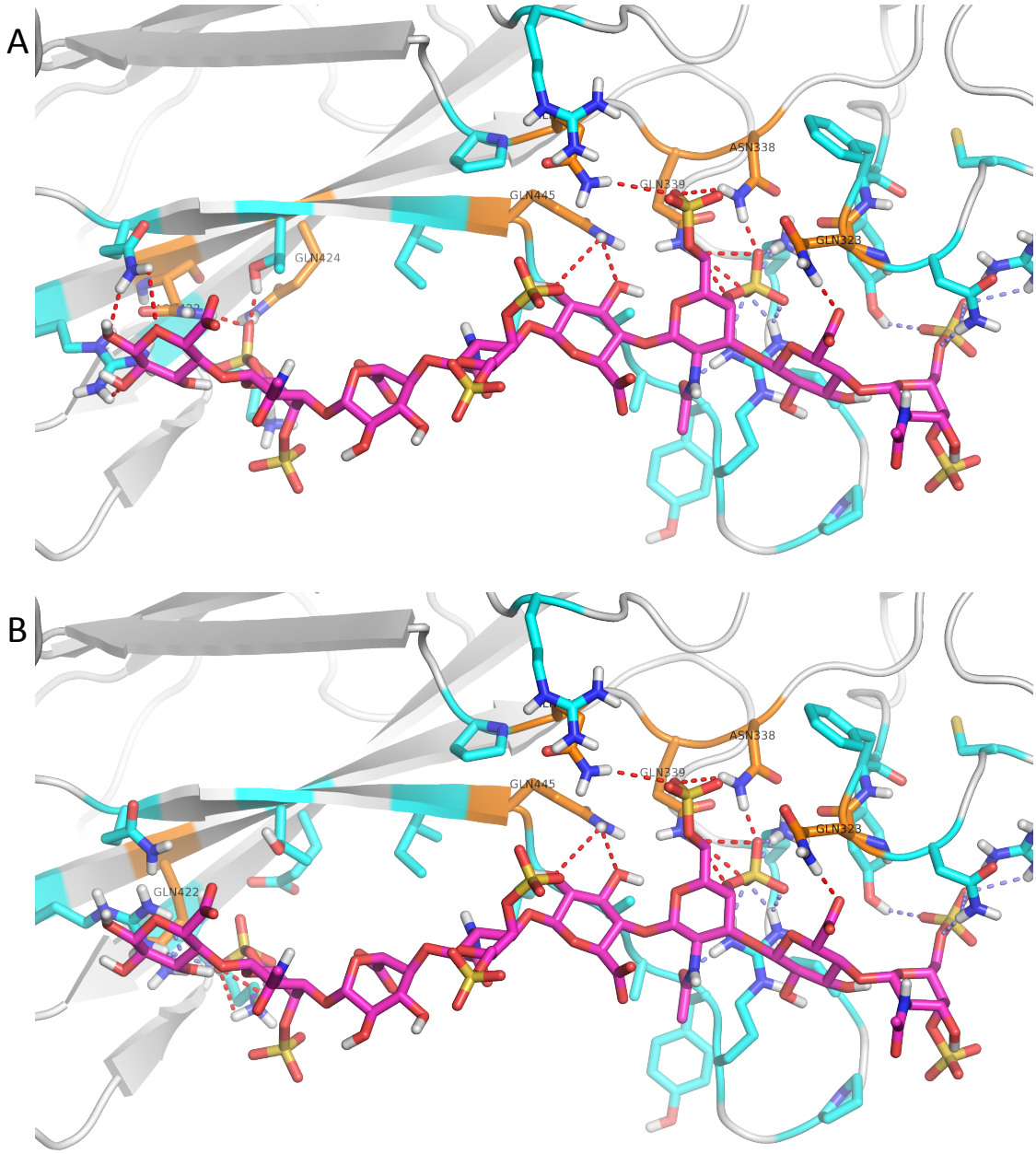

Figure S4-23 – Structures with proposed mutations for Mode 2. (A) Mutations: N323Q, T338N, V339Q, S341Q, T422Q, and N445Q. This improves binding energy by 46.64 kcal/mol or 11.4% better than the wild-type. (B) Adds the E424Q mutant, resulting in an improvement of 163.84 kcal/mol or 40%.

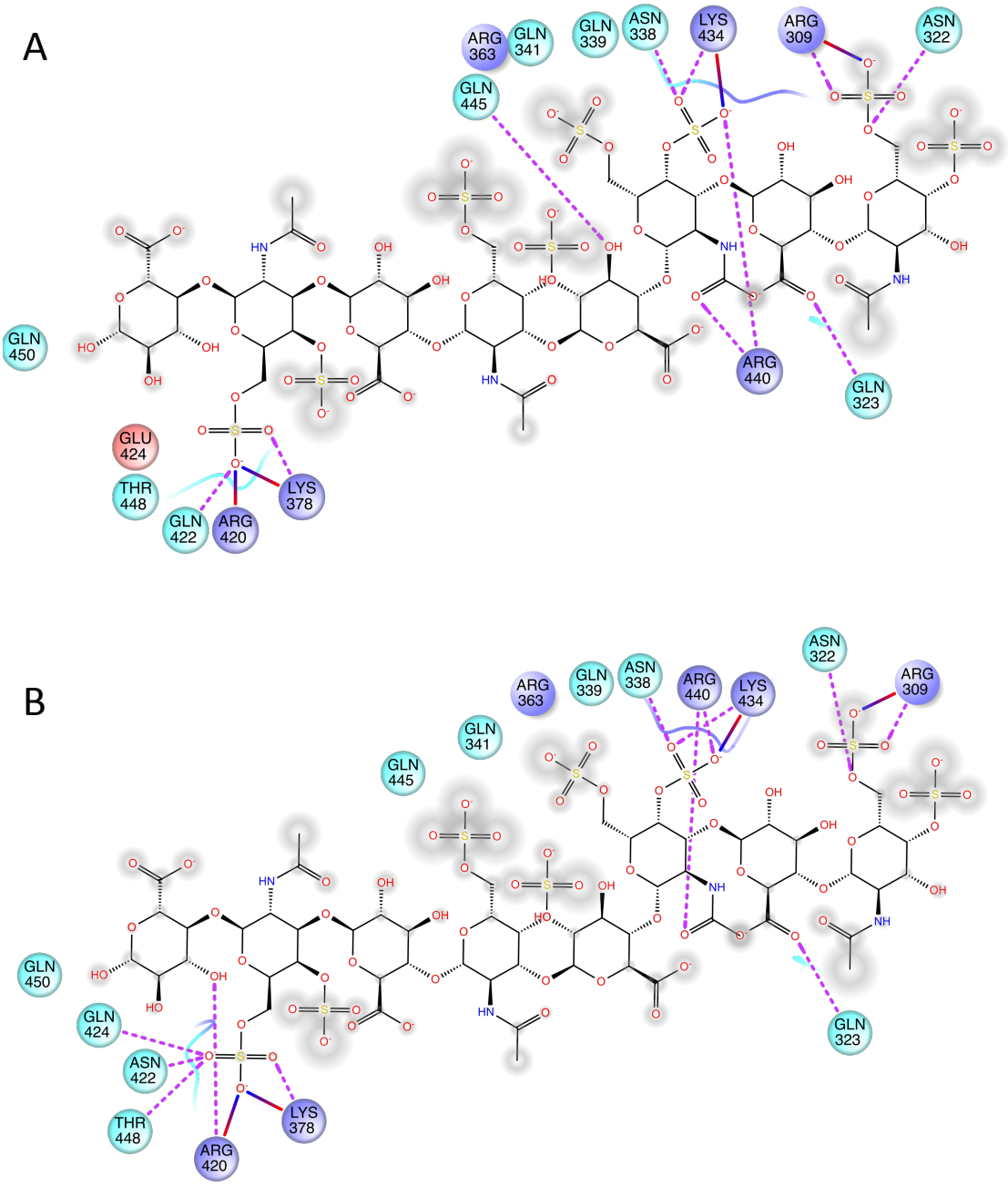

Figure S4-24 – Pharmacophores for proposed mutations for Mode 2. (A) Mutations: N323Q, T338N, V339Q, S341Q, T422Q, and N445Q. This improves binding energy by 46.64 kcal/mol or 11.4% better than the wild-type. (B) Adds the E424Q mutant, resulting in an improvement of 163.84 kcal/mol or 40%.

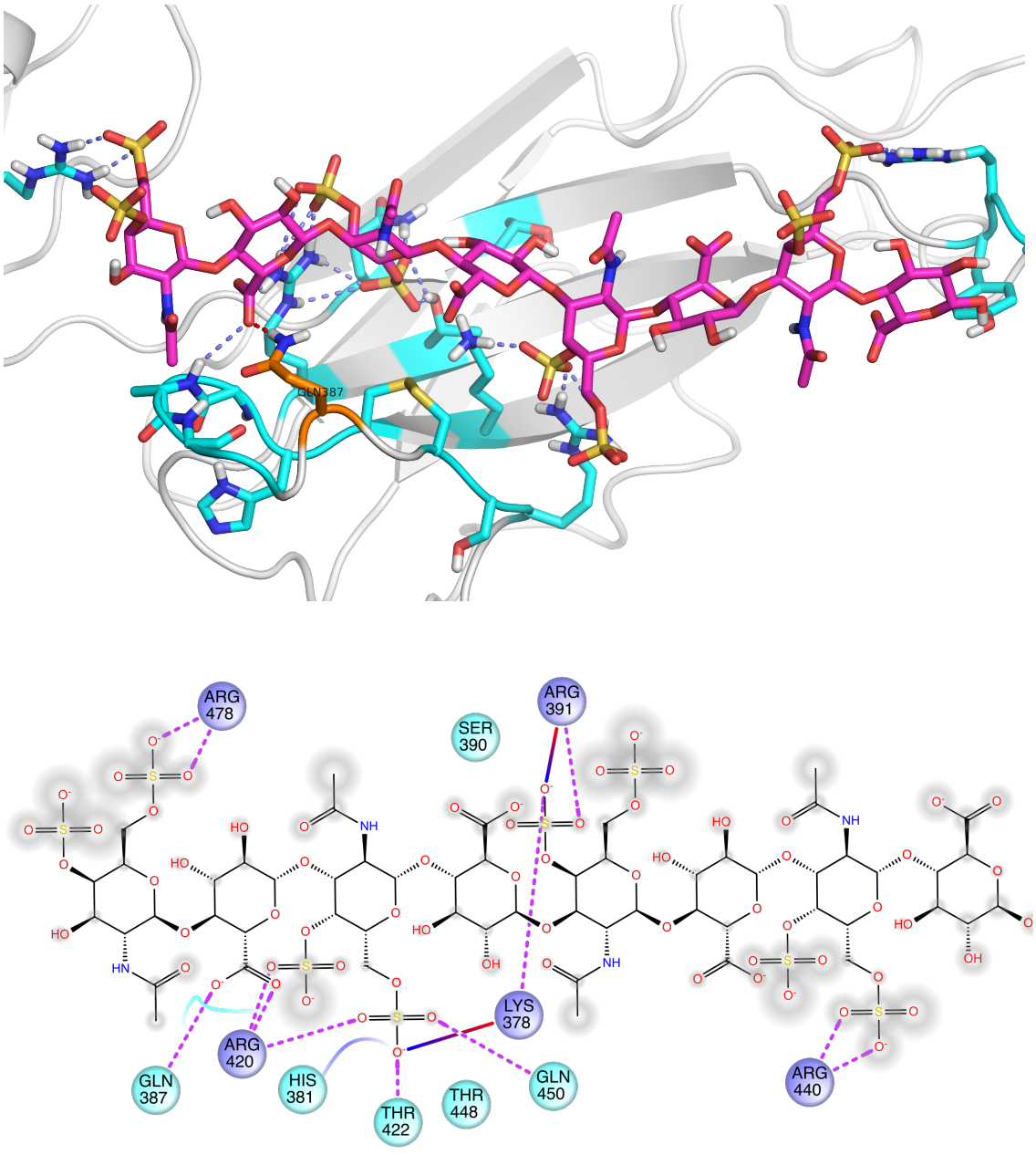

Figure S4-25 – Structure and pharmacophore for Mode 3 mutation: S387Q. This improves binding energy by 7.84 kcal/mol or 1.53% better than the wild-type.

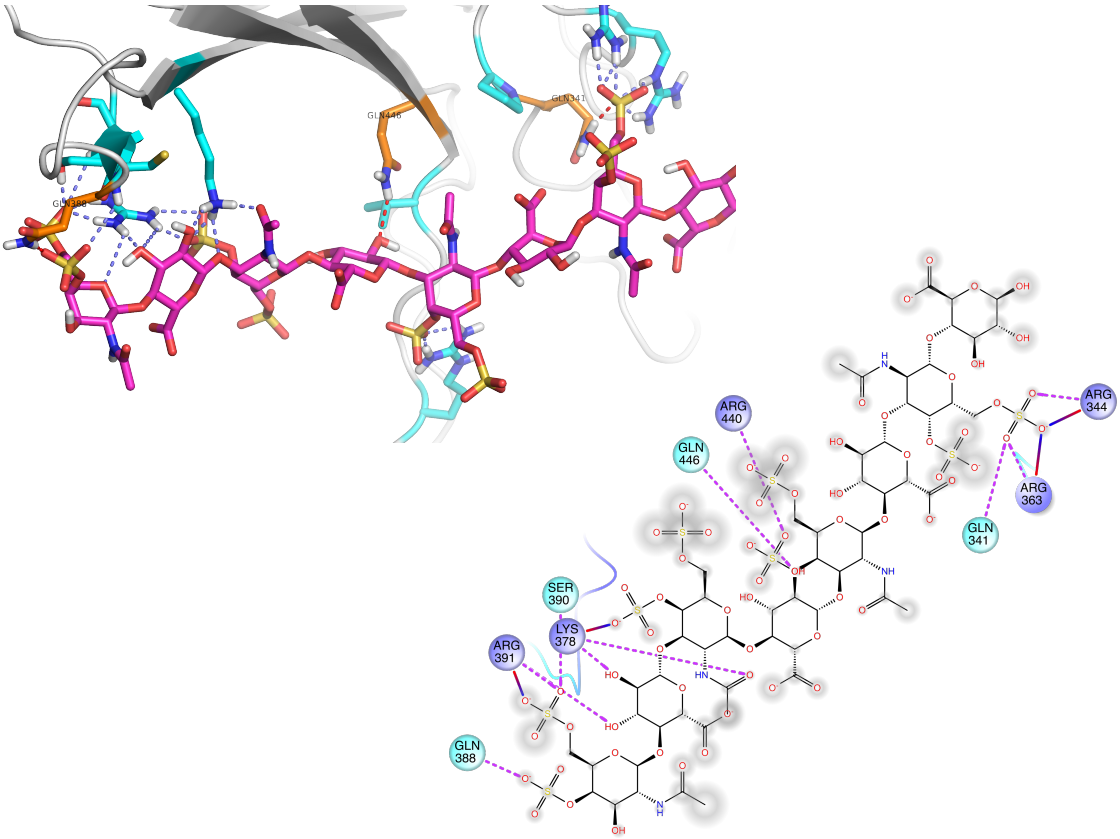

Figure S4-26 – Structure and pharmacophore for Mode 4 mutations: S341Q, A388Q, I446Q. This improves binding energy by 25.71 kcal/mol or 6.43% better than the wild-type.

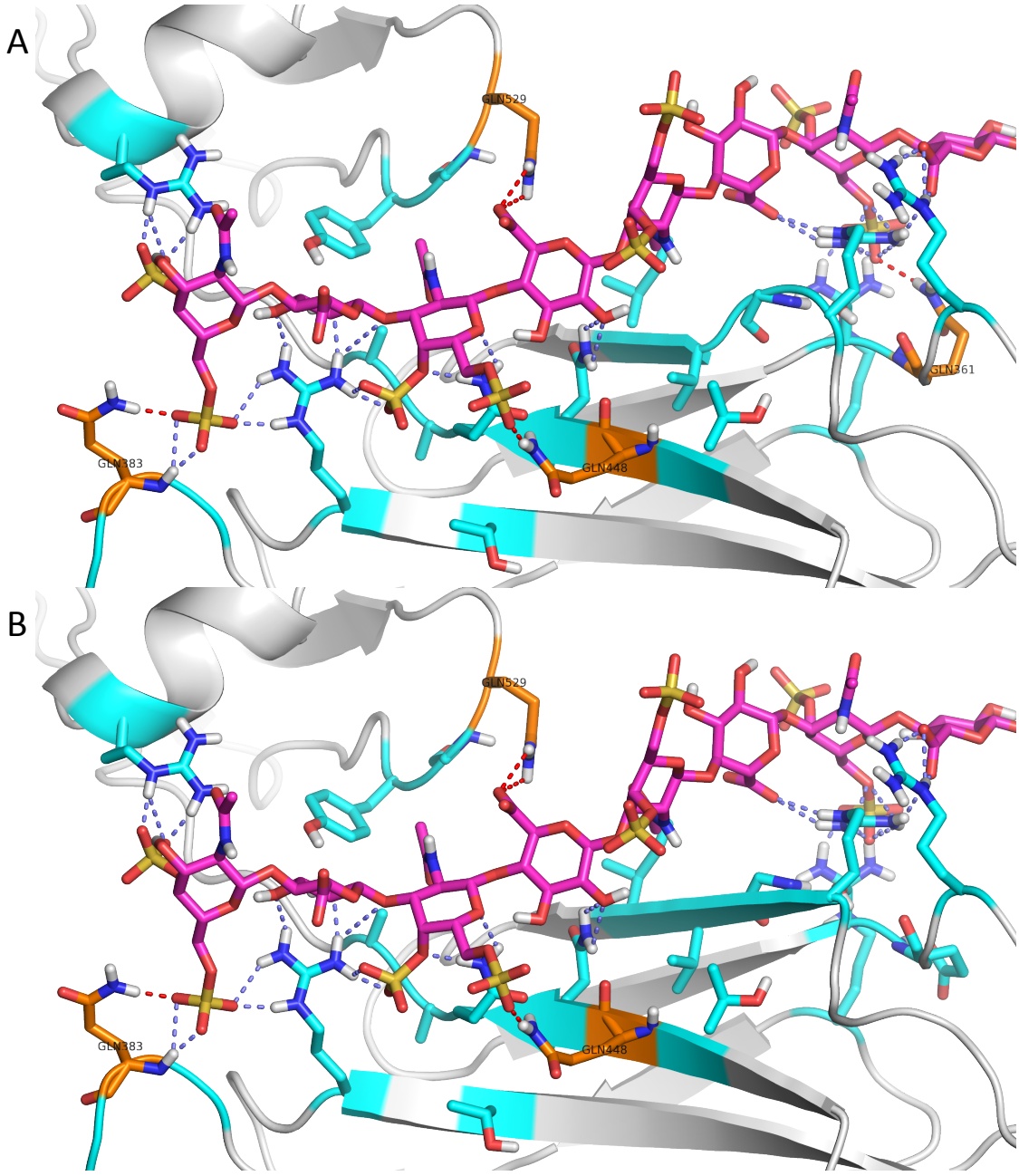

Figure S4-27 – Structures with proposed mutations for Mode 5. (A) Mutations: A383Q, T448Q, A529Q. This improves binding energy by 29.11 kcal/mol or 8% better than the wild-type. (B) Adds the E361Q mutant, resulting in an improvement of 133.82 kcal/mol or 37%.

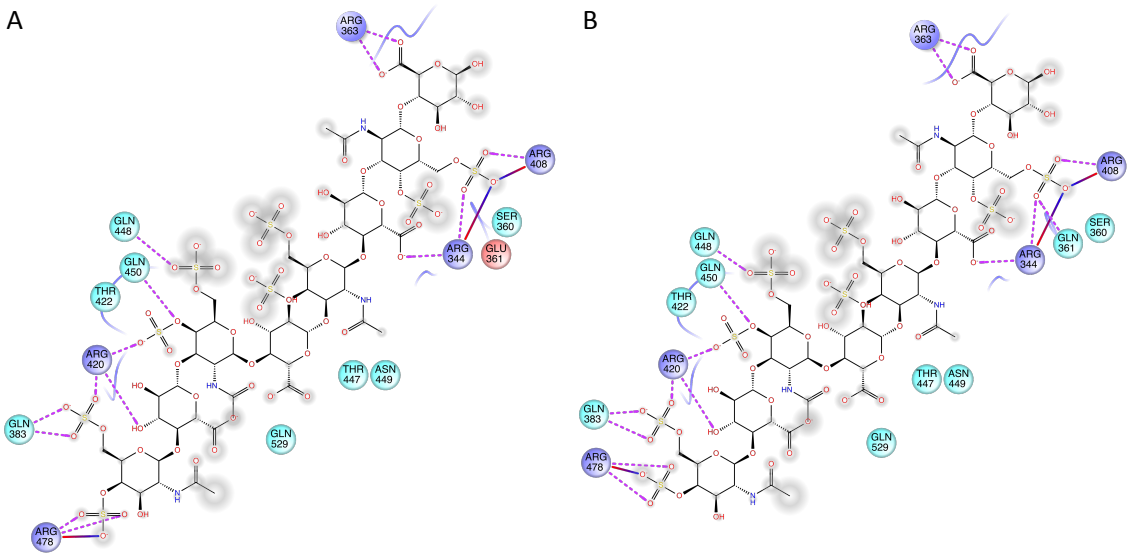

Figure S4-28 – Pharmacophores with proposed mutations for Mode 5. (A) Mutations: A383Q, T448Q, A529Q. This improves binding energy by 29.11 kcal/mol or 8% better than the wild-type. (B) Adds the E361Q mutant, resulting in an improvement of 133.82 kcal/mol or 37%.

Table S4-3 - Binding energies for all mutations to glutamine that improve binding energy for the best pose in mode 1. Note that most mutations do not make new hydrogen bonds (highlighted in red). The increase in binding energy for those mutations can be attributed to Coulomb energy. We only wish to use mutants that make new contacts with the ligand (highlighted in green).

residue	energy	rel. to wt	% incr.	new hbond?
E361	-477.71	-76.57	16.03	no
E424	-473.64	-72.49	15.18	no
T448	-419.91	-18.77	3.93	yes
V339	-414.79	-13.65	2.86	yes
I446	-413.62	-12.48	2.61	yes
A442	-413.11	-11.97	2.50	yes
N445	-412.56	-11.42	2.39	yes
T319	-412.34	-11.20	2.34	yes
L437	-408.68	-7.54	1.58	no
A529	-407.92	-6.78	1.42	no
A443	-406.67	-5.53	1.16	no
A452	-406.09	-4.95	1.04	no
N323	-405.37	-4.23	0.89	yes
A383	-405.12	-3.98	0.83	no
N449	-404.09	-2.95	0.62	no
T338	-403.90	-2.76	0.58	no
I347	-403.35	-2.20	0.46	no
S435	-403.08	-1.94	0.41	no
S341	-402.80	-1.66	0.35	no
T422	-402.79	-1.64	0.34	no
Y531	-402.41	-1.27	0.27	no
F324	-402.13	-0.99	0.21	no
H381	-401.95	-0.81	0.17	no
N322	-401.93	-0.78	0.16	yes
Y441	-401.87	-0.73	0.15	no
W359	-401.70	-0.56	0.12	no
S360	-401.61	-0.47	0.10	no
V444	-401.52	-0.38	0.08	no
wt	-401.14			

Table S4-4 – Binding energies in kcal/mol for different sets of mutations for best pose in binding mode 1. Sets are ranked by binding energy. Set 8.2 (66 kcal/mol or 16.5% improvement in binding energy) is the best set where all mutated residues make a new contact with the ligand.

set	round	energy	rel. to wt	% incr.	mutations								
set7	1	-476.91	-75.77	18.89	Q319	Q323	Q322	Q339	Q442	N443	Q445	Q446	Q448
set5	1	-472.46	-71.32	17.78	Q319	Q323	Q322	Q339	Q442		Q445	Q446	Q448
set8	1	-471.38	-70.24	17.51	Q319		Q322	Q339	Q442	N443	Q445	Q446	Q448
set3	1	-471.34	-70.20	17.50	Q319	Q323		Q339	Q442	N443	Q445	Q446	Q448
set8.2	2	-467.18	-66.03	16.46	Q319		Q322	Q339	Q442	N443		Q446	Q448
set2	1	-464.81	-63.66	15.87	Q319			Q339	Q442		Q445	Q446	Q448
set7.2	2	-463.32	-62.17	15.50	Q319	Q323	Q322	Q339	Q442	N443		Q446	Q448
set3.2	2	-462.70	-61.55	15.34	Q319	Q323		Q339	Q442	N443		Q446	Q448
set4	1	-462.00	-60.86	15.17	Q319			Q339	Q442	N443	Q445	Q446	Q448
set1	1	-460.38	-59.24	14.77	Q319	Q323		Q339	Q442		Q445	Q446	Q448
set4.2	2	-459.85	-58.70	14.63	Q319			Q339	Q442	N443		Q446	Q448
set6	1	-459.00	-57.86	14.42	Q319		Q322	Q339	Q442		Q445	Q446	Q448
set6.2	2	-448.17	-47.03	11.72	Q319		Q322		Q442			Q446	Q448
set5.2	2	-446.28	-45.14	11.25	Q319	Q323	Q322		Q442			Q446	Q448
set2.2	2	-444.95	-43.81	10.92	Q319				Q442			Q446	Q448
set1.2	2	-435.31	-34.17	8.52	Q319	Q323			Q442			Q446	Q448
wt		-401.14			T319	N323	N322	V339	A442	A443	N445	I446	T448

Table S4-5 - Binding energies for all mutations to glutamine that improve binding energy for the best pose in mode 2. Note that most mutations do not make new hydrogen bonds (highlighted in red). The increase in binding energy for those mutations can be attributed to Coulomb energy. We only wish to use mutants that make new contacts with the ligand (highlighted in green).

residue	energy	rel. to wt	% incr.	new hbond?
E424	-525.91	-117.08	22.26	yes
E361	-468.11	-59.28	11.27	no
V339	-424.97	-16.14	3.07	yes
T422	-424.27	-15.44	2.94	yes
T338	-422.96	-14.13	2.69	yes
A442	-417.47	-8.64	1.64	no
N445	-416.27	-7.44	1.42	yes
S435	-415.01	-6.18	1.17	no
Y325	-414.43	-5.60	1.06	no
I446	-412.76	-3.93	0.75	no
T319	-412.71	-3.88	0.74	no
N323	-411.80	-2.97	0.56	yes
Y441	-411.07	-2.24	0.43	no
I347	-410.22	-1.39	0.26	no
S341	-409.75	-0.92	0.17	yes
A388	-409.66	-0.83	0.16	no
Y531	-409.30	-0.47	0.09	no
S360	-408.96	-0.13	0.02	no
A452	-408.86	-0.03	0.00	no
wt	-408.83			

Table S4-6 - Binding energies in kcal/mol for different sets of mutations to best pose from binding mode 2. Sets are ranked by binding energy. Set 4 (163.5 kcal/mol or 40% improvement in binding energy) and Set 11 (46.6 kcal/mol or 11.4% improvement in binding energy) are both selected for mode 2 due to the presence of residue E424. It is interesting to find E424 in such close proximity to a negatively charged ligand like CS-E and we are unsure of what other role E424 may be playing in the protein. Thus, mutations to E424 may have unexpected consequences, even with a relatively close mutant such as glutamine.

set	energy	rel. to wt	%incr	mutations						
set4	-572.31	-163.48	39.99	Q323	N338	Q339	Q341	N422	Q424	Q445
set8	-571.53	-162.70	39.80		N338	Q339	Q341	N422	Q424	Q445
set2	-571.27	-162.44	39.73	Q323	Q338	Q339	Q341	N422	Q424	Q445
set6	-568.59	-159.76	39.08		Q338	Q339	Q341	N422	Q424	Q445
set3	-546.79	-137.96	33.74	Q323	N338	Q339	Q341	Q422	Q424	Q445
set1	-544.24	-135.41	33.12	Q323	Q338	Q339	Q341	Q422	Q424	Q445
set7	-542.07	-133.24	32.59		N338	Q339	Q341	Q422	Q424	Q445
set5	-540.38	-131.54	32.18		Q338	Q339	Q341	Q422	Q424	Q445
set11	-455.47	-46.64	11.41	Q323	N338	Q339	Q341	Q422		Q445
set9	-455.03	-46.20	11.30	Q323	Q338	Q339	Q341	Q422		Q445
set10	-451.95	-43.12	10.55	Q323	Q338	Q339	Q341	N422		Q445
set13	-448.74	-39.91	9.76		Q338	Q339	Q341	Q422		Q445
set15	-448.39	-39.56	9.68		N338	Q339	Q341	Q422		Q445
set12	-447.20	-38.37	9.38	Q323	N338	Q339	Q341	N422		Q445
set14	-443.97	-35.14	8.59		Q338	Q339	Q341	N422		Q445
set16	-441.47	-32.64	7.98		N338	Q339	Q341	N422		Q445
wt	-408.83			N323	T338	V339	S341	T422	E424	N445

Table S4-7 - Binding energies for all mutations to glutamine that improve binding energy for the best pose in mode 3. Note that only *one* residue was able to make a new contact with the ligand. This is a strong indicator that binding mode 3 is not a reliable result. The increase in binding energy for those mutations can be attributed to Coulomb energy.

residue	energy	rel. to wt	% incr.	new hbond?
E424	-511.04	-104.87	20.52	no
E361	-471.32	-65.16	12.75	no
I446	-419.60	-13.44	2.63	no
T338	-415.66	-9.50	1.86	no
A388	-415.12	-8.96	1.75	no
V346	-415.10	-8.94	1.75	no
S435	-414.11	-7.95	1.56	no
S387	-414.00	-7.84	1.53	yes
T447	-413.63	-7.47	1.46	no
S390	-413.47	-7.31	1.43	no
I347	-413.13	-6.97	1.36	no
Y325	-412.65	-6.48	1.27	no
N445	-412.47	-6.31	1.23	no
Y531	-411.85	-5.69	1.11	no
A529	-411.72	-5.56	1.09	no
S360	-411.18	-5.02	0.98	no
F324	-410.95	-4.78	0.94	no
L437	-410.48	-4.32	0.85	no
S341	-408.40	-2.24	0.44	no
T319	-408.29	-2.12	0.42	no
A443	-408.18	-2.02	0.39	no
N323	-408.08	-1.91	0.37	no
V339	-407.77	-1.61	0.31	no
V444	-407.03	-0.87	0.17	no
W359	-406.77	-0.61	0.12	no
N322	-406.33	-0.16	0.03	no
Y441	-406.20	-0.04	0.01	no
wt	-406.16			

Table S4-8 - Binding energies for all mutations to glutamine that improve binding energy for the only pose in mode 4. Note that most mutations do not make new hydrogen bonds (highlighted in red). The increase in binding energy for those mutations can be attributed to Coulomb energy. We only wish to use mutants that make new contacts with the ligand (highlighted in green).

residue	energy	rel. to wt	% incr.	new hbond?
E424	-495.76	-95.77	19.32	no
E361	-465.02	-65.03	13.12	no
A443	-419.46	-19.46	3.93	no
A388	-412.03	-12.04	2.43	yes
S341	-411.02	-11.03	2.22	yes
N449	-410.33	-10.33	2.08	no
Y531	-410.20	-10.21	2.06	no
I446	-407.81	-7.82	1.58	yes
T338	-407.64	-7.64	1.54	no
I347	-405.08	-5.08	1.03	no
W359	-402.47	-2.48	0.50	no
T422	-402.32	-2.33	0.47	no
Y441	-402.25	-2.25	0.45	no
S435	-401.01	-1.01	0.20	no
wt	-399.99			

Table S4-9 - Binding energy in kcal/mol for the only set of mutations to the only pose from binding mode 4. Mutation results in 25.7 kcal/mol or 6.4% improvement in binding energy.

set	energy	rel. to wt	% incr.	mutations		
set1	-425.70	-25.71	6.43	Q341	Q388	Q446
wt	-399.99			S341	A388	I446

Table S4-10 - Binding energies for all mutations to glutamine that improve binding energy for the only pose in mode 5. Note that most mutations do not make new hydrogen bonds (highlighted in red). The increase in binding energy for those mutations can be attributed to Coulomb energy. We only wish to use mutants that make new contacts with the ligand (highlighted in green).

residue	energy	rel. to wt	% incr.	new hbond?
E361	-464.85	-103.36	22.23	yes
E424	-462.29	-100.79	21.68	no
A383	-378.48	-16.98	3.65	yes
T448	-369.95	-8.46	1.82	yes
A388	-369.66	-8.17	1.76	no
A529	-369.62	-8.13	1.75	yes
S390	-369.57	-8.08	1.74	no
I347	-368.36	-6.87	1.48	no
H381	-366.63	-5.13	1.10	no
V444	-365.92	-4.42	0.95	no
S435	-365.33	-3.83	0.82	no
T422	-364.86	-3.37	0.72	no
Y325	-364.47	-2.98	0.64	no
T338	-362.17	-0.67	0.14	no
N323	-361.95	-0.45	0.10	no
I446	-361.70	-0.20	0.04	no
wt	-361.50			

Table S4-11 - Binding energy in kcal/mol for mutations to the only pose from binding mode 5. Set 1, which includes a mutation to E361, improves binding energy by 133.8 kcal/mol or 37%. Set 2, which skips the mutation to E361, improves binding by 29.1 kcal/mol or 8.1%. Similarly to binding mode 2, it is interesting to find E361 (a different glutamic acid) in such close proximity to a negatively charged ligand like CS-E and we are unsure of what other role E361 may be playing in the protein. Thus, mutations to E361 may have unexpected consequences, even with a relatively close mutant such as glutamine.

set	energy	rel. to wt	% incr.	mutations			
set1	-495.32	-133.82	37.02	Q361	Q383	Q448	Q529
set2	-390.61	-29.11	8.05		Q383	Q448	Q529
wt	-361.50			E361	A383	T448	A529

List of Figures

List of Tables