

Analyzing Machupo virus-receptor binding by molecular dynamics simulations

Austin G. Meyer^{1,2,3*}, Sara L. Sawyer², Andrew D. Ellington²,
and Claus O. Wilke¹

Address:

¹Department of Integrative Biology, Institute for Cellular and Molecular Biology, and Center for Computational Biology and Bioinformatics. The University of Texas at Austin, Austin, TX 78712, USA.

²Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA.

³School of Medicine, Texas Tech University Health Sciences Center, Lubbock, TX 79430, USA.

*Corresponding author

Email: austin.meyer@utexas.edu

Phone: +1 512 971 0123

Manuscript type: research article

Keywords: protein–protein interaction, molecular dynamics, arenavirus, machupo

1 Introduction

2 The computational prediction of mutational effects on protein–protein interactions remains a chal-
3 lenging problem. Several methods are available to perform an energy difference calculation from an
4 experimentally determined co-crystal structure. For example, end point methods can be performed
5 rapidly, with relatively low computational cost (Gront et al. 2011; Kortemme et al. 2004). How-
6 ever, such methods can suffer from various simplifying assumptions. For example, they generally
7 use an implicit solvent approximation and assume the end state difference with minimal structural
8 rearrangement is sufficient to discriminate energetic differences (Gront et al. 2011; Kortemme et al.
9 2004). Alternative approaches have been developed using machine learning, training coefficients
10 in a weighted equation containing geometric and energetic parameters (Vreven et al. 2011, 2012;
11 Bajaj et al. 2011; Hwang et al. 2010). Unfortunately, such machine-learning approaches often suf-
12 fer in novel applications, for which available training sets are small or non-existent. As such, these
13 methods are poorly suited for most host-virus protein–protein systems. By contrast, first principles
14 methods can forgo training, but currently available methods such as free energy perturbation (FEP)
15 and thermodynamic integration (TI) rely on a transitional model (where one state may be wild-type
16 and the other may be a mutant) to make rigorous free energy calculations (Gilson et al. 1997; Lu
17 et al. 2004; Chodera et al. 2011; Gumbart et al. 2013a). While these may be considered two of the
18 gold standard techniques for calculating affinity differences, there are a huge number of theoreti-
19 cal and technical complexities that must all be properly managed to ensure a converged solution
20 (Gumbart et al. 2013b). Such considerations quickly come to dominate the protocol, and the nec-
21 essary book keeping introduces the possibility of human error (Gumbart et al. 2013b). Moreover,
22 as the two ending states look ever more dissimilar the chances of convergence fall rapidly. To en-
23 sure convergence, these techniques are typically limited to small differences (such as point mutant
24 comparisons) with a few, very impressive exceptions (Wang et al. 2006; Gumbart et al. 2013a,b).
25 For most investigators, larger differences quickly become intractable as the number of intermedi-
26 ate steps required to compute a converged solution grows or the complexity of adding restraining
27 potentials and computing approximations expands (Wang et al. 2006; Gumbart et al. 2013a,b).

28 Here we propose that much of these complexities can be avoided if all we are interested in is a
29 relative comparison of the effects of different mutations on protein-protein interactions, rather than

30 measuring an absolute or relative binding affinity with experimentally realistic units. We impart
31 a pulling force within an all-atom molecular dynamics simulation on one member of the complex
32 while the other is held in place. Then, we measure the force required for dissociation (Lu and
33 Schulten 1999; Isralewitz et al. 2001b,a; Park and Schulten 2004; Gumbart et al. 2012; Miño et al.
34 2013). Although such biasing techniques are commonly used in protein-ligand binding problems,
35 they are less commonly applied to protein–protein interactions, and almost never to mutational
36 analysis in a protein–protein system. This is largely the result of free energy convergence dif-
37 ficulties and computational limitations (Cuendet and Michelin 2008; Cuendet and Zoete 2011).
38 Using a proxy for relative binding affinity rather than caluclating absolute affinities can solve these
39 problems. Here, as proxies, we use the maximum applied force required for separation and the
40 area under the force-versus-distance curve (AUC). For comparison, we also calculate relative free
41 energy differences using the traditional dual topology FEP paradigm, and we show that the two
42 approaches yield congruent results.

43 We used SMD and FEP to interrogate the interaction between machupo virus (MACV) spike
44 glycoprotein (GP1) and the human transferrin receptor (hTfR1) (Abraham et al. 2010; Charrel and
45 de Lamballerie 2003). Machupo virus is an ambisense RNA virus of the arenavirus family (Char-
46 rel and de Lamballerie 2003). Worldwide, arenaviruses represent a significant source of emerging
47 zoonotic diseases for the human population (Charrel and de Lamballerie 2003). Members of the
48 arenavirus family include the Lassa fever virus endemic to West Africa, the lymphochoriomening-
49 gitis virus (LCMV) endemic to rodents in several areas of the United States, and the Guanarito,
50 Junin, and Machupo viruses endemic to rodents in South America (Charrel and de Lamballerie
51 2003). The South American arenaviruses typically infect humans after rodent contamination and
52 can cause a devastating hemorrhagic fever with high mortality (Charrel and de Lamballerie 2003).

53 The hTfR1 is the primary receptor used by MACV for binding its host cell prior to infection.
54 The primary role of hTfR1 *in vivo* is to bind transferrin for cellular iron uptake. The hTfR1 protein
55 contains three extracellular domains: two basilar domains and an apical domain. The two basilar
56 domains serve most of the transferrin-binding function (Abraham et al. 2010; Radoshitsky et al.
57 2011). Viral entry is initiated by GP1 binding to the apical domain of hTfR1. Previous work has
58 indicated that the GP1/hTfR1 binding interaction is the primary determinant of MACV host range
59 variation (Choe et al. 2011; Radoshitsky et al. 2011). The co-crystal structure shows that the high

60 affinity interaction between GP1 and hTfR1 forces the normally flexible loop in the apical domain
61 of hTfR1 into a rigid β -pleated sheet domain. For GP1, several extended loops mediate binding to
62 hTfR1 (Abraham et al. 2010; Radoshitzky et al. 2011), and many of the interface interactions are
63 mediated by extensive hydrogen-bonding networks (Abraham et al. 2010). Experimental alanine-
64 scanning and whole-cell infectivity assays have identified several sites in both GP1 and hTfR1 that
65 are probably critical for establishing infection (Choe et al. 2011; Radoshitzky et al. 2011).

66 We applied our computational method to wild type (WT) and mutant complexes, and found
67 that we could resolve relative differences in unbinding and predict significant affinity changes.
68 Importantly, the affinity changes predicted using only max force or AUC show a strong correlation
69 with rigorous relative free energy differences computed by FEP. At sites known to be important for
70 successful viral entry, we found that the biochemical cause of reduced infectivity may not be as
71 simple as the static structure suggests. For example, the static structure shows a hydrogen-bonding
72 network connected to site N348 in hTfR1. According to our simulations, this network may not
73 affect binding affinity directly. In addition, our study offers an all-atom steered molecular dynamic
74 approach to avoid some of the pitfalls of several existing methods used to evaluate mutations in
75 protein–protein interfaces.

76 2 Materials and Methods

77 2.1 System Modeling

78 For our experiments, we used the experimentally determined GP1/hTfR1 structure (PDB-ID: 3KAS)
79 (Abraham et al. 2010). The apical domain of hTfR1 interacts directly with GP1 while the other two
80 domains are closer to the cell membrane and have essentially no interaction with GP1. The bio-
81 physical independence of the apical domain allowed us to isolate it without significantly affecting
82 the GP1/hTfR1 interaction.

83 We used the protein visualization software PyMOL (Schrödinger 2010) to remove residues
84 121-190, 301-329, and 383-756 in the hTfR1. No residues were removed from the viral protein.
85 Figure 1 shows a model of the initial structure and that of the pared structure. Although GP1 has
86 several glycosylatable residues, we opted to use the de-glycosylated protein for this study. The

87 complexity of correctly parameterizing diverse sugar moieties is outside of the scope of this paper.
88 Furthermore, although it is known that GP1 is glycosylated, and some of those sugars contact
89 hTfR1, the sugars in the available PDB structure are not physiological for mammals (Abraham
90 et al. 2010). In total we removed 10 sugars from the crystal structure for this study.

91 After system reduction, the Visual Molecular Dynamics (VMD) (Humphrey et al. 1996) pack-
92 age along with its system of back-ends was used for all subsequent modeling. The Orient add-on
93 package allowed us to rotate the system axis such that the direction of steering was oriented di-
94 rectly down the z-axis. De-glycosylation simplified the system such that Autopsf could easily find
95 the chain terminations and patch them appropriately. The Solvate package was used to generate a
96 TIP3P water model with a 5 Ångstrom buffer (relative to the maximum dimensions of the proteins)
97 on all sides except down the positive z-axis where a 20 Ångstrom buffer was created. Finally, we
98 used the Autoionize package to place 150 millimolar NaCl and neutralize the total system charge.
99 In the end, each modeled system had approximately 28,000 atoms.

100 **2.2 Equilibration**

101 NAMD was used for all simulations in this study (Phillips et al. 2005). In addition to the modeled
102 system, for equilibration we generated a configuration file that fixed the α -carbon backbone. This
103 was accomplished by setting the B-factor column to 1 for the fixed atoms and to zero for all other
104 atoms. Further, we generated a configuration file with fixed α -carbon atoms at residues 41-92 (num-
105 bered linearly, in this case, starting at 1 for the first amino acid as was required for NAMD) in the
106 hTfR1. The second file was used to affix a harmonic restraint, thus preventing any unfolding due to
107 system reduction. More importantly, the harmonic restraint allowed the protein complex to equili-
108 brate while preventing any drift from its predefined position; the restraint did not constrain the struc-
109 ture of each protein, or the relative position or orientation of the two proteins to each other. Finally,
110 we calculated the system center and dimensions for use in molecular dynamics settings. The exact
111 NAMD configuration files are available on github (https://github.com/clauswilke/MACV_SMD).

112 We used the Charmm27 (Brooks et al. 1983) all-atom force field. The initial system temperature
113 was set to 310K. Several typical MD settings were used including switching and cutoff distances
114 (see provided configuration files). In addition, we used a 2 femtosecond time step with rigid bonds.
115 We used periodic boundary conditions with the particle mesh ewald (PME) method of computing

116 full system electrostatics outside of the explicit box. Furthermore, we used a group pressure cell,
117 flexible box, langevin barostat, and lavegin thermostat during equilibration. A harmonic restraint
118 (called harmonic constraint in VMD) was set as stated previously.

119 To start the simulation, the barostat was switched off and the system was minimized for 1000
120 steps. Next, the fixed backbone was released, and the system was minimized for an additional 1000
121 time steps. Subsequently, the system was released into all-atom molecular dynamics for 3000 steps.
122 Finally, the langevin barostat was turned on and the system was simulated for 2 ns (1,000,000 steps)
123 of chemical time. For each mutant, twenty independent equilibration replicates were run with an
124 identical protocol.

125 **2.3 Steered Molecular Dynamics**

126 We used the final state from each equilibrated system to restart another MD simulation. Our steer-
127 ing protocol is fundamentally similar to Cuendet and Michelin (2008) with slightly different pa-
128 rameter choices. Perhaps the one significant difference lies in our choosing to not use a thermostat
129 or barostat. We can make this choice because we are not trying to calculate the binding free en-
130 ergy by any physically rigorous approach (the Jarzynski inequality being one example). Following
131 equilibration, the final state of each simulation was used to generate a configuration file fixing the
132 α -carbon on residues 1, 58, 73-83, 96, 136, 137, 138, and 161 (again with linear numbering) in
133 the hTfR1. These residues were selected as they are far from the binding interface and sufficiently
134 distributed to prevent any orientational motion of the receptor relative to the viral spike protein.
135 The center of mass of the α -carbons of all residues (163-318 in linear numbering) in GP1 received
136 an applied force during the simulation. The NAMD convention does not actually apply a force to
137 all α -carbon atoms but rather uses the selection to compute an initial center of mass. Then, during
138 the steering run, the single center of mass point is pulled with the parameters described below.
139 We used the same force field parameters (exclude, cutoff, switching, etc.), the same integrator pa-
140 rameters (time step, rigidbonds on, all molecular being wrapped, etc.), and the same particle mesh
141 ewald parameters as in equilibration. Periodic boundary conditions were incorporated as part of
142 the system (as is the convention in NAMD restart) and PME was again used to approximate full
143 system electrostatics.

144 We ran test simulations at several force constants and visually inspected the results. A force

145 constant of 5 kcal/mol/Å² was chosen due to its relatively low signal-to-noise ratio. This constant
146 is slightly lower than the more common 7 kcal/mol/Å² found in several recent studies; that value is
147 commonly selected primarily because it is the force constant found in the SMD tutorial available
148 through the NAMD developers. Moreover, the force constant could very likely be set to a range of
149 nearby values with little loss in predictive power.

150 In SMD experiments the pulling velocity should be as low as possible for the available com-
151 putational time (Cuendet and Michielin 2008; Cuendet and Zoete 2011). We choose a velocity of
152 0.000001 Å/fs = 1 Å/ns, and direction down the positive *z*-axis. One could use faster pulling if
153 the computing time must be reduced, but slower than necessary pulling speeds are not typically
154 considered problematic.

155 SMD was run for 15 ns (7,500,000 time steps) of chemical time. For each simulation, we
156 randomly selected one of the equilibration runs for restart. We ran 50 replicate simulations per
157 mutant for a total of 550 SMD simulations. All GP1/hTfR1 complexes separated by greater than 4
158 Å and many separated to 10 or more.

159 To leave the final trajectory of a tractable size, only 1000 evenly spaced frames were retained
160 from each simulation, leaving a final trajectory size of 323 MB. See the supplemental movie for
161 a representative unbinding trajectory. Initial development of the SMD protocol was carried out
162 on the Lonestar cluster at the Texas Advanced Computing Center (TACC). All production SMD
163 simulations were performed on the Hrothgar cluster at Texas Tech University, using NAMD 2.9.
164 Each simulation was parallelized over 60 computational cores and utilized approximately 20 hours
165 of computing time. The total chemical time simulated for this project was nearly 10 μs, requiring
166 slightly over 1 million cpu-hours.

167 2.4 Free Energy Perturbation

168 Briefly, we used the traditional dual topology approach to FEP (Gao et al. 1989; Pearlman 1989).
169 This involves a thermodynamic cycle where a set of atoms are progressively decoupled from the
170 environment while another set of atoms are progressively coupled. To compute the relative free
171 energy difference requires knowing the free energy change when the transformation is carried out
172 for the bound complex and the individual protein. Then, one can compute the relative free energy
173 difference between a WT and mutant complex by taking the difference between the energy required

174 to decouple/couple the atoms in solution from the energy required to decouple/couple the atoms in
175 the bound complex (Gao et al. 1989; Pearlman 1989).

176 Again, the NAMD configuration file is made available via github (https://github.com/clauswilke/MACV_SMD)
177 We used a similar configuration to that in equilibration. One significant difference was to make a
178 cubic water box with a side length equal to the long axis of the complex plus a 10 Å buffer on either
179 side, and simply restrict center of mass motion with the NAMD setting. This was done to avoid
180 affecting the system energy while calculating free energy differences.

181 The transition protocol for bound and free protein systems were identical. They started with
182 1000 steps of minimization and 250,000 steps of equilibration in the starting state for the forward
183 and reverse directions. Phase transitions were carried out in steps of $\lambda=0.05$. Each transition
184 was carried out for 250,000 steps. The first 100,000 steps after phase transition were reserved for
185 equilibration and the final 150,000 steps were used for data collection.

186 The VMD mutator tool was used to generate the necessary topology file and the parseFEP tool
187 (Liu et al. 2012) in VMD was used for subsequent analysis. We used it to perform error analysis
188 and compute the Bennett acceptance ratio as the maximum likelihood free energy difference of the
189 two states under consideration. Though the larger transitions presented difficulty in a small number
190 of windows, forward and reverse hysteresis was generally in good agreement for all complexes.
191 The double mutants were performed by first doing the Y211A mutation followed by the other of
192 the two mutants. Then, the ΔG 's were simply added together to get the total energetic difference.

193 2.5 Post-processing

194 The python packages MDAnalysis (Michaud-Agrawal et al. 2011) and ProDy (Bakan et al. 2011)
195 were both used at various points in post-processing. The molecular trajectory (comprising the
196 atomic coordinates per time) was parsed to compute the center-of-mass for each of the two com-
197 plexes. The starting center-of-mass distance was set to zero and the distance was re-computed at
198 each time step relative to the starting distance.

199 The statistical package R was used for all further analysis and visualization. Each of the 50
200 independent trajectories per mutant produced a fairly noisy force curve. The force curves for each
201 mutant were smoothed over all replicates by using the smooth.spline() and predict() functions in R
202 with default settings. The two primary descriptive statistics we used were maximum interpolated

203 applied force and total area under the interpolated curve (AUC). We tested for significant differences
204 in maximum force or AUC by carrying out t tests for all pairwise combinations (each mutant com-
205 pared to each other mutant), using the pairwise.t.test() function in R. We adjusted p values to cor-
206 rect for multiple testing using the False-Discovery-Rate (FDR) method (Benjamini and Hochberg
207 1995). The ggplot (Wickham 2009) package was used to generate most of the figures.

208 Analysis scripts and final data (except MD trajectories) are available on the github repository
209 accompanying this publication (https://github.com/clauswilke/MACV_SMD).

210 3 Results

211 3.1 The GP1/hTfR1 system

212 The GP1/hTfR1 interface (Figure 2) marks a particularly important and useful test system. There
213 are several sites on both the human and viral protein known to affect the infectivity phenotype
214 of MACV. Many of the important sites have been mapped by *in vitro* flow-cytometry based entry
215 assays. The GP1/hTfR1 interface appears not to be dominated by one particular type of interaction
216 (electrostatics, hydrogen-bonding, or van der Waals). In addition, much of the binding domain
217 on hTfR1 is on a loop that is flexible prior to viral binding, but organizes to become a strand
218 of a β -sheet on binding. As a result, many other computational techniques (Gront et al. 2011;
219 Kortemme et al. 2004) are only marginally useful. The complex nature of this interface represents
220 a particularly difficult challenge for traditional computational analysis.

221 In total, we tested 7 point mutants and 3 double mutants in addition to the WT complex (Ta-
222 ble 1). All of the mutations are within 5 Å of the protein–protein interface. Mutations in hTfR1
223 at site 211 have proven capable of causing loss-of-entry according to *in vitro* flow-cytometry in-
224 fection assays or known host-range limitations (Radoshitzky et al. 2008; Choe et al. 2011; Ra-
225 doshitzky et al. 2011). Most likely, this effect is caused by the destruction of a critical hydrogen
226 bond to Ser113 or Ser111 in GP1. The lost hydrogen bond would lead to the subsequent loss of a
227 large hydrogen-bonding network seen in the crystal structure (Table 1) (Abraham et al. 2010). In a
228 manner similar to site 211, N348 appears to be important for binding by participating in a critical
229 hydrogen bonding network (Radoshitzky et al. 2008; Abraham et al. 2010) to GP1. In particular,

230 N348Lys is reported in the literature to cause significantly reduced viral entry *in vivo* (Table 1)
231 (Radoshitsky et al. 2008; Abraham et al. 2010). Finally, an alanine mutation at site 111 in GP1
232 (mutation vR111A) has also been shown to cause decreased entry (Table 1) (Radoshitsky et al.
233 2011). For notation purposes, the viral site is always referred to with a preceding ‘v’.

234 Despite the fact that viral binding occurs at the site of a flexible loop in the free hTfR structure,
235 our data shows after binding the strand is extremely rigid. In the bound conformation, only two
236 sites of the loop have root mean squared fluctuation (RMSF) values in the top half of all receptor
237 sites during equilibration (Figure 3), and those are almost completely exposed to solvent. This is
238 unsurprising considering the high degree of burial that occurs as a result of viral binding. Com-
239 puting the root mean squared deviation (RMSD) of the entire structure over the trajectory shows
240 that none of the mutations are so deleterious as to cause rapid unbinding. In fact, the RMSD over
241 trajectory looks highly invariant across mutants (Figure 4). In the unbound state, calculated near
242 the end of the SMD trajectory, all of the residues in the WT receptor interfacial strand are in the
243 top half of RMSF over all receptor sites (Figure 5). Thus, if sufficient simulation time is not ded-
244 icated to allowing this unfolding process, standard free energy techniques may miss the energetic
245 contributions that result from ordering the flexible loop in the hTfR apical domain.

246 3.2 Molecular dynamics simulations

247 We analyzed the GP1/hTfR1 system using two molecular dynamics techniques. First, by carrying
248 out SMD using a known force constant and pulling with a constant velocity, we could calculate
249 the applied force during protein–protein dissociation (Cuendet and Michelin 2008; Cuendet and
250 Zoete 2011). A typical averaged force curve comparison can be seen in Figure 6, and individual
251 images of all averaged force curves are available in the associated github repository, in folder
252 figures/force_curves. As seen in Figure 6, the dissociation distance was relatively consistent among
253 mutants. The supplementary movie visually illustrates the separation distance between peptide
254 domains. The quantities maximum applied force and AUC were derived from the force-versus-
255 distances curves. Their summary statistics are reported in Table 2. As we are more interested in
256 the phenotypic impact of interface mutations we avoided many of the more physically rigorous, but
257 technically complicated calculations that are possible with SMD (Isralewitz et al. 2001b,a).

258 Before systematically applying SMD to the GP1/hTfR1 interaction, we needed to ensure the

method was sufficiently sensitive to distinguish between relatively minor point mutations. While SMD has been applied previously to measure the binding energy of high-affinity T-cell receptor interactions (Cuendet and Michelin 2008; Cuendet and Zoete 2011), it is rarely used to parse small energy differences in a protein–protein interaction energy landscape. For this initial sensitivity analysis, we tested alanine substitutions congruent with the traditional experimental and computational approach.

We proceeded to compare our SMD results to that of the standard dual topology FEP approach to calculate relative free energy differences. The correlation between the energetically rigorous FEP and our statistical approach is high. For all 11 complexes tested, the correlation between max force and FEP was $r = -0.795$ at $p = 0.0034$ (Figure 7), and the correlation between AUC and FEP was $r = -0.593$ at $p = 0.055$. Because of the strong correlation, we refer exclusively to the SMD results for the remainder of this work, focusing primarily on max force.

We found that relative to WT, one alanine mutation (Y211A) produced a very large and statistically significant difference in the maximum applied force and AUC (Figure 6, Table 3), while the other two did not (Table 3). When considering additional mutants (also discussed below), we found that maximum applied force was generally sufficient to distinguish mutants (Tables 3 and 4), and AUC was able to add a few more statistically significant differences (Table 5). In general, however, and consistent with the FEP results, maximum applied force seemed to be the more sensitive statistic than AUC.

3.3 Comparative analysis of the GP1/hTfR1 interface

Considering the involvement of extended hydrogen-bonding networks in the GP1/hTfR1 interface (Figure 2), it was not clear that individual alanine mutations, even those that should destroy such networks, would significantly change the strength of interaction. One major advantage of first principles simulations is the ability to test mutations other than alanine without additional underlying assumptions in the energy function. As shown in Table 1, we made additional mutations based on biochemical intuition or available experimental data to chemically diverse amino acids including tryptophan, lysine, aspartate, and threonine. Several mutations caused significant relative affinity changes. In addition, to detect synergistic effects, we tested several double mutants where both mutations appeared to cause similar changes in binding. Then, we compared the size of those

288 differences to single mutants (Figure 8 and 9).

289 Although Y211A appears to have a large impact on binding affinity, no single mutant can pro-
290 vide enough evidence to understand the biochemical difference in binding mechanism. Since ala-
291 nine is both smaller than tyrosine and also incapable of participating in hydrogen-bond interactions,
292 we tested further mutations to identify the critical biochemical difference responsible for change in
293 binding affinity. In particular, we substituted smaller side chains that, like tyrosine, were capable
294 of hydrogen bonding. We chose Y211D and Y211T, two mutations that have been discussed in the
295 context of selection pressure on hosts in rodent populations (Radoshitzky et al. 2008; Choe et al.
296 2011; Radoshitzky et al. 2011). Both mutations proved capable of causing a significant change
297 in binding affinity in our simulations, but the change appeared to be increased affinity (Figures 8
298 and 9, and Table 4).

299 We also simulated several point mutations at N348 in the hTfR1. As discussed above, the
300 alanine mutation at this site showed no significant difference in maximum applied force or AUC
301 from WT (Tables 4 and 5). In addition, neither the N348Lys nor the N348W mutation showed
302 a significant difference from WT. For both of these mutations, however, mean maximum applied
303 force and mean AUC was lower than for WT (See Table 2). On the other hand, there was a de-
304 tectable difference between N348A and N348Lys (Tables 4 and 5), with N348Lys being a weaker
305 binder. Moreover, N348W showed nearly identical results to N348Lys. The mutations to large
306 amino acids (N348W and N348Lys) produced nearly identical affinity changes, whereas the muta-
307 tions to amino acids not capable of hydrogen bonding (N348A and N348W) produced significantly
308 different affinity changes (Table 3). To check the consistency of our results, we hypothesized that
309 the combination of Y211A and N348W, being chemically disconnected in two different hydrogen-
310 bonding networks, would lead to a synergistic loss-of-binding. As expected, the double mutant
311 was the weakest binding mutant tested ($p < 10^{-6}$, Tables 4 and 5) in this study. Further, according
312 to maximum applied force (but not AUC), the combination of Y211A and N348W also showed
313 significantly weaker binding than Y211A by itself (Tables 4 and 5). We suspect that the effect
314 of N348W alone is near the limit of detection using our method. A larger number of replicates
315 would possibly have resolved affinity differences between N348W and WT or other mutants more
316 consistently.

317 Last, we further analyzed a single mutation in GP1, vR111A. As mentioned previously, in our

318 simulations this mutant showed no significant change in either maximum applied force or AUC
319 (Tables 4 and 5), even though both quantities were, on average, lower than in WT (Table 2). This
320 result was somewhat surprising, since Y211A, presumably disrupting the same hydrogen-bonding
321 network as vR111A, displayed a significant reduction in affinity. To probe the interaction between
322 position 111 in the GP1 and position 211 in the hTfR1 further, we also tested the double mutant
323 vR111A/Y211A. This double mutant showed affinity indistinguishable from WT and significantly
324 higher than Y211A alone (Table 3). This result shows that the two sites do indeed interact, and that
325 replacing the hydrogen-bonding network at these sites with a hydrophobic interaction could lead to
326 comparable binding affinity.

327 4 Discussion

328 We have applied a method utilizing steering forces in all-atom molecular dynamics simulations to
329 evaluate the effects of mutations at the GP1/hTfR1 interface. We modeled mutations at several sites
330 in the GP1/hTfR1 interface, and verified that our computational protocol was sensitive enough to
331 distinguish point mutants in hTfR1. Further, we identified two test statistics, maximum applied
332 force and AUC, that can be used as proxies for binding affinity. Both of these statistics correlate
333 well with FEP, but offer the simplicity of not requiring a large commitment to planning for the
334 theoretical issues inherent to free energy methods. We systematically tested several point mutations
335 to understand their contribution to the binding interaction. In the case of N348Lys, we have shown
336 that the static structure provides little insight into why this mutation causes loss-of-infectivity *in*
337 *vivo*. While N348 appears to be involved in a hydrogen-bonding network in the static structure,
338 change in binding at that site may actually be caused by size and charge restriction. We also found
339 that a negatively polar residue at site 211 in hTfR1 seem critical for a tight binding interaction.
340 Any non-polar mutation at Y211 in hTfR1 is likely to completely halt viral entry and dramatically
341 decrease the chances of MACV infection.

342 Traditionally SMD has been either applied to compute equilibrium free energies via a non-
343 equilibrium approximation (Park et al. 2003; Park and Schulten 2004; Giorgino and Fabritiis 2011),
344 used to estimate protein stability through unfolding (Lu and Schulten 1999), or used to calculate the
345 absolute free energy of small molecule ligand binding (Dixit and Chipot 2001). Likewise, others

346 have used SMD to understand the process of binding and unbinding at a resolution unmatched by
347 experiment (Cuendet and Zoete 2011; Giorgino and Fabritiis 2011). Here, we have shown that
348 SMD can provide insight into the *relative* strength of protein–protein interactions. Via SMD, one
349 can separate mutations whose likely effect is altered binding affinity with simple statistics like
350 maximum force of separation. Thus, SMD may open avenues for subsequent experimental work in
351 some situations where FEP may be prohibitively difficult.

352 Our findings rationalize several effects observed in both infectivity data and rodent populations
353 (Radoshitzky et al. 2008; Choe et al. 2011). First, we found that some substitutions at positions
354 211 and 348 did affect the strength of receptor binding. However, the computational data suggest
355 that the reason and nature of the effects at these two sites are very different. At position 211,
356 mutations to non-polar residues cause a large change in binding. This is congruent with what is
357 known from viral entry data (Radoshitzky et al. 2008; Choe et al. 2011). By contrast, mutations at
358 position 348 need only be small to maintain WT binding. The ability to hydrogen bond appears to
359 be insignificant. This can be inferred from the fact that Y211A paired with large (W) and positively
360 charged (Lys) substitutions at position 348 results in a larger than expected synergistic difference.
361 That is, the double mutant Y211A/N348W caused a much larger decrease in binding than we
362 expected from either mutation individually. Third, the GP1 mutation vR111A causes a loss-of-
363 infection during *in vitro* infectivity assays (Radoshitzky et al. 2011), yet it was indistinguishable
364 from the WT complex in our simulations. Although Y211A was the most disruptive single mutant
365 we tested, vR111A in the GP1 was able to restore mean maximum applied force to WT levels
366 (Table 2), and to levels significantly higher than observed for Y211A alone.

367 We would like to emphasize here that we cannot expect perfect agreement between our simula-
368 tions and the available experimental data, but the correspondence to a well established free energy
369 method bolsters our conclusions. While we have shown that our method can distinguish individual
370 point mutations, we do not know the limit of detection with our method. First, it is possible that
371 some mutants display measurable phenotypic effects in experiments yet appear identical in simula-
372 tion. More extensive sampling or refinement of the simulation protocol could help to differentiate
373 such mutants (see also next paragraph). Second, the SMD method is fundamentally limited by the
374 accuracy of our starting structure. Third, the available experimental data for the GP1/hTfR1 sys-
375 tem were generally obtained from entry assays or whole-cell binding assays rather than molecular

376 binding assays. A mutant may cause a phenotypic difference in infectivity without generating a
377 signal by our method. For example, entry could be lost in the experimental system because the
378 protein is grossly or partially misfolded. An additional analytical step with circular dichroism or an
379 analogous technique could clarify such large-scale folding differences. Further, since our simula-
380 tions start with a bound structure, any changes that may dramatically affect the rate of association
381 (different folds, trafficking issues, etc.) or relative orientation of the two proteins would be under-
382 estimated by our method.

383 There are a few additional challenges for investigating host-virus interactions via molecular
384 dynamics simulation. As with any atomistic simulation, there is going to be a fairly large noise-to-
385 signal ratio. To reduce noise, one could further customize each simulation, e.g. by determining the
386 optimal pulling speed. Furthermore, larger amounts of computational resources will have a direct
387 and powerful impact on the strength of any atomistic study (Jensen et al. 2012). Such resources
388 could come in the form of increased compute time, improved code, or customized hardware for
389 floating point operations (Shaw et al. 2011). With improved resources, we could investigate thou-
390 sands of individual permutations in the GP1/hTfR1 binding interface. In addition, with additional
391 compute time it would be possible to incorporate equilibrium sampling approaches (Buch et al.
392 2011) or use brute force equilibrium approaches (Giorgino et al. 2012) to improve resolution.

393 For future studies, although our approach offers the simplicity of not requiring prior knowledge
394 about a system of interest (other than a bound model), at this point SMD may not be the best approach
395 for many relative affinity calculations. To ensure one's results are independent of the dissociation
396 path one selects would require computing the work of separation for all likely paths. Such an
397 approach eventually requires using the Jarzynski inequality (Jarzynski 1997) to establish a lower
398 limit for binding energy and would quickly become computationally inefficient for evaluating a
399 large number of mutations in most systems. However, considering the strong correlation between
400 FEP and SMD in this system, it may not be important to ensure one's results are path independent
401 for relative affinity calculations, as long as the same path is used for all complexes.

402 More importantly, with no *a priori* knowledge of the appropriate number of equilibration sam-
403 ples, the best duration of equilibration, the appropriate number of pulling runs, or the best pulling
404 speed means the computational expense in our SMD protocol may not be commensurate with the
405 information provided. For example, another all atom approach that makes calculations via short

406 simulations of spatially restrained complexes has proven capable of generating relatively accurate
407 binding affinities with less compute time than is required from our steering strategy (Gumbart et al.
408 2013a,b). That being said, there is no reason to believe this SMD approach to mutagenic stud-
409 ies could not be optimized to reduce computational expense. Further analysis will be needed to
410 understand the lower limits of resources required for accurate predictions.

411 5 Acknowledgements

412 This work was carried out using high-performance computing resources provided by the High Per-
413 formance Computing Center (HPCC) at Texas Tech University at Lubbock (<http://www.hpcc.ttu.edu>)
414 and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (<http://www.tacc.utexas.edu>).
415 We would like to thank Bryan Sutton for opening access to the Hrothgar cluster and the reviewers
416 Ilan Samish and Matteo Masetti for their helpful comments on this work.

417 References

- 418 **Abraham J, Corbett K, Farzan M, Choe H, Harrison S. 2010.** Structural basis for receptor
419 recognition by new world hemorrhagic fever arenaviruses. *Nat. Struct. Mol. Bio.* **17**:438–444.
- 420 **Bajaj C, Chowdhury R, Siddhanavalli V. 2011.** *F²Dock*: Fast fourier protein-protein docking.
421 *IEEE Transactions on Computational Biology and Bioinformatics* **8**:45–58.
- 422 **Bakan A, Meireles L, Bahar I. 2011.** ProDy: Protein dynamics inferred from theory and experi-
423 ments. *Bioinformatics* **27**:1575–1577.
- 424 **Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful
425 approach to multiple testing. *J. Royal Stat. Soc., Series B* **57**:289–300.
- 426 **Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. 1983.** CHARMM:
427 A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput.*
428 *Chem.* **4**:187–217.

- 429 **Buch I, Sadiq S, Fabritiis GD.** 2011. Optimized potential of mean force calculations for standard
430 binding free energies. *J. Chem. Theory Comput.* **7**:1765–1772.
- 431 **Charrel RN, de Lamballerie X.** 2003. Arenaviruses other than Lassa virus. *Antiviral Res.* **57**:89–
432 100.
- 433 **Chodera J, Mobley D, Shirts M, Dixon R, Branson K, Pande V.** 2011. Alchemical free energy
434 methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology*
435 **21**:150–160.
- 436 **Choe H, Jemielity S, Abraham J, Radoshitzky S, Farzan M.** 2011. Transferrin receptor 1 in the
437 zoonosis and pathogenesis of new world hemorrhagic fever arenaviruses. *Current Opinion in*
438 *Microbiology* **12**:467–482.
- 439 **Cuendet M, Michelin O.** 2008. Protein-protein interaction investigated by steered molecular
440 dynamics: The TCR-PMHC complex. *Biophysical J.* **95**:3575–3590.
- 441 **Cuendet M, Zoete V.** 2011. How T-cell receptors interact with peptide-MHCs: A multiple steered
442 molecular dynamics study. *Proteins* **79**:3007–3024.
- 443 **Dixit S, Chipot C.** 2001. Can absolute free energies of association be estimated from molecular
444 mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **105**:9795–
445 9799.
- 446 **Gao J, Kuczera K, Tidor B, Karplus M.** 1989. Hidden thermodynamics of mutant proteins: A
447 molecular dynamics analysis. *Science* **244**:1069–1072.
- 448 **Gilson MK, Given JA, Bush BL, McCammon JA.** 1997. The statistical-thermodynamic basis for
449 computation of binding affinities: A critical review. *Biophys. J* **72**:11047–1069.
- 450 **Giorgino T, Buch I, Fabritiis GD.** 2012. Visualizing the induced binding of SH2-phosphopeptide.
451 *J. Chem. Theory Comput.* **8**:1171–1175.
- 452 **Giorgino T, Fabritiis GD.** 2011. A high-throughput steered molecular dynamics study on the free
453 energy profile of ion permeation through Gramicidin A. *J. Chem. Theory Comput.* **7**:1943–1950.

- 454 **Gront D, Kulp DW, Vernon RM, Strauss CEM, Baker D.** 2011. Generalized fragment picking
455 in Rosetta: Design, protocols and applications. *PLoS ONE* **6**.
- 456 **Gumbart J, Roux B, Chipot C.** 2013a. Standard binding free energies from computer simulations:
457 What is the best strategy? *J. Chem. Theory Comput.* **9**:794–802.
- 458 **Gumbart J, Schreiner E, Wilson D, Beckmann R, Schulten K.** 2012. Mechanisms of SecM-
459 mediated stalling in the ribosome. *Biophys. J.* **203**:331–341.
- 460 **Gumbart JC, Roux B, Chipot C.** 2013b. Efficient determination of protein-protein standard
461 binding free energies from first principles. *J. Chem. Theory Comput.* **9**:3780–3798.
- 462 **Humphrey W, Dalke A, Schulten K.** 1996. VMD - Visual Molecular Dynamics. *J. Molec.*
463 *Graphics* **14**:33–38.
- 464 **Hwang H, Vreven T, Janin J, Weng Z.** 2010. Protein-protein docking benchmark version 4.0.
465 *Proteins* **78**:3111–3114.
- 466 **Isralewitz B, Baudrya J, Gullingsruda J, Kosztinav D, Schulten K.** 2001a. Steered molecular
467 dynamics investigations of protein function. *J. Molec. Graphics* **19**:13–25.
- 468 **Isralewitz B, Gao M, Schulten K.** 2001b. Steered molecular dynamics and mechanical functions
469 of proteins. *Current Opinion in Structural Biology* **25**:225–230.
- 470 **Jarzynski C.** 1997. Equilibrium free-energy differences from nonequilibrium measurements: A
471 master-equation approach. *Physical Review E* **56**:5018–5035.
- 472 **Jensen M, Jogini V, Borhani D, Leffler A, Dror R, Shaw D.** 2012. Mechanism of voltage gating
473 in potassium channels. *Science* **336**:229–233.
- 474 **Kortemme T, Kim DE, Baker D.** 2004. Computational alanine scanning of protein-protein inter-
475 faces. *Sci. STKE* **2004**:12.
- 476 **Liu P, Dehez F, Cai W, Chipot C.** 2012. A toolkit for the analysis of free-energy perturbation
477 calculations. *J. Chem. Theor. Comput.* **8**:2606–2616.

- 478 **Lu H, Schulten K. 1999.** Steered molecular dynamics simulations of force-induced protein domain
479 unfolding. *Protein Struct. Funct. Genet.* **35**:453–463.
- 480 **Lu N, Kofke DA, Woolf TB. 2004.** Improving the efficiency and reliability of free energy pertur-
481 bation calculations using overlap sampling methods. *J. Comput. Chem.* **25**:28–39.
- 482 **Miño G, Baez M, Gutierrez G. 2013.** Effect of mutation at the interface of Trp-repressor dimeric
483 protein: a steered molecular dynamics simulation. *Eur. Biophys. J.* **42**:683–690.
- 484 **Michaud-Agrawal N, Denning E, Woolf T, Beckstein O. 2011.** MDAnalysis: A toolkit for the
485 analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**:2319–2327.
- 486 **Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. 2003.** Free energy calculation from steered
487 molecular dynamics simulations using Jarzynski’s equality. *J. Chem. Phys.* **119**:3559–3567.
- 488 **Park S, Schulten K. 2004.** Calculating potentials of mean force from steered molecular dynamics
489 simulations. *J. Chem. Phys.* **120**:5946–5961.
- 490 **Pearlman DA. 1989.** A comparison of alternative approaches to free energy calculations. *J. Phys.*
491 *Chem.* **98**:1487–1493.
- 492 **Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L,**
493 **Schulten K. 2005.** Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**:1781–1802.
- 494 **Radoshitsky S, Kuhn J, Spiropoulou C, Albarino C, Nguyen D, Salazar-Bravo J, Dorfman
495 T, Lee A, Wang E, Ross S, Choe H, Farzan M. 2008.** Receptor determinants of zoonotic
496 transmission of new world hemorrhagic fever arenaviruses. *Proc. Natl. Acad. Sci.* **19**:2664–
497 2669.
- 498 **Radoshitsky S, Longobardi L, Kuhn J, Retterer C, Clester J, K Kota JC, Bavari S. 2011.**
499 Machupo virus glycoprotein determinants for human transferrin receptor 1 binding and cell entry.
500 *PLoS ONE* **6**.
- 501 **Schrödinger. 2010.** The PyMOL molecular graphics system, version 1.3r1. Schrödinger, LLC.

- 502 Shaw D, Dror R, Salmon J, Grossman J, Mackenzie K, Bank J, Young C, Deneroff M, Batson
503 B, Bowers K, Chow E, Eastwood M, Ierardi D, Klepeis J, Kuskin J, Larson R, Lindorff-
504 Larsen K, Maragakis P, Moraes M, Piana S, Shan Y, Towles B. 2011. Millisecond-scale
505 molecular dynamics simulations on anton. *Proceedings of the Conference on High Performance*
506 *Computing, Networking, Storage and Analysis* SC09:39.
- 507 Vreven T, Hwang H, Pierce B, Weng Z. 2012. Prediction of protein-protein binding free energies.
508 *Protein Science* 21:396–404.
- 509 Vreven T, Hwang H, Weng Z. 2011. Integrating atom-based and residue-based scoring functions
510 for protein-protein docking. *Protein Science* 20:1576–1586.
- 511 Wang J, Deng Y, Roux B. 2006. Absolute binding free energy calculations using molecular dy-
512 namics simulations with restraining potentials. *Biophys. J.* 91:2798–2814.
- 513 Wickham H. 2009. *ggplot2: Elegant graphics for data analysis.* Springer New York. URL
514 <http://had.co.nz/ggplot2/book>.

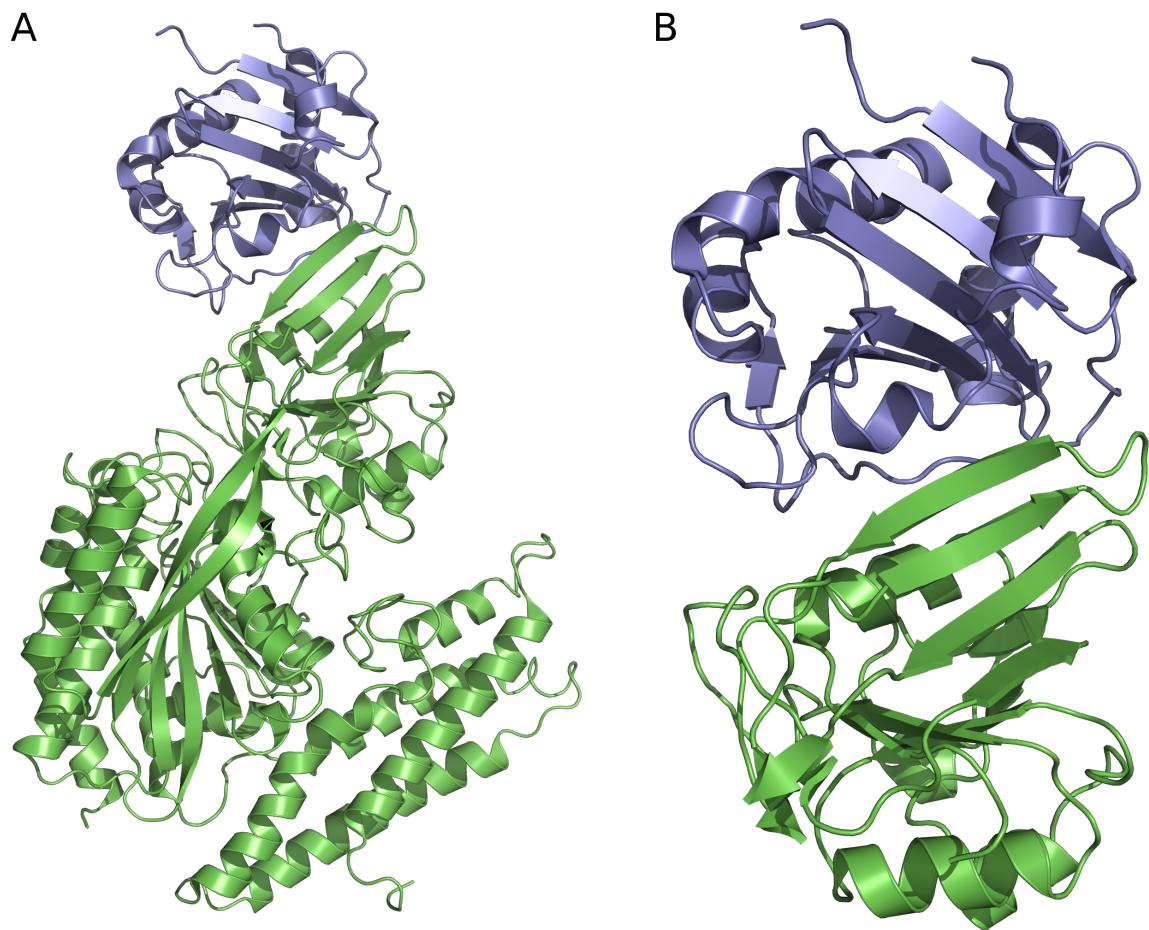


Figure 1: The GP1/hTfR1 complex. GP1 is shown in blue and hTfR1 is shown in green. (A) The full, de-glycosylated GP1/hTfR1 co-crystal structure. (B) The reduced structure used in SMD simulations.

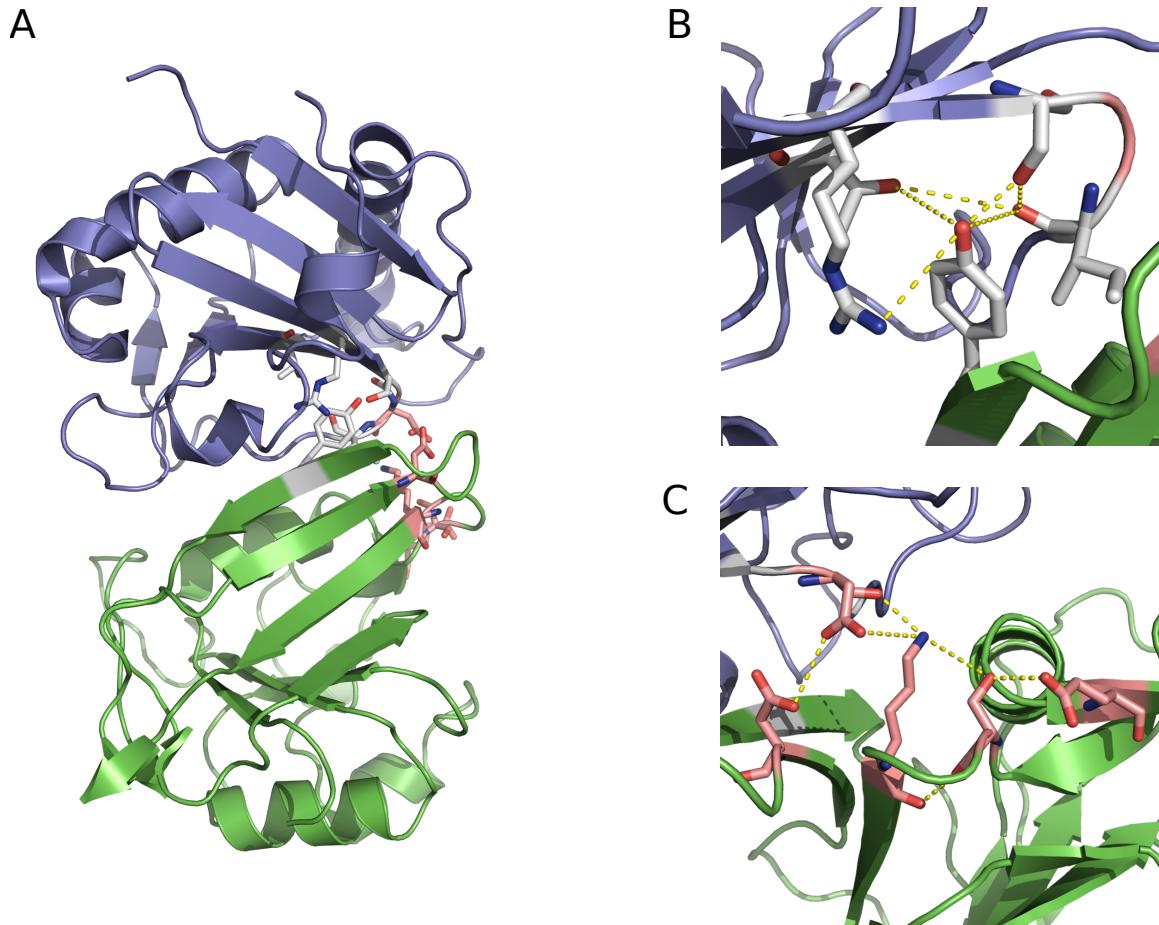


Figure 2: The two hydrogen bonding networks. GP1 is shown in blue and hTfR1 is shown in green. (A) The first network including Y211 and R111 is shown in white, and the second network containing N348 is shown in pink. (B) Near view of the first network with contacts in yellow. (C) Near view of the second network with contacts in yellow.

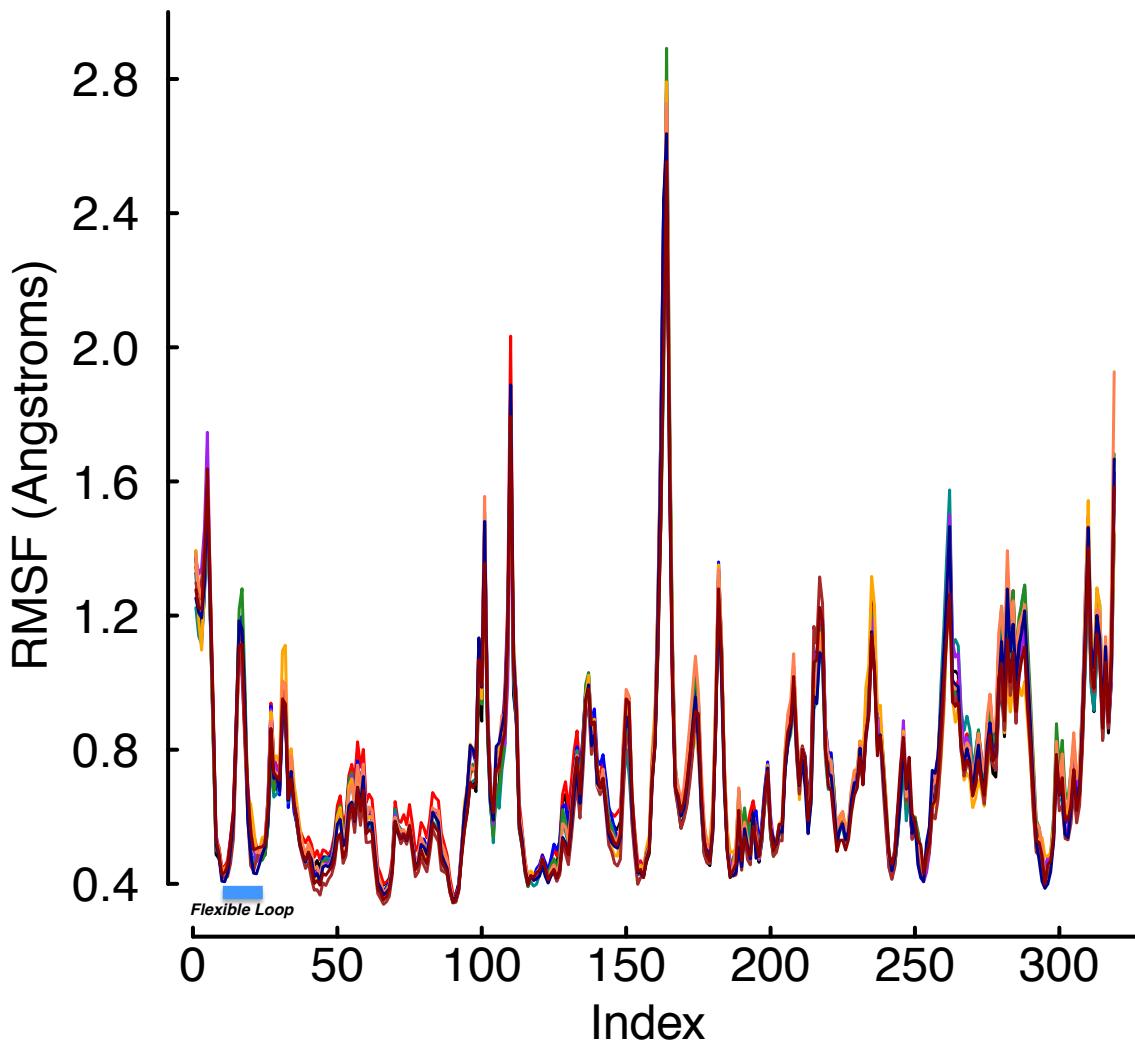


Figure 3: RMSF values during equilibration. The RMSF values for every site in the bound complex computed during the equilibration phase of the protocol. Each color represents the average over 20 trajectories of a single mutant. Indices 17–25 are the hTfR flexible loop. The plot shows the flexibility of each site is essentially independent of mutation, and two sites (indices 17 and 18) above 0.72 Å are a part of the flexible loop in the free receptor. However, these two residues are not actually found in the protein–protein interface, but rather are almost completely solvent exposed with the virus bound.

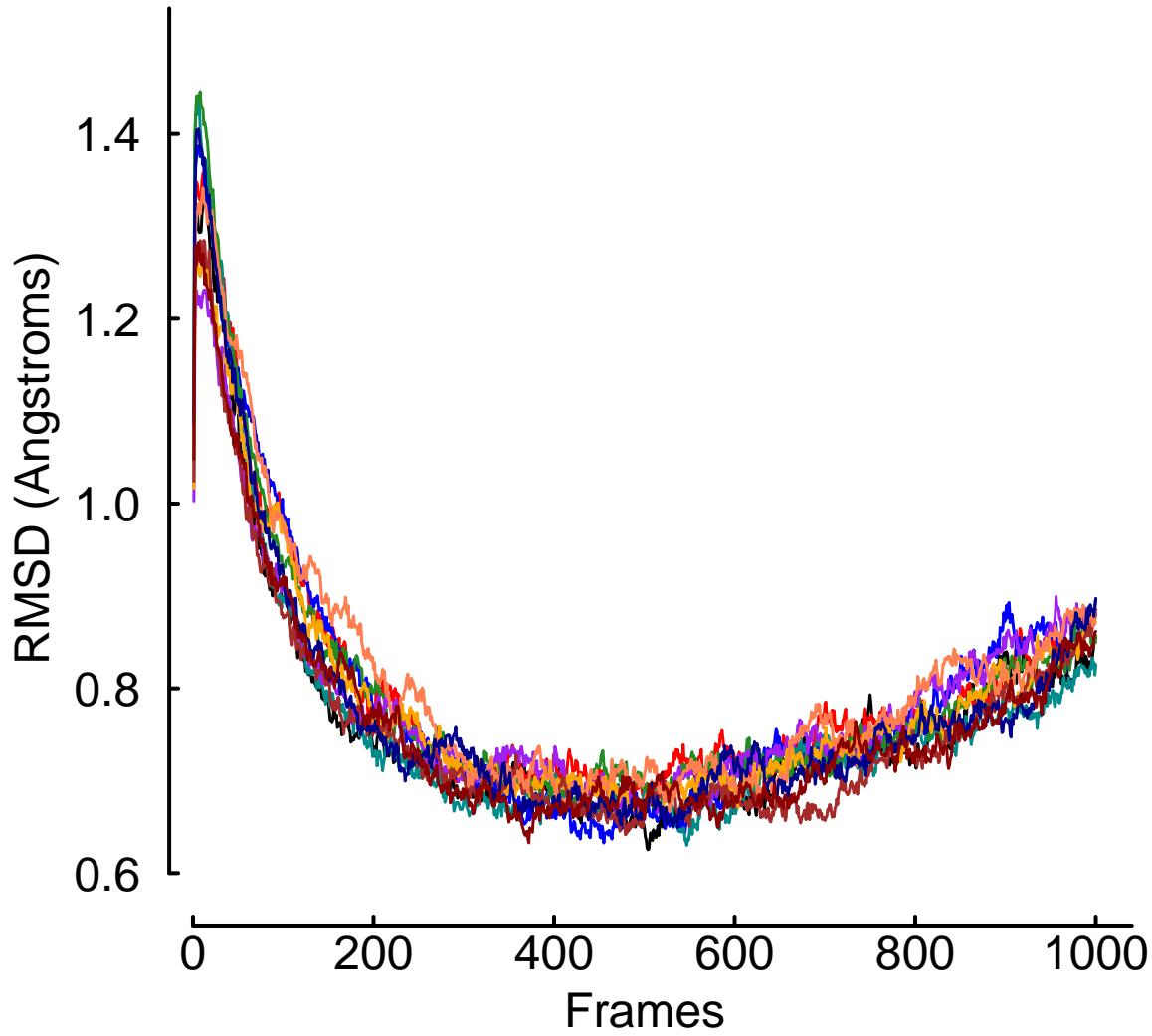


Figure 4: RMSD values during equilibration. The RMSD values over the time of the trajectory computed during the equilibration phase of the protocol. Each color represents the average over 20 trajectories of a single mutant. The plot shows none of the mutants causes immediate unbinding of the protein–protein complex. In addition, the universal upward trend near the end of the equilibration trajectories may indicate the crystal is more tightly packed than would normally occur in solution.

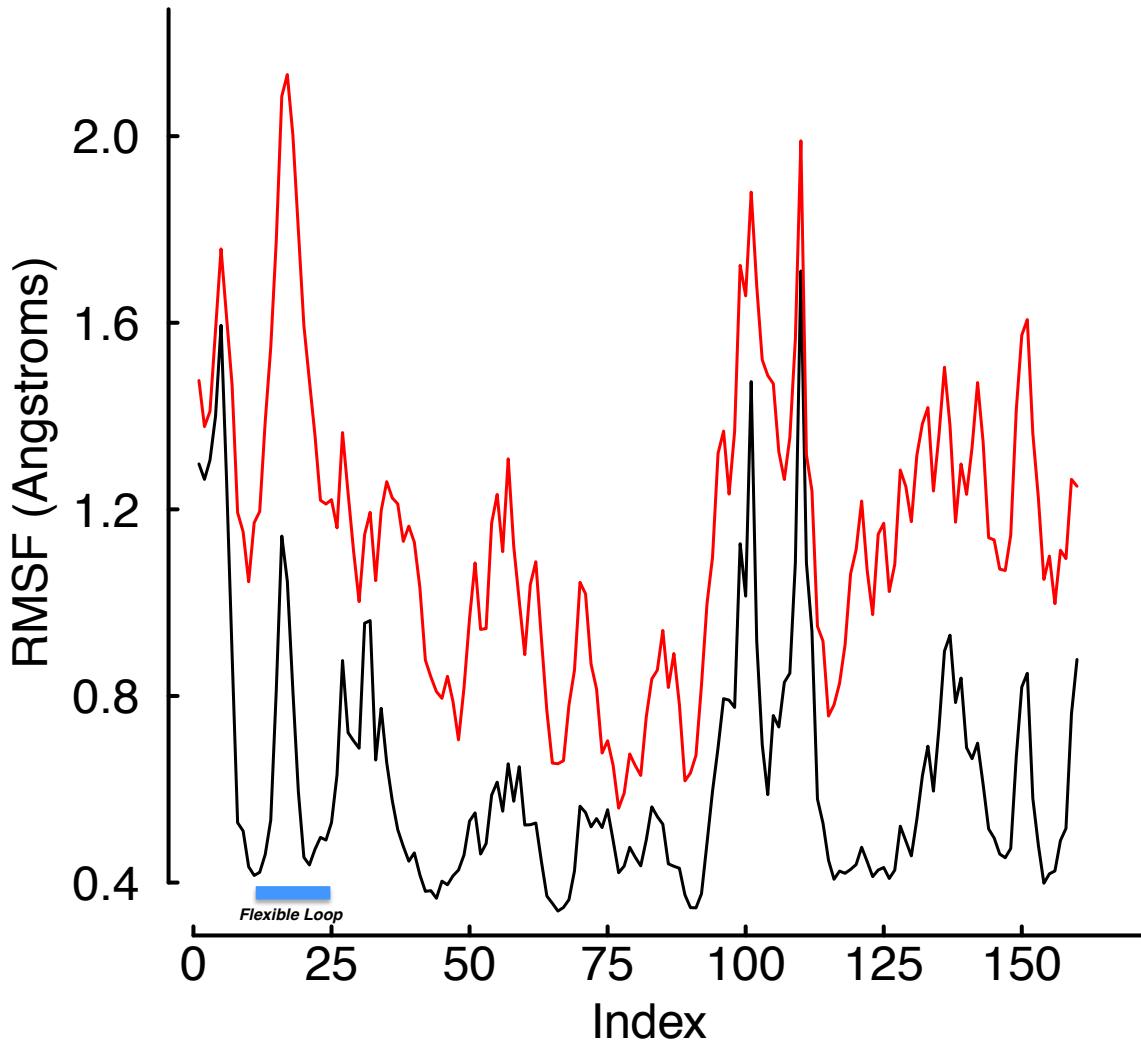


Figure 5: RMSF values of WT hTfR in equilibration and SMD. The RMSF values for every site in the WT receptor were computed during the equilibration phase and during final 50 frames of the SMD trajectories. The black line was computed over equilibration and the red line during SMD. The plot shows the solution mobility of the hTfR flexible loop increases more than the average during the unbinding process.

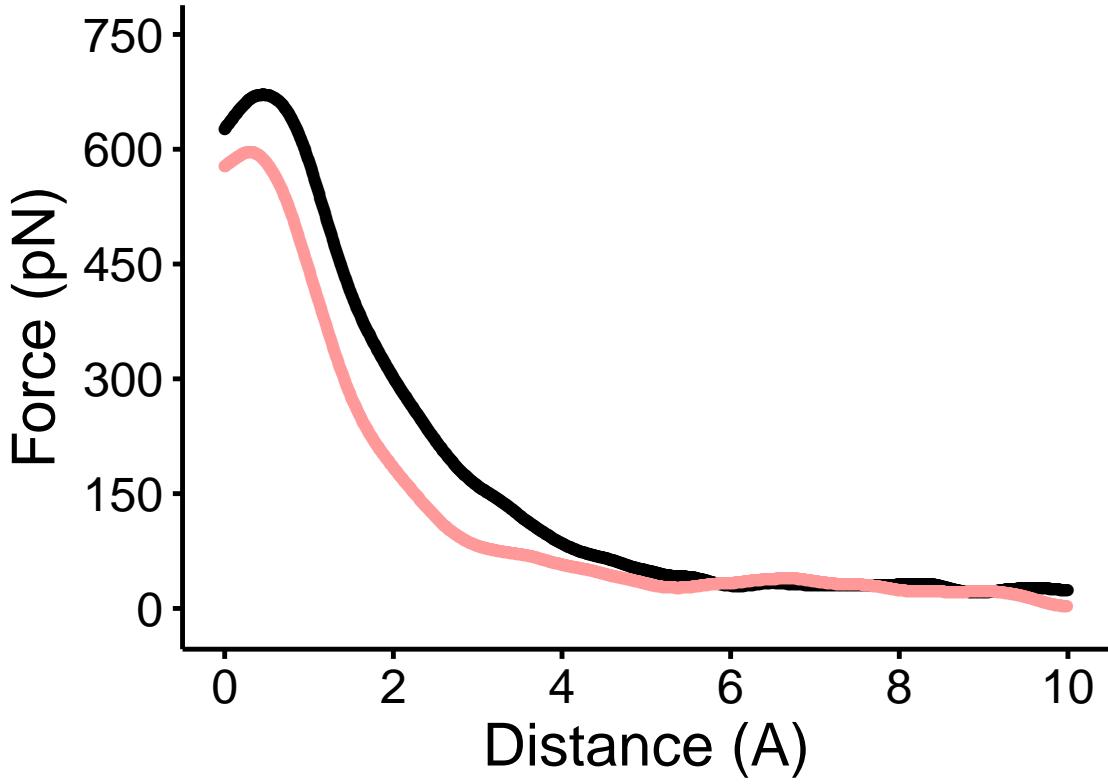


Figure 6: Force versus distance curve of WT and the Y211A mutant. The average force curve for 50 replicates of the WT complex is shown in black, and the average of 50 replicates of the Y211A mutant is shown in red. There is a large difference in both maximum applied force and AUC between the two complexes.

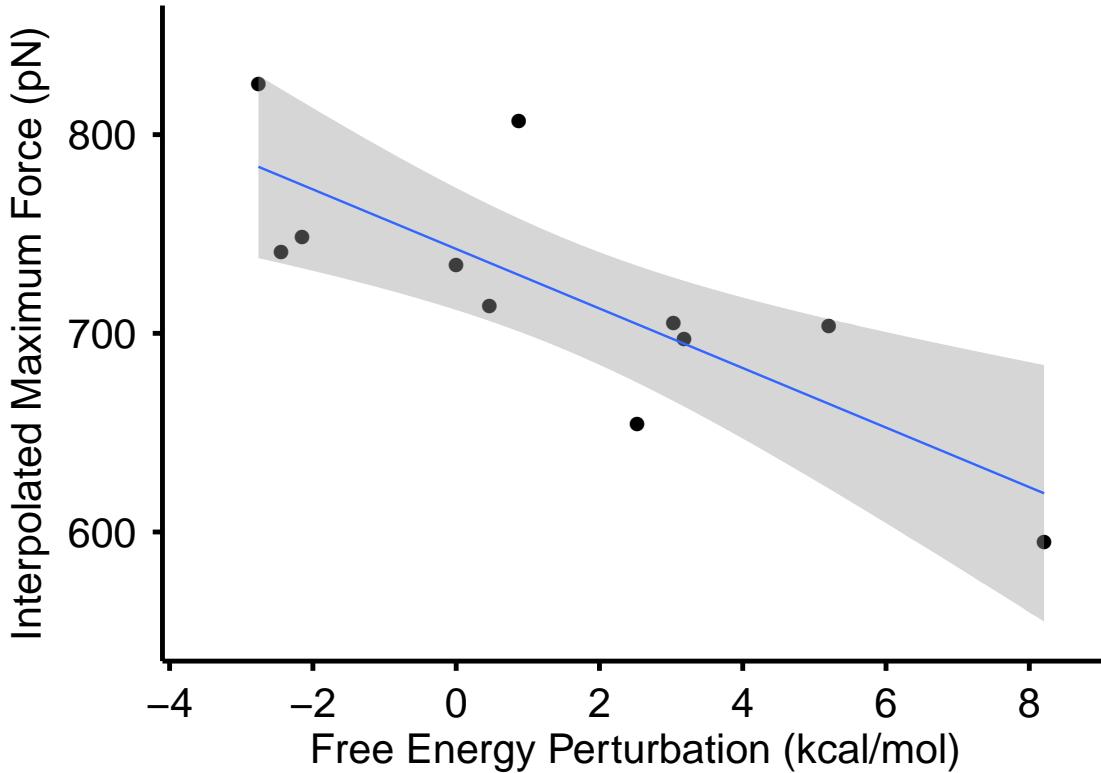


Figure 7: Max force versus free energy perturbation. Scatter plot of maximum force in SMD versus the relative free energy difference calculated by FEP for all 10 mutants tested plus the WT complex. The WT complex for FEP was simply set to 0.0. The correlation between the two is $r = -0.795$ with $p = 0.0034$.

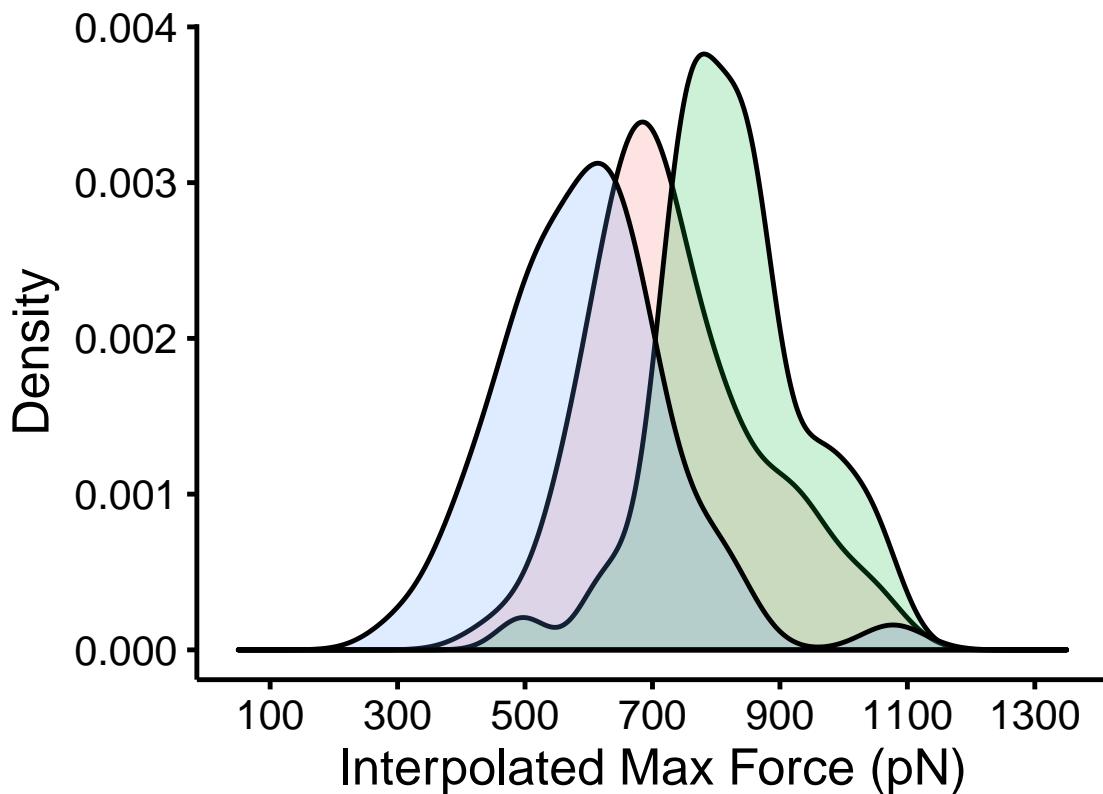


Figure 8: Distribution of interpolated maximum force for three different GP1/hTfR1 complexes. The WT GP1-hTfR1 complex in the middle is flanked by the tighter binding mutant Y211D on the right and the weaker binding double mutant N348W/Y211A on the left. The large non-overlapping areas indicate a large and statistically significant difference in these three complexes.

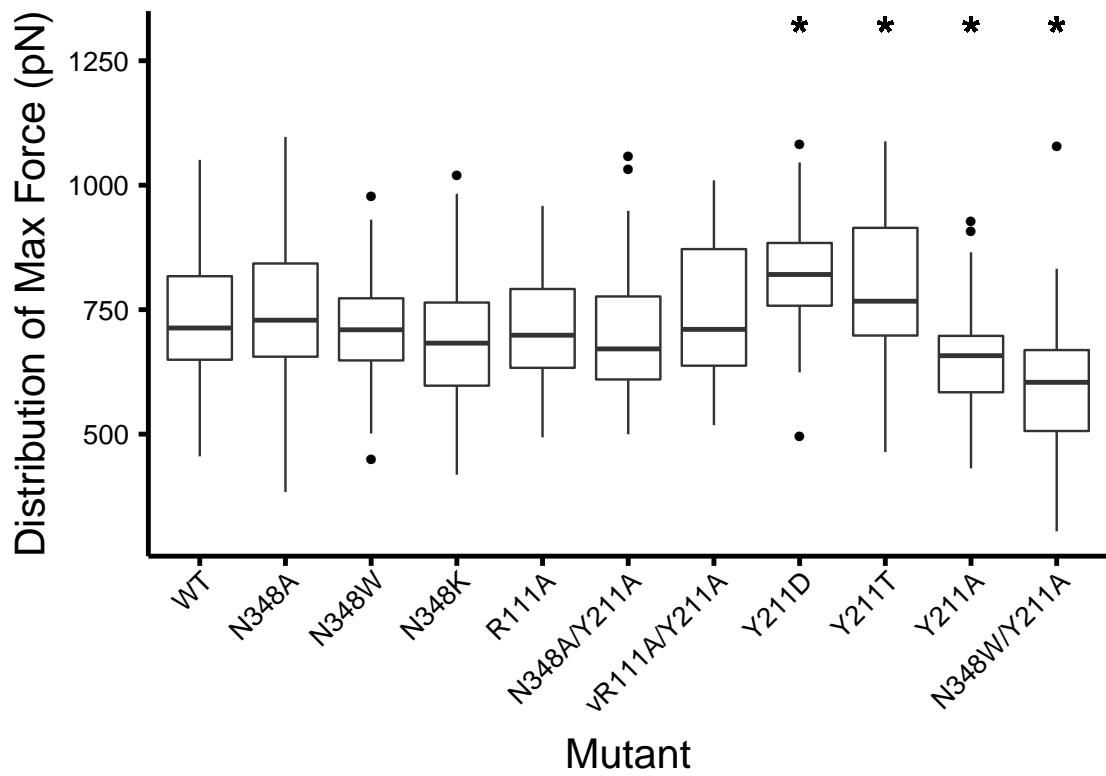


Figure 9: Distribution of interpolated maximum force for all bound complexes tested. Stars above the boxplots indicate a statistically significant difference in mean maximum force relative to the WT complex.

Table 1: Summary of prior information available for each mutation tested. Observed *in vivo* refers to mutations that have been observed in rodent populations. Phenotype *in vitro* refers to the observed phenotype in *in vitro* viral entry assays.

Mutation	Observed <i>in vivo</i>	Phenotype <i>in vitro</i>
WT	Yes	Normal Entry
N348A	No	-
N348K	Yes	Diminished Entry
N348W	No	-
vR111A	No	Diminished Entry
N348A/Y211A	No	-
vR111A/Y211A	No	-
Y211D	Yes	No Expression
Y211T	No	Diminished Entry
Y211A	No	No Expression
N348W/Y211A	No	-

Table 2: Summary statistics for each mutation tested. μ_{MAF} is the mean in piconewtons and σ_{MAF} is the standard deviation of maximum applied force over all simulations. μ_{AUC} is the mean and σ_{AUC} is the standard deviation of AUC over all simulations. ΔG is the free energy difference in kcal/mol calculated via FEP by the dual topology paradigm.

Mutation	μ_{MAF} (pN)	σ_{MAF}	μ_{AUC}	σ_{AUC}	ΔG (kcal/mol)
WT	734.4856	131.6513	145460.4	60232.26	0.000
N348A	748.5217	137.4864	133913.9	51078.64	-2.149
N348K	705.0707	108.5079	141084.4	54450.28	+3.184
N348W	697.3642	132.6436	136886.0	53796.44	+3.033
vR111A	713.8081	106.7374	136103.2	52070.85	+0.466
N348A/Y211A	703.7027	128.5866	113464.2	57451.62	+5.203
vR111A/Y211A	741.0642	131.6287	130070.6	47665.56	-2.440
Y211D	825.2586	115.4343	158878.7	63039.08	+2.760
Y211T	806.8593	136.5648	167110.7	78849.29	+0.875
Y211A	654.1138	108.5343	108090.0	43661.09	+2.526
N348W/Y211A	594.9044	134.8233	108984.2	45451.00	+8.206

Table 3: Pairwise differences (row variable minus column variable) in mean maximum applied force. Bolded values are statistically significant at $p < 0.05$.

	WT	N348A	N348W	N348K	vR111A	N348A/Y211A	vR111A/Y211A	Y211D	Y211T	Y211A
N348A	+14.036									
N348W	-29.414	-43.451								
N348K	-37.121	-51.157	-7.7060							
vR111A	-20.677	-34.713	+8.7370	+16.443						
N348A/Y211A	-30.782	-44.819	-1.3670	+6.3380	-10.105					
vR111A/Y211A	+6.5790	-7.4570	+35.993	+43.700	+27.256	+37.361				
Y211D	+90.772	+76.736	+120.19	+127.89	+111.45	+121.56	+84.194			
Y211T	+72.373	+58.337	+101.79	+109.50	+93.051	+103.16	+65.795	-18.399		
Y211A	-80.371	-94.407	-50.956	-43.250	-59.694	-49.588	-86.950	-171.14	-152.75	
N348W/Y211A	-139.58	-153.62	-110.17	+102.46	-118.903	-108.80	+146.16	+230.35	-211.95	-59.209

Table 4: Pairwise difference p -values for maximum applied force. Bolded values are statistically significant at $p < 0.05$.

	WT	N348A	N348W	N348K	vR111A	N348A/Y211A	vR111A/Y211A	Y211D	Y211T	Y211A
N348A	0.60									
N348W	0.31	0.077								
N348K	0.20	0.038	0.81							
vR111A	0.51	0.16	0.79	0.60						
N348A/Y211A	0.29	0.07	0.95	0.81	0.77					
vR111A/Y211A	0.82	0.79	0.21	0.13	0.35	0.20				
Y211D	0.00093	0.0012	1.4x10⁻⁵	5.0x10⁻⁶	5.6x10⁻⁵	1.2x10⁻⁵	0.0022			
Y211T	0.01	0.018	0.00022	8.7x10⁻⁵	0.0008	0.0002	0.021	0.56		
Y211A	0.0034	7.2x10⁻⁵	0.074	0.13	0.035	0.079	0.0016	4.2x10⁻¹⁰	4.2x10⁻⁸	
N348W/Y211A	3.9x10⁻⁷	1.1x10⁻¹⁰	6.5x10⁻⁵	0.00021	1.6x10⁻⁵	7.2x10⁻⁵	1.3x10⁻⁷	1.2x10⁻¹⁶	2.0x10⁻¹⁴	0.036

Table 5: Pairwise difference p -values for interpolated AUC. Bolded values are statistically significant at $p < 0.05$.

	WT	N348A	N348W	N348K	vR111A	N348A/Y211A	vR111A/Y211A	Y211D	Y211T	Y211A
N348A	0.33									
N348W	0.76	0.59								
N348K	0.59	0.80	0.76							
vR111A	0.55	0.85	0.76	0.94						
N348A/Y211A	0.017	0.07	0.031	0.076	0.08					
vR111A/Y211A	0.26	0.76	0.46	0.68	0.72	0.22				
Y211D	0.33	0.029	0.18	0.09	0.08	0.00046	0.029			
Y211T	0.09	0.0056	0.046	0.027	0.023	4.1x10⁻⁵	0.006	0.59		
Y211A	0.0056	0.027	0.016	0.029	0.031	0.75	0.09	8.2x10⁻⁵	8.5x10⁻⁶	
N348W/Y211A	0.006	0.029	0.017	0.032	0.034	0.76	0.1	9.4x10⁻⁵	8.5x10⁻⁶	0.94