

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/41967713>

Origins of Resistance to the HIVgp41 Viral Entry Inhibitor T20

ARTICLE *in* BIOCHEMISTRY · MARCH 2010

Impact Factor: 3.02 · DOI: 10.1021/bi901915g · Source: PubMed

CITATIONS

20

READS

37

4 AUTHORS, INCLUDING:



Brian Mcgillick

Brookhaven National Laboratory

2 PUBLICATIONS 20 CITATIONS

[SEE PROFILE](#)



Sudipto Mukherjee

Temple University

9 PUBLICATIONS 389 CITATIONS

[SEE PROFILE](#)



NIH Public Access

Author Manuscript

Biochemistry. Author manuscript; available in PMC 2011 May 4.

Published in final edited form as:
Biochemistry. 2010 May 4; 49(17): 3575–3592. doi:10.1021/bi901915g.

Origins of Resistance to the HIVgp41 Viral Entry Inhibitor T20 †

Brian E. McGillick^{|,§}, Trent E. Balias^{‡,§}, Sudipto Mukherjee[‡], and Robert C. Rizzo^{*,‡,||}

Department of Biomedical Engineering, Department of Applied Mathematics and Statistics, and Institute of Chemical Biology and Drug Discovery, Stony Brook University, Stony Brook, New York, 11794

Abstract

The peptide T20, which targets the HIV protein gp41, represents the first approved member in the HIV drug class known as membrane fusion inhibitors. However, mechanisms which lead to resistance through clinical use of T20 are not well-understood because the structure of the bound complex remains undetermined. In this report, an atomic-level model of a T20-gp41 complex embedded in an explicit DOPC membrane was constructed and molecular dynamics simulations, followed by binding energy analysis (MM-GBSA method) was performed to delineate structural and energetic features that contribute to drug resistance. Per-residue binding footprints for T20 with wildtype gp41 reveal strong intermolecular van der Waals, Coulombic, and H-bond interactions in striking agreement with clinically observed resistance patterns. In addition, seven deleterious gp41 point mutations (L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K) were simulated and all correctly showed decreases in binding, including the fact that L33Q and Q40K are most detrimental. Six out of the seven simulations yield good quantitative agreement ($r^2 = 0.72$, N=6) with available experimental fold resistance data. Results from energy decomposition, heat-map analysis, and differential (mutant-wildtype) footprinting indicate: (1) Mutations disrupt intermolecular H-bonding and reduce favorable contact with gp41 at M19. (2) Charged mutations (I37K, Q40K, V38E) lead to significant Coulombic changes that reduce favorable van der Waals interactions. (3) Q40K is more detrimental than I37K due to interaction differences with a polar/charged patch on T20 in the initial (wildtype) state. (4) Resistance for L33S vs L33Q likely involves side-chain packing differences in the final (mutated) state. A valuable finding of the work involves identification of favorable interactions between the C-terminal end of T20 (WNWF motif), residues on gp41 (including the fusion peptide), and head groups in the adjacent membrane. The results suggests a complete T20 binding site would contribute to a stable complex, which could help to explain why prior studies, that employed truncated gp41 constructs, reported that C-terminal T20 residues may not interact with gp41. A hypothesis resulting from this study is that peptides could be designed to increase favorable contact with both the membrane and gp41 which would lead to enhanced activity.

An estimated 30 to 36 million people are currently infected with the deadly human immunodeficiency virus (HIV), the causative agent of AIDS, and 7,400 persons become newly infected each day (1). Drug cocktails, comprised of inhibitors discovered in large-part using structure-based methods and molecular modeling, have provided major therapeutic impact (2). However, since escape mutants arise through clinical use of anti-HIV drugs, development of new therapies with complementary inhibition mechanisms is paramount. Inhibiting viral-

†This work was funded in part by the Stony Brook University Office of the Vice President for Research, the New York State Office of Science Technology and Academic Research (NYSTAR), and NIH grants R01GM083669 (to RCR) and F31CA134201 (to TEB).

*Corresponding author rizzorc@gmail.com, phone: 631-944-2891, fax: 631-632-8490.

[|]Department of Biomedical Engineering

[‡]Department of Applied Mathematics and Statistics

^{||}Institute of Chemical Biology and Drug Discovery

[§]These authors contributed equally to this work

host cell membrane fusion, which is a prerequisite for infection and viral replication, is a new antiviral design paradigm. Targeting the HIV glycoprotein gp41, which mediates this process, has resulted in several promising inhibitors which prevent fusion of the host cell and viral membranes. The first FDA-approved drug (3) in this new membrane fusion class is a gp41 inhibitor called T20 (enfuvirtide, FUZEON). A recent clinical trial (4) consisting of T20, in combination with the HIV integrase inhibitor (3) MK-0518 (raltegravir, ISENTRESS), achieved a drop in viral load to below-detectable levels in 90–95% of treatment-experienced patients compared with 60–70% using MK-0518 alone. However, despite the success of T20, unfortunately, escape mutants also arise which adversely affects long-term clinical use (5). No structures of T20 in complex with gp41 have yet been reported. Prior structural studies employed truncated gp41 domains without the complete T20 docking site. Thus, the molecular mechanisms which lead to resistance to this important anti-fusion drug are not well-understood. Construction, validation, and characterization of a complete binding model for T20 in complex with gp41 is the primary focus of this report.

HIV infection requires fusion of virus and host-cell membranes and is mediated by the viral glycoprotein complex gp160 composed of two proteins, gp41 and gp120 (6). Upon binding of gp160 to host cell receptors, large conformational changes occur in gp41 which lead to an intermediate fusion state as described by Kim and coworkers (6). Compounds which bind this intermediate are potential membrane fusion inhibitors. Structural studies of truncated gp41, thought to be representative of a the fusogenic state, have revealed extensive interhelical interactions (7) arranged in a coiled-coil trimer-of-dimers (8,9) hairpin motif common to Class I viral fusion proteins (10) as shown in Figure 1a. Each dimer is composed of an N-helix (red) and C-helix (green) separated by a flexible loop (Figure 1a).

Consistent with the proposed intermediate, in which the C and N-helices are transiently exposed, numerous peptides of various lengths, and a limited number of small molecules, have been reported which bind gp41 and prevent re-formation of the coiled-coil thereby inhibiting membrane fusion (11–13). While inhibitors have been designed based on both N and C-helix sequences, the most successful inhibitors are based on outer C-helix regions. A series of D-peptides have also been reported (14,15). Figure 1b shows the sequence relationships between the native gp41 outer C-helices, the approved peptide drug T20, and a related inhibitor C34. Both these peptide inhibitors (Figure 1b) are simply capped fragments of native C-helix sections (11–13). C34 and related peptides contain three sidechains (Figure 1, residues 117, 120, and 124) known to interact within a conserved hydrophobic pocket region on gp41 which has been proposed (16) to be an attractive target for design of small molecule inhibitors (17–20). Based on sequence, T20 is not believed to interact with this pocket.

As occurs with all anti-HIV drugs, a growing body of experimental data (Table 1) indicates that peptide fusion inhibitors are also susceptible to resistance mutations (5,21–27). Mutations which arise from use of T20 are primarily observed to occur on gp41 in a contiguous stretch of 10 amino acids between residues 33–43 (gp41 numbering) on the inner N-helices. Mutations at sites L33Q, V38E, Q40K, and Q41R are particularly detrimental (Table 1). Figure 2 shows the mutations from Table 1 drawn as red spheres and mapped to the gp41 ectodomain (from Figure 1a) with the outer C-helices removed for clarity. The spheres encompass the G36, I37, V38 region (Table 1) for which gp41 mutations are most common (21,28). The known binding pose of C34, including the three pocket binding residues, is also shown in Figure 2 (orange ribbon). Although T20 does not interact with the pocket, it is likely the peptide binds the gp41 inner coiled-coil given the fact that mutations which reduce activity map directly to the N-helical core (Figure 2, spheres). Based on the C-helix sequence (Figure 1b), T20 should be shifted by about 10 a.a. towards the N-terminus relative to C34 (Figure 2, dashed line).

Considerable experimental evidence indicates T20 interacts with the gp41 inner coiled-coil but the full extent of the peptide binding footprint is unknown. Further, although the C-terminal end of T20 is thought to interact in some fashion with the host cell membrane the extent of such interactions are as yet undescribed. We hypothesized T20 would interact favorably with both the gp41 inner coiled-coil and the lipid membrane. Our hypothesis was based on the fact that the C-terminal end of T20 would be butted up against the outer leaflet of the host cell lipid membrane provided that (i) similar sequence-footprint relationships between known C-helix derived peptides such as C34 with gp41 were maintained, (ii) T20 was primarily alpha-helical in the bound state, and (iii) the fusion peptide domain consisting of the first 16 a.a. (8) on gp41 were embedded into a lipid bilayer.

Based on this overall topology, atomic-level structural models of lipid-bound gp41 with T20 have been constructed to answer questions regarding the relative importance of specific interactions along the gp41 and lipid membrane interfaces in an effort to deduce and understand the origins of drug resistance for seven point mutations (Table 1). Specifically, all-atom MD simulations have been used to model T20 in complex with membrane-bound gp41 in explicit solvent and explicit lipid with the following goals: (i) to develop structural models consistent with experimental observation, (ii) to characterize what factors contribute most to experimental activities, and (iii) to delineate which structural and energetic features contribute to clinically reported resistance profiles for gp41 variants. Given the importance of this new class of drugs in the anti-AIDS arsenal, a greater understanding of what drives molecular recognition and resistance for gp41 peptide inhibitors is crucial.

METHODS

Model Construction and Energy Refinement

Using our previously validated computational binding model for six C34 inhibitors with gp41 as a starting point, (29) the inner-coil of gp41 was extended (Figure 3) to accommodate the proposed T20 footprint shown in Figure 2. As described in Results, the sections added to gp41 were alpha-helical however two alternative conformations for the FP regions were also explored. Importantly, the extensions had to be constructed in such a way as to match the twist which occurs along the inner coil axis of the gp41 heptad repeat motif. We used the GCN4-gp41 construct (pdb code 1ENV), (9) which contained a 47 residue N-helical section of gp41 that overlapped with the previous C34 binding model (constructed from pdb code 1IF3), (30) which contained 28 residues that extended beyond the C34-gp41 construct towards the N-terminus (Figure 3).

The two structures were aligned (gp41 numbering) by a best fit root-mean-square-deviation (rmsd) match of the 1ENV trimer to the 1IF3 trimer using residues 31–35@C,O and 53–57@C,O in common between the chains. Residue A47 was chosen as the optimal connection point for merging the structures into one set of coordinates given the nearly complete overlap of C-alpha atoms which occurred at this point. The new backbone template was then obtained by merging the two structures which required deleting residues from the C-terminal end of the inner coiled-coil of 1ENV up to A47 and from the N-terminal end of 1IF3 to A47. The resultant structure contained a seamless extension of the N-helical coiled-coil of the same length required to construct a gp41 sequence containing the complete fusion peptide (FP) domain. The added N-terminal residues from 1ENV (N=28) were mutated to the correct a.a. sequence for wildtype gp41 (LW12.3 isolate) (31) using an iterative combination of side-chain rotamer exploration and manual inspection with 3-D stereo visualization. Evaluation with side-chain energy minimization was used to resolve any steric clashes. This refined gp41 model of the complete inner coiled-coil with the wildtype sequence then served as a master template for construction of L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K point mutations (3 a.a. because of the trimer). This same iterative refinement procedure was used to construct these point mutations.

The program MOE (32) was used to construct all sequence modifications. The completed coiled-coil also served as a template for construction of two alternative models in which the FP region was in a beta-sheet or extended alpha-helix conformation (see Results).

To construct the T20 ligand, in the proposed binding mode with extended gp41, a second copy of an outer gp41 C-helix peptide was aligned onto the original C-helix coordinate frame in 1IF3 using backbone coordinates but shifted by 10 amino acids towards the N-terminus. The amino acid sequence was then mutated to match T20 and rotameric searches were performed to minimize any steric conflicts occurring at the binding interface with the gp41 trimer. A short gas-phase MD simulation using a 4r distance dependant dielectric constant was performed to allow the T20 peptide to relax in relationship to gp41 which was held restrained. An examination of initial versus final energies from additional short gas-phase energy minimizations indicated that reasonable side-chain placements, as well as rigid body rotations and translations, were obtained for T20 with wildtype gp41 and for mutants. The AMBER program, (33) in conjunction with the FF99SB force-field, (34) was used for the initial coarse refinements. The focus was to obtain energetically and structurally reasonable starting geometries, for a fully extended gp41 inner coiled-coil in complex with T20, for subsequent long MD equilibrations and production simulations of the entire solvated system.

Simulation Setup and MD Details

Following initial model building and coarse refinements, the NAMD and VMD (35,36) programs were used in combination with the CHARMM27 force-field (37,38) to setup and simulate T20 with gp41 embedded into an explicit lipid membrane including explicit solvent. The VMD program was used to place the peptide-gp41 complex in the center of a periodic box containing lipid and water with the first 16 residues of the gp41 fusion peptide fully inserted into the lipid bilayer in a perpendicular orientation. The systems contains ca. 30,000 TIP3P (39) explicit water molecules, ca. 160 DOPC (37,38) lipid molecules, 282 gp41 residues, and 36 T20 peptide ligand residues in a periodic box of ca. size $80 \times 80 \times 200 \text{ \AA}^3$. The bilayer was constructed using coordinates of a pre-equilibrated DOPC membrane kindly provided for download by the Feller Lab (40). To accommodate gp41, it was necessary to increase the size of the original membrane which was done through replication and translation of the pre-equilibrated lipids using VMD. After insertion of gp41 into the center of the modified membrane all lipids having steric overlap with the protein were deleted. This procedure yielded a DOPC bilayer containing approximately 80 lipid chains on each leaflet with a ca $30\text{--}35 \text{ \AA}$ buffer between the edge of the periodic box in the x-y plane and the inserted protein (see Figure 4). Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) (41) method with a 12.0 \AA direct-space non-bonded cutoff. MD simulations were carried out in the NPT ensemble at 310 K using Langevin dynamics with a damping coefficient of 1/ps and the Langevin piston method (42) with a piston target pressure of 1.01325 bar. The Shake algorithm (43) as implemented in NAMD (35) was used to constrain bond lengths for all hydrogens. Simulations employed a 1 fs time-step for equilibration and 2 fs for production.

Consistent with our laboratories earlier modeling of C34 peptides with wildtype gp41, (29) an extensive energy and MD equilibration protocol (six steps total) was used to relax the structures in a systematic way prior to production runs. Three short rounds of energy minimization for 1000 cycles each were performed on each complex using a harmonic restraint weight of 5.0 kcal/mol \AA^2 in the following order: all protein atoms and all lipid heavy atoms (step 1), all protein atoms (step 2), and only protein heavy atoms (step 3). This was followed by restrained MD for 500 ps using a 5.0 kcal/mol \AA^2 weight (step 4) applied to protein heavy atoms only plus an additional 10 ps with a reduced weight of 1.0 kcal/mol \AA^2 (step 5). An additional 10 ps of MD was performed with a 10-fold lower weight of 0.1 kcal/mol \AA^2 on protein heavy atoms except for 12 alpha carbons on gp41 near the C-terminal end of the trimer at positions

82 to 94, 176 to 188, and 270 to 282 which retained the 1.0 kcal/mol Å² restraint (step 6). Production MD followed for a total of 10 ns and was unrestrained with the exception of the aforementioned 12 a.a. C-terminal end of the model which retained the weak 1.0 kcal/mol Å² restraints. Coordinates and energies were saved every 1 ps (N=10,000) for subsequent post-processing and analysis.

Binding Energy Calculations

Free energies of binding were estimated using the single trajectory Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method which has been well described (44,45). The ability to study large structural changes and compare binding energies between ligands with diverse topologies, combined with relative ease of setup and implementation, makes use of MM-GBSA as an attractive alternative to other free energy calculation methods. Despite the approximations made in MM-GBSA, our laboratory has used the method with good overall success to study a variety of problems, including gp41 (29,46–48). Explicit solvent MD simulations was used to generate ensembles of low energy structures (snapshots) which was post-processed to obtain the individual terms used to estimate the overall free energy of binding $\Delta G_{\text{bind calc}} = \Delta E_{\text{coul}} + \Delta E_{\text{vdw}} + \Delta \Delta G_{\text{polar}} + \Delta \Delta G_{\text{nonpolar}}$. Following production MD, explicit solvent (water and lipid) was removed and intermolecular non-bonded Coulombic (ΔE_{coul}) and Lennard-Jones (ΔE_{vdw}) interactions between T20 and gp41 were computed using the pair interaction routine in NAMD. The remaining terms were computed using the Hawkins-Cramer-Truhlar (49,50) Generalized Born (GB) pair-wise descreening algorithm ($\Delta \Delta G_{\text{polar}}$ terms) and the AMBER8 (33) icosahedra solvent-accessible-surface-area algorithm (SASA terms) as implemented into the program DOCK (51). Nonpolar terms were computed as $\Delta G_{\text{nonpolar}} = \gamma \text{SASA} + \beta$ with SASA in Å² using standard values of $\gamma = 0.00542$ kcal/mol Å² and $\beta = 0.92$ kcal/mol (52). No additional entropic terms were added. It should be emphasized that all MD simulations employed explicit solvent and explicit lipid thus the ensemble of conformations is representative of that generated in the condensed phase. Implicit solvent GBSA calculations were used only for estimation of the desolvation terms, $\Delta \Delta G_{\text{polar}}$ and $\Delta \Delta G_{\text{nonpolar}}$, after stripping off explicit solvent. Autocorrelation functions (ACF) in conjunction with block averaged standard errors of the mean (BASEM), using varying block sizes, were used to estimate errors and error convergence using the ensemble of snapshots (N=10,000) obtained from coordinates saved every ps during the 10 ns production phase. For comparison with simulation results, experimental fold resistance values (FR = mutant/wild-type activities) for inhibitors with gp41 (Table 1) were estimated as experimental free energies of binding using $\Delta \Delta G_{\text{FR exptl}} \approx RT \ln(\text{FR})$ at 25 °C. Hydrogen bonds were defined as a structural interaction between a donor (H_D) and acceptor (X_A) with a distance of 2.5 Å or less, and an angle between X_D–H_D–X_A of between 120 and 180°.

Molecular Footprint Calculations

To determine how specific amino acids may contribute to binding, molecular "footprints" for all inhibitors with gp41 were computed. Footprints represent a residue-based decomposition of intermolecular interactions between two bound species and may be thought of as a molecular signature. Our laboratory has successfully used footprints to characterize binding for C34 inhibitors with HIVgp41, (29) sialic acid inhibitors with neuraminidase, (48) and ATP competitive inhibitors with EGFR (46). MD trajectory coordinates were post-processed to obtain per-residue Coulombic, van der Waals, and H-bond footprints in the following two ways: (i) Target-based, each individual gp41 residue with the entire ligand; (ii) Ligand-based, each individual T20 residue with gp41 as a whole. The net sum of energies along a given footprint is equal to the total number of non-bonded intermolecular interactions while the per-residue decomposition reflects the relative strength (i.e. importance) of each specific residue. Difference (Δ) footprints were also computed and represent the change in energy or H-bond counts from mutation – wildtype results at any given a.a. position. In addition, for the wildtype

simulation of T20 with gp41, the van der Waals energy between each T20 residue with each gp41 residue was computed and plotted as an interaction energy matrix (heat map). Summation down the rows or the columns in the heat map yield the target-based or the ligand-based footprints, respectively. Footprint and matrix elements were averaged over the course of each production trajectory ($N=10,000$ snapshots) and plotted as a function of residue number or numbers. To aid in data representation, footprint and matrix results representing gp41's three identical monomers (3×94 residues) were mapped onto a single monomer (94 residues).

RESULTS AND DISCUSSION

T20-gp41 Binding Model in Explicit Lipid

T20 was modeled as bound along the inner trimeric groove of gp41 but shifted by 10 a.a. towards the N-terminus relative to C34 as shown in Figure 4. This required a concomitant extension of gp41 (described in Figure 3) to accommodate T20 and our laboratories prior computational set-up for C34 (29) was used as a starting point (see Methods). Although other starting configurations are possible, the conformation for the complex was based on sequence considerations of T20 with respect to gp41 (see Figure 1) which provide a physically reasonable constraint for initial rigid body placement. Complete structural information is lacking for the domain located between the known N-helical regions and the fusion peptide (termed here TD_{FP}, Figure 4a) however recent studies by Sackett et al (53) strongly indicate that TD_{FP} is α -helical. The terminal FP trimer was also modeled as an extension of the inner gp41 α -helical coiled-coil (Figure 4a) however it should be noted there is debate surrounding secondary structure in this region with various studies reporting the FP region is α -helical as modeled here, in-register parallel β -strand, out-of-register parallel β -strand, anti parallel β -strand, or a mixture of parallel and anti-parallel strands (54–59). Although the results indicate that membrane-bound FP trimers modeled as α -helical are stable, this does not rule out the existence of other conformational possibilities (see additional discussion below). The FP domain consists of the last 16 N-terminal gp41 residues, is primarily hydrophobic, and is normally associated with the host-cell lipid bilayer during fusion. In the present model, insertion of the FP trimer (a.a 1–16) into an explicit DOPC lipid bilayer yielded a binding geometry particularly well-accommodated to T20 with peptide inhibitor residues poised to interact with the outer leaflet of the membrane as shown in Figure 4b. As discussed further below, favorable pi-stacking and head group interactions are observed during the MD simulations similar to those reported in other lipid-bound systems (60,61).

Simulation Stability

Simulation protocols (see Methods) followed the same general approach recently employed in our laboratory to successfully characterize origins of SAR for C34 peptides with gp41 (29). However, the earlier gp41 studies (without FP or membrane) revealed fraying near the N-terminus of the trimer and weak backbone restraints were ultimately required as a pseudo-mimic for forces which would normally hold gp41 membrane-bound in the biologically active system. We hypothesized that explicit membrane simulations would not require extensive restraints. With the exception of a small 12 a.a. loop region (Figure 4b, red tube), restrained to account for the stabilizing presence of the viral envelope, the present simulations are unrestrained. As described below, the overall good energetic and structural behavior of the present simulations provides compelling evidence that the hypothesis formed as a result of the prior C34 studies without membrane was correct.

Root mean square deviations (rmsd) over the course of the MD trajectory reveal the membrane-bound model is very stable as shown in Figure 5. Here, two types of rmsd fits are presented for each individual region representing the fusion peptide (FP, a.a. 1 to 16), transition domain (TD_{FP}, a.a. 17 to 29), N-helices (residues 30 to 81), or T20. (1) Black lines in Figure 5 represent

"fit-to-self" results which are defined as rmsds computed by fitting each region to itself. (2) Gray lines represent "fit-to-gp41" results which represent rmsds for each region computed after an initial best fit match of the gp41 inner coiled-coil (excluding the FP). All fits were relative to the first MD frame of the production runs and employed Ca, C, N, O backbone atoms in the match criteria. In all cases, fit-to-self results for the various regions (Figure 5a black lines) show very low rmsds < 2 Å which indicates the internal geometries are stable and minimally changing conformation over the course of the trajectories. Rmsds computed using the fit-to-gp41 criteria (Figure 5a gray lines) show larger fluctuation due to the inclusion of rigid body motion as a result of the fitting criteria. Importantly, T20 rmsds yields both low internal (fit-to-self) and rigid body (fit-to-gp41) backbone rmsds (ca 1–2 Å) which is a strong indication the proposed binding interface is physically reasonable (Figure 5a black and gray lines).

Energetically, the T20-wildtype gp41 simulations mirror the overall good behavior observed in the structural rmsd plots (Figure 5b). Here, the sum of energy terms used to estimate the free energy of binding (ΔG_{bind} calcd), and individual van der Waals (ΔE_{vdw}), Coulombic energies (ΔE_{coul}), $\Delta \Delta G_{\text{polar}}$ (GB energy), and $\Delta \Delta G_{\text{nonpolar}}$ (SASA) components are in general well-behaved. Polar desolvation and intermolecular Coulombic energies show in general larger fluctuations, and as previously observed in our laboratory, (29, 48) are ca. equal and opposite in magnitude (Figure 5b middle panels). This effect is especially dramatic when all 10,000 instantaneous values are plotted together (Figure 6) which yields an almost perfect anti-correlation of $r = -0.97$ in addition to $r^2 = 0.94$. The net result is that the sum of all four energy terms yields an overall free energy for ΔG_{bind} calcd showing good stability with both instantaneous and block smoothed (N=100 blocks) fluctuations oscillating about a reasonable converged mean value (Figure 5b bottom, gray vs smooth black lines).

Binding Free Energy Convergence: Autocorrelation and Block Averaging

As shown in Figure 7 simulation convergence was additionally assessed by plotting autocorrelation functions (ACF) and block averaged standard errors of the mean (BASEM) for the ΔG_{bind} calcd time-series from T20 with wildtype gp41, as well as seven gp41 mutants (L33Q, L33S, G36V, I37K, V38E, Q40H, Q40K). For a given time series, ACF functions provide a means to assess how data are correlated and BASEM analysis allows error convergence to be estimated by allowing block length to vary. (62,63) Ideally, reasonably uncorrelated data should be used for block averaging. Figure 7a reveals that all ACF curves for ΔG_{bind} calcd drop quickly from 1 (100% correlated) at lag time = 0 ps, to less than ca. 25% correlated at 250 ps, before showing fluctuations which oscillate about 0% which is indicative of largely uncorrelated data.

Interestingly, the ACF results for T20 with I37K show a distinctly different trend (green solid arrow), in comparison to other trajectories, which is an indication this one simulation is not in reasonable equilibrium or as well-behaved. BASEM curves for similarly reveal I37K as an outlier in terms of poorly converged error estimates. An additional MD simulation was run for I37K which employed a different random seed number in an attempt to yield a trajectory with improved equilibrium/convergence statistics. As shown in Figure 7 (dashed green arrows), ACS and BASEM curves for the new MD run yielded results more consistent with the other converged simulations and this latter trajectory was used for all further analysis.

Error estimates for ΔG_{bind} calcd were obtained using the BASEM plots in Figure 7b. Here, block averaged standard errors of the mean in kcal/mol were computed as a function of block size which ranges from 1 ps to 5000 ps (1 to 5000 snapshots). The BASEM results show the expected exponential increase as block averaging size becomes larger that begins to reach a plateau which is indicative of reasonable equilibrium. BASEM errors computed at N=1 likely underestimate the true error in the simulations while at N=5000 (1/2 the simulation length) the errors may not be reliable due to the fact that only two blocks are used. For the present 10 ns

trajectories, Table 2 plots BASEM errors, along with the % of uncorrelated data from ACF curves, for the eight simulations at block sizes of 1 ps, 100 ps, 250 ps, 500 ps, and 1000 ps. Individual system results and the overall average for these five block sizes are both reported. The regime between 250 and 500 ps yields data which is largely uncorrelated (ca 75 – 85%), and having a sufficiently large numbers of independent blocks (250 ps = 40 blocks and 500 ps = 20 blocks), thus the errors of 0.68 – 0.91 kcal/mol may be taken as a reasonable estimate of the statistical noise for ΔG_{bind} calcd for these simulations. Although the overall structural, energetic, and convergence metrics monitored over the course of the MD trajectories (Figures 5,7 and Table 2) indicate the T20-gp41 models are physically reasonable and well-behaved, ideally, longer time-series for each simulation would be desirable.

Molecular Footprints: Key gp41 Residues for T20 Binding

To gauge the importance that specific residues may contribute to binding, per-residue interaction breakdowns were computed between T20 and wildtype gp41 as shown in Figure 8 for the gp41 range Leu12 - Trp60. Here, footprints for van der Waals (top, ΔE_{vdw}), Coulombic (middle, ΔE_{coul}), and hydrogen bonds (bottom, H-bonds) are defined as the pair-wise interaction sum between all T20 residues and each individual gp41 residue. For simplicity, contributions from the three gp41 N-helical chains were summed together and mapped to a single monomer. Mutations which adversely affect binding (Table 1) are colored red in Figure 8.

Notably, the van der Waals footprint (Figure 8 top, ΔE_{vdw}) for T20 shows a binding motif consistent with an alpha-helix binding interface in which side chains are packed into gp41 in ca. alternate and repeating manner. Compellingly, there are strong favorable ΔE_{vdw} interactions which occur in the gp41 range 33–43 at energetically significant mutation sites (Table 1), in particular at L33, Q40, and N43 (ca. –5 to –6 kcal/mol). The importance of this region for binding is reinforced by the shape of the Coulombic energy footprint (Figure 8 middle, ΔE_{coul}) which also predicts Q40, N41, and N43 (ca. –12 to –15 kcal/mol) as residues making significant electrostatic interactions with T20. Significant H-bonding in the region is also observed at positions V38, Q40, and N43 (Figure 8 bottom). Here, the most populated H-bond position occurs on gp41 at Q40. Mutations at or near energetically favorable residues would be expected to adversely affect binding. Thus, the footprint results for T20 with gp41 in Figure 8 are strikingly consistent with known experimental FR data in Table 1.

Experimental vs Calculated Fold Resistance

To further validate the model seven additional MD simulations of the same length were performed to compute the energetic effects of L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K for comparison with experiment (Figure 9, Table 3). Table 3 shows the individual energetic and H-bonding contributions to binding in both an absolute (ΔG_{bind} calcd) and relative ($\Delta \Delta G_{\text{FR}}$ calcd) sense while Figure 9 plots the experimental vs computational FR results. Computationally FR energies ($\Delta \Delta G_{\text{FR}}$ calcd) represent the difference between mutant and wildtype free energies of binding (ΔG_{bind} calcd) from two separate MD simulations. Estimated uncertainties for values in Table 3 were computed using the BASEM procedure with a block length size of 250 ps (see discussion above).

Importantly, the simulation results correctly show that all mutations lead to reduced activity for T20 with gp41 and in a relative sense six of the seven datapoints (Figure 9, $r^2 = 0.72$) well-reproduce the experimental FR energies ($\Delta \Delta G_{\text{FR}}$ exptl). Notably, the calculations correctly predict that L33Q and Q40K have the largest loss of binding. Although the correlation is high, the magnitudes of the computed FR values are, with the exception of L33S, systematically overpredicted by ca. 3 to 4-fold. As all mutations except L33S result in a larger sidechain the offset could be related to size and associated entropic terms not included in the calculations.

Changes in solute configurational entropies are usually assumed to be minimal (i.e. cancel) when relative $\Delta\Delta G_{\text{FR}}$ computational values are compared (64). However, as each gp41 point mutation actually changes 3 a.a. in the trimer, and unbound C-helical peptides such as T20 are known to be relatively unstructured in solution, (65,66) complete cancellation of entropy terms are unlikely. These effects could play a role in systematic overpredictions for the mutations to larger sidechains. Nevertheless, despite the L33S offset in comparison with the other results in Figure 9, use of the gp41 model to predict relative FR energies leads to good overall agreement with experiment.

Energy decomposition reveals significant intermolecular van der Waals energies (Table 3 column B, ΔE_{vdw}), due to the large C-peptide binding interface, that are about 2–4 times greater than that observed in previous binding studies from our laboratory for small organic molecules (46–48). Computed gas-phase Coulombic interactions (Table 3 column C, ΔE_{coul}) are also large and in this case reflect the fact that T20 has numerous charged residues. However, as we have previously observed, (29,48) desolvation penalties (Table 3 column D, ΔG_{polar}) of ca. equal in magnitude, but opposite in sign, oppose binding thus the overall electrostatic contribution is small (Table 3 column E, $\Delta G_{\text{elec}} = \Delta E_{\text{coul}} + \Delta G_{\text{polar}}$). Given that the ΔG_{elec} values are positive (unfavorable) the overall peptide-gp41 recognition event appears to be dominated by favorable ΔE_{vdw} interactions (29). The final $\Delta G_{\text{nonpolar}}$ terms (Table 3 column E), which are always negative values, reflect favorable binding contributions as a result of burial of solvent accessible surface (hydrophobic effect).

For the mutations, the origin of the large changes in Coulombic ($\Delta\Delta E_{\text{coul}}$) and accompanying desolvation ($\Delta\Delta G_{\text{polar}}$) terms in Table 3 become clear when viewed in context of the predicted T20 binding pose in the geometry of the complex as shown in Figure 10. Here, V38E introduces a negative charge on gp41 (3 red spheres), which is proximal to a total of seven Glu and Asp residues on T20, (Figure 10a red sticks) thus leading to repulsion between negatively charged sidechains (red line, Figure 10b). I37K and Q40K mutations show the corresponding opposite effect (dark and light blue lines, Figure 10b) by which introduction of positively charged gp41 residues (Figure 10a blue spheres) leads to dramatic favorable energetic increases with nearby negative T20 residues (Figure 10a red sticks). The net effect is positively charged mutations (Q40K, I37K) yield the most negative $\Delta\Delta G_{\text{elec}}$ values in contrast to the negatively charged mutation (V38E) which yields the most positive $\Delta\Delta G_{\text{elec}}$ value (Table 3 column F). Notably, the two curves for both positively charged mutations are nearly superimposable (dark vs blue lines, Figure 10b) suggesting not only good convergence in the computational results but that from T20's point of view these two different mutations lead to very similar changes in intermolecular Coulombic interactions. Neutral mutations in contrast introduce much smaller changes in $\Delta\Delta E_{\text{coul}}$ relative to charged mutations (Table 3).

The simulations also correctly predict that changes in nonpolar energy are smallest for L33S which is the only mutation that results in a less bulky sidechain relative to wildtype (Table 3 column E, $\Delta\Delta G_{\text{nonpolar}} = 0.06 \text{ kcal/mol}$). For six of the seven mutations, loss of intermolecular H-bonds with gp41 is also observed (Table 3 column G). In particular, the two mutations yielding the largest losses (Table 3 column G, $\Delta H\text{-bond}$) are for Q40H and Q40K which is consistent with the Q40 being the strongest H-bonding residue in the peptide binding interface (see Figure 8). Finally, six of the seven mutations show a loss in van der Waals energy (Table 3 column B, ΔE_{vdw}) with greatest losses being for the two positively charged lysine mutations (see discussion below). Only the L33S-wildtype change does not lead to a loss in steric packing. This latter case is probably related to the fact noted above that L33S is the only mutation which leads to a smaller sidechain.

Origins of Resistance: Charged Mutations

Q40K and I37K are analogous in that both result in a replacement of a neutral residue with a positively charged residue and, due to the residue per turn ratio of the gp41 coiled-coil alpha helices, the mutations are in similar locations (see Figure 10a). However, experimentally (see Table 1) the Q40K mutation is much more detrimental than I37K. Further, despite the opposite charge, V38E shows a FR change which minimally differs by only ca. 0.3 kcal/mol compared with Q40K (Table 1). In an attempt to identify origins of these experimental observations, per-residue changes in van der Waal ($\Delta\Delta E_{vdw}$) and Coulombic ($\Delta\Delta E_{coul}$) energies were computed by subtracting mutant from wildtype per-residue footprints as shown in Figure 11. To simplify analysis, charged (Figure 11a-b) and neutral (Figure 11c-d) mutations are grouped together.

Comparison of $\Delta\Delta E_{vdw}$ for charged mutations reveals that the shapes of all three curves are generally similar with the exception that relative to V38E, mutations I37K and Q40K (Figure 11b green, blue, vs red lines) both show significant energetic losses in the contiguous region between residues L33 and L45 (Figure 11b). Interestingly, all of the experimental gp41 mutations listed in Table 1 are contained within this 13 residue range. To aid in identifying the gp41 a.a's. that interact strongly with specific T20 residues, a van der Waals interaction energy matrix (heat map) was calculated using wildtype simulation results as shown in Figure 12. Here, the sum of any row or column in the matrix equals the magnitude for the respective peak shown in the accompanying gp41 (column) or T20 (row) footprints. Compellingly, the gp41 residues in the L33–L45 mutation region (Figure 12 x-axis, shaded region) map nearly perfectly (black rectangle) onto the predominantly charged/polar patch on T20 in the range E136 – E148 (Figure 12 y-axis, shaded region).

Focusing in on the gp41 footprint peaks, T20 is observed to make stronger more favorable ΔE_{vdw} contacts with the wildtype receptor at Q40 vs I37 positions (Figure 12 x-axis, green vs orange ellipsoids). Based on the interaction matrix, the underlying physical reason for these stronger interactions is that the more hydrophobic I37 interacts less favorably with residues in the T20 charged/polar patch compared with Q40 (Figure 12 y-axis, green vs orange ellipsoids). Thus, it is reasonable to propose that mutations occurring at I37 would affect binding less than mutations at Q40. In support of this argument, differential changes in $\Delta\Delta E_{vdw}$ localized to the mutations sites show no loss in energy at I37K (Figure 11a green line) compared with Q40K (Figure 11a blue line). The ability of the computational models to correctly capture these subtle differences in two otherwise highly similar mutations, and predict their relative binding affinities in good agreement with experiment, provides strong support for the proposed T20 binding geometry. For V38E, the differential footprint for van der Waals energy is much flatter (Figure 11a red line) which leads to small (1.06 kcal/mol) unfavorable changes (Table 3 column B) in steric packing compared with I37K (13.85 kcal/mol) or Q40K (20.18 kcal/mol). Although V38E share peaks with I37K and Q40K within the L33–L45 region (Figure 11a S35, V38, N42, and L45 positions) the magnitude of unfavorable changes at these sites is smaller. A likely physical explanation for this observation involves changes in $\Delta\Delta E_{coul}$ as discussed below.

As highlighted in Table 3 (column C), charged mutations lead to dramatic overall changes in Coulombic energy which are clearly evident in the $\Delta\Delta E_{coul}$ footprints (Figure 11b). Again, similar to $\Delta\Delta E_{vdw}$, the most variation amongst differential footprints occurs within the L33 and L45 mutation region. T20 has an overall net formal charge of -5 and in the case of I37K and Q40K the resultant added positive charge leads to large Coulombic attraction directly at the site of the mutations (Figure 11b green and blue lines). In contrast, the profile for V38E shows the opposite effect (Figure 11b red line). The significant Coulombic attractions for I37K and Q40K seen in Figure 11b appear to be coupled to the large losses in favorable van der Waals energy seen in the $\Delta\Delta E_{vdw}$ footprints (Figure 11a). A physical explanation may involve the fact that as a result of the shape of the 6–12 Lennard-Jones potential, Coulombic attractions (I37K, Q40K) lead to greater losses in $\Delta\Delta E_{vdw}$ than Coulombic repulsions (V38E). Stated

another way, negatively charged T20 is pulled towards positive mutations on gp41 (I37K and Q40K) which affects $\Delta\Delta E_{vdw}$ more significantly (more unfavorable changes to $\Delta\Delta E_{vdw}$) than repulsion as a result of the negatively charged mutation (V38E). Consistent with this interpretation, favorable Coulombic attractions in this system are observed to lead in some cases to unfavorable van der Waals energies at the site of the interaction as shown by the two matrix entries colored blue in Figure 12 which represent a intermolecular R31-D153 salt bridge and a R46-Q139 H-bond. For V38E, intermolecular repulsions which occurs as a result of the change to a negative residue leads to large losses in favorable $\Delta\Delta E_{coul}$ (Table 3 column C) and a corresponding unfavorable solvent-mediated electrostatic energy $\Delta\Delta G_{elec}$ (Table 3 column F). Overall, the simulation results suggest that $\Delta\Delta G_{FR}$ is dominated by losses in favorable steric packing for I37K and Q40K vs losses in favorable electrostatic energy for V38E as a result of large changes in Coulombic attraction or repulsion. This analysis is consistent with a recent study by Eggink et al (67) who grouped charged gp41 mutations into similar mechanistic classes. In a related study, Gochin et al (68) found long-range electrostatic forces to be important for C-peptide peptide binding.

Origins of Resistance: Neutral Mutations

L33 is situated at the start of the gp41 mutation region. Examination of the interaction heatmap in Figure 12 shows strong wildtype ΔE_{vdw} interactions for L33 on gp41 primarily with two polar residues within the charged/polar patch on T20 at N145 and E148 but also with two hydrophobic residues at the end of the patch at positions L149 and L152. The mutations L33S and L33Q both represent nonpolar to polar mutations occurring at the same location on gp41 and originating from the same wildtype residue. Therefore, changes in binding as a result of mutations at L33 are due solely to differences in the final state in contrast to Q40K and I37K for which FR appears to be primarily a function of differences originating in the initial state. Although L33S and L33Q are neutral there are significant changes both in size and polarity. Interestingly, differential footprints show significant energy loss for L33S at position 33 while L33Q shows only minimal change (Figure 11c orange vs yellow line). However, additional losses for L33Q relative to L33S are observed in a cluster of four hydrophobic residues at locations A22, M24, T25, and L26, likely related to suboptimal packing of the more bulky glutamine side chain relative to serine. Losses localized to position 33 are consistent with L33S leading to reduction in size for serine relative to either glutamine or leucine. Likewise, gains at position 36 for van der Waals as a result of G36V (Figure 11c magenta line) are physically reasonable given the increase in size and hydrophobicity. In contrast, Q40H reveals no change in $\Delta\Delta E_{vdw}$ at position 40.

It should also be noted that all mutations (neutral and charged) lead to reduced van der Waals interactions at M19 on gp41 (Figure 11a,c), and to a lesser extent A15. Examination of the heat map matrix reveals M19 and A15 (Figure 12 x-axis) makes favorable van der Waals contacts primarily with T20 residues W159 and F162 (Figure 12, y-axis). W159 in particular makes the most favorable ΔE_{vdw} interactions overall across the gp41 interface thus all mutations compellingly appear to involve one of the most important T20 residues in terms of steric packing. It should be noted that there is considerable discussion in the literature regarding C-terminal residues on T20 (WNWF motif), and how they may interact with gp41 and/or the lipid bilayer (69–74). As discussed further below, the present computational results clearly indicate that both types of interactions are significant.

Comparison of the key Coulombic changes for L33S and L33Q (Figure 11d orange vs yellow lines) reveal that in general, positions which show losses in $\Delta\Delta E_{coul}$ are those which originally demonstrated strong interactions in wildtype gp41 (see Figure 8). Here, glutamine is more polar than serine which could explain the accentuated unfavorable $\Delta\Delta E_{coul}$ effects of the L33Q mutation as compared to L33S at positions Q40, Q41, R46 (Figure 11d orange vs yellow lines).

Otherwise, similar effects on the $\Delta\Delta E_{\text{coul}}$ landscapes are observed with both L33 mutations leading to favorable changes occurring at positions Q29, R31, L33, Q39, N42, and E49 (Figure 11d). As L33S and L33Q are both polar mutations similarities in $\Delta\Delta E_{\text{coul}}$ profiles are not unexpected. In general, Coulombic profiles for Q40H and G36V (Figure 11d magenta and purple lines) show fewer per-residue losses and are also somewhat flatter compared to L33S and L33Q (Figure 11d orange and yellow lines) which could play a role in these mutations being less detrimental.

Comparison with T20-analog Studies

The binding geometry presented here for T20 agrees well with additional experimental results for other mutations which affect C-peptide binding with gp41 (67,74–76). Champagne et al (74) reported four T20 substitutions L152A, D153A, W155A, and A156Q each of which resulted in a loss of binding with wildtype gp41. As illustrated in Figure 12 (y-axis magenta ellipsoids), three of the four T20 residues are predicted to make significant van der Waals contact with wildtype gp41 thus suggesting that mutations which decrease packing at these sites would be detrimental. For position D153, although favorable packing is not significant (Figure 12 y-axis), the strong salt bridge with R31 noted earlier would be disrupted by D153A, and, as postulated by Champagne et al, (74) affinity would likely be lost as a result of the mutation. The same study (74) also reported that a T128I substitution on T20 resulted in an increase in binding. Given the relatively weak initial interaction observed between T128 and gp41 (Figure 12 y-axis blue ellipsoid), the experimentally observed increase in binding energy for T128I analog is likely a function of increased favorable packing with adjacent hydrophobic Leu residues in the gp41 range 54–57.

Izumi et al. (75) reported a series of T20 analogs with mutations at position S138 for which non-aromatic hydrophobic substitutions were observed to enhance binding with wildtype gp41 (75). Figure 12 reveals that in addition to strong interactions with N43, T20 also interacts favorably with residues Q40, L44, and L45. Favorable interactions here involving Leu residues are consistent with the authors' suggestion that clinically observed secondary S138A mutation (or other non-aromatic hydrophobic substitutions) compensates for losses resulting from the commonly observed N43D mutation through increased favorable packing with L44 and L45 (75,76).

Membrane Interactions: The importance of the WNWF motif

A growing body of experimental evidence (69–74) suggests that T20 activity requires both N-helical binding and membrane binding. In particular, the four C-terminal residues on T20 termed the WNWF motif (residues 159–162, Figure 1) are thought to interact with the host cell membrane during fusion (72,73). A side-by-side comparison of van der Waals and Coulombic energy footprints (Figure 13) clearly indicate that in addition to gp41 (dashed lines), the C-terminal end of T20 also makes significant interactions with the lipid bilayer (black lines). T20 packing interactions with membrane first appear to be important starting around W155, followed by L158, W159, W161, and F162. Importantly, the two terminal T20 residues (Trp161, and Phe162) show strikingly enhanced lipid-packing compared with other residues and in a relative sense the interactions are much stronger than those which occur with gp41 (Figure 13a solid vs dashed lines). L158 also makes stronger van der Waals interactions with the membrane than with gp41. Interestingly, the side-by-side comparison reveals that when C-terminal T20 residues make strong packing interactions with gp41, weaker interactions are observed with the lipid and vice versa.

The Coulombic plot reveals that the seven charged T20 residues interact with gp41 and the membrane (Figure 13b dashed vs black lines) in an overall similar manner (i.e. peak location) which results in a net favorable interaction energy with the membrane. Interestingly, despite

the fact that the T20 is bound along the full length of the inner coiled-coil, more favorable Coulombic energies are observed to occur with the membrane, as opposed to gp41, for three out of the four residues in the C-terminal WNWF motif (Figure 13b dashed vs black lines). The regularly repeating pattern in the van der Waals footprint indicates which T20 residues are in close contact with gp41 (Figure 13a dashed lines) and include W159, N160, and F162 in the WNWF motif. In contrast, prior experimental studies (72, 74) have suggested T20 C-terminal residues may not interact with gp41. Champagne et al. (74) recently reported that the mutation WNWF→ANAA led to no change in T20 binding affinity using a gp41 construct termed 5H-ex (N-helix a.a. 19–71). And, Wexler-Cohen et al. (72) has reported that D-configuration substitutions on T20 at positions L158 and W159 did not significantly change binding to a construct termed N54 (N-helix a.a. 17–70). However, the van der Waals heat map results (Figure 12 x-axis) clearly indicate that T20 also makes favorable contacts with the gp41 inner coiled-coil down to residues 16, 15, and 12. Thus, the shorter truncated N-helical sites employed in these prior experimental studies (72, 74) may not have provided a complete binding interface for T20. Consistent with this explanation, Champagne et al. (74) noted that, compared to the peptide C37, substantial differences in affinity for 5H-ex were observed with T20 which might suggest the construct does not fully mimic a complete gp41 binding site.

From a structural perspective, several types of interesting interactions were identified with the aid of stereo 3-D visualization between residues in the WNWF motif and the membrane which could contribute to stability of the overall complex (Figure 14). Specifically, lipid head groups appeared to be interacting favorably with aromatic rings and with the T20 backbone. In an effort to quantify such interactions, radial distribution functions (rdfs) were computed for several types of interactions identified visually as being potentially long-lived. Distinct structural features, representative of a "first solvation shell" for T20 with lipid were observed in several rdfs as shown in Figure 14a. Representative examples of favorable interactions (d_2-d_4 distances $\leq 5 \text{ \AA}$) are visually illustrated in the accompanying graphic (Figure 14b).

Plots in Figure 14a show the rdf which yielded the largest number of interactions (distance $\leq 5 \text{ \AA}$) out of several atom selections evaluated for each WNWF residue with the lipid head groups. Although numerous distance definitions (and thus rdfs) are possible, summations up to the peaks at around 5 \AA strongly suggest that the aromatic ring of W161 is particularly well solvated by lipid head groups with 0.85 interactions followed by the backbone O of N160 with 0.58 interactions (Figure 14a). For W161, this highly populated pi-type interaction likely corresponds with the distinct Coulombic energy peak seen at this position in the membrane vs gp41 footprint plots for T20 shown in Figure 13. In sharp contrast, the rdf for the aromatic ring of W159 shows essentially no first solvation shell peak (Figure 14a top) indicating an absence of specific interactions with the polar lipid head groups. This result is consistent with W159 being particularly well packed on the gp41 surface in comparison with other residues in the WNWF region (Figure 13 dashed vs solid lines). For F162, although the backbone O does show weak lipid interactions (Figure 14, d4), surprisingly, the side chain is not solvated by head groups as seen with W161, despite the fact that the aromatic ring might also have participated in pi-type bonding. Examination of the MD trajectories reveals that the phenyl group on F162 is too well buried in the lipid bilayer to make direct contact with polar head groups. Supporting this explanation, results in Figure 13a show that F162, the terminal residue on T20, makes stronger more favorable van der Waals interactions with the membrane than any other residue in the WNWF motif.

Alternative Fusion Peptide Conformations: Alpha-helix vs beta-sheet

As noted earlier, secondary structure preferences for the FP region have been reported as α -helical, β -strand, or various mixtures (54–59) however these prior studies have generally employed FP monomers and not trimers which is the more biologically relevant form. A more

recent experimental study by Qiang and Weliky (77), which used covalently linked FP trimers, revealed conformational preferences which were consistent with a β -sheet or a mixture of α -helical and β -sheet structure depending on the presence or absence, respectively, of cholesterol in lipid membranes. To examine if the essential binding features of the T20-gp41 model are strongly dependant on the initial choice of FP conformation we constructed two additional models representing a antiparallel β -sheet conformation (Figure 15a middle), an alternative alpha-helix construct in an extended form (Figure 15a right), for comparison with the original alpha-helix coiled-coil (Figure 15a left).

For construction of a FP trimer with beta-sheet character, a beta-structure motif of suitable length was taken from pdb entry 3D58 (residues 70–88) and mapped onto the initial model through rmsd matching of backbone atoms on 3D58 (residue 88) with gp41 (residue 18). This allowed for replacement of the initial FP trimer (first 16 a.a.) with a backbone having a beta-sheet conformation. The alternative alpha-helix extended model (Figure 15a right) was similarly constructed by replacing residues 1–17 of the coiled-coil. In this case, with coordinates taken from the solution structure of monomeric g41 FP bound to DPC (pdb entry 2PJY) reported by Li and Tamm.(59) Here, the two structures were merged after a best-fit rmsd match using Ca, C, N backbone atoms at residue Ser 17. Analogous to construction of the initial coiled-coil (see Methods), added motifs were mutated to the correct HIV sequence if required followed by side-chain rotamer exploration, 3-D stereo visualization, and energy minimization to resolve any steric conflicts as a result of model building prior to beginning MD.

As shown in Figure 15b, MD simulations using the three different FP conformations (Figure 15a) yield ΔE_{vdw} , ΔE_{coul} , and HBond footprints for T20 with wildtype gp41 which are remarkably similar. Such agreement strongly suggests that the previously discussed quantitative and qualitative arguments made regarding origins of resistance for T20 are likely not critically dependent on the choice of FP starting structure. However, future studies would be desirable to more fully investigate this issue. Notably, all three simulations yield essentially superimposable ΔE_{vdw} profiles (Figure 15b top), particularly in the mutation hotspot region of interest (a.a. 33–43, Table 1). Footprint profiles for ΔE_{coul} (Figure 15b middle) also show good overlap in the mutation region although there is more variation than for ΔE_{vdw} which likely involves inherently larger fluctuations for gas-phase Coulombic energies (see Figure 5 and Table 3 BASEM values). Interestingly, results in the mutation region from the alpha-helical coiled-coil (red, Figure 15b middle) and beta-sheet (green, Figure 15b middle) simulations are more similar than from the extended alpha-helix (blue, Figure 15b middle). Similar to ΔE_{vdw} , the Hbond footprints show nearly perfect overlap in the key mutation region, including the large peaks observed at V38, Q40, and N43. The overall agreement in the footprints derived from different simulations provides additional assurance that reasonable convergence and reproducibility has been achieved, specifically with regards to these intermolecular interactions.

CONCLUSION

In this study, explicit solvent, explicit lipid, and all-atom molecular dynamics (MD) computer simulations and free energy calculations were used to estimate fold resistance energies ($\Delta\Delta G_{FR}$) for the viral entry inhibitor T20 in complex with its target HIVgp41. The primary goals were to develop structural models consistent with experimental observation in an effort to understand which structural and energetic features contribute to drug resistance. Notably, this study has resulted in development and validation of the first complete model for T20 in complex with gp41 that includes the fusion peptide region embedded into a host-cell lipid bilayer (Figure 4). Simulations stability was carefully assessed by monitoring fluctuations in root-mean-square-deviation (rmsd) and in the energy components used to estimate free energies of binding. Well behaved instantaneous energies, block-averaged energies, and

structural rmsds (Figure 5) indicate stable simulation behavior and autocorrelation functions (ACF) and block averaged standard errors of the mean (BASEM) for each ΔG_{bind} calcd time-series (Figure 7, Table 2) indicate reasonably converged results.

The results from simulations of T20 with wildtype gp41 reveal the inhibitor makes particularly strong intermolecular ΔE_{vdw} , ΔE_{coul} , and H-bond interactions (Figure 8) with gp41 residues known to mutate (Table 1). Thus, the proposed T20 footprint is in good qualitative agreement with clinically observed resistance patterns. Interestingly, simulations which employed a beta-sheet structure or an extended alpha-helix conformation for the gp41 FP region yielded remarkably similar ΔE_{vdw} , ΔE_{coul} , and Hbond footprints (Figure 15) suggesting the binding interface for T20 is not critically influenced by the FP structure. From more quantitative points of view, additional simulations of T20 in complex with the gp41 coiled-coil containing L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K point mutations correctly lead to decreases in binding (Tables 3). And, the calculations correctly predict L33Q and Q40K mutations will be most detrimental. Compellingly, for six out of the seven mutations studied, the overall trend in the experimental data is well-reproduced with calculated vs experimental fold resistance energies yielding a strong $r^2 = 0.72$ (Figure 9). Interestingly, results for L33S are offset from the other six datapoints although in an absolute sense the magnitude $\Delta\Delta G_{\text{FR}}$ calcd is in good agreement. As L33S is the only mutation leading to a smaller sidechain the offset is probably linked to differences in size and associated entropic effects not accounted for in the calculations.

Decomposition of the energy components used to estimate $\Delta\Delta G_{\text{FR}}$ revealed that charged mutations lead to very large changes in Coulombic energy ($\Delta\Delta E_{\text{coul}}$) and compensating desolvation ($\Delta\Delta G_{\text{polar}}$) terms (Table 3). Structurally, these effects are traced to repulsion and attraction involving gp41 and seven Asp/Glu and two Lys residues on T20 (net change -5) proximal to the mutation sites (Figure 10). As a result, unfavorable $\Delta\Delta E_{\text{coul}}$ and solvent-mediated electrostatics ($\Delta\Delta G_{\text{elec}}$) for the negatively charged V38E mutation are significantly more unfavorable than other mutations (Table 3). Interestingly, all mutations but one lead to reduced van der Waals interactions (Table 3) and the greatest losses occur for the two positively charged mutations (I37K, Q40K). A reasonable mechanism to explain these results is that negatively charged T20 is pulled towards positive gp41 mutations which lead to changes in $\Delta\Delta E_{\text{vdw}}$ more unfavorable than those arising from repulsion due to gp41 mutations of the opposite charge. The mutations are also observed to disrupt H-bonding in all cases but one (Table 3) and the largest losses occur for mutations involving position Q40 which originally showed the greatest number of interactions in the wildtype system (Figure 8).

Origins of the difference in experimentally observed fold resistance (FR) for I37K (less detrimental) vs Q40K (more detrimental) mutations was traced to initial wildtype interactions (Figure 12) involving a polar/charged patch on T20. In contrast, FR trends for the neutral L33S and L33Q mutations appear to be due to differences in the final state (Figure 11) as a result of suboptimal packing of glutamine relative to the smaller serine sidechain. Compellingly, all mutations lead to less favorable van der Waals interactions with M19 on gp41 (Figure 11) which is in strong contact with the most important T20 residue (W159) in terms of inhibitor steric packing (Figure 12). The model geometry is also consistent with recent experimental results (74) reporting T20 mutations which enhanced (T128I) or reduced (L152A, D153A, W155A, and A156Q) binding. T128 is observed to make only weak initial interactions with gp41 (Figure 12) thus a mutation to Ile would likely increase favorable packing with adjacent Leu residues. On the other hand, T20 mutations which reduce binding likely involve disruption of initially favorable van der Waals (L152, W155, A156) or Coulombic (D153) interactions (Figures 10, 12).

C-terminal T20 residues are known to be critical for antiviral activity (69–71) and thought to interact with the host cell membrane (72,73) although some studies have concluded they may

not contribute to gp41 binding (72,74). In contrast, results from the present work indicate the C-terminal region makes significant interactions with both the lipid bilayer and with gp41 (Figures 4,13, and 14). F162 on T20 was found to interact most favorably with the membrane (Figure 13) and radial distribution functions (rdfs) reveal that W161 (aromatic rings) and N160 (backbone) are particularly well-solvated by lipid headgroups. Importantly, the van der Waals heat map results indicate favorable interactions occur between T20 and gp41 which extend down the inner coiled-coil to residues 16, 15, and 12 (Figure 12). Based on these results, prior studies may not have employed complete enough N-helical complements for T20 which could partially explain why C-terminal residues did not appear to be important in gp41 binding (72,74). Overall, results from the present computational study are remarkably consistent with suggestions by Champagne et al. (74) that T20 binding could be stabilized through interactions of the WNWF motif with residues in the gp41 fusion peptide region (Figure 12) and that favorable interactions with membrane (Figures 13–14) could also contribute to stability of the complex. A reasonable hypothesis suggested by the present work is that peptide inhibitors designed to increase favorable interactions with the membrane could stabilize binding to gp41 and improve affinity. Such computational evaluations are currently underway in our laboratory. Given the importance of T20 as the first-in-class membrane fusion inhibitor, parallel experimental studies should also be pursued.

Acknowledgments

Gratitude is expressed to David Green, Carlos Simmerling, and Haipeng Xing for helpful discussions and to Bentley Strockbine, Noel Carrascal, and Patrick M. Holden for computational assistance. This research utilized resources at the New York Center for Computational Sciences at Stony Brook University/Brookhaven National Laboratory which is supported by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886 and by the State of New York.

Abbreviations

ACF	autocorrelation function
BASEM	block averaged standard errors of the mean
DOPC	dioleoylphosphatidylcholine
FP	fusion peptide
FR	fold resistance
HIVgp41	human immunodeficiency virus glycoprotein 41
MD	Molecular Dynamics
MM-GBSA	Molecular Mechanics Generalized Born Surface Area
rmsd	root-mean-square-deviation

References

1. Report on the global AIDS epidemic: Epidemiology slides, Joint United Nations Programme on HIV/AIDS (UNAIDS) and The World Health Organization (WHO). 2008 [accessed Aug 3, 2009]. http://data.unaids.org/pub/GlobalReport/2008/2008_globalreport_core_en.ppt
2. NIGMS-Supported Structure-Based Drug Design Saves Lives. NIH/NIGMS Fact Sheet. [accessed Aug 3, 2009]. http://www.nigms.nih.gov/Publications/structure_drugs.htm
3. Drugs@FDA website. <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm> accessed Aug 3, 2009
4. Grinsztejn B, Nguyen BY, Katlama C, Gatell JM, Lazzarin A, Vittecoq D, Gonzalez CJ, Chen J, Harvey CM, Isaacs RD. Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in

- treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial. *Lancet* 2007;369:1261–1269. [PubMed: 17434401]
5. Davison, DK.; Medinas, RJ.; Mosier, SM.; Bowling, TS.; Delmedico, MK.; Dwyer, JJ.; Cammack, N.; Greenberg, ML. New fusion inhibitor peptides, TRI-999 and TRI-1144, are potent inhibitors of enfuvirtide and T-1249 resistant isolates. XVI International AIDS Conference: Conference Reports for NATAP, Poster THPE0021; 2006.
 6. Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* 2001;70:777–810. [PubMed: 11395423]
 7. Suntoko TR, Chan DC. The Fusion Activity of HIV-1 gp41 Depends on Interhelical Interactions. *J Biol Chem* 2005;280:19852–19857. [PubMed: 15772068]
 8. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997;89:263–273. [PubMed: 9108481]
 9. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 1997;387:426–430. [PubMed: 9163431]
 10. Schibli, Weissenhorn W. Class I and class II viral fusion protein structures reveal similar principles in membrane fusion. *Mol Membr Biol* 2004;21:361–371. [PubMed: 15764366]
 11. Eckert DM, Kim PS. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc Natl Acad Sci USA* 2001;98:11187–11192. [PubMed: 11572974]
 12. Jiang SB, Lin K, Strick N, Neurath AR. Hiv-1 Inhibition by a Peptide. *Nature* 1993;365:113–113. [PubMed: 8371754]
 13. Wild C, Oas T, McDanal C, Bolognesi D, Matthews T. A Synthetic Peptide Inhibitor of Human Immunodeficiency Virus Replication: Correlation between Solution Structure and Viral Inhibition. *Proc Natl Acad Sci USA* 1992;89:10537–10541. [PubMed: 1438243]
 14. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* 1999;99:103–115. [PubMed: 10520998]
 15. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS. Potent D-peptide inhibitors of HIV-1 entry. *Proc Natl Acad Sci USA* 2007;104:16828–16833. [PubMed: 17942675]
 16. Chan DC, Chutkowski CT, Kim PS. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc Natl Acad Sci USA* 1998;95:15613–15617. [PubMed: 9861018]
 17. Debnath AK, Radigan L, Jiang SB. Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. *J Med Chem* 1999;42:3203–3209. [PubMed: 10464007]
 18. Jiang SB, Lu H, Liu SW, Zhao Q, He YX, Debnath AK. N-substituted pyrrole derivatives as novel human immunodeficiency virus type 1 entry inhibitors that interfere with the gp41 six-helix bundle formation and block virus fusion. *Antimicrob Agents Chemother* 2004;48:4349–4359. [PubMed: 15504864]
 19. Frey G, Rits-Volloch S, Zhang XQ, Schooley RT, Chen B, Harrison SC. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proc Natl Acad Sci USA* 2006;103:13938–13943. [PubMed: 16963566]
 20. Cai L, Gochin M. A Novel Fluorescence Intensity Screening Assay Identifies New Low Molecular Weight Inhibitors of the gp41 Coiled Coil Domain of HIV-1. *Antimicrob Agents Chemother* 2007;23:23.
 21. Chinnadurai R, Rajan D, Munch J, Kirchhoff F. HIV-1 Variants Resistant against 1st and 2nd Generation Fusion Inhibitors and Cytopathic in ex vivo Human Lymphoid Tissue. *J Virol* 2007;11:11.
 22. Armand-Ugon M, Gutierrez A, Clotet B, Este JA. HIV-1 resistance to the gp41-dependent fusion inhibitor C-34. *Antiviral Res* 2003;59:137–142. [PubMed: 12895697]
 23. Armand-Ugon M, Clotet-Codina I, Tintori C, Manetti F, Clotet B, Botta M, Este JA. The anti-HIV activity of ADS-J1 targets the HIV-1 gp120. *Virology* 2005;343:141–149. [PubMed: 16168454]
 24. Greenberg ML, Cammack N. Resistance to enfuvirtide, the first HIV fusion inhibitor. *Journal of Antimicrobial Chemotherapy* 2004;54:333–340. [PubMed: 15231762]
 25. Mink M, Mosier SM, Janumpalli S, Davison D, Jin L, Melby T, Sista P, Erickson J, Lambert D, Stanfield-Oakley SA, Salgo M, Cammack N, Matthews T, Greenberg ML. Impact of human

- immunodeficiency virus type 1 gp41 amino acid substitutions selected during enfuvirtide treatment on gp41 binding and antiviral potency of enfuvirtide in vitro. *J Virol* 2005;79:12447–12454. [PubMed: 16160172]
26. Melby T, Sista P, DeMasi R, Kirkland T, Roberts N, Salgo M, Heilek-Snyder G, Cammack N, Matthews TJ, Greenberg ML. Characterization of Envelope Glycoprotein gp41 Genotype and Phenotypic Susceptibility to Enfuvirtide at Baseline and on Treatment in the Phase III Clinical Trials TORO-1 and TORO-2. *AIDS Research and Human Retroviruses* 2006;22:375–385. [PubMed: 16706613]
27. Nameki D, Kodama E, Ikeuchi M, Mabuchi N, Otaka A, Tamamura H, Ohno M, Fujii N, Matsuoka M. Mutations conferring resistance to human immunodeficiency virus type 1 fusion inhibitors are restricted by gp41 and rev-responsive element functions. *J Virol* 2005;79:764–770. [PubMed: 15613304]
28. Rimsky LT, Shugars DC, Matthews TJ. Determinants of Human Immunodeficiency Virus Type 1 Resistance to gp41-Derived Inhibitory Peptides. *J Virol* 1998;72:986–993. [PubMed: 9444991]
29. Strockbine B, Rizzo RC. Binding of Anti-fusion Peptides with HIVgp41 from Molecular Dynamics Simulations: Quantitative Correlation with Experiment. *Proteins: Struct Funct Bioinf* 2007;63:630–642.
30. Caffrey M. Model for the structure of the HIV gp41 ectodomain: insight into the intermolecular interactions of the gp41 loop. *BBA-Molecular Basis of Disease* 2001;1536:116–122. [PubMed: 11406346]
31. Lori F, Hall L, Lusso P, Popovic M, Markham P, Franchini G, Reitz MS Jr. Effect of reciprocal complementation of two defective human immunodeficiency virus type 1 (HIV-1) molecular clones on HIV-1 cell tropism and virulence. *J Virol* 1992;66:5553–5560. [PubMed: 1501290]
32. MOE Version 2005. 06. Chemical Computing Group Inc; Montreal, Canada:
33. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comp Chem* 2005;26:1668–1688. [PubMed: 16200636]
34. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712–725. [PubMed: 16981200]
35. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem* 2005;26:1781–1802. [PubMed: 16222654]
36. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33. [PubMed: 8744570]
37. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.
38. Feller SE, MacKerell AD. An improved empirical potential energy function for molecular simulations of phospholipids. *J Phys Chem B* 2000;104:7510–7515.
39. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 1983;79:926–935.
40. Feller, Scott E. Wabash College Chemistry Department; [accessed December 22, 2008]. <http://persweb.wabash.edu/facstaff/fellers>
41. Darden T, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–10092.
42. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 1995;103:4613–4621.
43. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* 1977;23:327–341.
44. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc* 1998;120:9401–9409.

45. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;33:889–897. [PubMed: 11123888]
46. Baliaus TE, Rizzo RC. Quantitative Prediction of Fold Resistance for Inhibitors of EGFR. *Biochemistry* 2009;48:8435–8448. [PubMed: 19627157]
47. Carrascal N, Rizzo RC. Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13. *Bioorg Med Chem Lett* 2009;19:47–50. [PubMed: 19042129]
48. Chachra R, Rizzo RC. Origins of Resistance Conferred by the R292K Neuraminidase Mutation via Molecular Dynamics and Free Energy Calculations. *J Chem Theory Comput* 2008;4:1526–1540.
49. Hawkins GD, Cramer CJ, Truhlar DG. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem Phys Lett* 1995;246:122–129.
50. Hawkins GD, Cramer CJ, Truhlar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* 1996;100:19824–19839.
51. Moustakas DT, Therese Lang PT, Pegg S, Pettersen E, Kuntz ID, Broojimans N, Rizzo RC. Development and Validation of a Modular, Extensible Docking program: DOCK 5. *J Comput Aided Mol Des* 2006;20:601–619. [PubMed: 17149653]
52. Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J Phys Chem* 1994;98:1978–1988.
53. Sackett K, Wexler-Cohen Y, Shai Y. Characterization of the HIV N-terminal fusion peptide-containing region in context of key gp41 fusion conformations. *J Biol Chem* 2006;281:21755–21762. [PubMed: 16751188]
54. Kamath S, Wong TC. Membrane structure of the human immunodeficiency virus gp41 fusion domain by molecular dynamics simulation. *Biophys J* 2002;83:135–143. [PubMed: 12080106]
55. Gordon LM, Mobley PW, Pilpa R, Sherman MA, Waring AJ. Conformational mapping of the N-terminal peptide of HIV-1 gp41 in membrane environments using C-13-enhanced Fourier transform infrared spectroscopy. *BBA-Biomembranes* 2002;1559:96–120. [PubMed: 11853678]
56. Gordon LM, Mobley PW, Lee W, Eskandari S, Kaznessis YN, Sherman MA, Waring AJ. Conformational mapping of the N-terminal peptide of HIV-1 gp41 in lipid detergent and aqueous environments using C-13-enhanced Fourier transform infrared spectroscopy. *Protein Sci* 2004;13:1012–1030. [PubMed: 15044732]
57. Sackett K, Shai Y. The HIV fusion peptide adopts intermolecular parallel beta-sheet structure in membranes when stabilized by the adjacent N-terminal heptad repeat: A C-13 FTIR study. *J Mol Biol* 2005;350:790–805. [PubMed: 15964015]
58. Zheng Z, Yang R, Bodner ML, Weliky DP. Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry* 2006;45:12960–12975. [PubMed: 17059213]
59. Li Y, Tamm LK. Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers. *Biophys J* 2007;93:876–885. [PubMed: 17513369]
60. Yau WM, Wimley WC, Gawrisch K, White SH. The preference of tryptophan for membrane interfaces. *Biochemistry* 1998;37:14713–14718. [PubMed: 9778346]
61. Scarlata S, Gruner SM. Role of phosphatidylethanolamine lipids in the stabilization of protein-lipid contacts. *Biophys Chem* 1997;67:269–279. [PubMed: 9397529]
62. Grossfield, A.; Zuckerman, DM.; Ralph, AW. Annual Reports in Computational Chemistry. Vol. 5. Elsevier; 2009. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations; p. 23–48.
63. Hess B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *The Journal of chemical physics* 2002;116:209–217.
64. Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov Design* 2000;18:113–135.
65. Dwyer JJ, Wilson KL, Davison DK, Freel SA, Seedorff JE, Wring SA, Tvermoes NA, Matthews TJ, Greenberg ML, Delmedico MK. Design of helical, oligomeric HIV-1 fusion inhibitor peptides with

- potent activity against enfuvirtide-resistant virus. *Proc Natl Acad Sci USA* 2007;104:12772–12777. [PubMed: 17640899]
66. Lu M, Blacklow SC, Kim PS. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat Struct Biol* 1995;2:1075–1082. [PubMed: 8846219]
67. Eggink D, Langedijk JP, Bonvin AM, Deng Y, Lu M, Berkhouit B, Sanders RW. Detailed mechanistic insights into HIV-1 sensitivity to three generations of fusion inhibitors. *J Biol Chem* 2009;284:26941–26950. [PubMed: 19617355]
68. Gochin M, Cai L. The role of amphiphilicity and negative charge in glycoprotein 41 interactions in the hydrophobic pocket. *J Med Chem* 2009;52:4338–4344. [PubMed: 19534533]
69. Lawless MK, Barney S, Guthrie KI, Bucy TB, Petteway SR Jr, Merutka G. HIV-1 membrane fusion mechanism: structural studies of the interactions between biologically-active peptides from gp41. *Biochemistry* 1996;35:13697–13708. [PubMed: 8885850]
70. Hildinger M, Dittmar MT, Schult-Dietrich P, Fehse B, Schnierle BS, Thaler S, Stiegler G, Welker R, von Laer D. Membrane-anchored peptide inhibits human immunodeficiency virus entry. *J Virol* 2001;75:3038–3042. [PubMed: 11222732]
71. Peisajovich SG, Gallo SA, Blumenthal R, Shai Y. C-terminal octylation rescues an inactive T20 mutant: implications for the mechanism of HIV/SIMIAN immunodeficiency virus-induced membrane fusion. *J Biol Chem* 2003;278:21012–21017. [PubMed: 12646555]
72. Wexler-Cohen Y, Johnson BT, Puri A, Blumenthal R, Shai Y. Structurally altered peptides reveal an important role for N-terminal heptad repeat binding and stability in the inhibitory action of HIV-1 peptide DP178. *J Biol Chem* 2006;281:9005–9010. [PubMed: 16455666]
73. Liu S, Jing W, Cheung B, Lu H, Sun J, Yan X, Niu J, Farmar J, Wu S, Jiang S. HIV gp41 C-terminal heptad repeat contains multifunctional domains. Relation to mechanisms of action of anti-HIV peptides. *J Biol Chem* 2007;282:9612–9620. [PubMed: 17276993]
74. Champagne K, Shishido A, Root MJ. Interactions of HIV-1 inhibitory peptide T20 with the gp41 N-HR coiled coil. *J Biol Chem* 2009;284:3619–3627. [PubMed: 19073602]
75. Izumi K, Kodama E, Shimura K, Sakagami Y, Watanabe K, Ito S, Watabe T, Terakawa Y, Nishikawa H, Sarafianos SG, Kitaura K, Oishi S, Fujii N, Matsuoka M. Design of peptide-based inhibitors for human immunodeficiency virus type 1 strains resistant to T-20. *J Biol Chem* 2009;284:4914–4920. [PubMed: 19073606]
76. Watabe T, Terakawa Y, Watanabe K, Ohno H, Nakano H, Nakatsu T, Kato H, Izumi K, Kodama E, Matsuoka M, Kitaura K, Oishi S, Fujii N. X-ray crystallographic study of an HIV-1 fusion inhibitor with the gp41 S138A substitution. *J Mol Biol* 2009;392:657–665. [PubMed: 19616557]
77. Qiang W, Weliky DP. HIV fusion peptide and its cross-linked oligomers: efficient syntheses, significance of the trimer in fusion activity, correlation of beta strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry* 2009;48:289–301. [PubMed: 19093835]

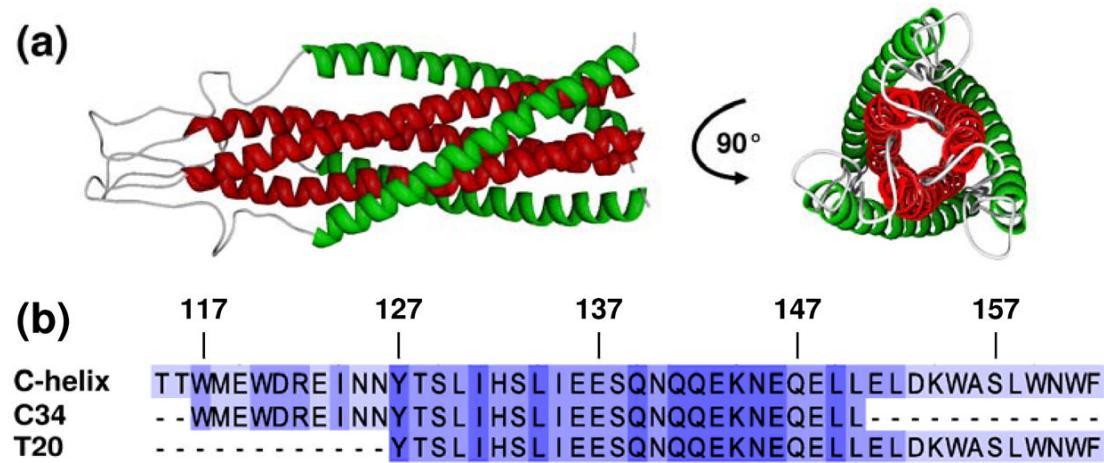


Figure 1.

(1a) HIVgp41 ectodomain from Caffery (pdb entry 1IF3) with outer C-helices colored green and inner N-helices colored red. (1b) Sequence alignment showing fusion inhibitors T20 and C34 in comparison with a region of the native outer HIVgp41 C-helix (top sequence).

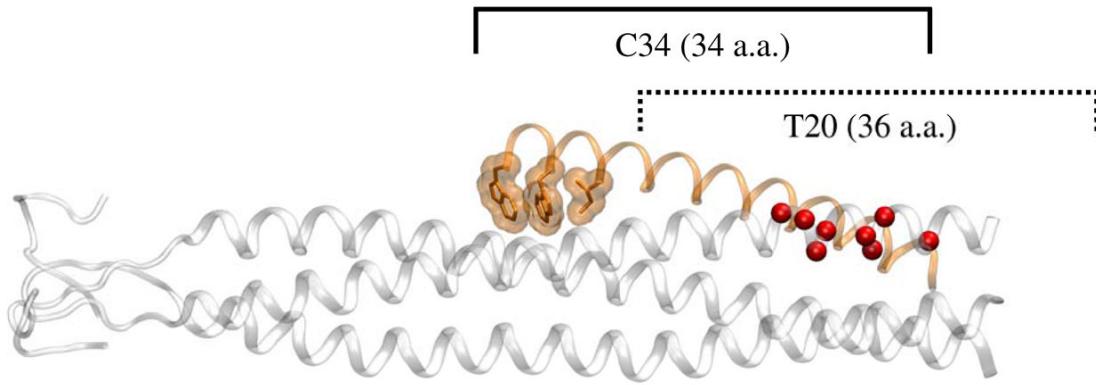


Figure 2.

Primary mutations (red spheres) which arise from use of T20 (5,21–27) map to the inner gp41 N-helices (gray ribbon). Inhibitor C34 depicted in orange ribbon with hydrophobic pocket-binding residues in molecular surface. The proposed T20 binding footprint is indicated by a dashed line. Coordinates from 1IF3.

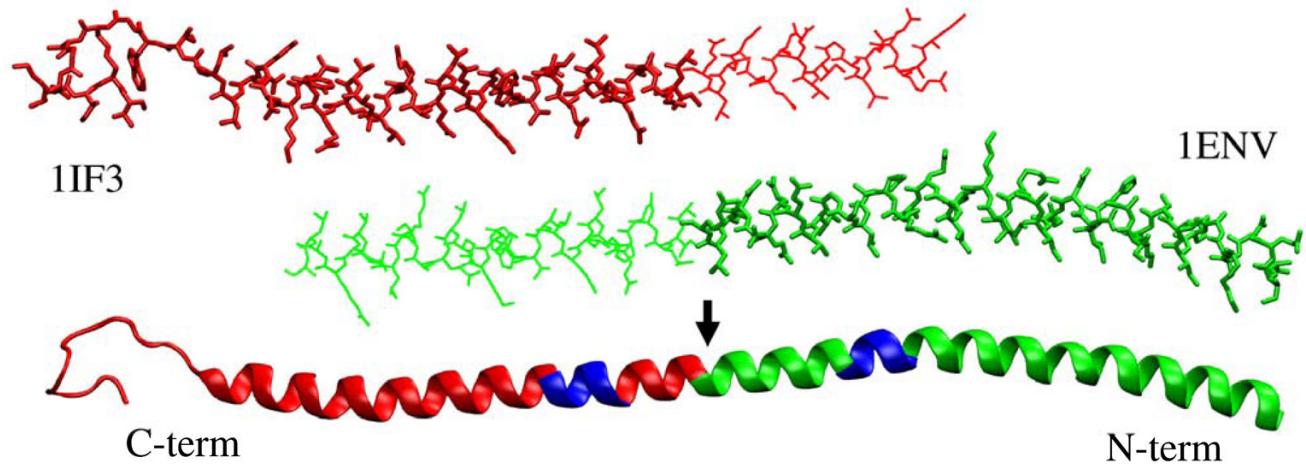
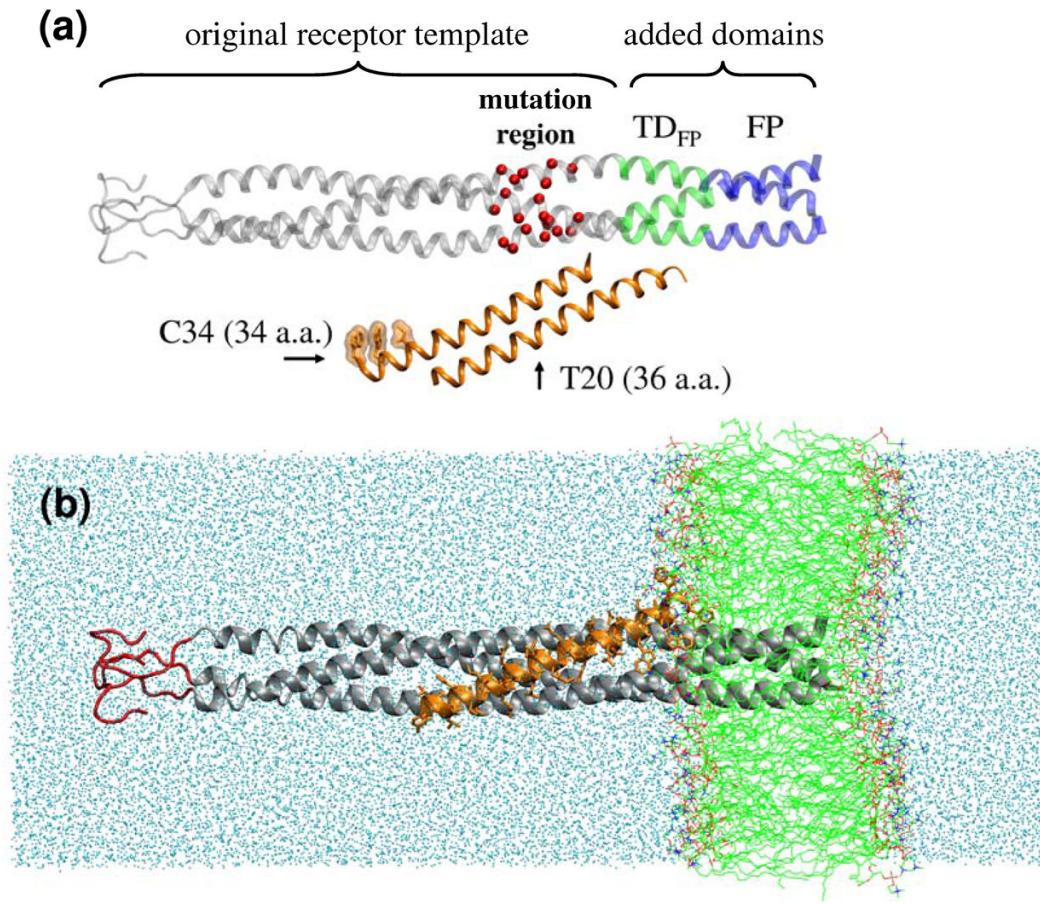
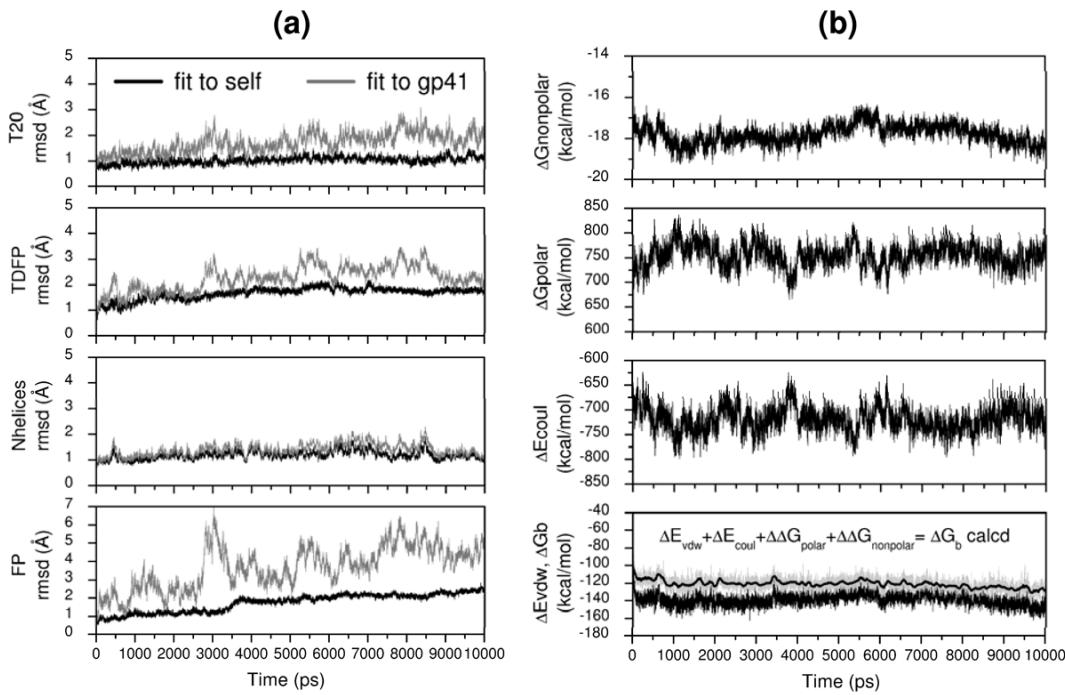


Figure 3.

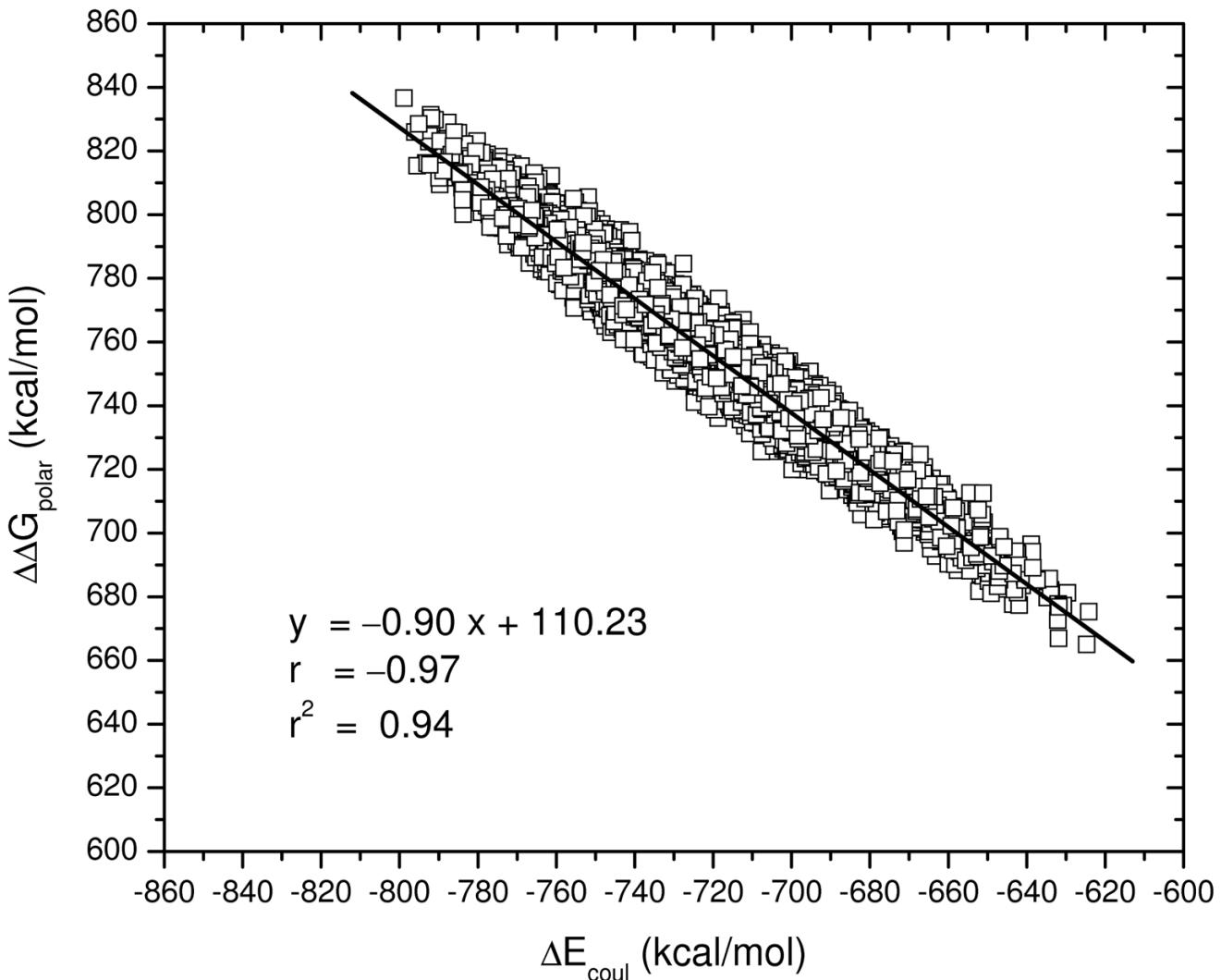
Construction of an extended gp41 structural model to accommodate T20. Pdb entries 1IF3 (top, red) and 1ENV (middle, green) were aligned using residues 31–35@C,O and 53–57@C,O in common between the two chains (blue region, bottom) and the structures were merged (arrow) to create a single continuous chain.

**Figure 4.**

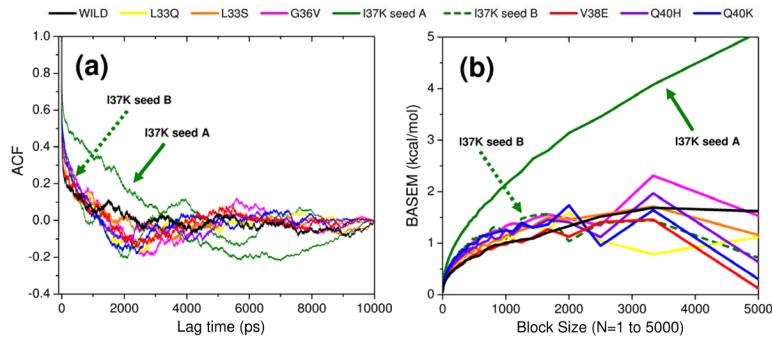
Structure of the complete gp41 model. **(4a)** Sections previously employed by Strockbine and Rizzo (29) to characterize C34 peptide binding shown in gray ribbon. Sections added include the transition domain adjacent to the fusion peptide (TD_{FP}, a.a. 17 to 29, green ribbon) and the fusion peptide (FP, a.a. 1 to 16, blue ribbon). Location of clinically relevant point mutations shown as red spheres. C34 and T20 are shown in orange ribbon offset from gp41 in the vertical direction to emphasize the two different binding footprints. **(4b)** Fully solvated, membrane-bound gp41-ligand complex with gp41 in gray, T20 in orange, lipid in green, and water in cyan. Loop region (red) restrained during simulations.

**Figure 5.**

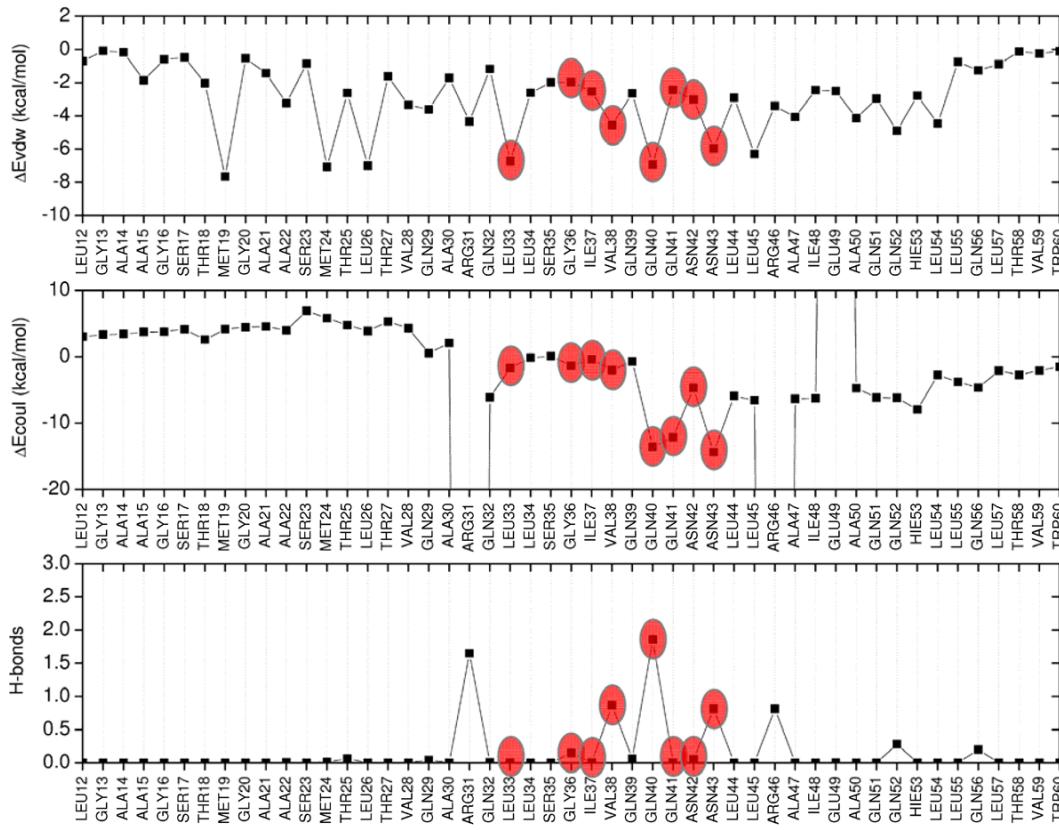
Instantaneous results from simulations of the peptide inhibitor T20 with wildtype HIVgp41 plotted vs time. (5a) Root mean square deviations (rmsd) for fusion peptide (FP, a.a. 1 to 16), transition domain adjacent to fusion peptide (TD_{FP}, a.a. 17 to 29), N-helical (residues 30 to 81), and T20 regions. Black lines represent rmsds after fitting each region to itself (fit-to-self). Gray lines represent rmsds after an initial fit to the gp41 inner coiled-coil which excluding the FP region (fit-to-gp41). Fits computed using backbone atoms (C_α, C, N, O) as the match criteria relative to the first MD frame. (5b.) Individual energy components for ΔE_{vdw} , ΔE_{coul} , $\Delta \Delta G_{\text{polar}}$, $\Delta \Delta G_{\text{nonpolar}}$, and the total estimated binding energy ($\Delta G_{\text{bind}} \text{ calcd}$, light gray) from the sum of the four energy terms. The solid black line represents block smoothed (N=100 blocks) averaging for $\Delta G_{\text{bind}} \text{ calcd}$.

**Figure 6.**

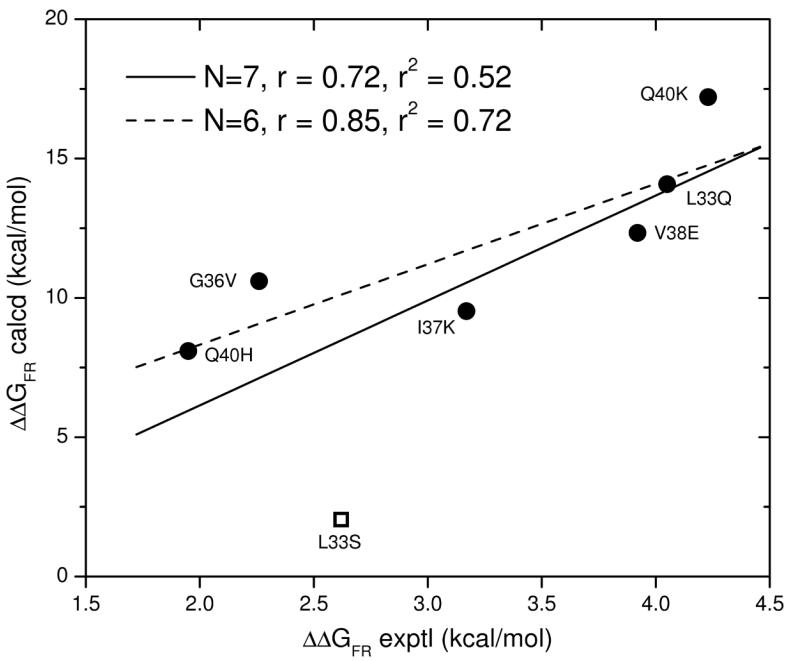
Intermolecular protein-ligand Coulombic energies (ΔE_{coul}) versus opposing change in polar desolvation penalties ($\Delta \Delta G_{\text{polar}}$). Symbols represent the instantaneous energies obtained from all frames saved during the 9 ns MD trajectory (N=9000) of T20 with wildtype gp41 and the black line is the best fit to the data.

**Figure 7.**

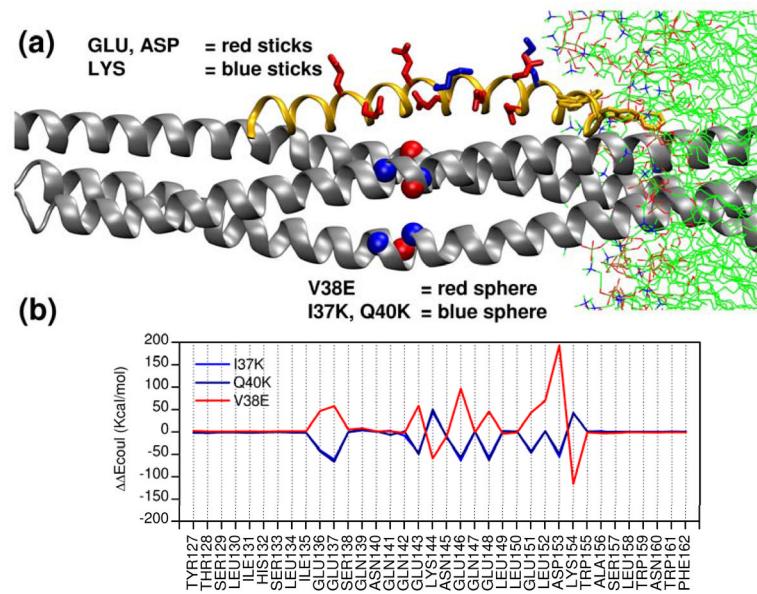
(a) Autocorrelation functions (ACF) of calculated binding energies ($\Delta G_{\text{bind}} \text{ calcd}$) for T20 with wild HIVgp41 and seven mutants. (b) Block average standard error of the mean (BASEM) in kcal/mol as a function of block size from 1 to 5000. Solid arrows indicate initial I37K results while dashed arrows indicate I37K results obtained using a different random seed.

**Figure 8.**

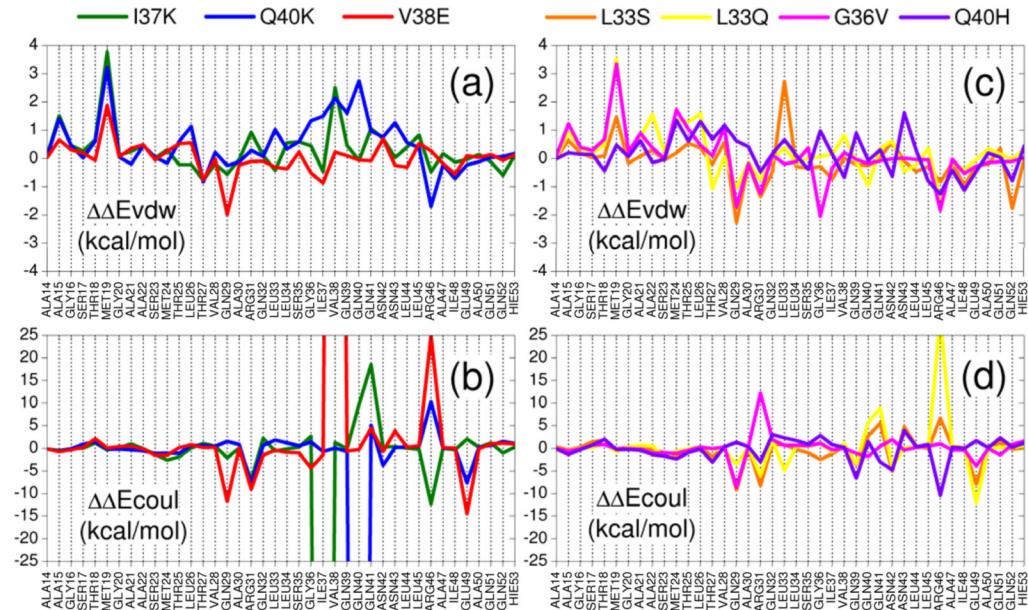
Mutations from Table 1 (red) mapped to molecular footprints obtained from simulations of T20 with wildtype HIVgp41. Footprints represent average per-residue van der Waals (top, ΔE_{vdw}), Coulombic (middle, ΔE_{coul}), and H-bonds (bottom) interactions between T20 (all residues) with each individual gp41 residue for the range Leu12-Trp60 from 10,000 MD simulation snapshots.

**Figure 9.**

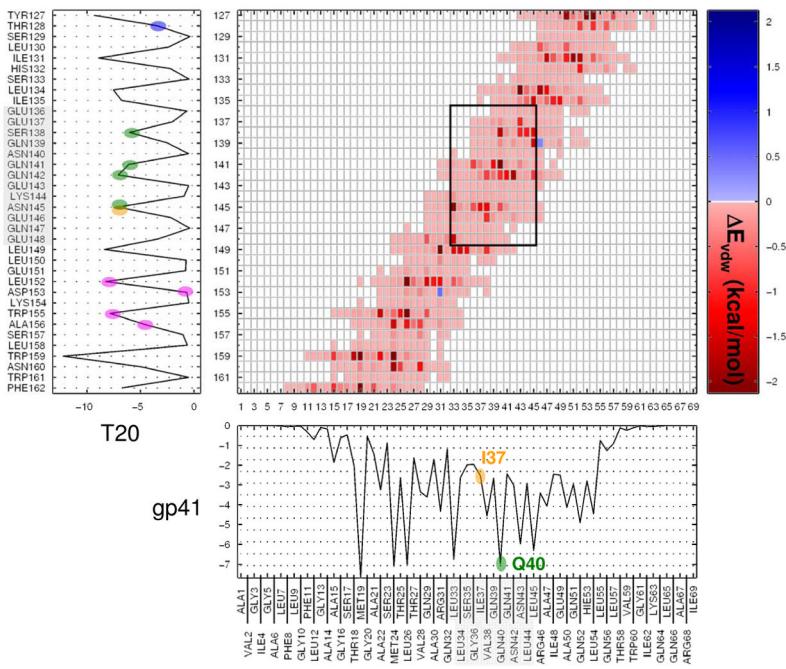
Predicted FR energies ($\Delta\Delta G_{FR}$ calcd) vs experimental FR energies ($\Delta\Delta G_{FR}$ exptl) for inhibitors with HIVgp41. Each computed datapoint is the difference in results from two independent MD simulations. Experimental values from Table 1.

**Figure 10.**

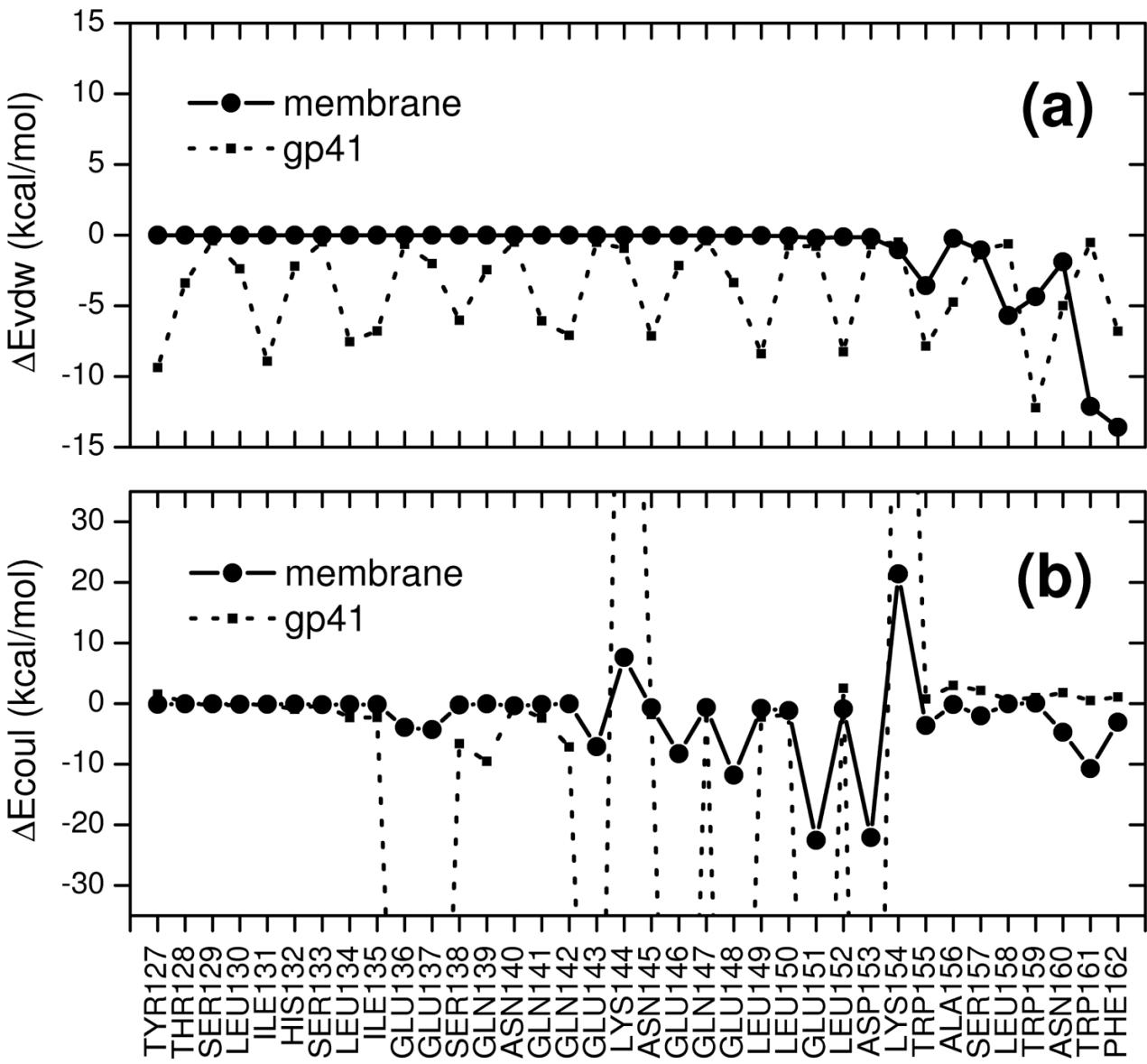
(a) Charged mutations (colored spheres) on HIVgp41 (gray ribbon) for V38E (red), I37K (blue), and Q40K (blue) relative to negative (red = GLU, ASP) and positive (blue = LYS) charged residues on T20 (orange ribbon). Lipid bilayer in green. (b) Per-residue Coulombic energy changes for each T20 residue with all of gp41 for three charged mutations.

**Figure 11.**

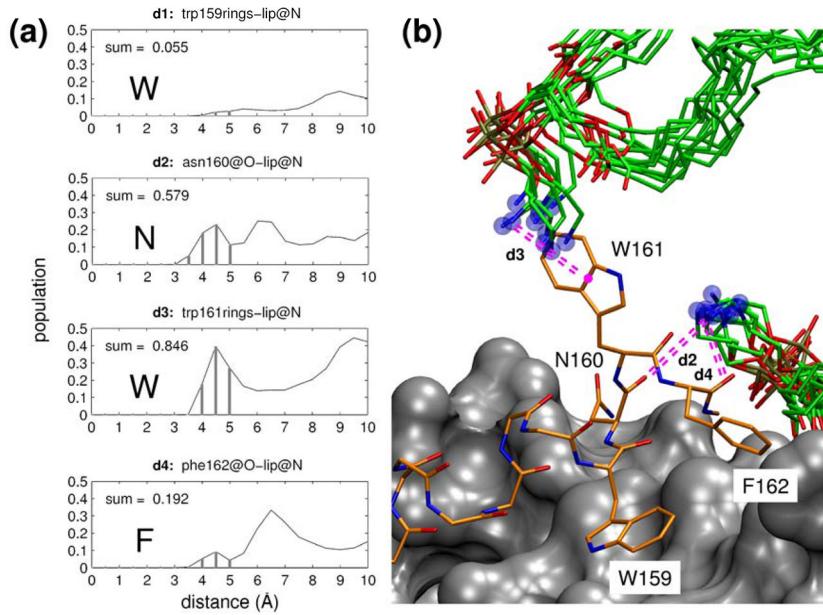
Per-residue differential (mutant-wildtype) footprints for T20 with HIVgp41 representing van der Waals (top **a,c**) and Coulombic (bottom **b,d**) and energy changes due to charged (left **a,b**) and neutral (right **c,d**) mutations.

**Figure 12.**

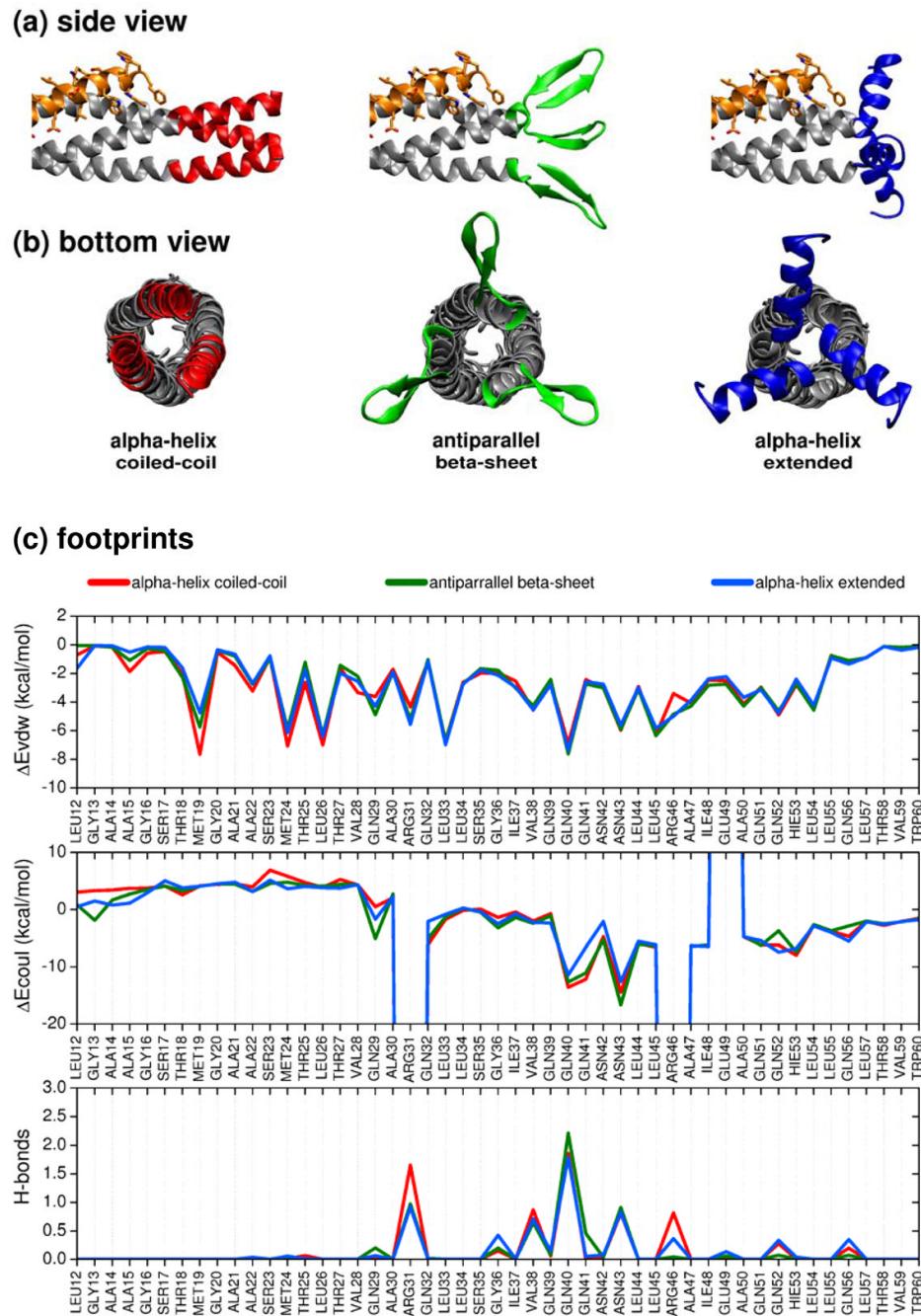
Intermolecular van der Waals interaction energy matrix (heat map) for all gp41 with T20 residues from the wildtype simulation. Footprint peak magnitudes represent summation ΔE_{vdw} (kcal/mol) along each row (T20) or column (gp41). Gray shaded regions indicate mapping of the gp41 mutation region(L33–L45) to a charged/polar patch (E136–E148) on T20 defined by the black box in the ca. center of the matrix. Heat map gradient colors indicate the magnitude of unfavorable (blue) or favorable (red) interactions with dark red being most favorable.

**Figure 13.**

Comparison of the per-residue van der Waals (a) and Coulombic (b) interaction energies between gp41 (dashed line, small squares) and lipid membrane (solid line, filled circles) for each T20 residue from wildtype simulations. Each datapoint represents the average value obtained from 10,000 MD snapshots saved every 1 ps.

**Figure 14.**

Interactions of the T20 WNWF motif with membrane. **14a** shows radial distribution functions (rdfs) for the average distances (N=10,000 snapshots) between all lipid headgroup N atoms and (d1) the center of ring atoms at W159, (d2) the backbone O at N160, (d3) the center of ring atoms at W161, and (d4) the backbone O at F162 (d4). Summation of bins (panel a gray vertical lines) out to a distance $\leq 5 \text{ \AA}$ approximates the number of interactions occurring in the first T20-lipid solvation shell represented by the first peaks in the rdfs. **14b** graphically illustrates representative favorable interactions (distances $\leq 5 \text{ \AA}$) for the T20-lipid interaction pairs defined as d2–d4 (magenta dashed lines). The graphic shows a single simulation snapshot of T20 (orange sticks) complexed with gp41 (gray surface) interacting with ten representative conformations of two nearby lipids (green sticks). Lipid head group carbons omitted for clarity.

**Figure 15.**

FP conformations for gp41 (a.a. 1 to 16) compatible with the proposed T20 binding interface. **15a** and **15b** show side and bottom views respectively for FP as an extension of the N-helical coiled-coil (red), an antiparallel beta-sheet (green), or an extended alpha-helix (blue). Other regions of gp41 are colored gray and T20 is colored orange. **15c** compares per-residue footprints for van der Waals (top), Coulombic (middle), and H-bond (bottom) interactions from averaging over 10ns simulations of T20 in complex with gp41 using each of the three FP conformations.

Table 1

Experimental fold resistance (FR) values for T20 with HIVgp41.

mutation	FR = mutant/wildtype	$\Delta\Delta G_{FR} = RT \ln FR^f$
* L33Q	924 ^a	4.05
* L33S	64, ^a 108 ^a	2.62 ^g 2.46, 2.77
G36D	8, ^b 10, ^c 46 ^d	1.62 ^g 1.23, 1.36, 2.27
G36E	39 ^c	2.17
G36S	7, ^b 7 ^c	1.15
* G36V	45 ^d	2.26
* I37K	212 ^e	3.17
I37T	13 ^e	1.15
V38A	16, ^b 18, ^c 54 ^d	1.90 ^g 1.64, 1.71, 2.36
* V38E	494, ^d 1100 ^c	3.92 ^g 3.68, 4.15
V38M	17, ^d 26 ^e	1.81 ^g 1.68, 1.93
* Q40H	21, ^b 28, ^c 34 ^d	1.95 ^g 1.80, 1.97, 2.09
* Q40K	1268 ^a	4.23
Q41R	983, ^a 1137, ^a 1433, ^a	4.19 ^g 4.08, 4.17, 4.31
N42T	4, ^b 26 ^d	1.38 ^g 0.82, 1.93
N43D	18, ^b 23, ^c 37 ^d	1.90 ^g 1.71, 1.86, 2.14

^aDavison, D. K.; et al.(5)

^bGreenburg, M.L.; et al.(24)

^cMink, M.; et al.(25)

^dMelby, T.; et al.(26)

^eNameki, D.; et al.(27) using NL4-3 strain with DIV sequence (residue numbers 36–38).

^f $\Delta\Delta G_{FR} = RT \ln (FR) at 25 ^\circ C in kcal/mol.$

^gAverage $\Delta\Delta G_{FR}$ values.

* Mutations studied in this report.

Table 2

Autocorrelation function percent uncorrelated data (ACF %) and block average standard error of the mean (BASEM) for ΔG_{bind} calc results from simulations of T20 with wildtype HIVgp41 and seven mutants for various block lengths.

mutation	N = 1 ps			N = 100 ps			N = 250 ps			N = 500 ps			N = 1000 ps		
	ACF % ^a	BASEM ^b	ACF %	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM
WT	46.13	0.06	79.44	0.38	82.93	0.56	87.91	0.73	95.10	1.01					
L33Q	41.45	0.07	65.54	0.44	75.20	0.65	84.34	0.88	88.44	1.13					
L33S	39.96	0.07	69.46	0.45	81.34	0.67	88.43	0.88	90.89	1.15					
G36V	34.54	0.08	64.29	0.52	71.87	0.76	78.96	1.03	86.10	1.34					
I37K	33.64	0.08	59.43	0.54	69.77	0.80	85.24	1.07	100.39	1.09					
V38E	44.79	0.07	74.14	0.39	79.68	0.56	86.96	0.78	95.13	0.97					
Q40H	36.55	0.07	66.64	0.48	70.09	0.68	81.25	0.91	94.81	1.27					
Q40K	32.04	0.07	61.20	0.53	75.26	0.79	83.38	0.99	96.64	1.24					
average	38.64	0.07	67.52	0.47	75.77	0.68	84.56	0.91	93.44	1.15					

^a ACF in % of uncorrelated data.

^b BASEM energies in kcal/mol.

Table 3

Calculated free energies of binding (ΔG_{bind} calcd^a), fold resistance ($\Delta\Delta G_{\text{FR}}$ calcd), and component breakdown from simulations of T20 with HIV gp41.

mutation	ΔG_{bind} calcd ^a	ΔE_{vdw}	ΔE_{coul}	ΔG_{polar}	$\Delta G_{\text{nonpolar}}$	ΔG_{elec}	#H-bond ^b
	A = B+C+D+E	B	C	D	E	F = C+D	G
WT	-120.83 ± 0.56	-138.84 ± 0.65	-718.87 ± 2.43	754.68 ± 2.32	-17.80 ± 0.06	35.81	6.85
L33Q	-106.75 ± 0.65	-129.23 ± 0.80	-705.12 ± 2.66	744.26 ± 2.43	-16.66 ± 0.08	39.14	6.59
L33S	-118.79 ± 0.67	-142.58 ± 0.59	-732.76 ± 3.54	774.28 ± 3.30	-17.74 ± 0.05	41.53	6.65
G36V	-110.23 ± 0.76	-134.76 ± 0.73	-716.63 ± 4.79	758.19 ± 4.55	-17.02 ± 0.07	41.56	6.28
I37K	-111.31 ± 0.80	-124.99 ± 0.67	-1015.91 ± 2.66	1046.58 ± 2.59	-16.99 ± 0.05	30.67	7.33
V38E	-108.51 ± 0.56	-137.78 ± 0.59	-419.85 ± 2.25	466.52 ± 2.10	-17.40 ± 0.06	46.68	6.36
Q40H	-112.74 ± 0.68	-132.49 ± 0.50	-735.65 ± 2.98	772.88 ± 2.91	-17.49 ± 0.05	37.23	5.68
Q40K	-103.63 ± 0.79	-118.66 ± 0.94	-1023.13 ± 3.11	1054.92 ± 2.93	-16.77 ± 0.09	31.80	4.97
difference ^c		$\Delta\Delta G_{\text{FR}}$					
		calcd	\exp^{d}	$\Delta\Delta E_{\text{vdw}}$	$\Delta\Delta E_{\text{coul}}$	$\Delta\Delta G_{\text{polar}}$	$\Delta\Delta G_{\text{nonpolar}}$
L33Q-WT	14.08	4.05	9.61	13.75	-10.42	1.14	3.33
L33S-WT	2.04	2.62	-3.74	-13.89	19.60	0.06	5.72
G36V-WT	10.60	2.26	4.08	2.24	3.51	0.78	5.75
I37K-WT	9.52	3.17	13.85	-297.04	291.90	0.81	-5.14
V38E-WT	12.32	3.92	1.06	299.02	-288.16	0.40	10.87
Q40H-WT	8.09	1.95	6.35	-16.78	18.20	0.30	1.42
Q40K-WT	17.20	4.23	20.18	-304.26	300.24	1.04	-4.01
							-1.88

^a All energies in kcal/mol ± block averaged standard errors of the mean (block size = 250 ps, N = 40 evenly spaced blocks).

^b Average number of interactions.

^c Difference values ($\Delta\Delta$) represent mutant–wildtype.

^d Experimental average values from Table 1.