BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 3575–3592 3575
DOI: 10.1021/bi901915g

Origins of Resistance to the HIVgp41 Viral Entry Inhibitor T20[†]

Brian E. McGillick,[‡,⊥] Trent E. Balius,[§,⊥] 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.*
[⊥]*These authors contributed equally to this work.*

Received November 6, 2009; Revised Manuscript Received March 14, 2010

ABSTRACT: Peptide T20, which targets the HIV protein gp41, represents the first approved member of the class of HIV drugs 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), were performed to delineate structural and energetic features that contribute to drug resistance. Per-residue binding footprints for T20 with wild-type 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 exhibited decreases in the level of binding, including the fact that L33Q and Q40K are most detrimental. Six 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 minus wild-type) footprinting indicate the following. (1) Mutations disrupt intermolecular H-bonding and reduce the level of favorable contact with gp41 at M19. (2) Charged mutations (I37K, Q40K, and V38E) lead to significant Coulombic changes that weaken favorable van der Waals interactions. (3) Q40K is more detrimental than I37K because of interaction differences with a polar/charged patch on T20 in the initial (wild-type) state. (4) Resistance for L33S versus L33Q likely involves side chain packing differences in the final (mutated) state. A valuable finding of the work involves identification of favorable interactions among the C-terminal end of T20 (WNWF motif), residues on gp41 (including the fusion peptide), and headgroups in the adjacent membrane. The results suggest a complete T20 binding site would contribute to a stable complex, which could help to explain why prior studies, which 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 the level of favorable contact with both the membrane and gp41 which would lead to enhanced activity.

An estimated 30−36 million people are currently infected with the deadly human immunodeficiency virus (HIV), the causative agent of AIDS, and 7400 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−host cell membrane fusion, which is a prerequisite for infection and viral replication, is a new antiviral design paradigm. Targeting the HIV glycoprotein gp41 (HIVgp41),[1] which mediates this process, has resulted in several promising inhibitors that 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 that lead to resistance to this important antifusion 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 work.

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*). When gp120 binds to host cell receptors, large conformational changes occur

---

[†] 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, and National Institutes of Health Grants R01GM083669 (to R.C.R.) and F31CA134201 (to T.E.B.).

*To whom correspondence should be addressed. E-mail: rizzorc@gmail.com. Phone: (631) 944-2891. Fax: (631) 632-8490.

[1] 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; PDB, Protein Data Bank; rmsd, root-mean-square deviation.

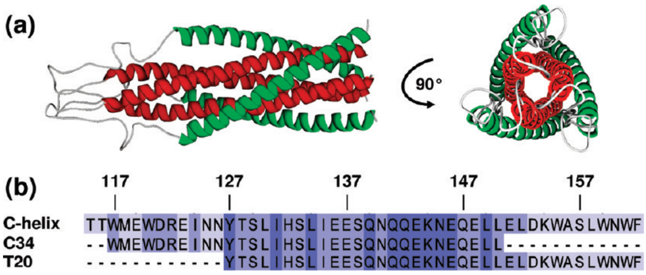

Figure 1: (a) HIVgp41 ectodomain from Caffery (PDB entry 1IF3) with outer C-helices colored green and inner N-helices colored red. (b) Sequence alignment showing fusion inhibitors T20 and C34 in comparison with a region of the native outer HIVgp41 C-helix (top sequence).

Table 1: Experimental Fold Resistance (FR) Values for T20 with HIVgp41

mutation	FR = mutant/wild type	$\Delta\Delta G_{FR}$ = *RT* ln FR[f]	
L33Q[h]	924[a]	4.05	
L33S[h]	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[h]	45[d]	2.26	
I37K[h]	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[h]	494,[d] 1100[c]	3.92[g]	3.68, 4.15
V38M	17,[d] 26[e]	1.81[g]	1.68, 1.93
Q40H[h]	21,[b] 28,[c] 34[d]	1.95[g]	1.80, 1.97, 2.09
Q40K[h]	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

[a] From ref *5*. [b] From ref *24*. [c] From ref *25*. [d] From ref *26*. [e] From ref *27* using strain NL4-3 with a DIV sequence (residues 36−38). [f] $\Delta\Delta G_{FR}$ = *RT* ln FR at 25 °C in kilocalories per mole. [g] Average $\Delta\Delta G_{FR}$ values. [h] Mutations studied in this report.

in gp41 which lead to an intermediate fusion state as described by Kim and co-workers (*6*). Compounds that bind this intermediate are potential membrane fusion inhibitors. Structural studies of truncated gp41, thought to be representative of 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 on the basis of 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 among 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 side chains (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 the design of small molecule inhibitors (*17*−*20*). On the basis of sequence, T20 is not believed to interact with this pocket.

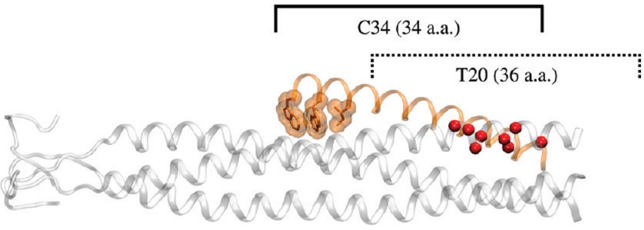

Figure 2: Primary mutations (red spheres) that arise from the use of T20 (*5*, *21*−*27*) map to the inner gp41 N-helices (gray ribbon). Inhibitor C34 is depicted as an orange ribbon with hydrophobic pocket-binding residues shown as molecular surface. The proposed T20 binding footprint is indicated by a dashed line. Coordinates are from PDB entry 1IF3.

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 that arise from the use of T20 are primarily observed to occur on gp41 in a contiguous stretch of 10 amino acids between residues 33 and 43 (gp41 numbering) on the inner N-helices. Mutations 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 the sake of clarity. The spheres encompass the region of G36, I37, and V38 (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 that reduce activity map directly to the N-helical core (Figure 2, spheres). On the basis of the C-helix sequence (Figure 1b), T20 should be shifted by ~10 amino acids toward 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 has not yet been described. 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 α-helical in the bound state, and (iii) the fusion peptide domain consisting of the first 16 amino acids (*8*) on gp41 was embedded in a lipid bilayer.

On the basis of 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

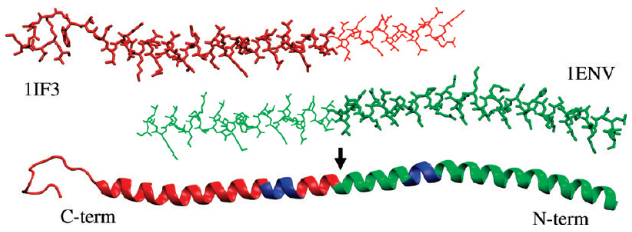

FIGURE 3: Construction of an extended gp41 structural model to accommodate T20. PDB entries 1IF3 (top, red) and 1ENV (middle, green) were aligned using residues 31−35@C,O and 53−57@C,O in common between the two chains (blue region, bottom), and the structures were merged (arrow) to create a single continuous chain.

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, an improved 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 the Results, the sections added to gp41 were α-helical; however, two alternative conformations for the FP regions were also explored. Importantly, the extensions had to be constructed in such a way that they matched the twist which occurs along the inner coil axis of the gp41 heptad repeat motif. We used the GCN4−gp41 construct (PDB entry 1ENV) (*9*), which contained a 47-residue N-helical section of gp41 that overlapped with the previous C34 binding model (constructed from PDB entry 1IF3) (*30*), which contained 28 residues that extended beyond the C34−gp41 construct toward 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 Cα atoms which occurred at this point. The new backbone template was then obtained by merging the two structures which required deletion of residues from the C-terminal end of the inner coiled coil of 1ENV 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 amino acid sequence for wild-type gp41 (LW12.3 isolate) (*31*) using an iterative combination of side chain rotamer exploration and manual inspection with three-dimensional 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 wild-type sequence then served as a master template for construction of L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K point mutations (three amino acids because of the trimer). This same iterative refinement procedure was used to construct these point mutations. 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 β-sheet or extended α-helix conformation (see the 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 with the original C-helix coordinate frame in 1IF3 using backbone coordinates but shifted by 10 amino acids toward 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-dependent 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 wild-type gp41 and for mutants. AMBER (*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, NAMD and VMD (*35*, *36*) were used in combination with the CHARMM27 force field (*37*, *38*) to set up and simulate T20 with gp41 embedded into an explicit lipid membrane including explicit solvent. VMD 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. 30000 TIP3P (*39*) explicit water molecules, ca. 160 DOPC (*37*, *38*) lipid molecules, 282 gp41 residues, and 36 T20 peptide ligand residues in an 80 Å × 80 Å × 200 Å periodic box. 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. 30−35 Å 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 Å direct-space nonbonded cutoff. MD simulations were conducted in the NPT ensemble at 310 K using Langevin dynamics with a damping coefficient of 1 ps^{-1} 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 a 2 fs time step for production.

Consistent with our laboratory's earlier modeling of C34 peptides with wild-type 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^{-1} Å^{-2} in the following order: all protein atoms and all lipid heavy atoms (step 1), all protein atoms (step 2), and only

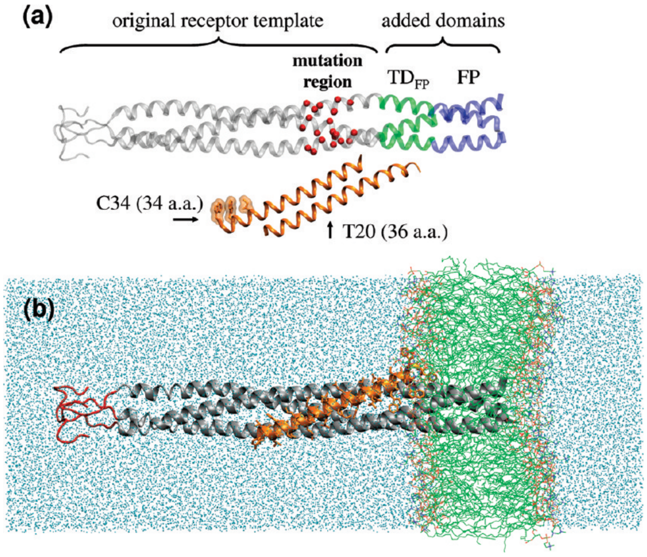

FIGURE 4: Structure of the complete gp41 model. (a) Sections previously employed by Strockbine and Rizzo (*29*) to characterize C34 peptide binding shown as a gray ribbon. Sections added include the transition domain adjacent to the fusion peptide (TD_{FP}, amino acids 17−29, green ribbon) and the fusion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

protein heavy atoms (step 3). This was followed by restrained MD for 500 ps using a 5.0 kcal mol^{-1} $Å^{-2}$ weight (step 4) applied to protein heavy atoms only plus an additional 10 ps with a reduced weight of 1.0 kcal mol^{-1} $Å^{-2}$ (step 5). An additional 10 ps of MD was performed with a 10-fold lower weight of 0.1 kcal mol^{-1} $Å^{-2}$ on protein heavy atoms except for 12 α-carbons on gp41 near the C-terminal end of the trimer at positions 82−94, 176−188, and 270−282 which retained the 1.0 kcal mol^{-1} $Å^{-2}$ restraint (step 6). Production MD followed for a total of 10 ns and was unrestrained with the exception of the aforementioned 12-amino acid C-terminal end of the model which retained the weak 1.0 kcal mol^{-1} $Å^{2}$ restraints. Coordinates and energies were saved every 1 ps (N = 10000) for subsequent postprocessing 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 described well (*44*, *45*). The ability to study large structural changes and compare binding energies between ligands with diverse topologies, combined with the relative ease of set up 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 were used to generate ensembles of low-energy structures (snapshots) which were postprocessed to yield the individual terms used to estimate the overall free energy of binding [$\Delta G_{\mathrm{bind}}(\mathrm{calcd}) = \Delta E_{\mathrm{coul}} + \Delta E_{\mathrm{vdw}} + \Delta\Delta G_{\mathrm{polar}} + \Delta\Delta G_{\mathrm{nonpolar}}$]. Following production MD, explicit solvent (water and lipid) was removed and intermolecular nonbonded 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) pairwise descreening algorithm ($\Delta\Delta G_{\mathrm{polar}}$ terms) and the AMBER8 (*33*) icosahedra solvent-accessible-surface-area algorithm (SASA terms) as implemented in DOCK (*51*). Nonpolar terms were computed as $\Delta G_{\mathrm{nonpolar}} = \gamma \mathrm{SASA} + \beta$, with SASA in square angstroms using standard values of γ (0.00542 kcal mol^{-1} $Å^{-2}$) and β (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, $\Delta\Delta G_{\mathrm{polar}}$ and $\Delta\Delta G_{\mathrm{nonpolar}}$, after explicit solvent had been stripped off. 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 = 10000) obtained from coordinates saved every picosecond during the 10 ns production phase. For comparison with simulation results, experimental fold resistance values (FR = mutant activity/wild-type activity) for inhibitors with gp41 (Table 1) were estimated as experimental free energies of binding using the relationship $\Delta\Delta G_{\mathrm{FR}}(\mathrm{exptl}) \approx RT \ln(\mathrm{FR})$ at 25 °C. Hydrogen bonds were defined as a structural interaction between a donor (H_D) and acceptor (X_A) with a distance of $\leq$2.5 Å, and an $X_D{-}H_D{-}X_A$ angle 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 considered 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 postprocessed to yield 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; and (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 nonbonded 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 minus wild type results at any given amino acid position. In addition, for the wild-type 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 yields the target-based or ligand-based footprints, respectively. Footprint and matrix elements were averaged over the course of each production trajectory (N = 10000 snapshots) and plotted as a function of residue number or numbers. To aid in the representation of data, 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

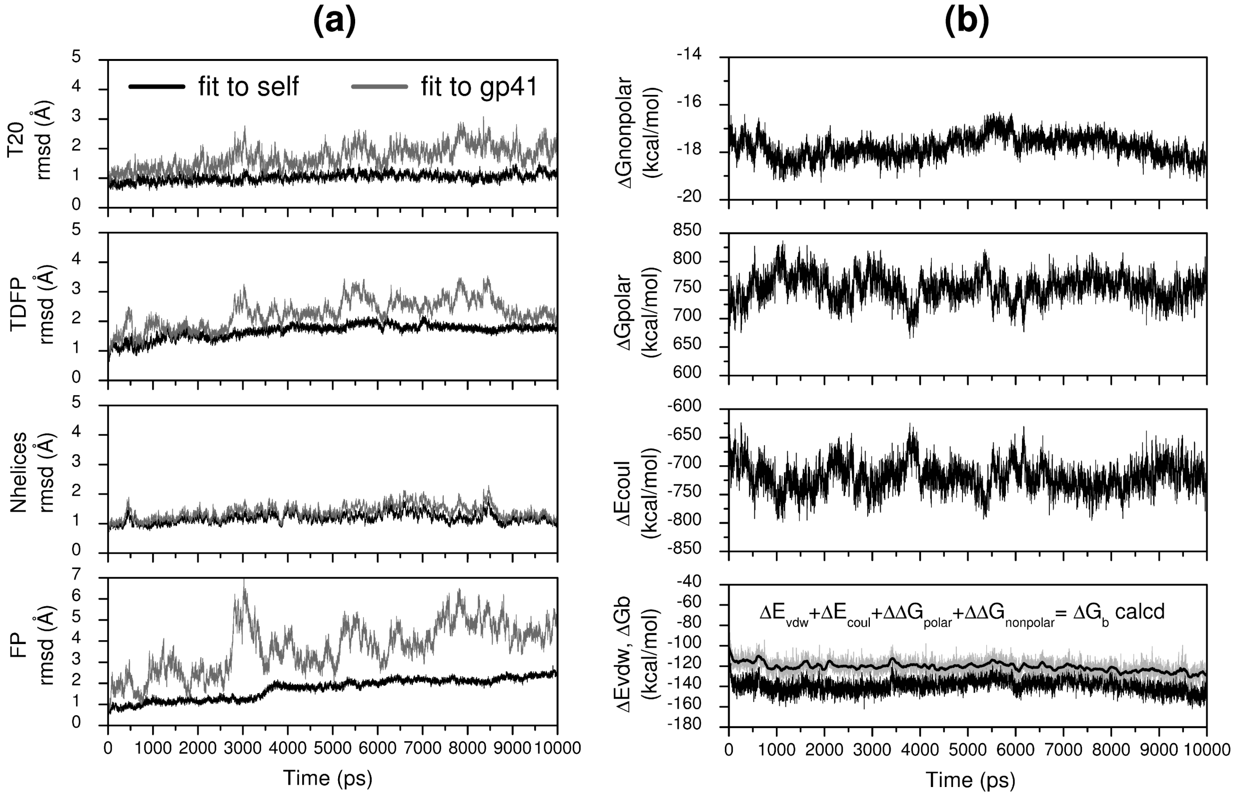

Figure 5: Instantaneous results from simulations of the peptide inhibitor T20 with wild-type HIVgp41 plotted vs time. (a) Root-mean-square deviations for fusion peptide (FP, amino acids 1−16), transition domain adjacent to fusion peptide (TD$_{FP}$, amino acids 17−29), N-helical (residues 30−81), and 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 excludes the FP region (fit-to-gp41). Fits computed using backbone atoms (Cα, C, N, and O) as the match criteria relative to the first MD frame. (b) Individual energy components for ΔE_{vdw}, ΔE_{coul}, $\Delta\Delta G_{\mathrm{polar}}$, $\Delta\Delta G_{\mathrm{nonpolar}}$, and the total estimated binding energy [ΔG_{bind}(calcd) light gray] from the sum of the four energy terms. The solid black line represents block-smoothed (N = 100 blocks) averaging for ΔG_{bind}(calcd).

shifted by 10 amino acids toward the N-terminus relative to C34 as shown in Figure 4. This required a concomitant extension of gp41 (described in the legend of Figure 3) to accommodate T20, and our laboratory's prior computational setup 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 [here termed 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, antiparallel β-strand, or a mixture of parallel and antiparallel 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 this model, insertion of the FP trimer (amino acids 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 π-stacking and headgroup 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 pseudomimic for forces that 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-amino acid loop region (Figure 4b, red tube), restrained to account for the stabilizing presence of the viral envelope, these simulations are unrestrained. As described below, the overall good energetic and structural behavior of these simulations provides compelling evidence that the hypothesis formed as a result of the prior C34 studies without a membrane was correct.

Root-mean-square deviations 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, amino acids 1−16), transition domain (TD$_{FP}$, amino acids 17−29), N-helices (residues 30−81), or T20. (1) Black lines in Figure 5 represent "fit-to-self" results that are defined as rmsds computed by fitting each region to itself. (2) Gray lines represent "fit-to-gp41" results that 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 Cα, C, N, and 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 of <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 yield 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−wild-type 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}), $\Delta\Delta G_{\mathrm{polar}}$ (GB energy), and $\Delta\Delta G_{\mathrm{nonpolar}}$ (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 approximately equal and opposite in magnitude (Figure 5b, middle panels). This effect is especially dramatic when all 10000 instantaneous values are plotted together (Figure 6) which yields an almost perfect anticorrelation 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).

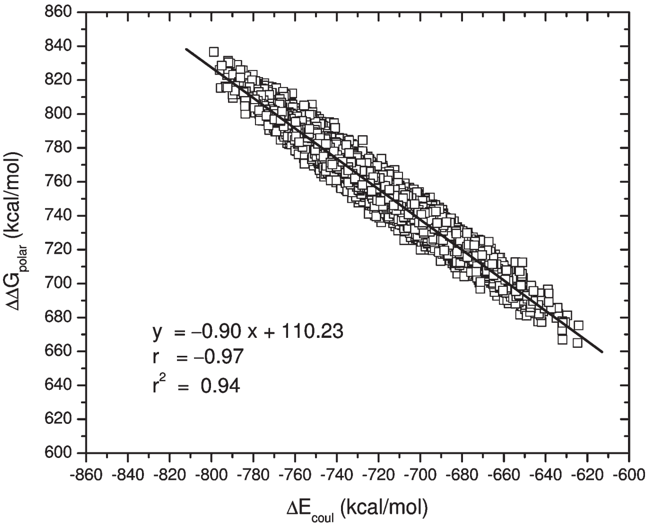

FIGURE 6: Intermolecular protein−ligand Coulombic energies (ΔE_{coul}) vs the opposing change in polar desolvation penalties ($\Delta\Delta G_{\mathrm{polar}}$). Symbols represent the instantaneous energies obtained from all frames saved during the 9 ns MD trajectory ($N = 10000$) of T20 with wild-type gp41, and the black line is the best fit to the data.

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 wild-type gp41, as well as seven gp41 mutants (L33Q, L33S, G36V, I37K, V38E, Q40H, and Q40K). For a given time series, ACF 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 a lag time of 0 ps to $<25\%$ correlated at 250 ps, before showing fluctuations which oscillate ∼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 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 versus 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 kilocalories per mole were computed as a

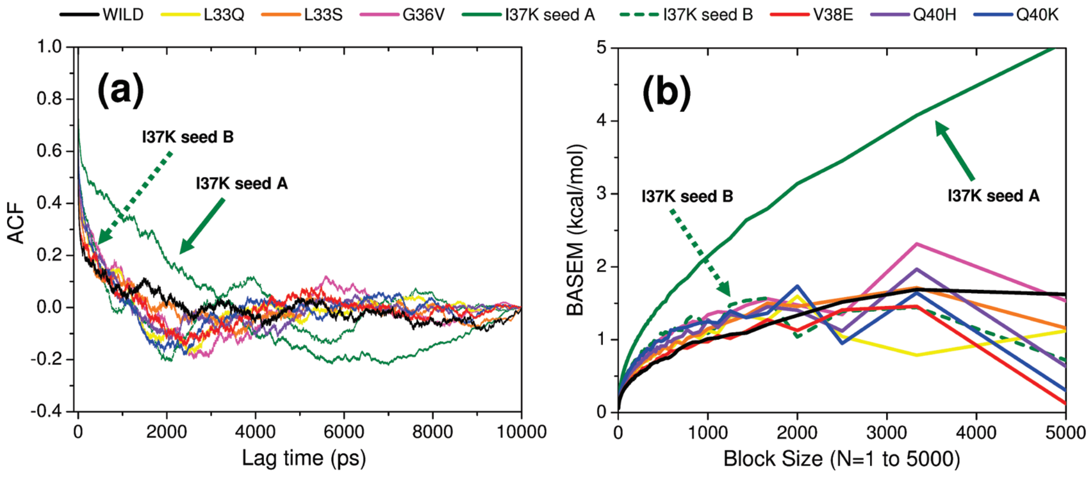

FIGURE 7: (a) Autocorrelation functions (ACF) of calculated binding energies [ΔG_{bind}(calcd)] for T20 with wild-type HIVgp41 and seven mutants. (b) Block-averaged standard error of the mean (BASEM) in kilocalories per mole as a function of block size from 1 to 5000. Solid arrows indicate initial I37K results, while dashed arrows indicate I37K results obtained using a different random seed.

Table 2: Autocorrelation Function Percent Uncorrelated Data (ACF %) and Block-Averaged Standard Errors of the Mean (BASEM) for ΔG_{bind}(calcd) Results from Simulations of T20 with Wild-Type HIVgp41 and Seven Mutants for Various Block Lengths

	N = 1 ps		N = 100 ps		N = 250 ps		N = 500 ps		N = 1000 ps	
	ACF %[a]	BASEM[b]	ACF %[a]	BASEM[b]	ACF %[a]	BASEM[b]	ACF %[a]	BASEM[b]	ACF %[a]	BASEM[b]
wild type	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 percent of uncorrelated data. [b] BASEM energies in kilocalories per mole.

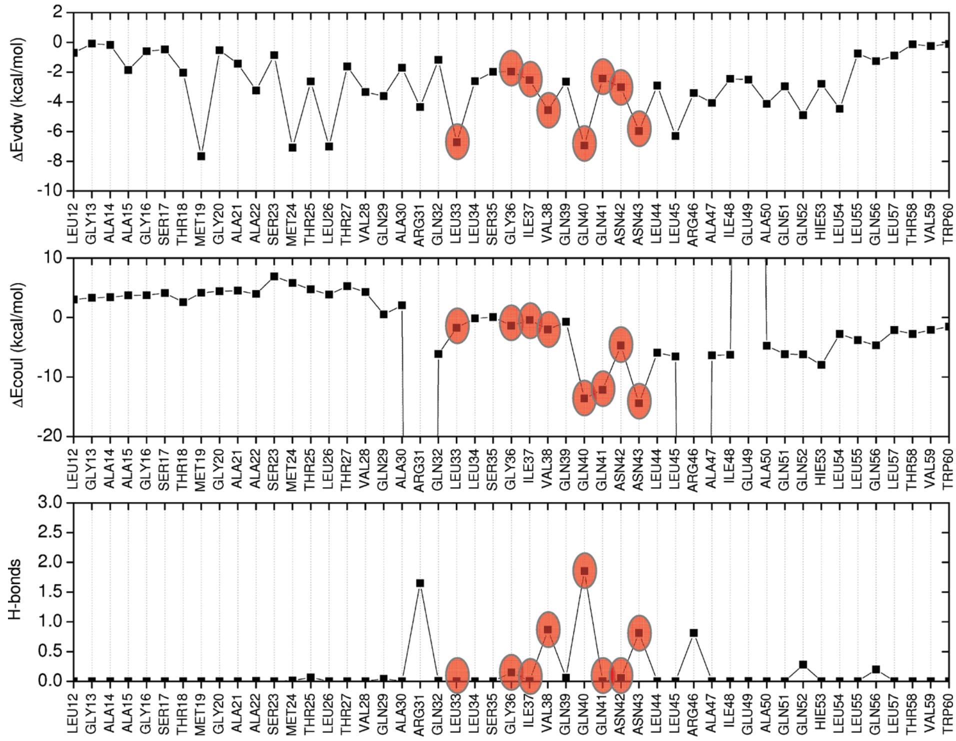

FIGURE 8: Mutations from Table 1 (red) mapped to molecular footprints obtained from simulations of T20 with wild-type HIVgp41. Footprints represent average per-residue van der Waals (top, ΔE_{vdw}), Coulombic (middle, ΔE_{coul}), and H-bond (bottom) interactions between T20 (all residues) and each individual gp41 residue for the range of Leu12−Trp60 from 10000 MD simulation snapshots.

function of block size which ranges from 1 to 5000 ps (1 to 5000 snapshots). The BASEM results show the expected exponential increase as the block averaging size becomes larger and then begins to reach a plateau which is indicative of reasonable equilibrium. BASEM errors computed at the N = 1 level likely underestimate the true error in the simulations, while at the N = 5000 level (half the simulation length), the errors may not be reliable due to the fact that only two blocks are used. For our 10 ns trajectories, Table 2 plots BASEM errors, along with the percentage of uncorrelated data from ACF curves, for the eight simulations at block sizes of 1, 100, 250, 500, 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 that are largely uncorrelated (ca. 75−85%) and have a sufficiently large number 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 and 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 wild-type gp41 as shown in Figure 8 for the gp41 range of Leu12−Trp60. Here, footprints for van der Waals (top, ΔE_{vdw}), Coulombic (middle, ΔE_{coul}), and hydrogen bond (bottom, H-bonds) interactions are defined as the pairwise interaction sum between all T20 residues and each individual gp41 residue. For the sake of simplicity, contributions from the three gp41 N-helical chains were summed together and mapped to a single monomer. Mutations that 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 α-helix binding interface in which side chains are packed into gp41 in an approximately alternate and repeating manner. Compellingly, there are strong favorable ΔE_{vdw} interactions that occur in the gp41 range of residues 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 identifies Q40, N41, and N43 (ca. −12 to −15 kcal/mol) as residues engaging in 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.

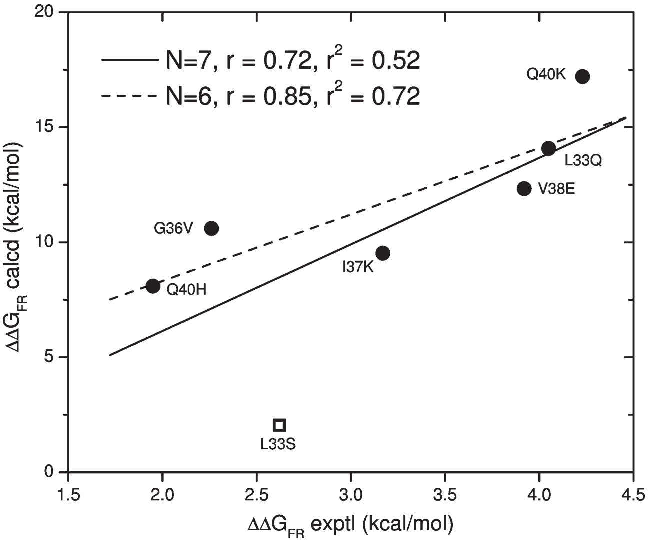

FIGURE 9: Predicted FR energies [$\Delta\Delta G_{\mathrm{FR}}$(calcd)] vs experimental FR energies [$\Delta\Delta G_{\mathrm{FR}}$(exptl)] for inhibitors with HIVgp41. Each computed data point is the difference in results from two independent MD simulations. Experimental values are listed in Table 1.

Experimental versus 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 and Table 3). Table 3 shows the individual energetic and H-bonding contributions to binding in both an absolute [ΔG_{bind}(calcd)] and relative [$\Delta\Delta G_{\mathrm{FR}}$(calcd)] sense, while Figure 9 plots the experimental versus computational FR results. Computationally, FR energies [$\Delta\Delta G_{\mathrm{FR}}$(calcd)] represent the difference between mutant and wild-type 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 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 data points (Figure 9; r^2 = 0.72) reproduce well the experimental FR energies [$\Delta\Delta G_{\mathrm{FR}}$(exptl)]. Notably, the calculations correctly predict that L33Q and Q40K undergo the largest loss of binding. Although the correlation is high, the magnitudes of the computed FR values are, with the exception of that of L33S, systematically overpredicted by ∼3−4-fold. As all mutations except L33S result in a larger side chain, 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 $\Delta\Delta G_{\mathrm{FR}}$ computational values are compared (*64*). However, as each gp41 point mutation actually changes three amino acids 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 is unlikely. These effects could play a role in systematic overpredictions for the mutations to larger side chains. Nevertheless, despite the L33S offset in

Table 3: Calculated Free Energies of Binding [ΔG_{bind}(calcd)], Fold Resistance [$\Delta\Delta G_{\mathrm{FR}}$(calcd)], and Component Breakdown from Simulations of T20 with HIVgp41

	ΔG_{bind}(calcd)[a] (A = B + C + D + E)	ΔE_{vdw} (B)	ΔE_{coul} (C)	$\Delta G_{\mathrm{polar}}$ (D)	$\Delta G_{\mathrm{nonpolar}}$ (E)	ΔG_{elec} (F = C + D)	no. of H-bonds[b] (G)
wild type (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

	$\Delta\Delta G_{\mathrm{FR}}$							
difference[c]	calcd	exptl[d]	$\Delta\Delta E_{\mathrm{vdw}}$	$\Delta\Delta E_{\mathrm{coul}}$	$\Delta\Delta G_{\mathrm{polar}}$	$\Delta\Delta G_{\mathrm{nonpolar}}$	$\Delta\Delta G_{\mathrm{elec}}$	ΔH-bond
L33Q − WT	14.08	4.05	9.61	13.75	−10.42	1.14	3.33	−0.26
L33S − WT	2.04	2.62	−3.74	−13.89	19.60	0.06	5.72	−0.20
G36V − WT	10.60	2.26	4.08	2.24	3.51	0.78	5.75	−0.57
I37K − WT	9.52	3.17	13.85	−297.04	291.90	0.81	−5.14	+0.48
V38E − WT	12.32	3.92	1.06	299.02	−288.16	0.40	10.87	−0.49
Q40H − WT	8.09	1.95	6.35	−16.78	18.20	0.30	1.42	−1.17
Q40K − WT	17.20	4.23	20.18	−304.26	300.24	1.04	−4.01	−1.88

[a] All energies in kilocalories per mole ± block-averaged standard errors of the mean (block size of 250 ps; N = 40 evenly spaced blocks). [b] Average number of interactions. [c] Difference values (ΔΔ) represent mutant minus wild type. [d] Experimental average values from Table 1.

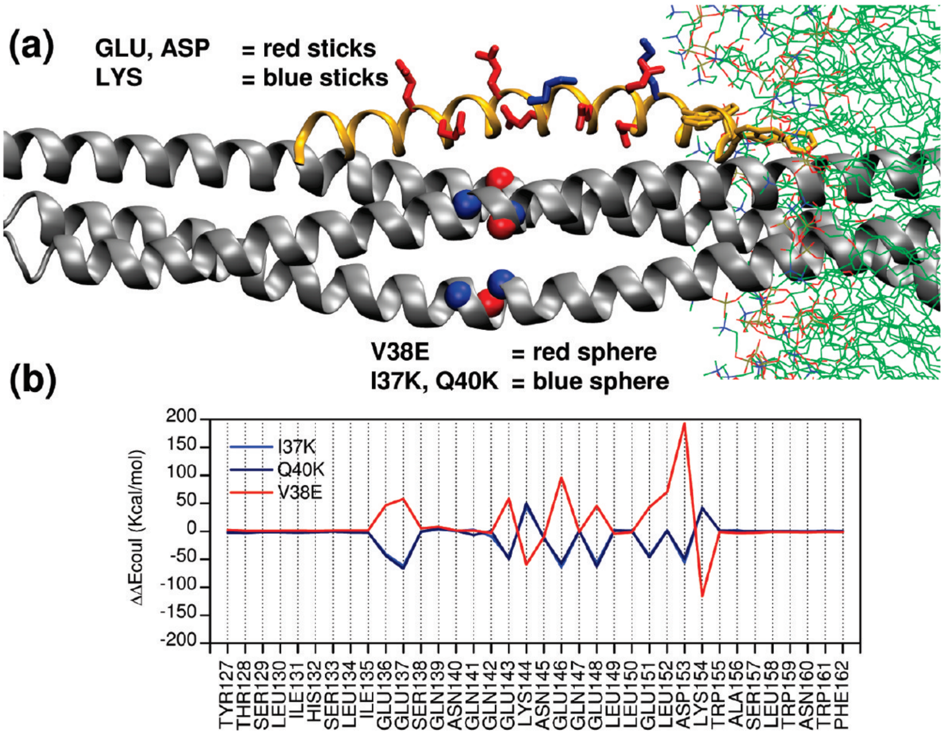

FIGURE 10: (a) Charged mutations (colored spheres) on HIVgp41 (gray ribbon) for V38E (red), I37K (blue), and Q40K (blue) relative to negative (red for Glu and Asp) and positive (blue for Lys) charged residues on T20 (orange ribbon). The lipid bilayer is colored green. (b) Per-residue Coulombic energy changes for each T20 residue with all of gp41 for three charged mutations.

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 ~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 strong 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, $\Delta G_{\mathrm{polar}}$) that are approximately equal in magnitude, but opposite in sign, oppose binding; thus, the overall electrostatic contribution is small (Table 3, column E, $\Delta G_{\mathrm{elec}} = \Delta E_{\mathrm{coul}} + \Delta G_{\mathrm{polar}}$). 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 $\Delta G_{\mathrm{nonpolar}}$ 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 ($\Delta\Delta E_{\mathrm{coul}}$) and accompanying desolvation ($\Delta\Delta G_{\mathrm{polar}}$) terms in Table 3 becomes clear when they are viewed in the 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 (three 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 side chains (red line, Figure 10b). I37K and Q40K mutations have 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 and I37K) yield the most negative $\Delta\Delta G_{\mathrm{elec}}$ values in contrast to the negatively charged mutation (V38E) which yields the most positive $\Delta\Delta G_{\mathrm{elec}}$ 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 also from T20's point of view the fact that these two different mutations lead to very similar changes in intermolecular Coulombic interactions. Neutral mutations in contrast introduce much smaller changes in $\Delta\Delta E_{\mathrm{coul}}$ 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 side chain relative to that of the wild type (Table 3, column E, $\Delta\Delta G_{\mathrm{nonpolar}} = 0.06$ 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, ΔH-bond) are Q40H and Q40K which is consistent with Q40 being the strongest H-bonding residue in the peptide binding interface (see Figure 8). Finally, six of the seven mutations show a loss of van der Waals energy (Table 3, column B, $\Delta\Delta E_{\mathrm{vdw}}$), with the largest losses being for the two positively charged lysine mutations (see below). Only the L33S−wild-type change does not lead to a loss of steric packing. This latter case is probably related to the fact noted above that L33S is the only mutation that leads to a smaller side chain.

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 α-helices, the mutations are in similar locations (see Figure 10a). However, experimentally (see Table 1), the Q40K mutation is much more detrimental than the I37K mutation. Further, despite the opposite

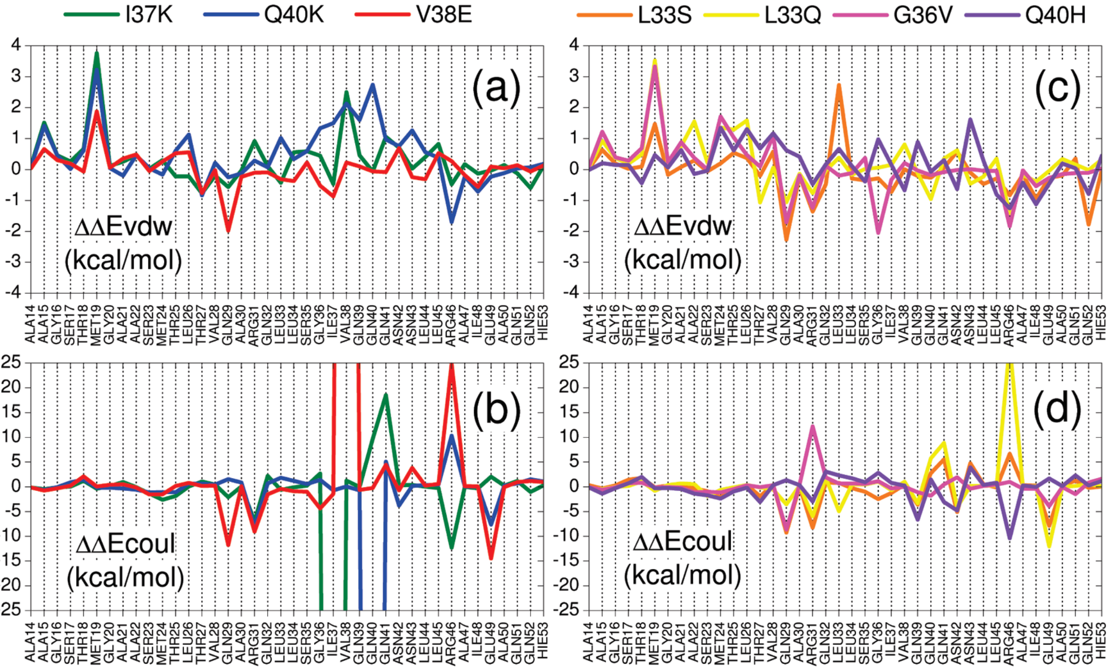

FIGURE 11: Per-residue differential (mutant minus wild type) footprints for T20 with HIVgp41 representing van der Waals (a and c) and Coulombic (b and d) energy changes due to charged (a and b) and neutral (c and d) mutations.

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 the origins of these experimental observations, per-residue changes in van der Waals ($\Delta\Delta E_{\mathrm{vdw}}$) and Coulombic ($\Delta\Delta E_{\mathrm{coul}}$) energies were computed by subtracting mutant from wild-type 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 $\Delta\Delta E_{\mathrm{vdw}}$ values 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, and 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 amino acids of gp41 that interact strongly with specific T20 residues, a van der Waals interaction energy matrix (heat map) was calculated using wild-type 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 of E136−E148 (Figure 12, *y*-axis, shaded region).

If we focus on the gp41 footprint peaks, T20 is observed to make stronger, more favorable ΔE_{vdw} contacts with the wild-type receptor at Q40 than at I37 (Figure 12, *x*-axis, green vs orange ellipsoids). On the basis of 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 $\Delta\Delta E_{\mathrm{vdw}}$ localized to the mutations sites show no loss of 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 that of I37K (13.85 kcal/mol) or Q40K (20.18 kcal/mol). Although V38E shares peaks with I37K and Q40K within the L33−L45 region (Figure11a, 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 $\Delta\Delta E_{\mathrm{coul}}$ 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 $\Delta\Delta \mathrm{E}_{\mathrm{coul}}$ footprints (Figure 11b). Again, similar to $\Delta\Delta E_{\mathrm{vdw}}$, the most variation among 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 $\Delta\Delta E_{\mathrm{vdw}}$ 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 and Q40K) lead to greater losses in $\Delta\Delta E_{\mathrm{vdw}}$ than Coulombic repulsions (V38E). In other words, negatively charged T20 is pulled toward positive mutations on gp41 (I37K and Q40K) which affects $\Delta\Delta E_{\mathrm{vdw}}$ more significantly (more unfavorable changes to $\Delta\Delta E_{\mathrm{vdw}}$) 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

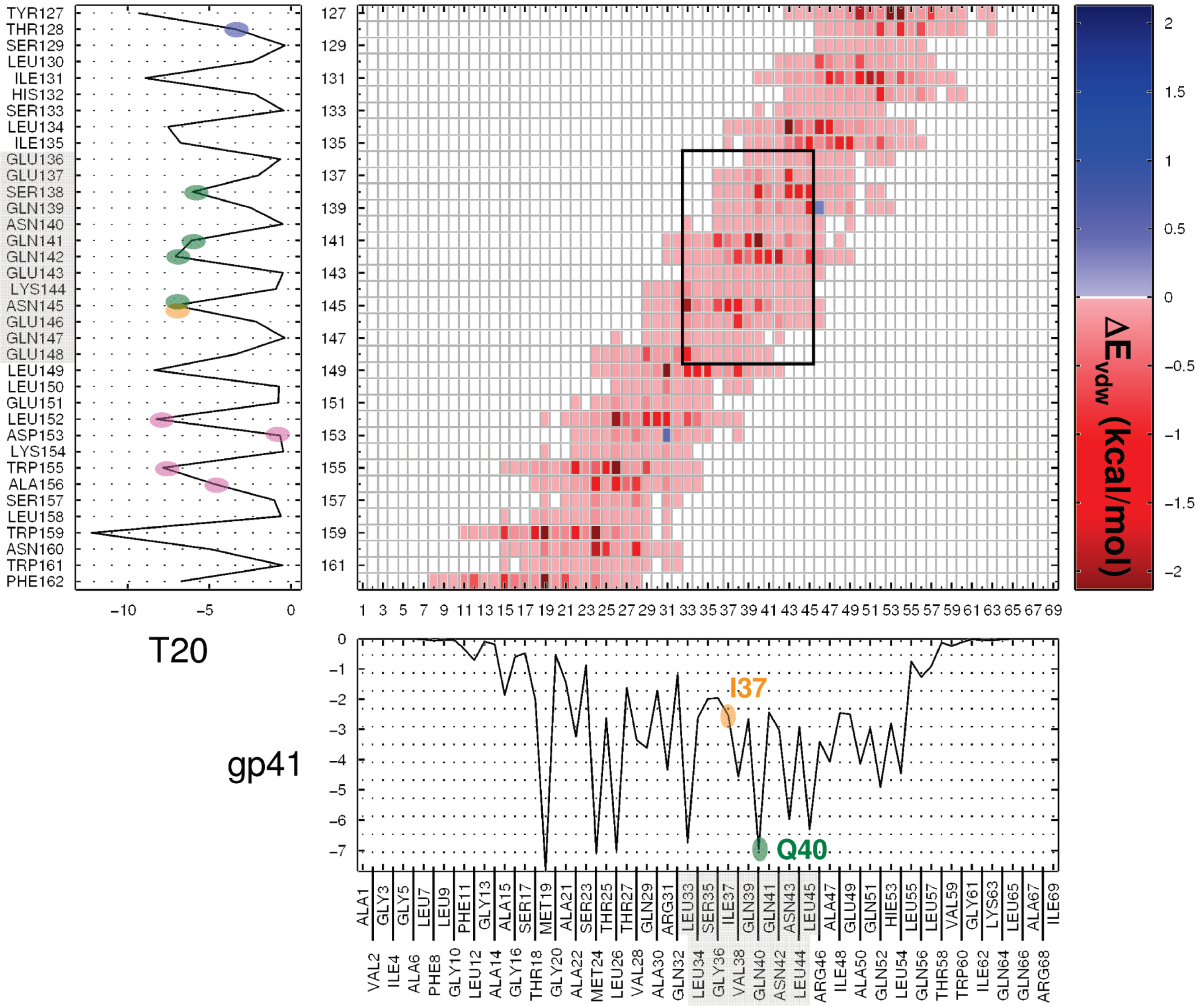

FIGURE 12: Intermolecular van der Waals interaction energy matrix (heat map) for all gp41 with T20 residues from the wild-type simulation. Footprint peak magnitudes represent summation of ΔE_{vdw} (kilocalories per mole) along each row (T20) or column (gp41). Gray shaded regions indicate mapping of the gp41 mutation region (L33−L45) to a charged/polar patch (E136−E148) on T20 defined by the black box in the approximate center of the matrix. Heat map gradient colors indicate the magnitude of unfavorable (blue) or favorable (red) interactions, with dark red being the most favorable.

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 that represent an intermolecular R31−D153 salt bridge and a R46−Q139 H-bond. For V38E, intermolecular repulsions that occur as a result of the change to a negative residue lead to large losses in favorable $\Delta\Delta E_{\mathrm{coul}}$ (Table 3, column C) and a corresponding unfavorable solvent-mediated electrostatic energy $\Delta\Delta G_{\mathrm{elec}}$ (Table 3, column F). Overall, the simulation results suggest that $\Delta\Delta G_{\mathrm{FR}}$ is dominated by losses in favorable steric packing for I37K and Q40K versus 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 binding.

Origins of Resistance: Neutral Mutations. L33 is situated at the start of the gp41 mutation region. Examination of the interaction heat map in Figure 12 shows strong wild-type 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. Mutations L33S and L33Q both represent nonpolar to polar mutations occurring at the same location on gp41 and originating from the same wild-type 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 in both 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 (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 a reduction in size for serine relative to either glutamine or leucine. Likewise, gains at position 36 for van der Waals interaction as a result of the G36V mutation (Figure 11c, magenta line) are physically reasonable given the increase in size and hydrophobicity. In contrast, Q40H reveals no change in $\Delta\Delta E_{\mathrm{vdw}}$ at position 40.

It should also be noted that all mutations (neutral and charged) lead to a reduced level of 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) make 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, these 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) reveals that in general, positions which show losses in $\Delta\Delta E_{\mathrm{coul}}$ are those which originally demonstrated strong interactions in wild-type gp41 (see Figure 8). Here, glutamine is more polar than serine which could explain the accentuated unfavorable $\Delta\Delta E_{\mathrm{coul}}$ effects of the L33Q mutation as compared to the effects of the L33S mutation at positions Q40, Q41, and R46 (Figure 11d, orange vs yellow lines). Otherwise, similar effects on the $\Delta\Delta E_{\mathrm{coul}}$ 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 $\Delta\Delta E_{\mathrm{coul}}$ 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 than L33S and L33Q (Figure 11d, orange and yellow lines) which could play a role in these mutations being less detrimental.

Comparison with T20 Analogue Studies. The binding geometry presented here for T20 agrees well with additional experimental results for other mutations that 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 wild-type 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 wild-type gp41, thus suggesting that mutations that weaken 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 in T20 resulted in an increase in the level of 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 the T128I analogue is likely a function of enhanced favorable packing with adjacent hydrophobic Leu residues in the gp41 range of residues 54−57.

Izumi et al. (*75*) reported a series of T20 analogues with mutations at position S138 for which nonaromatic hydrophobic substitutions were observed to enhance binding with wild-type 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 the clinically observed secondary S138A mutation (or other nonaromatic hydrophobic substitutions) compensates for losses resulting from the commonly observed N43D mutation through strengthened favorable packing with L44 and L45 (*75*, *76*).

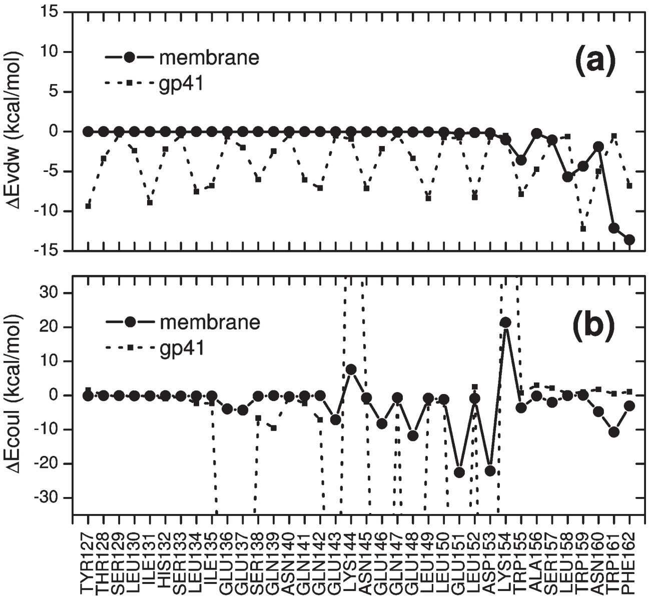

FIGURE 13: Comparison of the per-residue van der Waals (a) and Coulombic (b) interaction energies between gp41 (dashed line, small squares) and lipid membrane (solid line, filled circles) for each T20 residue from wild-type simulations. Each data point represents the average value obtained from 10000 MD snapshots saved every picosecond.

The Importance of the WNWF Motif to Membrane Interactions. 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 indicates 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 the 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 that of other residues, and in a relative sense, the interactions are much stronger than those that 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 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 of the four residues in the C-terminal WNWF motif (Figure 13b, dashed vs black lines). The regularly

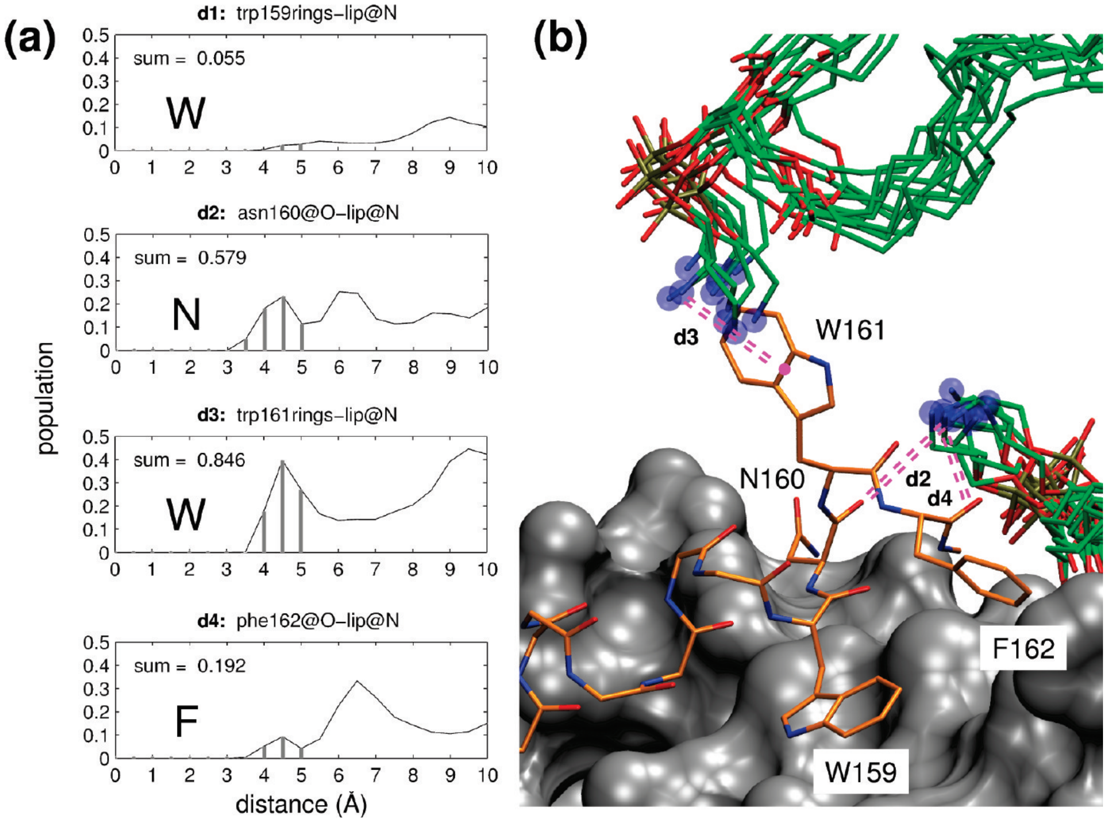

FIGURE 14: Interactions of the T20 WNWF motif with the membrane. Panel a shows radial distribution functions (rdfs) for the average distances (N = 10000 snapshots) between all lipid headgroup N atoms and (d1) the center of ring atoms at W159, (d2) the backbone O at N160, (d3) the center of ring atoms at W161, and (d4) the backbone O at F162. Summation of bins (panel a, gray vertical lines) out to a distance of ≤5 Å approximates the number of interactions occurring in the first T20−lipid solvation shell represented by the first peaks in the rdfs. Panel b graphically illustrates representative favorable interactions (distances of ≤5 Å) for the T20−lipid interaction pairs defined as d2−d4 (magenta dashed lines). The graphic shows a single simulation snapshot of T20 (orange sticks) in a complex with gp41 (gray surface) interacting with 10 representative conformations of two nearby lipids (green sticks). Lipid headgroup carbons have been omitted for the sake of clarity.

repeating pattern in the van der Waals footprint indicates which T20 residues are in close contact with gp41 (Figure 13a, dashed lines) and includes 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 WNWF → ANAA mutation led to no change in T20 binding affinity using a gp41 construct termed 5H-ex (N-helix amino acids 19−71), and Wexler-Cohen et al. (*72*) have reported that D-configuration substitutions in T20 at positions L158 and W159 did not significantly change binding to a construct termed N54 (N-helix amino acids 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 three-dimensional visualization between residues in the WNWF motif and the membrane which could contribute to the stability of the overall complex (Figure 14). Specifically, lipid headgroups 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 (d2−d4 distances of ≤5 Å) are visually illustrated in the accompanying graphic (Figure 14b).

Plots in Figure 14a show the rdf which yielded the largest number of interactions (distance of ≤5 Å) of several atom selections evaluated for each WNWF residue with the lipid headgroups. Although numerous distance definitions (and thus rdfs) are possible, summations up to the peaks at ~5 Å strongly suggest that the aromatic ring of W161 is particularly well solvated by lipid headgroups with 0.85 interaction followed by the backbone O of N160 with 0.58 interaction (Figure 14a). For W161, this highly populated π-type interaction likely corresponds with the distinct Coulombic energy peak seen at this position in the membrane versus 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 headgroups. 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 headgroups as seen with W161, despite the fact that the aromatic ring might also have participated in

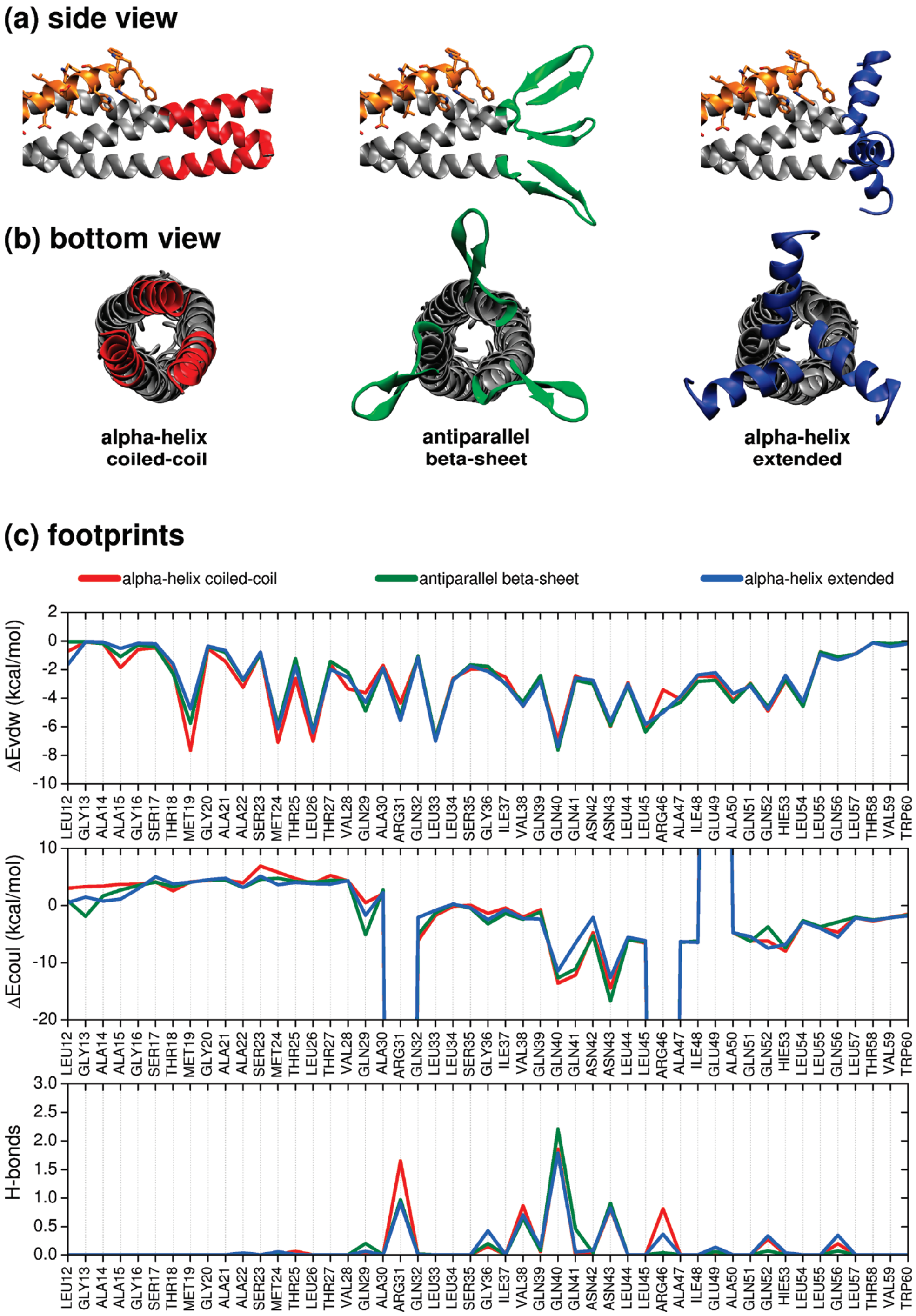

Figure 15: FP conformations for gp41 (amino acids 1−16) compatible with the proposed T20 binding interface. Panels a and b show side and bottom views, respectively, of FP as an extension of the N-helical coiled coil (red), an antiparallel β-sheet (green), or an extended α-helix (blue). Other regions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

π-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 headgroups. 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: α-Helix versus β-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 that 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 dependent on the initial choice of FP conformation, we constructed two additional models representing an antiparallel β-sheet conformation (Figure 15a,b, middle), an alternative α-helix construct in an extended form (Figure 15a,b, right), for comparison with the original α-helix coiled coil (Figure 15a,b, left).

For construction of a FP trimer with β-sheet character, a β-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 amino acids) with a backbone having a β-sheet conformation. The alternative α-helix extended model (Figure 15a,b, right) was similarly constructed via replacement of 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 2PJV) reported by Li and Tamm (*59*). Here, the two structures were merged after a best-fit rmsd match using Cα, C, and N backbone atoms at residue Ser17. 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, three-dimensional stereo visualization, and energy minimization to resolve any steric conflicts as a result of model building prior to beginning MD.

As shown in Figure 15c, MD simulations using the three different FP conformations (Figure 15a,b) yield ΔE_{vdw}, ΔE_{coul}, and H-bond footprints for T20 with wild-type gp41 that 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 with the goal of more fully investigating this issue are desirable. Notably, all three simulations yield essentially superimposable ΔE_{vdw} profiles (Figure 15c, top), particularly in the mutation hotspot region of interest [amino acids 33−43 (Table 1)]. Footprint profiles for ΔE_{coul} (Figure 15c, 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 α-helical coiled-coil (red, Figure 15c, middle) and β-sheet (green, Figure 15c, middle) simulations are more similar than from the extended α-helix (blue, Figure 15c, middle). Similar to ΔE_{vdw}, the H-bond footprints show nearly perfect overlap in the key mutation region, including the large peaks observed at V38, Q40, and N43 (Figure 15c, bottom). The overall agreement in the footprints derived from different simulations provides additional assurance that reasonable convergence and reproducibility have been achieved, specifically with regard 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 ($\Delta\Delta G_{\mathrm{FR}}$) 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 the development and validation of the first complete model for T20 in complex with gp41 that includes the fusion peptide region embedded in a host cell lipid bilayer (Figure 4). Simulation 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 and Table 2) indicate reasonably converged results.

The results from simulations of T20 with wild-type 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 that employed a β-sheet structure or an extended α-helix conformation for the gp41 FP region yielded remarkably similar ΔE_{vdw}, ΔE_{coul}, and H-bond 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 the level of binding (Table 3), and the calculations correctly predict L33Q and Q40K mutations will be most detrimental. Compellingly, for six of the seven mutations studied, the overall trend in the experimental data is well-reproduced with calculated versus experimental fold resistance energies yielding a strong r^2 of 0.72 (Figure 9). Interestingly, results for L33S are offset from the other six data points, although in an absolute sense the magnitude $\Delta\Delta G_{\mathrm{FR}}$(calcd) is in good agreement. As L33S is the only mutation leading to a smaller side chain, the offset is probably linked to differences in size and associated entropic effects not included in the calculations.

Decomposition of the energy components used to estimate $\Delta\Delta G_{\mathrm{FR}}$ revealed that charged mutations lead to very large changes in Coulombic energy ($\Delta\Delta E_{\mathrm{coul}}$) and compensating desolvation ($\Delta\Delta G_{\mathrm{polar}}$) 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 of −5) proximal to the mutation sites (Figure 10). As a result, unfavorable $\Delta\Delta E_{\mathrm{coul}}$ and solvent-mediated electrostatics ($\Delta\Delta G_{\mathrm{elec}}$) for the negatively charged V38E mutation are significantly more unfavorable than those for other mutations (Table 3). Interestingly, all mutations but one lead to weakened van der Waals interactions (Table 3), and the greatest losses occur for the two positively charged mutations (I37K and Q40K). A reasonable mechanism for explaining these results is that negatively charged T20 is pulled toward positive gp41 mutations which lead to changes in $\Delta\Delta E_{\mathrm{vdw}}$ 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 wild-type system (Figure 8).

The origin of the difference in experimentally observed fold resistance (FR) for I37K (less detrimental) versus Q40K (more detrimental) mutations was traced to initial wild-type 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 side chain. Compellingly, all mutations lead to less favorable van der Waals interactions with M19 on gp41

(Figure 11) which is in close 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 weakened (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 strengthen favorable packing with adjacent Leu residues. On the other hand, T20 mutations that weaken binding likely involve disruption of initially favorable van der Waals (L152, W155, and A156) or Coulombic (D153) interactions (Figures 10 and 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 this work indicate the C-terminal region makes significant interactions with both the lipid bilayer and 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). On the basis of these results, prior studies may not have employed sufficiently complete 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 our 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 a membrane (Figures 13 and 14) could also contribute to the stability of the complex. A reasonable hypothesis suggested by this work is that peptide inhibitors designed to strengthen 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.

ACKNOWLEDGMENT

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 DE-AC02-98CH10886 and by the State of New York.

REFERENCES

1. 2008 Report on the global AIDS epidemic: Epidemiology slides, Joint United Nations Programme on HIV/AIDS (UNAIDS) and The World Health Organization (WHO). http://data.unaids.org/pub/GlobalReport/2008/2008_globalreport_core_en.ppt (accessed August 3, 2009).
2. NIGMS-Supported Structure-Based Drug Design Saves Lives, NIH/NIGMS Fact Sheet. http://www.nigms.nih.gov/Publications/structure_drugs.htm (accessed August 3, 2009).
3. Drugs@FDA website. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm (accessed August 3, 2009).
4. Grinsztejn, B., Nguyen, B. Y., Katlama, C., Gatell, J. M., Lazzarin, A., Vittecoq, D., Gonzalez, C. J., Chen, J., Harvey, C. M., and Isaacs, R. D. (2007) Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multi-drug-resistant virus: A phase II randomised controlled trial. *Lancet 369*, 1261–1269.
5. Davison, D. K., Medinas, R. J., Mosier, S. M., Bowling, T. S., Delmedico, M. K., Dwyer, J. J., Cammack, N., and Greenberg, M. L. (2006) 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.
6. Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem. 70*, 777–810.
7. Suntoke, T. R., and Chan, D. C. (2005) The Fusion Activity of HIV-1 gp41 Depends on Interhelical Interactions. *J. Biol. Chem. 280*, 19852–19857.
8. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell 89*, 263–273.
9. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature 387*, 426–430.
10. Schibli, D. J., and Weissenhorn, W. (2004) Class I and class II viral fusion protein structures reveal similar principles in membrane fusion. *Mol. Membr. Biol. 21*, 361–371.
11. Eckert, D. M., and Kim, P. S. (2001) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. U.S.A. 98*, 11187–11192.
12. Jiang, S. B., Lin, K., Strick, N., and Neurath, A. R. (1993) Hiv-1 Inhibition by a Peptide. *Nature 365*, 113.
13. Wild, C., Oas, T., McDanal, C., Bolognesi, D., and Matthews, T. (1992) A Synthetic Peptide Inhibitor of Human Immunodeficiency Virus Replication: Correlation between Solution Structure and Viral Inhibition. *Proc. Natl. Acad. Sci. U.S.A. 89*, 10537–10541.
14. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999) Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell 99*, 103–115.
15. Welch, B. D., VanDemark, A. P., Heroux, A., Hill, C. P., and Kay, M. S. (2007) Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl. Acad. Sci. U.S.A. 104*, 16828–16833.
16. Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998) Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc. Natl. Acad. Sci. U.S.A. 95*, 15613–15617.
17. Debnath, A. K., Radigan, L., and Jiang, S. B. (1999) Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. *J. Med. Chem. 42*, 3203–3209.
18. Jiang, S. B., Lu, H., Liu, S. W., Zhao, Q., He, Y. X., and Debnath, A. K. (2004) 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. 48*, 4349–4359.
19. Frey, G., Rits-Volloch, S., Zhang, X. Q., Schooley, R. T., Chen, B., and Harrison, S. C. (2006) Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proc. Natl. Acad. Sci. U.S.A. 103*, 13938–13943.
20. Cai, L., and Gochin, M. (2007) A Novel Fluorescence Intensity Screening Assay Identifies New Low Molecular Weight Inhibitors of the gp41 Coiled Coil Domain of HIV-1. *Antimicrob. Agents Chemother. 51*, 2388–2395.
21. Chinnadurai, R., Rajan, D., Munch, J., and Kirchhoff, F. (2007) HIV-1 Variants Resistant against 1st and 2nd Generation Fusion Inhibitors and Cytopathic in ex vivo Human Lymphoid Tissue. *J. Virol.* DOI: 10.1128/JVI.02546-06.
22. Armand-Ugon, M., Gutierrez, A., Clotet, B., and Este, J. A. (2003) HIV-1 resistance to the gp41-dependent fusion inhibitor C-34. *Antiviral Res. 59*, 137–142.
23. Armand-Ugon, M., Clotet-Codina, I., Tintori, C., Manetti, F., Clotet, B., Botta, M., and Este, J. A. (2005) The anti-HIV activity of ADS-J1 targets the HIV-1 gp120. *Virology 343*, 141–149.
24. Greenberg, M. L., and Cammack, N. (2004) Resistance to enfuvirtide, the first HIV fusion inhibitor. *J. Antimicrob. Chemother. 54*, 333–340.
25. Mink, M., Mosier, S. M., Janumpalli, S., Davison, D., Jin, L., Melby, T., Sista, P., Erickson, J., Lambert, D., Stanfield-Oakley, S. A., Salgo, M., Cammack, N., Matthews, T., and Greenberg, M. L. (2005) 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. 79*, 12447–12454.
26. Melby, T., Sista, P., DeMasi, R., Kirkland, T., Roberts, N., Salgo, M., Heilek-Snyder, G., Cammack, N., Matthews, T. J., and Greenberg,

M. L. (2006) 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 Res. Hum. Retroviruses 22*, 375–385.

27. Nameki, D., Kodama, E., Ikeuchi, M., Mabuchi, N., Otaka, A., Tamamura, H., Ohno, M., Fujii, N., and Matsuoka, M. (2005) Mutations conferring resistance to human immunodeficiency virus type 1 fusion inhibitors are restricted by gp4l and rev-responsive element functions. *J. Virol. 79*, 764–770.
28. Rimsky, L. T., Shugars, D. C., and Matthews, T. J. (1998) Determinants of Human Immunodeficiency Virus Type 1 Resistance to gp41-Derived Inhibitory Peptides. *J. Virol. 72*, 986–993.
29. Strockbine, B., and Rizzo, R. C. (2007) Binding of Anti-fusion Peptides with HIVgp41 from Molecular Dynamics Simulations: Quantitative Correlation with Experiment. *Proteins: Struct., Funct., Bioinf. 63*, 630–642.
30. Caffrey, M. (2001) Model for the structure of the HIV gp41 ectodomain: Insight into the intermolecular interactions of the gp41 loop. *Biochim. Biophys. Acta 1536*, 116–122.
31. Lori, F., Hall, L., Lusso, P., Popovic, M., Markham, P., Franchini, G., and Reitz, M. S., Jr. (1992) 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. 66*, 5553–5560.
32. MOE, version 2005.06 (2005) Chemical Computing Group Inc., Montreal.
33. Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.
34. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65*, 712–725.
35. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.
36. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33.
37. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B 102*, 3586–3616.
38. Feller, S. E., and MacKerell, A. D. (2000) An improved empirical potential energy function for molecular simulations of phospholipids. *J. Phys. Chem. B 104*, 7510–7515.
39. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926–935.
40. Feller, S. E. Wabash College Chemistry Department. http://persweb.wabash.edu/facstaff/fellers (accessed December 22, 2008).
41. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089–10092.
42. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys. 103*, 4613–4621.
43. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints: Molecular-Dynamics of N-Alkanes. *J. Comput. Phys. 23*, 327–341.
44. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc. 120*, 9401–9409.
45. Kollman, P. A., 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, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res. 33*, 889–897.
46. Balius, T. E., and Rizzo, R. C. (2009) Quantitative Prediction of Fold Resistance for Inhibitors of EGFR. *Biochemistry 48*, 8435–8448.
47. Carrascal, N., and Rizzo, R. C. (2009) Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13. *Bioorg. Med. Chem. Lett. 19*, 47–50.
48. Chachra, R., and Rizzo, R. C. (2008) Origins of Resistance Conferred by the R292K Neuraminidase Mutation via Molecular Dynamics and Free Energy Calculations. *J. Chem. Theory Comput. 4*, 1526–1540.
49. Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995) Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett. 246*, 122–129.
50. Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem. 100*, 19824–19839.
51. Moustakas, D. T., Therese Lang, P. T., Pegg, S., Pettersen, E., Kuntz, I. D., Broojimans, N., and Rizzo, R. C. (2006) Development and Validation of a Modular, Extensible Docking program: DOCK 5. *J. Comput.-Aided Mol. Des. 20*, 601–619.
52. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem. 98*, 1978–1988.
53. Sackett, K., Wexler-Cohen, Y., and Shai, Y. (2006) Characterization of the HIV N-terminal fusion peptide-containing region in context of key gp41 fusion conformations. *J. Biol. Chem. 281*, 21755–21762.
54. Kamath, S., and Wong, T. C. (2002) Membrane structure of the human immunodeficiency virus gp41 fusion domain by molecular dynamics simulation. *Biophys. J. 83*, 135–143.
55. Gordon, L. M., Mobley, P. W., Pilpa, R., Sherman, M. A., and Waring, A. J. (2002) Conformational mapping of the N-terminal peptide of HIV-1 gp41 in membrane environments using C-13-enhanced Fourier transform infrared spectroscopy. *Biochim. Biophys. Acta 1559*, 96–120.
56. Gordon, L. M., Mobley, P. W., Lee, W., Eskandari, S., Kaznessis, Y. N., Sherman, M. A., and Waring, A. J. (2004) 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. 13*, 1012–1030.
57. Sackett, K., and Shai, Y. (2005) The HIV fusion peptide adopts intermolecular parallel β-sheet structure in membranes when stabilized by the adjacent N-terminal heptad repeat: A C-13 FTIR study. *J. Mol. Biol. 350*, 790–805.
58. Zheng, Z., Yang, R., Bodner, M. L., and Weliky, D. P. (2006) Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry 45*, 12960–12975.
59. Li, Y., and Tamm, L. K. (2007) Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers. *Biophys. J. 93*, 876–885.
60. Yau, W. M., Wimley, W. C., Gawrisch, K., and White, S. H. (1998) The preference of tryptophan for membrane interfaces. *Biochemistry 37*, 14713–14718.
61. Scarlata, S., and Gruner, S. M. (1997) Role of phosphatidylethanolamine lipids in the stabilization of protein-lipid contacts. *Biophys. Chem. 67*, 269–279.
62. Grossfield, A., Zuckerman, D. M., and Ralph, A. W. (2009) Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. In Annual Reports in Computational Chemistry, Vol. *5*, pp 23−48, Elsevier, Amsterdam.
63. Hess, B. (2002) Determining the shear viscosity of model liquids from molecular dynamics simulations. *J. Chem. Phys. 116*, 209–217.
64. Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des. 18*, 113–135.
65. Dwyer, J. J., Wilson, K. L., Davison, D. K., Freel, S. A., Seedorff, J. E., Wring, S. A., Tvermoes, N. A., Matthews, T. J., Greenberg, M. L., and Delmedico, M. K. (2007) Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant virus. *Proc. Natl. Acad. Sci. U.S.A. 104*, 12772–12777.
66. Lu, M., Blacklow, S. C., and Kim, P. S. (1995) A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat. Struct. Biol. 2*, 1075–1082.
67. Eggink, D., Langedijk, J. P., Bonvin, A. M., Deng, Y., Lu, M., Berkhout, B., and Sanders, R. W. (2009) Detailed mechanistic insights into HIV-1 sensitivity to three generations of fusion inhibitors. *J. Biol. Chem. 284*, 26941–26950.
68. Gochin, M., and Cai, L. (2009) The role of amphiphilicity and negative charge in glycoprotein 41 interactions in the hydrophobic pocket. *J. Med. Chem. 52*, 4338–4344.
69. Lawless, M. K., Barney, S., Guthrie, K. I., Bucy, T. B., Petteway, S. R., Jr., and Merutka, G. (1996) HIV-1 membrane fusion mechanism:

Structural studies of the interactions between biologically-active peptides from gp41. *Biochemistry 35*, 13697–13708.

70. Hildinger, M., Dittmar, M. T., Schult-Dietrich, P., Fehse, B., Schnierle, B. S., Thaler, S., Stiegler, G., Welker, R., and von Laer, D. (2001) Membrane-anchored peptide inhibits human immunodeficiency virus entry. *J. Virol. 75*, 3038–3042.
71. Peisajovich, S. G., Gallo, S. A., Blumenthal, R., and Shai, Y. (2003) C-Terminal octylation rescues an inactive T20 mutant: Implications for the mechanism of HIV/SIMIAN immunodeficiency virus-induced membrane fusion. *J. Biol. Chem. 278*, 21012–21017.
72. Wexler-Cohen, Y., Johnson, B. T., Puri, A., Blumenthal, R., and Shai, Y. (2006) 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. 281*, 9005–9010.
73. Liu, S., Jing, W., Cheung, B., Lu, H., Sun, J., Yan, X., Niu, J., Farmar, J., Wu, S., and Jiang, S. (2007) HIV gp41 C-terminal heptad repeat contains multifunctional domains. Relation to mechanisms of action of anti-HIV peptides. *J. Biol. Chem. 282*, 9612–9620.
74. Champagne, K., Shishido, A., and Root, M. J. (2009) Interactions of HIV-1 inhibitory peptide T20 with the gp41 N-HR coiled coil. *J. Biol. Chem. 284*, 3619–3627.
75. Izumi, K., Kodama, E., Shimura, K., Sakagami, Y., Watanabe, K., Ito, S., Watabe, T., Terakawa, Y., Nishikawa, H., Sarafianos, S. G., Kitaura, K., Oishi, S., Fujii, N., and Matsuoka, M. (2009) Design of peptide-based inhibitors for human immunodeficiency virus type 1 strains resistant to T-20. *J. Biol. Chem. 284*, 4914–4920.
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., and Fujii, N. (2009) X-ray crystallographic study of an HIV-1 fusion inhibitor with the gp41 S138A substitution. *J. Mol. Biol. 392*, 657–665.
77. Qiang, W., and Weliky, D. P. (2009) HIV fusion peptide and its cross-linked oligomers: Efficient syntheses, significance of the trimer in fusion activity, correlation of β strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry 48*, 289–301.