Research Article

# Optimization and Computational Evaluation of a Series of Potential Active Site Inhibitors of the V82F/I84V Drug-resistant Mutant of HIV-1 Protease: an Application of the Relaxed Complex Method of Structure-based Drug Design

**Alexander L. Perryman**[1,*], **Jung-Hsin Lin**[2] **and J. Andrew McCammon**[1,3]

[1]*Howard Hughes Medical Institute, Center for Theoretical Biological Physics and Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093-0365, USA*
[2]*School of Pharmacy, National Taiwan University, Taipei 100, Taiwan*
[3]*Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0365, USA*
*Corresponding author: Alexander L. Perryman, aperryma@caltech.edu, aperryma@mccammon.ucsd.edu*

The Relaxed Complex method, an approach to structure-based drug design that incorporates the flexibilities of both the ligand and target protein, was applied to the immunodeficiency virus protease system. The control cases used AutoDock3.0.5 to dock a fully flexible version of the prospective drug JE-2147 (aka SM-319777 or KNI-764) to large ensembles of conformations extracted from conventional, all atom, explicitly solvated molecular dynamic simulations of the wild type, and the V82F/I84V drug-resistant mutant of HIV-1 protease. The best set of run parameters from the control cases produced robust results when used against 2200 different conformations of the wild-type HIV-1 protease or against 2200 conformations of the mutant. The results of the control cases, the published advice from experts, and structural intuition were used to design a new series of 23 potential active site inhibitors. The compounds were evaluated by docking them against 700 different conformations of the V82F/I84V mutant. The results of this first round of lead optimization were quite promising. Approximately one-third of that series performed at least slightly better than the parent compound, and four of those compounds displayed significantly better binding affinities against that drug-resistant mutant (within our computational model).

**Key words:** AutoDock, computational biochemistry, computer-aided drug design, human immunodeficiency virus protease, JE-2147, KNI-764, lead optimization, Relaxed Complex method, SM-319777, structure-based drug design, targeting ensembles of protein conformations, V82F/I84V drug-resistant mutant of HIV-1 protease.

Received 6 April 2006, revised 25 April 2006 and accepted for publication 26 April 2006

One of the deadliest pathogens in existence is the human immunodeficiency virus (HIV), which has already killed more people than the bubonic plague. HIV protease is the enzyme that cleaves the polyprotein of HIV into its separate enzymatic and structural components, which is a critical step in the viral maturation process. Sufficient inhibition of HIV-1 protease results in the budding of morphologically immature, non-infectious viral particles that are rapidly cleared from the infected host. The importance of this drug target was demonstrated by the fact that the creation of HIV-1 protease inhibitors led to the development of the HAART 'cocktail' (the combination of HIV reverse transcriptase and HIV protease inhibitors that forms the highly active anti-retroviral therapy regime), which significantly reduced the mortality associated with HIV infection. However, because the frequency of high-level resistance to one or more drugs in that cocktail has been rapidly increasing during the past few years, the currently available drugs are losing their effectiveness (1). One of the worst drug-resistant mutants of HIV-1 protease is the V82F/I84V strain (valine82 mutated to phenylalanine and isoleucine84 changed to valine). Those two conservative mutations per monomer give that mutant protease molecule the ability to evade all of the currently prescribed HIV-1 protease inhibitors. That double mutation impedes the binding of those inhibitors by making their affinities for the mutant from 11- to 2000-fold worse than their affinities for the wild-type HIV-1 protease target (2–4).

The Relaxed Complex method is a new approach to drug design that incorporates the flexibilities of both the ligand and of the target protein. For a detailed description of the Relaxed Complex method of structure-based drug design, see our previous studies (5,6). The Relaxed Complex method involves, first, generating a large ensemble of conformations of the protein target [such as by running molecular dynamics (MD) or Monte Carlo calculations], and secondly, docking known or potential inhibitors to that ensemble of conformations in an automated and fully flexible fashion [such as by using AutoDock3.0.5 (7)]. By following such an approach, the

flexibilities of both the potential inhibitors and of the target protein can be included in the drug design process.

As this method utilizes a two-step approach, the computational costs of the preparatory phase are significant. Although this can be considered as a disadvantage, significant basic science information concerning the drug targets can be gleaned from the first phase. This was demonstrated by our recently published MD simulations of HIV-1 protease, from which we both discerned a plausible mechanism of drug resistance for that mutant and also suggested the location of a potential new binding site for allosteric inhibitors that could perhaps regulate flap opening and closing (8). That hypothesis of the mechanism of drug resistance was supported by our recent restrained MD simulations (9), and the utility of that potential new target for drug design was also demonstrated by those restrained MD simulations and by recent, independent, coarse-grained calculations of the HIV-1 protease system (9,10). Thus, the MD simulations that were performed to generate targets against which to dock in these Relaxed Complex experiments were quite useful on many levels.

Including the flexibilities of both the ligand and the protein target can enhance the drug design process by making it more realistic, and the inclusion of that flexibility could be critical when targeting a highly dynamic protein, such as HIV protease. Considering the huge conformational changes that HIV protease experiences (see Figure S16), and considering both the large size and the extensive flexibility displayed by the HIV-1 protease inhibitors that are currently used clinically, applying the Relaxed Complex method to this system was a formidable challenge.

To prepare for Relaxed Complex studies against HIV protease, a plethora of AutoDock jobs with different run parameters and prepar-
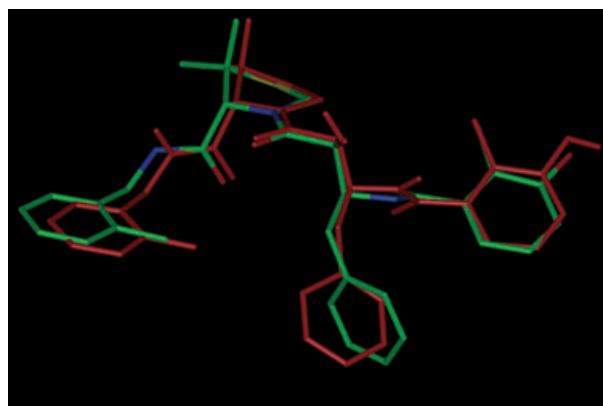


**Figure 2: The docked conformation of JE-2147 (from the highest ranked cluster in job II = 1-el) is displayed as the red stick figure, while the structure of JE-2147 from the 1KZK.pdb crystal structure is colored by atom type**. The protein target was hidden in this figure to allow one to focus on the accuracy of the docking calculations, when compared with the experimental evidence. For the purposes of hit to lead optimization, these docked results can be considered as an almost perfect reproduction of the crystallographic-binding mode.

ation protocols were performed in order to optimize the speed and the accuracy of the control experiments. The control case involved calculating the structure of the complex of wild-type HIV-1 protease with the second generation inhibitor JE-2147 (also known as the prospective drug SM-319777 or KNI-764), and the accuracy of the control experiments was judged by their ability to reproduce the 1KZK.pdb crystal structure of that complex (11; see Figures 1 and 2, Figures S12 and S13). Although a few of the initial sets of parameters did accurately reproduce the crystallographic binding mode of
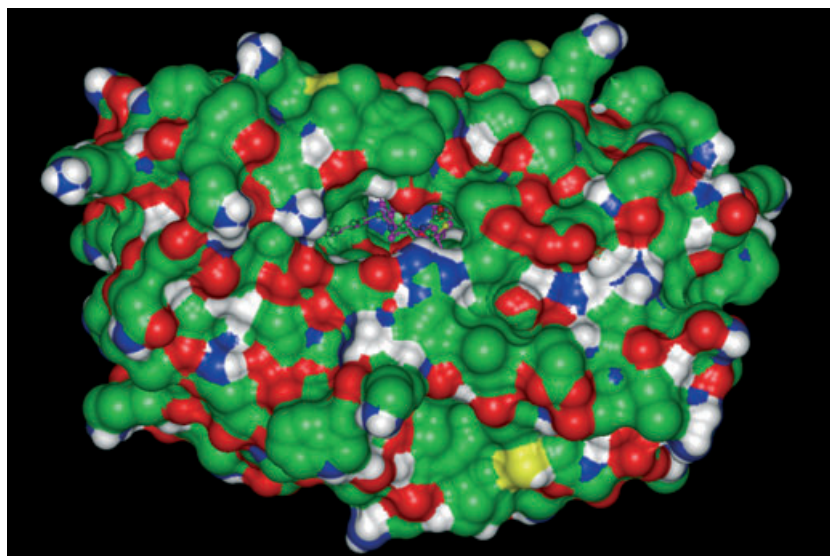


**Figure 1: The docked conformation of the prospective drug JE-2147 is shown as the magenta ball and stick structure in the center of the image, while the ball and stick structure that is colored by atom type displays JE-2147 from the 1KZK.pdb crystal structure**. The van der Waals surface of HIV-1 protease from that 1KZK.pdb crystal structure is shown, colored by atom type (red = oxygen, green = carbon, white = polar hydrogen, blue = nitrogen, and yellow = sulfur). The close superimposition of the docked and crystallographic structures will be highlighted in subsequent figures that zoom in on the active site; the purpose of this image is to highlight the location, the accessibility, and the general shape of the active site.

that inhibitor (see Chart I), subsequent control experiments were carried out to enrich the percentage of runs that produced the correct result and to decrease the amount of CPU time required per job (see Figure 2, Chart II).

The best preparation protocol and set of AutoDock3.0.5 parameters for docking to the crystal structure of this particular system (which involved a fully flexible ligand with 11 active torsions that is docked into a tunnel composed of the active site and the flaps that enclose it; see Figures 1 and 2, Figure S13, Chart II) were as follows: the united atom representation for both protein and ligand was used (i.e. the non-polar hydrogen atoms were merged onto their respective heavy atoms, which produced the set of atom types composed of CANOSH), random values for the initial dihedral angles of the ligand were generated, a focused docking approach was used (the center of the grid was located at the OD1 oxygen atom of the carboxylate group of the catalytic residue Asp25, and the grid size was $70 \times 70 \times 70$ points, with one point corresponding to 0.375 Å), only 1000 generations were used (instead of the default value of 27 000) in this Lamarckian Genetic Algorithm, a population size of 150 was selected (rather than the default value of 50 members), and an ls_search_frequency of 0.07 was used (as opposed to the default value of 0.06). The ls_search_frequency controls the probability of performing a local search (i.e. a pseudo-Solis and Wets minimization) on each individual member within the population, during each generation.

Many different sets of run parameters did reproduce the correct binding mode (see Charts I and II) – the previous parameter set just describes the most accurate job settings attempted when docking against that crystal structure. However, for Relaxed Complex purposes the best set of run parameters for docking to every 10th picosecond snapshot from those MDs trajectories (as judged by both their speed and reproducibility against a large and diverse ensemble of conformations) differed from the above set by using a population size of only 50 and by using just 750 generations (see Figures 2–4, Chart II).

When each snapshot was targeted in one Relaxed Complex experiment, 10 separate docking attempts were performed per snapshot, and the best run out of those 10 attempts was selected for energetic analysis. Each docking attempt involved hundreds of generations, and each generation involved hundreds of thousands of energy evaluations. Three independent Relaxed Complex experiments were performed on each system during the control studies (see Figures 3 and 4, Figures S14 and S15), but only one Relaxed Complex experiment was performed on each of the new compounds that were proposed (except for a few additional internal controls in which two or three experiments were performed on a few of the new compounds, see Table 1). The control cases involved targeting every 10th picosecond snapshot from 22 nanoseconds of classical, all atom, explicitly solvated MDs on the wild type and on the V82F/I84V drug-resistant mutant of HIV-1 protease (8). That is, 2200 different snapshots were targeted from each of the two systems (see Figures S14 and S15).

The robustness of this approach and of these run parameters was also confirmed by docking JE-2147 against several different
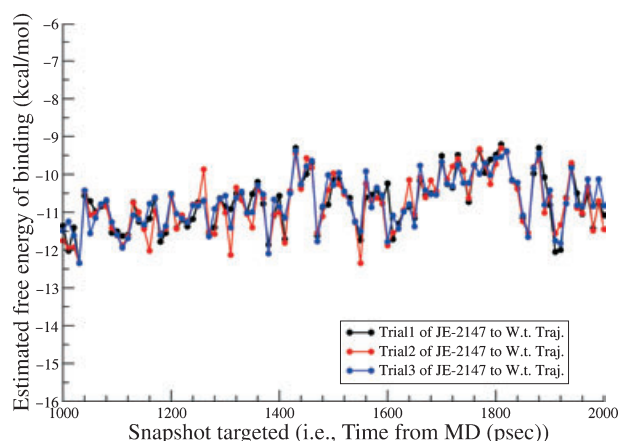


**Figure 3: The prospective drug JE-2147 was docked to 2200 different snapshots from the wild-type HIV-1 protease molecular dynamic (MD) trajectory, and the results of three independent trials were quite robust**. The results of targeting 100 of those conformations (every 10th snapshot from the second nanoseconds of the 22 nanoseconds wild-type trajectory) are shown above. Each docking run/snapshot involved hundreds to thousands of generations, and there were at least 50 members of the population within each generation. Each circle/trial signifies the best of 10 runs of docking to one particular snapshot of HIV protease, and each snapshot was targeted in three independent experiments. Therefore, a fully flexible version of JE-2147 was docked against each snapshot a total of 30 times, and 2200 different snapshots were targeted from each trajectory.
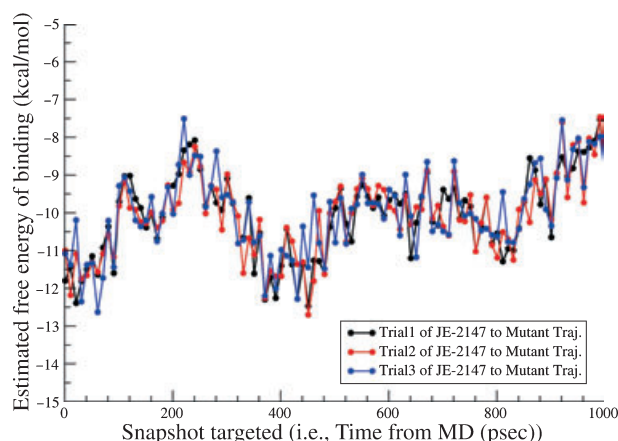


**Figure 4: The prospective drug JE-2147 was docked in three independent trials to 2200 snapshots of the V82F/I84V drug-resistant mutant HIV-1 protease in a reproducible manner, using the exact same optimized run parameters and procedures that were utilized when docking against the wild type's snapshots (see Figure 3)**. The results of targeting 100 of those snapshots (every 10th conformation from the first nanosecond of the 22 nanoseconds mutant trajectory) are shown above. To see the results of docking to the entire 22 nanoseconds from either the wild type or the mutant trajectory, see Figures S14 and S15, respectively. The few newly designed compounds that were utilized in multiple independent Relaxed Complex experiments as an internal control also displayed a similar level of robustness in their results (data not shown).

**Table 1:** Details of the structures and estimated binding free energies of the JE.D.I. series

| Ligand | $\langle \Delta G \rangle$ | Minimum | Maximum | P2 mod | P2′ mod | SD |
|---|---|---|---|---|---|---|
| 12 | −10.388 | −14.32 | −7.85 | CI** | CH3 | 1.03713 |
| 11 | −10.249 | −14 | −7.63 | CI** | CH3 | 1.03167 |
| 10 | −10.242 | −13.76 | −7.84 | CI** | | 1.03978 |
| 2–trial3 | −10.077 | −13.78 | −7.67 | CI** | | 1.04766 |
| 2 | −10.069 | −14 | −7.39 | CI** | | 1.04285 |
| 2–trial2 | −10.026 | −13.6 | −7.14 | CI** | | 1.04216 |
| 16 | −9.853 | −14.32 | −7.31 | OH & CH3 | CH3 & CH2 linker | 1.04620 |
| 19 | −9.691 | −13.63 | −7.09 | NH2 & CH3 | CH3 & CH2 linker | 1.01417 |
| 17 | −9.680 | −14.37 | −7.11 | OH | CH3 & CH2 linker | 0.99347 |
| 15 | −9.653 | −13.57 | −7.07 | OH & CH3 | CH2 linker | 0.99954 |
| 1–trial2 | −9.614 | −13.3 | −7.16 | NH2 & CH3 | CH3 | 1.04974 |
| 3 | −9.603 | −13.21 | −7.13 | CH3 | | 0.98879 |
| 1–trial3 | −9.602 | −13.16 | −7.14 | NH2 & CH3 | CH3 | 1.03912 |
| 1 | −9.583 | −12.47 | −7.04 | NH2 & CH3 | CH3 | 0.99851 |
| 18 | −9.563 | −14.11 | −7.13 | NH2 | CH3 & CH2 linker | 1.02511 |
| 8 | −9.526 | −13.69 | −7.09 | Arg tip | CH3 | 1.05497 |
| 24 | −9.525 | −13.03 | −6.71 | NH2 | CH3 | 1.02815 |
| 14 | −9.520 | −13.76 | −7.28 | OH | CH2 linker | 1.03586 |
| JE-2147–trial2 | −9.512 | −12.9 | −7.44 | Default | Default | 0.93706 |
| JE-2147 | −9.512 | −12.79 | −7.4 | Default | Default | 0.95553 |
| JE-2147–trial3 | −9.4735 | −12.63 | −6.85 | Default | Default | 0.93702 |
| 13 | −9.4730 | −13.16 | −6.96 | | CH2 linker | 1.01865 |
| 4 | −9.466 | −13.42 | −6.81 | NH2 & CH3 | | 1.01057 |
| 20 | −9.457 | −13.1 | −6.95 | NH2 & CH3 | CH2 linker | 1.02224 |
| 23 | −9.440 | −13.45 | −6.4 | NH2 & CH3 | | 1.00348 |
| 7 | −9.375 | −13.24 | −6.66 | CH3 & OH | | 1.06701 |
| 9 | −9.348 | −13.17 | −6.68 | Arg tip | | 1.03486 |
| 5 | −9.307 | −13.3 | −6.6 | NH2 | | 1.0091 |
| 22 = 5 | −9.306 | −12.8 | −6.58 | NH2 | | 1.00703 |
| 21 | −9.302 | −14.01 | −7.07 | NH2 | CH2 linker | 1.02525 |
| 6 | −9.211 | −12.62 | −6.36 | OH | | 0.98955 |

Although the entire series might not be synthetically accessible, all members can still be useful for computational QSAR (Quantitative Structure-Activity Relationships) studies.

$\langle \Delta G \rangle$ corresponds to the average value of the estimated free energy of binding that was calculated by docking to 700 different snapshots from the first seven nanoseconds of the V82F/I84V mutant's molecular dynamic (MD) trajectory. 'P2 mod' indicates the modifications that were made to the P2 side chain (see Figure 5), in a symmetrical manner. Likewise, 'P2′ mod' describes the modifications that were made to the P2′ side chain (also in a symmetrical manner). 'CH2 linker' signifies that an additional CH2 group was added between the backbone and the side chain of these peptidomimetic compounds, in order to increase the flexibility of that side chain. 'Arg tip' refers to the guanadinium group that is present at the end of an arginine side chain. 'Default' represents the side chain that is present in the parent compound, JE-2147. 'CI**' refers to 'confidential information' in a pending patent process (UCSD Docket No. SD2005-226).

restrained MD trajectories of the same wild type and mutant systems. This was underwent to demonstrate that the artificial restraints, which slightly expanded the peripheral surface of HIV protease (and which prevented flap opening behavior), did not have a significant deleterious effect on the dynamics of the active site (9). In fact, those restraints on the peripheral surface of the V82F/I84V drug-resistant mutant of HIV-1 protease actually improved the binding affinity of the active site inhibitor JE-2147 to the ensemble of conformations generated. Thus, that novel application of the Relaxed Complex method helped in the process of validating the peripheral surface of HIV protease as a potential new target for drug design (9).

As our new series of potential active site inhibitors were **d**esigned from and/or **i**nspired by **JE**-2147, we refer to them as the JE.D.I. series. Our series of 23 virtual compounds was docked against every 10th picosecond snapshot from the first seven nanoseconds of the V82F/I84V mutant's MD trajectory [i.e. against 700 different mutant conformations, which were generated by using the 1D4S.pdb crystal structure of the Tipranavir:mutant complex as an input (8,12)]. Compound JE.D.I.2 was also docked against 700 conformations from the wild-type ensemble (data not shown). The AutoDock3.0.5 calculated 'estimated free energy of binding' was used to evaluate the energies of the docked structures (i.e. no post-processing rescoring was performed).

## Methods and Materials

the apoprotease. Solvation of the protein (with its five chloride ions and 223 crystallographic water molecules present) with TIP3P water molecules was performed using LEaP of AMBER (16). The buffer distance chosen was 10 Å, which means that there was at least a 10 Å thick layer of water between the protein and each edge of the cubic box. To achieve electroneutrality of the system, one of the water molecules at a corner of the box was manually changed into an additional chloride counter-ion. Thus, the protease homodimer was surrounded by 8464 water molecules and six chloride ions during the minimization (for a total system size of 28 526 atoms).

A restrained minimization was performed on that protease system using the SANDER module of AMBER7 (16). The version of SANDER used was modified and recompiled to enable it to utilize the restraints in a more natural way, i.e. the value of the force constant for those restraints was applied and removed in a more gentle, linear ramping fashion, instead of being all-or-nothing. For both the minimization and the subsequent MD runs (8), the long-range Coulombic interactions were handled by the particle mesh Ewald method (17). The minimization was performed for 500 iterations, and a force constant of 30 kcal/mol/Å$^2$ was applied to the restrained atoms. During that minimization, the heavy atoms of the protein and the oxygen atoms of the 223 crystallographic waters were restrained, which allowed for the optimization of the hydrogen atoms added to both the protein and to the crystallographic waters as well as the optimization of the water molecules added by LEaP. The ligands were constructed using the molecular modelling platform SYBYL from Tripos Inc. (18) - See below.

After the restrained minimization, the minimization's restart file (in AMBER format) was used by our RST2PDB program to generate a pdb-formatted file. The minimized pdb-formatted file was read directly into AutoDock Tools (available at http://www.scripps.edu/mb/olson/doc/autodock), the non-polar hydrogen atoms were merged within AutoDock Tools (19), and Kollman united atom charges were loaded. AutoDock Tools was also used for the rest of the preparation phase, according to its standard protocol.

The exact same procedure was followed when preparing the mutant system (8), which was based upon the 1D4S.pdb crystal structure of the complex of Tipranavir bound to the V82F/I84V drug-resistant mutant of HIV-1 protease. For the mutant MD simulation, the inhibitor was also deleted prior to performing solvation, such that the MD experiment began with the closed conformation of the unbound protease.

For the first round of docking experiments, 12 different 'united atom' jobs and 15 different 'all atom' jobs were performed, with 10 runs attempted per job. The details of the first and second round of control experiments are presented in Charts I–III (as is a description of the less successful 'all atom' jobs).

The coordinates of the ligand JE-2147 were extracted from the 1KZK.pdb crystal structure of the wild type complex. The atom and bond types were manually corrected within SYBYL, all hydrogen atoms were added, Gasteiger–Marsili charges (20) were computed, and the hydrogen atoms had their atom names modified to make them sequentially numbered. Because the co-ordinates of the heavy atoms of JE-2147 were obtained from the crystal structure of the complex, the minimization of JE-2147 was performed with all of the heavy atoms constrained (i.e. they were 'aggregated' within SYBYL). The hydrogen atoms added to JE-2147 were minimized as follows: 500 iterations, an LS step size = 0.0005, the Tripos force field, and Gasteiger–Marsili (20) charges were used, and it was performed *in vacuo* (i.e. it used a distance-dependent dielectric function, with a constant = 1). For the AutoDock jobs involving 'JExtal' (the version of the ligand that used the crystallographic conformation as the AutoDock input), the minimized ligand was used directly within either AutoTors (a preparatory module of AutoDock) or AutoDock Tools. For the AutoDock jobs involving 'JEg98' (the version of the ligand for which quantum mechanical calculations were used to obtain the input conformation), the minimized version of the ligand was converted into a Gaussian98 input file format, and Gaussian98 (21) was used to optimize its geometry and to calculate the ESP charges. JExtal has 11 active torsions, while JEg98 has 15 active torsions (the difference was due to merging the non-polar hydrogen atoms in JExtal). An active torsion refers to a bond that is allowed to freely rotate during the AutoDock process.

The algorithmic details involved in applying the Relaxed Complex method to the HIV protease system included such things as converting all of the snapshots from all atom, parm99-formatted restart files into united atom, parm94-formatted pdb files. 'ParmXX' refers to the specific set of force field parameters that were used to describe the radii and charges for the atoms of the protein target, with parm94 being the AMBER force field from 1994 (22) and parm99 being the AMBER force field from 1999 (23). Those

pdb-formatted files were then converted into AutoDock's pdbqs format in a fully automated fashion, while maintaining the same relative position of the center of the grid. Significant trial-and-error was then involved in optimizing the run parameters for AutoDock3.0.5's Lamarckian Genetic Algorithm (7), in order to get reproducible results in an efficient manner.

In the control studies for the Relaxed Complex experiments that were performed, the completely flexible ligand JE-2147 was docked to every 10th picosecond snapshot extracted from both the 22 nanoseconds wild-type HIV-1 protease MD simulation as well as from the 22 nanoseconds V82F/I84V mutant MD simulation (8,9; see Figures 3 and 4, Figures S14 and S15). JE-2147 is the prospective drug that was crystallized with the wild-type HIV-1 protease in the 1KZK.pdb structure, which was the basis for the conventional MD simulation of the wild-type HIV protease (8). The exact same set of optimized run parameters (set I-I, see Figure 2, Chart II) performed very robustly when docking JE-2147 against all 22 nanoseconds of both the wild type and of the V82F/I84V mutant of HIV-1 protease (see Figures 3 and 4, Figures S14 and S15), even though the mutant was crystallized with Tipranavir (12), which is a member of a different class of HIV-1 protease inhibitors than JE-2147. When each Relaxed Complex experiment was repeated with that same set of optimized run parameters, the estimated free energy of binding that was calculated by docking against each particular snapshot had good agreement between the three independent trials that were performed on each of the two systems (see Figures 3 and 4, Figures S14 and S15). That is, the Relaxed Complex 'trajectories' of the estimated free energy of binding versus the snapshot that was targeted superimposed well when the experiments were repeated in triplicate on the wild type system and on the mutant system. After the reproducibility of this approach was established, an evaluation of our new virtual compounds was performed.

All of the control experiments were performed on the Keck II Center's linux cluster, which contains 14 IBM nodes with 512 MB RAM, dual 800 MHz Intel Pentium III processors and 24 Compaq nodes with 512 MB RAM, dual 800 or 933 MHz Intel Pentium III processors. A few of the control experiments were repeated on UCSD's NSF Center for Theoretical Biological Physics (CTBP) linux cluster, and all of the new compounds were evaluated on the CTBP cluster. That CTBP cluster has the following hardware: 145 nodes with PowerEdge 2650, dual 2.8 GHz Xeon processors, and 1 GB or 2 GB RAM/node.

## Results

Twenty-seven different jobs were performed in the first round of control experiments. For all 27 jobs, all of the runs docked the ligand into the active site (see Figures 1 and 2, Figures S12 and S13). For a qualitative summary of the results of the first round, see Chart I.

It is noted that some artifacts, which are obviously incorrect results with non-physical energy values (the estimated free energy of binding was in the negative thousands of kcal/mol, an error which was caused by maximizing electrostatics interactions at the expense of an impossible van der Waals overlap), were observed in the docking results, similar to what we had observed in the previous Relaxed Complex research (6). Artifacts, because of their extremely low energy values, were always ranked as the best cluster, even though they should be ignored. Therefore, when discussing these Relaxed Complex results, the 'best' cluster refers to the highest-ranked cluster that was not an artifact.

For the first round of control experiments, only jobs Ic, Ie, and IVc contained runs that calculated the correct answer (see the Supplementary Material for the details of the different jobs performed). Due to the large number of artifacts in round 1 (which seemed to be caused by having part of the protein target either threaded through one of the benzene rings of the ligand or by having a complete van der Waals overlap of a ligand atom with a protein atom), the jobs in

round 2 utilized a much smaller number of generations. By setting the output's level of detail, or the 'outlev', to 2, one can monitor the value of the estimated free energy of binding calculated for the best individual in each successive generation, which allows one to observe the generation number during which the artifact was first created. Although the number of artifacts was substantially decreased after the first round of control experiments, a few artifactual results were still observed. To draw attention to these artifacts when presenting the Relaxed Complex data, the artifacts were manually assigned an estimated free energy of binding value of zero; consequently, in the Relaxed Complex 'trajectories', each artifact is displayed as a prominent spike. Artifacts were removed before the statistical analyses were performed to generate the histograms illustrated in Figures 6–9, Figure S17. For the generation of those histograms, 10 bins were used to sort the energetic data.

The best set of optimized run parameters (set I-I, see Figure 2, Figure S12, Chart II) was then utilized in Relaxed Complex experiments that involved the evaluation of a newly designed series of potential active site inhibitors. These virtual compounds were docked to 700 different conformations of the V82F/I84V drug-resistant mutant of HIV-1 protease, in an attempt to develop compounds that should be more effective against that mutant. Structural intuition and guidance from both the literature and from the control experiments were used to design that series of 23 new, slightly different compounds. Thus, we are attempting to optimize a prospective drug to increase its affinity against one of the worst drug-resistant mutants of HIV protease. We followed the published advice from Ernesto Freire's group that the size, the flexibility, and the asymmetry of the P2/P2′ side chains should be increased when trying to design an HIV protease inhibitor that will be more effective against the drug-resistant mutants (4).

Details for the 19 compounds that will be disclosed are shown in Table 1. Because these compounds were **d**erived from and/or **i**nspired by **JE**-2147, they are known as the JE.D.I series. The structures of the parent compound JE-247 and of the sixth best virtual compound (i.e. JE.D.I.19) are illustrated in Figure 5. All 23 of those JE.D.I. compounds were utilized in independent Relaxed Complex experiments that targeted the first seven nanoseconds from the MD trajectory of the V82F/I84V mutant of HIV-1 protease (8). (Note: although Table 1 contains a compound '24', there are really only 23 unique compounds in this series, as JE.D.I.5 and JE.D.I.22 are identical.) This served as an accidental, blind, internal control, because those two compounds had almost identical $\langle \Delta G \rangle$ values (average, estimated changes in free energy upon binding) and were ranked adjacent to each other (see Table 1). Other internal controls are displayed on Table 1 as 'trial2' or 'trial3' of other compounds from that series (e.g. JE.D.I.1, JE.D.I.2, and JE-2147 were all used as internal controls). When those compounds were used in subsequent experiments during those internal controls, they still had the same relative ranking amongst the entire series of compounds. These Relaxed Complex experiments all used the same preparation protocols and run parameters that were used to generate the data illustrated in Figures 1–4 (i.e. the I-I set, see Chart II, Figures S12, S14 and S15). In Table 1, the $\langle \Delta G \rangle$ corresponds to the average value of the estimated free energy of binding that was calculated by docking that ligand to 700 different mutant conformations (as judged by AutoDock3.0.5's
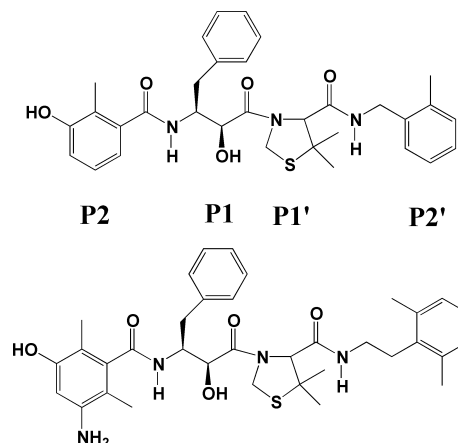


**Figure 5: The structure of the prospective drug, JE-2147 (11), is shown above, while the sixth best JE.D.I. compound, JE.D.I.19, is depicted below**. This parent compound has subnanomolar affinity against the wild-type HIV-1 protease. In the series of newly designed compounds proposed here, the size, the asymmetry, and/or the flexibility of the P2 and/or the P2′ side chains were systematically modified in an attempt to optimize the compound's binding affinity against the V82F/I84V drug-resistant mutant HIV-1 protease's ensemble of conformations [as per the suggestions published by Ernesto Freire's group (4)]. For example, methyl groups were added to either the P2 or the P2′ side chains (in a symmetrical manner, with respect to JE-2147) to give a dimethylated phenylalanine-like side chain, because some of the docking attempts in the control cases that were almost correct, displayed flipping of those rings. JE.D.I. compounds that contained a second methyl group on the P2′ ring (such as JE.D.I.19) displayed better affinities against the mutant ensemble (see Table 1), which agrees with the trend from recent experimental results regarding the optimization of both the affinity and the metabolic profile of this prospective drug (24).

scoring function), and all of the rows are sorted by that $\langle \Delta G \rangle$ value. The 'P2 mod' and 'P2′ mod' columns refer to the additional modifications that were made (in a symmetrical manner) to the P2 or the P2′ side chains of the parent compound, JE-2147, in order to obtain the new virtual compound (see Figure 5). 'Arg tip' refers to the addition of the guanadinium group that is normally found at the tip of an Arginine side chain. 'CH2 linker' means that an additional CH2 group was added between the backbone and the side chain, in order to increase the flexibility of that side chain.

The four best compounds in that JE.D.I. series (see Figures 6–10, Table 1, Figure S17) were structurally very similar to each other, but they were very different from the other 19 compounds in the series. And the three worst compounds were also structurally very similar to each other (see Table 1). As is apparent from Figures 6–9, approximately one-third of the JE.D.I. series (eight of the 23 compounds) performed at least slightly better than the prospective drug JE-2147 when docked against the first seven nanoseconds of the MD trajectory of the V82V/I84V drug-resistant mutant of HIV-1 protease. JE.D.I.2 also performed slightly better than JE-2147 when docked against 700 different wild type conformations (data not shown).

As can be illustrated in Figures 10 and 11, Figure S17, the trends in the estimated free energies of binding that were displayed in those histograms were also reflected in the graphs of the Relaxed
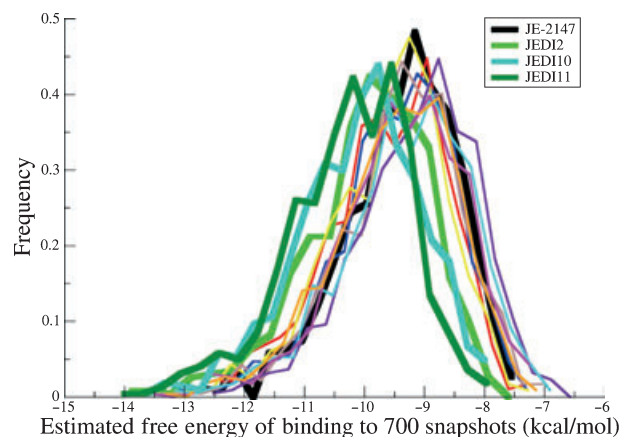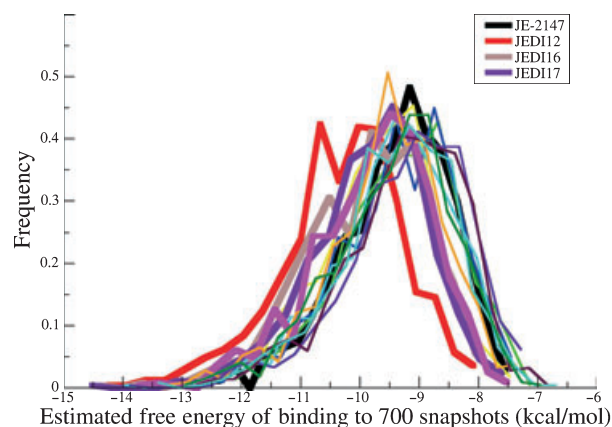
**Figure 6: Histograms displaying the results of docking the first half of the newly designed JE.D.I. series to the first seven nanoseconds of the molecular dynamic (MD) trajectory of the V82F/I84V mutant of HIV-1 protease. JE.D.I. stands for compounds that were derived from and/or inspired by JE-2147**. The thick black line signifies the histogram of the parent compound, JE-2147. The other three histograms with thicker lines represent three compounds, which performed better than JE-2147 when docked against that ensemble of the V82F/I84V mutant's conformations (as judged by both the global minima and the energetic profiles produced by docking against all 700 conformations). However, most of the first half of the series performed similarly to or slightly worse than the parent drug; those compounds are depicted by the histograms with thinner lines.



**Figure 7: Histograms displaying the results of docking the second half of our newly designed JE.D.I. series to the first seven nanoseconds of the molecular dynamic (MD) trajectory of the V82F/I84V mutant of HIV-1 protease**. The thick black line signifies the histogram of the prospective drug, JE-2147, which inspired the creation of this new series of compounds. The other three histograms with thicker lines represent three compounds, which performed better than JE-2147 when docking against 700 different conformations of that drug-resistant V82F/I84V mutant.

Complex 'trajectories'. That is, when examining the results of docking to each and every one of those 700 conformations that were targeted, the four best JE.D.I. compounds in the histograms tended to have a better free energy of binding to almost every conforma-
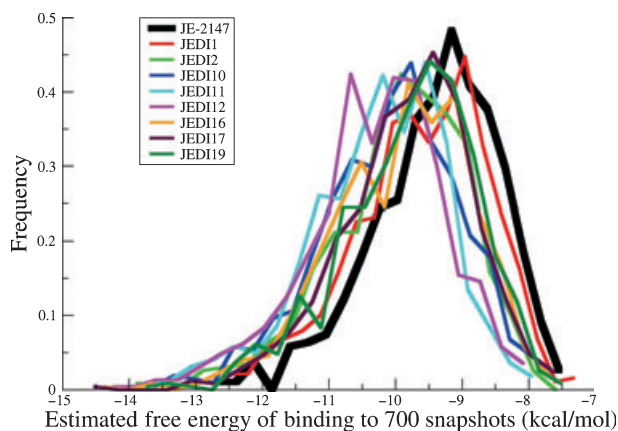


**Figure 8: Histograms displaying the results of all of the JE.D.I. compounds that performed better than the prospective drug, JE-2147, which inspired the creation of this new series of compounds and is shown in black**. Of the 23 compounds that were designed, eight performed at least slightly better than the parent compound, which was a surprising success in this first attempt at lead optimization (i.e. drug to 'super-drug' optimization).
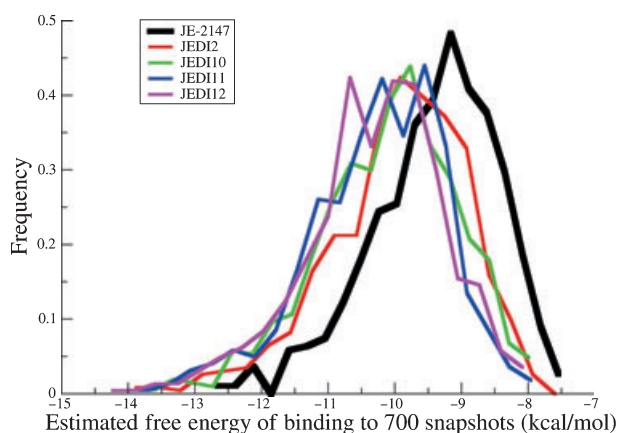


**Figure 9: Histograms comparing the Relaxed Complex results of the four best JE.D.I. compounds to the data of the parent compound, JE-2147 (shown in black), when docked against 700 different conformations from the first seven nanoseconds of the V82F/I84V mutant's molecular dynamic (MD) trajectory**. An improvement of at least 1–2 kcal/mol has been achieved in the free energy of binding (according to the computational model); consequently, these new compounds should have significantly higher affinity against that drug-resistant mutant (i.e. they should possess approximately five- to 30-fold better affinity against that mutant).

tion, while the three worst compounds in the histograms displayed binding free energies that were similar to or worse than JE-2147 for almost every snapshot that was targeted.

## Discussion

A few different sets of run parameters were discerned that consistently and quickly produced the correct structure of the complex of
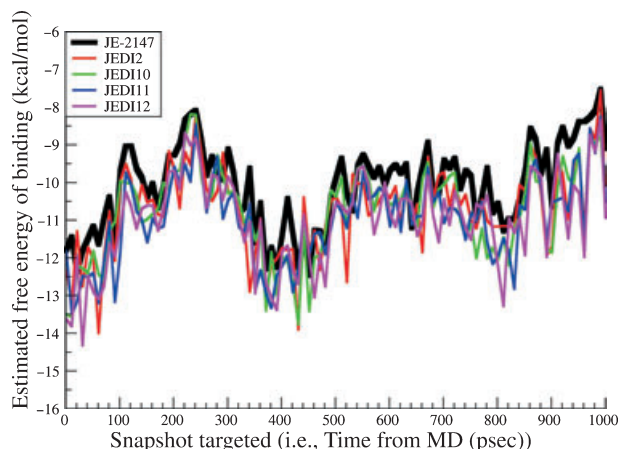
**Figure 10: 'Trajectories' of the Relaxed Complex results of the four best JE.D.I. compounds compared with the results of the prospective drug, JE-2147 (shown in black), when docked against conformations extracted from the molecular dynamic (MD) simulation of the V82F/I84V mutant.** The best JE.D.I. compounds (JE.D.I.2, 10, 11, and 12) had better free energies of binding than the parent compound for almost every snapshot that was targeted throughout the entire seven nanoseconds of the experiment – sometimes several kcal/mol better. Thus, the trends in the trajectories and in the histograms had excellent agreement. Only the first nanosecond is shown.

JE-2147 with HIV-1 protease, as judged by reproducing the 1KZK.pdb crystal structure (see the Supplementary Material for the description of the different sets of parameters that were investigated).

A Relaxed Complex protocol was developed, and the run parameters were optimized, such that reproducible results were obtained when docking the prospective drug JE-2147 to the structurally diverse, 22 nanoseconds long MD trajectories of either the wild type or of the V82F/I84V drug-resistant mutant of HIV-1 protease (see Figures 3 and 4, Figures S14–S16), even though the crystal structure of the mutant protease that was used as the input for its MD simulation contained a different class of inhibitor (Tipranavir). After establishing the reproducibility of this approach, a new series of potential active site inhibitors of HIV-1 protease was created and tested computationally.

The 'trajectories' of the Relaxed Complex results illustrated in Figures 10 and 11 agreed completely with the trends derived from the histograms of the free energies of binding to those 700 different conformations of the V82F/I84V drug-resistant mutant of HIV-1 protease (Figures 6–11). The four best JE.D.I. compounds (JE.D.I.2, 10, 11, and 12) had histograms that were shifted to more favorable binding free energies by 1–2 kcal/mol (with respect to the prospective drug JE-2147), and those four compounds had Relaxed Complex trajectories that were almost always lower in energy than the trajectory of the parent compound (sometimes by 3–4 kcal/mol). That is, those four compounds had a better free energy of binding than the prospective drug JE-2147 when docked against almost every conformation that was targeted from the first seven nanoseconds of the V82F/I84V drug-resistant mutant's MD trajectory (see
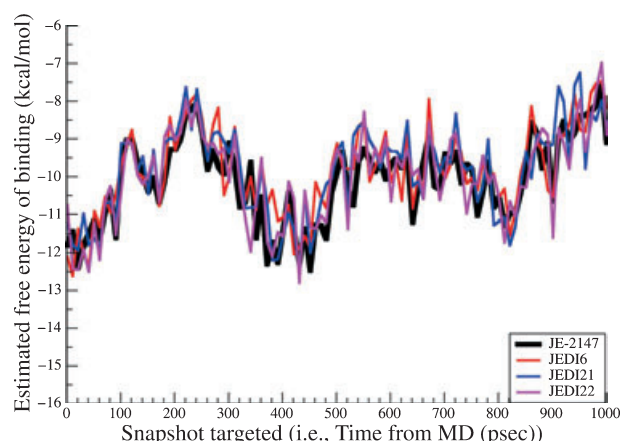


**Figure 11: Trajectories of the Relaxed Complex results of the three worst JE.D.I. compounds compared with the results of the prospective drug, JE-2147 (depicted in black), for the first nanosecond of the experiment.** These three compounds (JE.D.I.6, 21, and 22) performed similarly to or slightly worse than the parent compound for almost every snapshot that was targeted from the first seven nanoseconds of the V82F/I84V mutant's molecular dynamic (MD) simulation.

Figure 10). Substantial flap displacements were observed during those first seven nanoseconds of the mutant's MD trajectory (see Figure S16). A 1–2 kcal/mol improvement in the free energy of binding corresponds to a predicted five- to 30-fold improvement in the binding affinity of the inhibitors against that drug-resistant mutant. Conversely, the three worst compounds had Relaxed Complex trajectories that either overlapped or were higher in energy than the parent compound, JE-2147 (see Figure 11). From that series of 23 new compounds, four inhibitors have been proposed that should display significantly better binding affinities with the V82F/I84V drug-resistant mutant of HIV-1 protease. The full structural details of the four best compounds will be disclosed in the forthcoming study.

## Conclusions and Future Directions

An automated reformatting protocol and optimized run parameters were developed that allowed for the rapid and reproducible docking of the prospective drug JE-2147 (in a completely flexible manner) to every 10th picosecond snapshot from the entire 22 nanoseconds long MD trajectories of both the wild type and of the V82F/I84V drug-resistant mutant HIV-1 protease systems. That is, over 2200 snapshots were targeted from each system during the control studies. Even though the wild-type protease was crystallized with JE-2147 in that 1KZK.pdb structure, while the mutant was crystallized with a drug from a different class of active site inhibitors (Tipranavir), and even though the ligands were deleted before initiating either MD experiment, reproducible results for the value of the estimated free energy of binding calculated for each particular snapshot were obtained when docking the prospective drug JE-2147 to 2200 different wild type conformations or to 2200 different mutant conformations (using the exact same set of run parameters for both systems, set I-l; see the Supplementary Material).

Because the I-I set of run parameters gave very reproducible results when docking JE-2147 against a large and diverse ensemble of conformations from either the wild type or the mutant system, it was assumed that they would be useful for the relative evaluation of a series of new potential inhibitors that we created, i.e. the JE.D.I. series. The results of the internal control experiments performed on a few of the JE.D.I. compounds support that assumption.

We used structural intuition, information from the control cases that were close to being correct, and the published conclusions from Ernesto Freire's (4) group (that the size, the asymmetry, and the flexibility of the P2 and P2′ side chains should be increased when trying to design an inhibitor that will be effective against drug-resistant mutants of HIV-1 protease) to design and computationally create a series of 23 potential new active site inhibitors of HIV-1 protease (see Table 1). This new series of virtual compounds was then evaluated by using them in Relaxed Complex studies against every 10th picosecond snapshot from the first seven nanoseconds of the V82F/I84V mutant's classical, all atom, explicitly solvated MD trajectory. That is, each new compound was docked to 700 different mutant conformations, and each conformation was targeted 10 times per job. From that series of 23 new compounds that we designed, eight performed at least slightly better than the prospective parent drug, and four compounds look very promising. All of the new compounds contained a larger number of atoms than the parent compound, and some of the compounds in that series performed worse than JE-2147. Consequently, the observed improvement in affinity for those eight compounds was not due to a simple increase in van der Waals interactions resulting from a larger number of atoms in the new ligands. The compounds in the JE.D.I. series that contained a second methyl group on the P2′ side chain displayed better affinities against the drug-resistant mutant's ensemble of conformations, in agreement with the trend discussed in a recently published study on the optimization of the affinity and the metabolic profile of the prospective parent drug JE-2147 (24).

The four best JE.D.I. compounds are structurally very similar to each other, but they are quite different than the rest of the series. And the three worst compounds of the series are also structurally very similar to each other. Our internal controls all performed as predicted. Thus, we have now proposed four slightly different compounds, which should have significantly better affinity against the V82F/I84V drug-resistant mutant than the parent, compound JE-2147 displays (according to our computational model). As these four new compounds should possess at least a five- to 30-fold improvement in their ability to bind and inhibit that V82F/I84V mutant of HIV-1 protease, these compounds could be quite effective against that mutant (and perhaps against other drug-resistant mutants of HIV protease). Once ALP establishes his own laboratory, he will begin the construction and analysis of the next generation of this JE.D.I. series. When new drug-resistant mutations of HIV protease eventually arise, this Relaxed Complex approach could be utilized again in order to optimize the currently available drugs and produce new compounds that should hopefully be more effective. This approach should also be useful against drug-resistant mutants from other pathogenic systems – an assumption that ALP certainly plans to test in the future. Totally, novel leads can be investigated and optimized using this approach, as well. It is hoped that this Relaxed Complex method could potentially allow the drug design community to 'chase' HIV in sequence space until it can no longer mutate in a way that allows it to both evade the available drugs and still perform catalysis efficiently.

## Acknowledgments

## References

1. Little S.J., Holte S., Routy J.P., Daar E.S., Markowitz M., Collier A.C., Koup R.A. et al. (2002) Antiretroviral-drug resistance among patients recently infected with HIV. N Engl J Med;347:385–394.

2. Ala P.J., Huston E.E., Klabe R.M., McCabe D.D., Duke J.L., Rizzo C.J., Korant B.D. et al. (1997) Molecular basis of HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors. Biochemistry;36:1573–1580.

3. Klabe R.M., Bacheler L.T., Ala P.J., Erickson-Viitanen S., Meek J.L. (1998) Resistance to HIV protease inhibitors: a comparison of enzyme inhibition and antiviral potency. Biochemistry;37:8735–8742.

4. Ohtaka H., Velazquez-Campoy A., Xie D., Freire E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of HIV-1 protease. Protein Sci;11:1908–1916.

5. Lin J.-H., Perryman A.L., Schames J.R., McCammon J.A. (2002) Computational drug design accommodating receptor flexibility: the Relaxed Complex scheme. J Am Chem Soc;124:5632–5633.

6. Lin J.-H., Perryman A.L., Schames J.R., McCammon J.A. (2003) The Relaxed Complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers;68:47–62.

7. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comp Chem;19:1639–1662.

8. Perryman A.L., Lin J.H., McCammon J.A. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci;13:1108–1123.

9. Perryman A.L., Lin J.H., McCammon J.A. (2006) Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design. Biopolymers; 82(3):272–284.

10. Tozzini V., McCammon J.A. (2005) A coarse grained model for the dynamics of flap opening in HIV-1 protease. Chem Phys Lett;413:123–128.

11. Reiling K.K., Endres N.F., Dauber D.S., Craik C.S., Stroud R.M. (2002) Anisotropic dynamics of the JE-2147-HIV protease complex: drug resistance and thermodynamic binding mode examined in a 1.09 A structure. Biochemistry;41:4582–4594.

12. Thaisrivongs S., Skulnick H.I., Turner S.R., Strohback J.W., Tommasi R.A., Johnson P.D., Aristoff P.A. et al. (1996) Structure-based design of HIV protease inhibitors: sulfonamide-containing 5,6-dihydro-4-hydroxy-2-pyrones as non-peptidic inhibitors. J Med Chem;39:4349–4353.

13. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. et al. (2000) The protein data bank. Nucleic Acids Res;28:235–242.

14. Dolinksy T.J., Nielsen J.E., McCammon J.A., Baker N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res;32:W665–667.

15. Nielsen J.E., Andersen K.V., Honig B., Hooft R.W., Klebe G., Vriend G., Wade R.C. (1999) Improving macromolecular electrostatics calculations. Protein Eng;12:657–662.

16. Case D.A., Pearlman D.A., Caldwell J.W., Cheatham T.E. III, Wang J., Ross W.S., Simmerling C.L. et al. (2002) AMBER 7. San Francisco, USA: University of California.

17. Essmann U., Perera L., Berkowitz M.L., Darden T. (1995) A smooth particle mesh Ewald method. J Chem Phys;103:8577–8593.

18. Tripos Inc., (2001) SYBYL6.7.2. St Louis, MO, USA.

19. Huey R., Morris G.M. (2003) AutoDock Tools. La Jolla, CA, USA: The Scripps Research Institute.

20. Gasteiger J., Marsili M. (1980) Iterative partial equalization of orbital electronegativities – a rapid access to atomic charges. Tetrahedron;36:3219–3228.

21. Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Zakrzewski V.G. et al. (1998) Gaussian 98 (Revision A.2). Pittsburgh, PA, USA: Gaussian, Inc.

22. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Kenneth J., Merz M., Ferguson D.M. et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc;117:5179–5197.

23. Wang J.M., Cieplak P., Kollman P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem;21:1049–1074.

24. Mimoto T., Terashima K., Nojima S., Takaku H., Nakayama M., Shintani M., Yamaoka T. et al. (2004) Structure-activity and structure-metabolism relationships of HIV protease inhibitors containing the 3-hydroxy-2-methylbenzoyl-allophenylnorstatine structure. Bioorg Med Chem;12:281–293.

## Supplementary Material

Supporting information is available online at http://www.blackwell-synergy.com. The following information (and substantial additional text) supplements the article above.

**Chart I:** The first round of control studies: 27 different jobs that AutoDocked JE-2147 to the 1KZK structure of the wild-type HIV-1 protease.

**Chart II:** The second round of control studies displayed significant enrichment: 16 different jobs that AutoDocked JE-2147 to the 1KZK structure of the wild-type HIV-1 protease.

**Chart III:** A comparison of some of the CPU costs/job.

**Figure S12:** A ribbon view with a comparison of the computed and the crystallographic binding modes of JE-2147.

**Figure S13:** Close-up comparison of the predicted and experimental binding modes of the parent compound.

**Figure S14:** The Relaxed Complex results of targeting the entire 22 ns wild type MD trajectory.

**Figure S15:** The Relaxed Complex results of targeting the entire 22 ns MD trajectory of the V82F/I84V drug-resistant mutant.

**Figure S16:** Comparison of the flap motions displayed in the wild type and in the mutant MD simulations that were source of the target conformations.

**Figure S17:** Histograms comparing the Relaxed Complex results of the four best JE.D.I. compounds to the three worst compounds.