Chem Biol Drug Des 2007; **69:** 298–313

Research Article

doi: 10.1111/j.1747-0285.2007.00514.x

Design of Mutation-resistant HIV Protease Inhibitors with the Substrate Envelope Hypothesis

Sripriya Chellappan[1,†], G. S. Kiran Kumar Reddy[2,†], Akbar Ali[2], Madhavi N. L. Nalam[3], Saima Ghafoor Anjum[2], Hong Cao[2], Visvaldas Kairys[1,‡], Miguel X. Fernandes[1,‡], Michael D. Altman[4], Bruce Tidor[5,§], Tariq M. Rana[2,§], Celia A. Schiffer[3,§] and Michael K. Gilson[1,*,§]

[1]*Center for Advanced Research in Biotechnology, University of Maryland, Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA*
[2]*Chemical Biology Program, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA*
[3]*Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA*
[4]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[5]*Biological Engineering Division, Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
**Corresponding author: Michael K. Gilson, gilson@umbi.umd.edu; gilsonlab.umbi.umd.edu*
[†]*These authors contributed equally to this study.*
[‡]*Current address: Centro de Química da Madeira, Departamento de Química da Universidade da Madeira, Funchal, Portugal.*
[§]*These authors contributed equally to this study.*

There is a clinical need for HIV protease inhibitors that can evade resistance mutations. One possible approach to designing such inhibitors relies upon the crystallographic observation that the substrates of HIV protease occupy a rather constant region within the binding site. In particular, it has been hypothesized that inhibitors which lie within this region will tend to resist clinically relevant mutations. The present study offers the first prospective evaluation of this hypothesis, via computational design of inhibitors predicted to conform to the substrate envelope, followed by synthesis and evaluation against wild-type and mutant proteases, as well as structural studies of complexes of the designed inhibitors with HIV protease. The results support the utility of the substrate envelope hypothesis as a guide to the design of robust protease inhibitors.

Key words: AIDS, crystal, drug design, inhibitor, protease

Received 24 March 2007, revised and accepted for publication 17 April 2007

The value of HIV protease (HIVP) as a drug-target is powerfully validated by the fall in morbidity and mortality of HIV-positive individuals when the protease inhibitors were introduced to clinical practice (1). However, the utility of the first-generation protease inhibitors is challenged by the emergence of resistance and cross-resistance, which is associated primarily with mutations of protease that lead to decreased affinity of the inhibitors. There is thus a need for inhibitors that will retain affinity in the face of mutations. One approach to this problem is to seek inhibitors that bind wild-type protease so tightly that the losses in affinity produced by mutations can be tolerated. Another approach is to devise inhibitors with asymmetric, flexible chemical groups that can adapt conformationally to form stabilizing interactions with both wild-type and mutant proteases (2).

The present study explores a third approach which is based upon the availability of crystal structures of inactivated HIVP with substrate peptides whose sequences reflect the various sites in the Gag-Pol gene product that are cleaved by protease (3). These structures reveal that the substrates all occupy essentially the same region of the binding site despite their differing amino acid sequences. The border of this region has been termed the substrate envelope (4). It has furthermore been remarked that key resistance mutations of HIVP lie where inhibitors project outside the substrate envelope (5). These observations lead to the hypothesis that an inhibitor which fits within the substrate envelope will resist clinically relevant mutations. The rationale is that any mutation of HIVP that reduces the affinity of such an inhibitor should also reduce the affinity of substrate, and hence diminish the viability of the virus. This substrate envelope hypothesis is supported by a retrospective data analysis (6), but has not heretofore been tested prospectively.

This paper presents the first prospective evaluation of the substrate envelope hypothesis. Computational methods were used to design a focused combinatorial library of compounds that would fit within the substrate envelope and also bind with high affinity. The two nanomolar inhibitors that resulted were assayed against a panel of clinically relevant mutants of HIVP, and the resulting resistance profiles were compared with those of a panel of well-known HIVP inhibitors. The complexes of the two designed inhibitors were furthermore solved crystallographically and used to evaluate the accuracy of the computational designs.

Methods

Selection of scaffold and candidate substituents

The hydroxyethylaminosulfonamide scaffold shown in Figure 1 was chosen as a basis for combinatorial design based upon its proven suitability as a basis for high-affinity protease inhibitors, such as amprenavir (APV), and its synthetic accessibility. In addition, the scaffold is well anchored by hydrogen bonds to the aspartyl dyad and the flap water, so it was expected that a variety of substituents could be predictably positioned in the targeted subsites. The synthetic approach (see Section Chemistry) requires use of a carboxylic acid at R1, a primary amine at R2, and a sulfonyl halide at R3. Carboxylic acids and primary amines were collected from the 'all purchasable' subset of the Zinc database (7), and sulfonyl halides were collected from the Sigma-Aldrich (St. Louis, MO, USA) and Maybridge (Cornwall, UK) catalogs. Candidate substituents with more than 11 non-hydrogen atoms were discarded in order to favor the construction of small compounds that could fit the substrate envelope. This led to approximately 7000, 1200, and 350 candidate substituents at R1, R2, and R3, respectively, for an initial virtual library of about 3 billion compounds.

Computational methods

Combinatorial library design

Selection of compounds from this large virtual library was made tractable as follows. First, APV was used as a test bed for evaluating the candidate subsituents at each site. Thus, each candidate substituent at position R1 was substituted at R1 of APV, the resulting 7000 compounds were docked and their combined *Z*-scores (see Section Combined docking and substrate envelope scores) were evaluated, and all but the top-scoring 150 candidates were discarded. The same process was used to select the top-scoring 150 candidates out of the 1200 and 350 candidate subsituents at R2 and R3, respectively.

A Genetic Algorithm (GA) method was then used to seek compounds with optimal combinations of the top 150 substituents at each site. The GA describes a compound from the virtual library as a chromosome with three genes, one for each of the three substitution sites of the scaffold. A specific compound is defined by the allele of each gene, i.e. by the choice of substituent at each site. An initial generation of 40 chromosomes with random substituents is formed, and all 40 compounds are docked and their combined *Z*-scores (see Section Combined docking and substrate envelope scores) are evaluated. These 40 compounds are merged into the pool of all compounds encountered so far during the entire GA run. A new generation is then formed by drawing compounds from the entire pool, using roulette-wheel selection (8) based upon ranking of the combined *Z*-scores, and using genetic modifications to produce another 40 compounds. Single-site mutations, in which one substituent is replaced with another drawn at random from the substituent library, are used to generate 75% of the subsequent generation. Crossover exchanges, in which one or two substituents are exchanged between two compounds, are used to generate the remaining 25% of the subsequent generation. These steps are iterated for 12 generations and the best-scoring compounds found during the entire procedure are identified. Longer runs were not found to markedly improve the combined *Z*-scores of the resulting compounds.

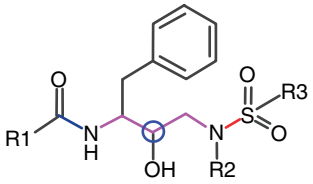


Figure 1: Combinatorial scaffold showing the translational center (blue circle) and the restrained torsions. Rotation about the blue, magenta, and red colored bonds is restrained to ±20°, ±30°, and ±90°, respectively, from their values in the structure of amprenavir in crystal structure 1HPV (9).

Ten GA runs with different random number seeds were carried out and substituents that occurred repeatedly among the top-scoring compounds were filtered to eliminate unsuitable substituents, such as symmetric amines and compounds that proved to be unavailable. Further substituents were eliminated in order to formulate fully combinatorial focused libraries. A fully combinatorial library of 27 compounds was arrived at and 26 of the compounds were synthesized and tested.

Docking and energy scoring

Candidate inhibitors were docked into an HIVP structure solved with APV [1HPV (9)], which embodies the same chemical scaffold used in the libraries studied in Section Selection of scaffold and candidate substituents. The protein structure was prepared by removing the bound APV molecule and other non-protein components except for the flap water, which is required for the intended hydrogen bonding interactions. The program QUANTA (Accelrys, Inc., San Diego, CA, USA) was used to add all polar hydrogen atoms to the protein and optimize their positions via energy minimization with only hydrogen atoms free to move. The program VDOCK (10,11) was used to dock and score candidate inhibitors. Energies were computed with a distance-dependent dielectric constant with a coefficient of 4. The Coulombic and van der Waals grids were assigned grid spacings of 0.2 Å, and the search parameters were $N_{hunt} = N_{fine} = 3000$ (11).

The initial three-dimensional (3D) structure of each candidate inhibitor to be docked was assembled by joining an initial 3D structure of the scaffold matching that of APV in 1HPV with precalculated 3D structures of the required substituents. The substituent structures were drawn as-is from the Zinc database or, when such a structure was not available, constructed with the program QUANTA. A substituent was joined to the scaffold by geometrically overlaying the appropriate bonds in the scaffold and substituent, deleting redundant atoms, forming a new bond between the scaffold and the substituent, and assigning force-field parameters to the resulting compound. Ab initio-like partial charges of the resulting compounds were generated with VCHARGE (12), an electronegativity equalization method parameterized to reproduce electrostatic potential fields computed at the 6-31G* level. Lennard-Jones parameters were assigned from CHARMM (13) atom types and torsional parameters were drawn from the Dreiding force field (14).

During docking, the ligands were restrained to poses (defined by position, orientation, and conformation) similar to that of APV in 1HPV. Translations, defined by the position of one scaffold atom (Figure 1), were limited to ±1 Å along each axis. Overall ligand rotations were limited to ±30° relative to an initial position with the scaffold atoms superimposed on the corresponding atoms of bound APV. Core torsions of the scaffold were sampled over reduced ranges around the corresponding values for bound APV (Figure 1).

For a given compound, 10 independent docking runs generating 20 full docked conformations were carried out, each run using a different random number seed. The lowest energy conformation of all 200 resulting conformations was taken as the predicted binding pose, and the energy of this pose was taken as the compound's score. Docking and scoring a compound with this protocol requires approximately 6 min of CPU time on a commodity processor. The calculations were performed in parallel on a loosely coupled cluster of Linux processors.

Substrate envelope scoring

Candidate inhibitors were scored for their fit within the substrate envelope with a grid-based method, detailed elsewhere (6). Briefly, six crystal structures of inactivated HIVP with bound substrate [1F7A, 1KJ4, 1KJ7, 1KJF, 1KJG, 1KJH (3)] were overlaid, based upon main-chain atomic coordinates. Operating the twofold rotational chemical symmetry of the protease yielded a total of 12 overlaid substrate structures. A cubic 3D grid with side length of 10 Å and grid spacing of 0.2 Å was centered on the active site and each grid point was assigned an initial value of $g_{ijk} = 0$. The grid values were then incremented by 1 for every substrate structure for which the grid point lay within the CHARMM (13) van der Waals radius of any non-hydrogen atom of the substrate. The resulting grid values range from $g_{ijk} = 0$ for points outside all 12 substrate structures to $g_{ijk} = 12$ for points inside all 12 substrate structures.

The fit of a candidate inhibitor to the substrate envelope is computed by overlaying the docked inhibitor on the substrate density grid and computing the effective volume of the inhibitor outside the substrate envelope, V_{out}, by summing the values of the grid points g_{ijk} that lie within the van der Waals volume of the inhibitor, normalizing the sum by 12, and converting to a volume by multiplying by the 0.008 Å^3 volume of a grid box:

$$V_{\text{out}} = \frac{0.008}{12} \sum_{ijk}^{\text{inside}} (12 - g_{ijk}) \qquad (1)$$

Here, 'inside' implies that the sum runs only over grid points *ijk* within the van der Waals volume of the inhibitor. A smaller value of V_{out} implies better fit to the substrate envelope.

Combined docking and substrate envelope scores

Each docked candidate inhibitor *i* is assigned a combined score S_i based upon its energy score E_i and its volume outside the substrate envelope $V_{\text{out},i}$:

$$S_i = \left(\frac{Z_{E,i} + Z_{V,i}}{2} \right) \qquad (2)$$

$$Z_{E,i} = \frac{E_i - \bar{E}}{\sigma_E} \qquad (3)$$

$$Z_{V,i} = \frac{V_{\text{out},i} - \overline{V_{\text{out}}}}{\sigma_V}, \qquad (4)$$

where $\bar{E}$ and $\overline{V_{\text{out}}}$ are the average values of energy and V_{out} and σ_E and σ_V are the standard deviations of energy and V_{out}. Each *Z*-value this is a *Z*-score, i.e. the number of standard deviations by which a quantity deviates from the mean of its distribution. A low value of S_i indicates a compound that combines low energy and good fit to the substrate envelope.

Global analysis of compound sets

The sets of compounds chosen at various stages of the design process are globally characterized and compared via the fast approximation that each subsituent makes an additive contribution to the computed docking energy *E*. This in effect assumes that each substituent contributes independently to the score, an approximation which we have found to be reasonably accurate for this system (S. Chellappan et al., paper in preparation). The primary results of the present study are not affected by the validity of this approximation, in any case. The docking energy of a compound with three substituents is denoted as E_{ijk}, where *i*, *j*, and *k* are the indices of the substituents at each site. Thus, *i* varies from 1 to 7000 in the full virtual library. Amprenavir is arbitrarily selected as a reference compound with energy E_{000}, and the docking energy obtained by replacing APV's R1 substituent with substituent *i* is denoted E_{i00}. These single-site replacement energies are all available from the initial docking calculations described in Section Combinatorial library design, and they allow the docking energy of any compound in the entire virtual library to be estimated by

$$\begin{aligned} E_{ijk} \approx & E_{000} + (E_{i00} - E_{000}) + (E_{0j0} - E_{000}) + (E_{00k} - E_{000}) \\ = & E_{i00} + E_{0j0} + E_{00k} - 2E_{000} \end{aligned} \qquad (5)$$

Analogous calculations provide estimates of V_{out} for each compound:

$$V_{ijk} \approx V_{i00} + V_{0j0} + V_{00k} - 2V_{000} \qquad (6)$$

Experimental methods

Chemistry

Figure 2 illustrates the synthetic route of the designed protease inhibitor library (15). The Boc-protected(R)-(hydroxyethylamino)sulfonamide intermediates **5** were prepared following reported procedures. Briefly, ring opening of the commercially available chiral epoxide (1*S*,2*S*)-(1-oxiranyl-2-phenylethyl)carbamic acid *t*-butyl ester **1** with primary amines **2** provided the amino alcohols **3**. Reactions of various sulfonyl chlorides **4** with **3** afforded the sulfonamide intermediates **5**. After removal of the Boc protection, the free amines **6** were reacted with various carboxylic acids **7** under

standard amide coupling conditions (EDCI/HOBt/DIPEA) to generate the designed protease inhibitors **8**.

HIV-1 protease inhibition assays

Inhibitors were tested against wild-type HIVP and a panel of three proteases with clinically relevant mutation sets: M1 (L10I/G48V/I54V/L63P/V82A), M2 (D30N/L63P/N88D), and M3 (L10I/L63P/A71V/G73S/I84V/L90M). The mutant protease genes were constructed with standard site-directed mutagenesis, and the proteins were overexpressed and purified (16). As previously discussed (6) the clinical relevance C of a protease mutant is quantified by its tendency to occur in patients treated with protease inhibitors, and in the absence of other major protease mutations that could by themselves account for clinical resistance:

$$C \equiv 100 \frac{N_{\text{only}}}{N_{\text{all}}} \qquad (7)$$

Here, N_{only} is the number of HIVP sequences with a given mutation set and no other major mutations, as defined at the Stanford HIV Drug Resistance Database (17), and N_{all} is the total number of *all* HIVP isolates with the given mutation set. By this measure, mutants M1, M2, and M3 are highly clinically relevant, with C values of 19, 69, and 38, respectively.

The HIV-1 protease inhibitory activities of all new designed inhibitors were determined by fluorescence resonance energy transfer (FRET; 15). Protease substrate [Arg-Glu(EDANS)-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-Lys(DABCYL)-Arg], was purchased from Molecular Probes (Eugene, OR, USA). The energy transfer donor (EDANS) and acceptor (DABCYL) dyes were labeled at two ends of the peptide, respectively, to perform FRET. Fluorescence measurements were carried out on a fluorescence spectrophotometer (Photon Technology International, Brimingham, NJ, USA) at 30 °C. Excitation and emission wavelengths were set at 340 nm and 490 nm, respectively. Each reaction was recorded for about 10 min. Wild-type HIV-1 protease (Q7K) and its MDR variants M1 (L10I, G48V, I54V, L63P, V82A), M2 (D30N, L63P, N88D), and M3 (L10I, L63P, A71V, G73S, I84V, L90M) were desalted through PD-10 columns (Amersham Biosciences, Uppsala, Sweden). Sodium acetate (20 mM, pH 5) was used as elution buffer. Apparent protease concentrations were around 50 nM estimated by UV spectrophotometry at 280 nm. All inhibitors were dissolved in dimethylsulfoxide (DMSO) and diluted to appropriate concentrations. Protease (2 μL) and inhibitor (2 μL) or DMSO were mixed and incubated for 20–30 min at room temperature before initializing the substrate cleavage reaction. For all experiments, 150 μL of 1 μM substrate was used in substrate buffer [0.1 M sodium acetate, 1 M sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 2% DMSO, and 1 mg/mL bovine serum albumin (BSA) with an adjusted pH of 4.7]. Inhibitor binding dissociation constant (K_i) values were obtained by nonlinear regression fitting (GRAFIT 5, Erithacus software Surrey, UK) to the plot of initial velocity as a function of inhibitor concentrations based on the Morrison equation. The initial velocities were derived from the linear range of reaction curves.

Protein crystallography

Protein expression, isolation, and purification were carried out as previously described (18). Crystals were set up with a threefold molar excess of inhibitors to protease, with the protease at a concentration of 1.6 mg/mL. The hanging drop vapor diffusion method was used for crystallization. The reservoir solution consisted of 126 mM phosphate buffer at pH 6.2, 63 mM sodium citrate and ammonium sulfate in a range of 24–29%.

Intensity data were collected with an in-house Rigaku X-ray (Rigaku, Corp., Tokyo, Japan) generator equipped with an R-axis IV image plate system. Data were collected at −80 °C and data were processed with the programs DENZO and SCALEPACK (19,20), respectively. Table 1 lists the data collection statistics. The CCP4i interface to the CCP4 suite (21) was used to refine the structures, and the molecular replacement package AMORE (22), with 1F7A (3) as the starting model, was used to

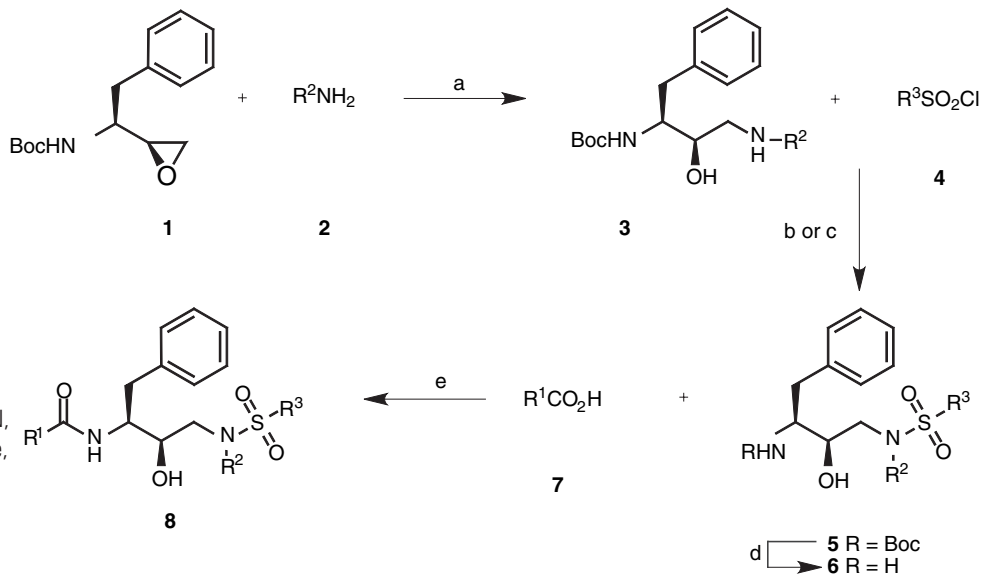


Figure 2: Synthetic scheme. (a) EtOH or iPrOH, 80 °C, 2–3 h; (b) aq. Na_2CO_3, CH_2Cl_2, 0 °C to room temperature, 4–8 h; (c) Et_3N, CH_2Cl_2, 0 °C to room temperature, 4–8 h; (d) TFA, CH_2Cl_2, room temperature, 1 h; (e) EDCI, HOBt, DIPEA, DMF-CH_2Cl_2 (1:1), 0 °C to room temperature, overnight.

obtain the structure solution. The molecular replacement phases were further improved by using the program ARP/WARP (23) to build solvent molecules into the remaining unaccounted regions of electron density. Model building was performed with the interactive graphics program O (24). Conjugate gradient refinement using REFMAC5 (25) was performed by incorporating the Schomaker and Trueblood tensor formulation of translation, libration, and screw-rotation parameters (26–28). The working R (R_{factor}) and its cross-validation (R_{free}) were monitored throughout the refinement. Table 1 provides the final refinement statistics.

Results

Computational design

The course of the design process can be understood through the score distributions of the compounds at various stages. The initial library of about 3 billion compounds is predicted to have very broad distributions of docking energies and substrate envelope fits, with modes of about −30 kcal/mol and 250 Å^3, respectively. (See solid black curves in Figure 3.) Eliminating all but the top-scoring 150 candidate subsituents at each site yields an initial focused library of about 3 million (150 × 150 × 150) compounds, whose predicted distributions of E and V_{out} are narrower and lower. (See dashed black curves in Figure 3.) This initial focused library contains some compounds with very low predicted energies (the best 100 are shown in red in Figure 3). However, this set of compounds has relatively large values of V_{out}, as shown in red in the bottom graph. Similarly, the 100 compounds with the lowest values of V_{out} (green bottom graph) have relatively poor energy scores (green, top graph).

The blue curves in both graphs show the distribution of docking scores and V_{out} computed for the 26 compounds that were actually synthesized and tested. This distribution lies between between those of the lowest energy compounds (red) and the lowest V_{out} compounds (green). Thus, the compounds that were synthesized represent a compromise between the requirements of achieving low energy and of achieving low V_{out}. Here, approximately equal weight is applied to both requirements because the combined score of a compound is a simple average of their energy and V_{out} Z-scores. The energy and V_{out} scores of the compounds that were synthesized (blue) also lie well to the left of most compounds in the initial focused library of approximately 3 million compounds (dashed black). For comparison, Figure 3 shows that the computed values for APV as well (vertical black line).

Table 1: X-ray crystallographic data collection and refinement statistics for complexes of CARB-AD37 and CARB-KB45 with the wild-type HIV-1 protease

	CARB-AD37	CARB-KB45
Resolution (Å)	1.93	1.75
Temperature (°C)	−80	−80
Space group	P212121	P212121
a (Å)	50.65	50.69
b (Å)	57.64	57.59
c (Å)	61.78	61.84
Z	4	4
R_{merge} (%)	6.5	4.5
Completeness (%)	99.4	98.4
Total reflections	92 428	114 039
Unique reflections	14 044	18 590
R_{free} (%)	19.8	20.8
R_{factor} (%)	16.1	16.5
Bond length RMSD (Å)	0.008	0.008
Bond angle RMSD (°)	1.221	1.266
Crystallographic waters	168	195
PDB ID	2PSV	2PSV

RMSD, root-mean-square deviation; PDB, Protein Data Bank.

Measured affinities

Seven of the 26 new compounds bind wild-type HIVP with submicromolar dissociation constants, and two compounds, CARB-AD37 and CARB-KB45, have low-nanomolar dissociation constants (Table 2). For comparison, the table also shows APV, the smallest of the first-generation clinical inhibitors, and ritonavir, one of the largest, along with their dissociation constants from the present assay. The new compounds are similar in size to APV.

Table 3 presents the affinities of CARB-AD37 and CARB-KB45 for three clinically relevant mutant proteases. These results are compared with those for six first-generation clinical inhibitors. The susceptibility of an inhibitor to a set of protease mutations can be defined as $\log \frac{K^{mut}}{K^{wt}}$, where K^{mut} and K^{wt} are the inhibition constants for the wild type and mutant, respectively. The susceptibilities are summarized in Figure 4, where each line represents one compound, and a lower, more level line indicates less susceptibility to mutation. The new compounds (red lines) are less susceptible to the mutations studied here than any of the clinical inhibitors (green) except APV (black). The new inhibitors also fit the substrate envelope better than any of the other inhibitors except for APV, according to the values of V_{out} in Table 3. On the other hand, the affinities of the new compounds are lower than those of the clinical inhibitors.

Structures of inhibitor complexes

The docked structures of bound CARB-AD37 and CARB-KB45 are illustrated in Figures 5 and 6, respectively. These compounds differ only in their R3 substituents, so their predicted interactions with HIVP are similar. In both cases, the scaffold forms the intended hydrogen bonds with the flap water and residue Asp 25 (Figures 5A and 6A), and the common phenyl and cyclopropyl groups form non-polar interactions with, respectively, Pro 81′ and Val 82′ at the S1′ subsite and with Leu 23 and Val 82 at the S1 subsite (Figures 5B,D and 6B,D). The amide R1 substituent occupies the S2 subsite, forming several non-polar interactions, and polar interactions with Asp 29, Asp 30, and Gly 48. However, the polar interactions do not meet standard geometric criteria for hydrogen bonding (Figures 5C and 6C). The R3 substituents are predicted to form non-polar contacts in the S2′ subsite, especially with Ala 28, and the methyoxy oxygen of CARB-AD37 is predicted in addition to accept a hydrogen bond from Asp 30′.

Overall, the crystallographic poses of these compounds agree well with the docked poses. Thus, the root-mean-square deviations

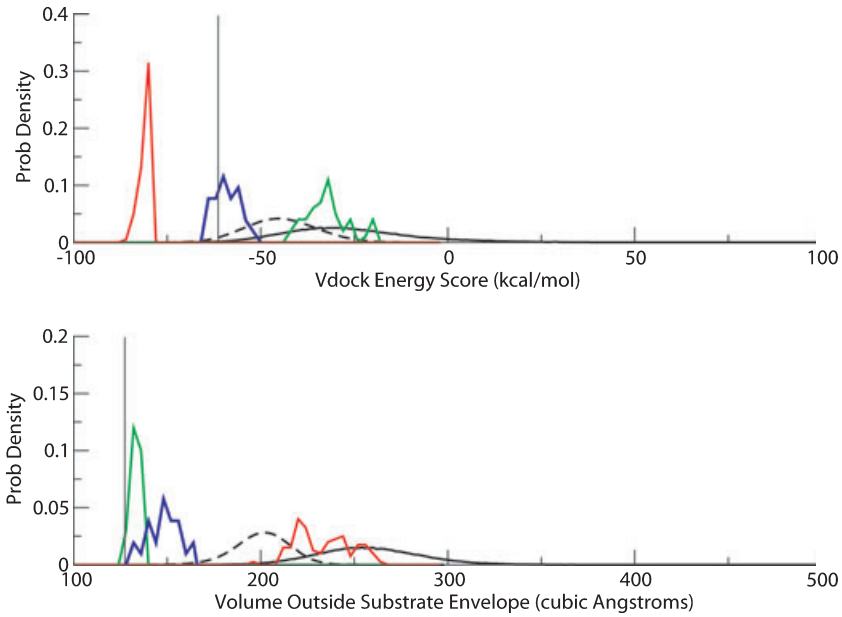


Figure 3: Distributions of VDOCK energy scores (top) and V_{out} substrate envelope scores (bottom) at various stages in the design process. Solid black: 10^8 compounds drawn randomly from the full virtual library of approximately 3×10^9 compounds. Dashed black: all compounds constructed from the 150 candidate substituents at each position that gave optimal combined *Z*-scores (see Section Combinatorial library design). Red: 100 compounds with the best (lowest) docking scores. Green: 100 compounds with the best (lowest) values of V_{out}. Blue: 26 synthesized compounds. Corresponding results for amprenavir are indicated by vertical black lines.

(RMSD) between the docked and crystallographic co-ordinates of non-hydrogen ligand atoms are 1.32 Å and 1.08 Å for CARB-AD37 and CARB-KB45, respectively. The observed hydrogen bonding interactions between the scaffold and the flap water are as predicted, but the scaffold in the crystal structures forms more hydrogen bonds with Asp 25′ than anticipated: compare Figures 5A with 7A and Figure 6A with 8A. The observed hydrophobic interactions of the scaffold's phenyl group are similar to those observed in the docked poses, involving Pro 81′, Val 82′, and Gly 49 in the S1′ subsite (Figures 5B versus 7B and 6B with 8B). However, the orientation and the interactions of the cyclopropyl group at R2 differ significantly from the predictions for both inhibitors. Also, the R1 substituents of both ligands meet geometric criteria for acceptance of a hydrogen bond from the backbone nitrogen of Asp 29 at the S2 subsite, although this interaction was not observed in the most stable computed poses (Figures 5C versus 7C and 6C with 8C). The R3 moiety of CARB-AD37 in the crystal structure is rotated relative to the docked conformation and forms more non-polar interactions from the phenyl moiety to Val 32. Similarly, more non-polar interactions are observed than predicted between the R3 substituent of CARB-KB45 and nearby non-polar groups (Figures 6E versus 8E).

Importantly, the geometry at the sulfonamide nitrogen in the crystal structures is inverted relative to that assumed in the docking calculations (Figures 9 and 10). This difference could account for some differences between the predicted and observed positions of the various substituents. The following subsection therefore analyzes its consequences for the docking calculations.

Consequences of sulfonamide nitrogen geometry

The sulfonamide geometry of the combinatorial scaffold can invert to produce two distinct conformations. VDOCK holds the geometry of this nitrogen in the conformation of the starting 3D structure of the scaffold, i.e. the S conformation observed in APV bound to 1HPV (Figure 9, top). However, the crystal structures of the designed ligands bound to HIVP show that the nitrogen adopts the R geometry (Figure 10, bottom), as noted above.

The geometry at the sulfonamide nitrogen could have influenced the VDOCK predictions directly, by affecting the positioning of the inhibitors' substituents; and/or indirectly by affecting the protein conformation, as the docking calculations used a crystal structure which had been solved with APV in the same sulfonamide geometry as used in the docking calculations. These possibilities were studied by additional docking calculations for CARB-AD37 and CARB-KB45 with both sulfonamide geometries into two receptor structures crystallized with ligands having different sulfonamide geometries, 1HPV (9) and UMASS-KB60 (unpublished results). The four resulting docked conformations of each ligand were superimposed and compared.

Tables 4 and 5 provide the RMSD of non-hydrogen atoms in the four docked conformations to the newly solved crystal structures for CARB-AD37 and CARB-KB45, respectively. Docking with the R conformation found in the new crystal structures yields lower RMSDs. However, the protein structure used in the docking calculations has an inconsistent influence on the RMSD values: docking into KB60 gives lower RMSDs for CARB-AD37, but higher RMSDs for CARB-KB45.

Table 2: Designed compounds and their measured inhibition constants for wild-type HIV protease. The first-generation drugs amprenavir (APV) and ritonavir (RTV) are included for comparison, with inhibition constants measured by the same assay.

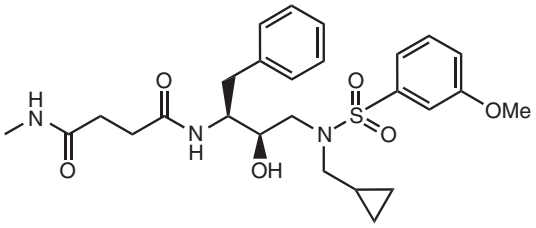
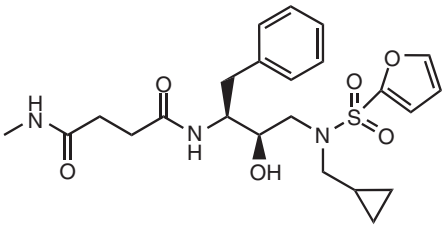
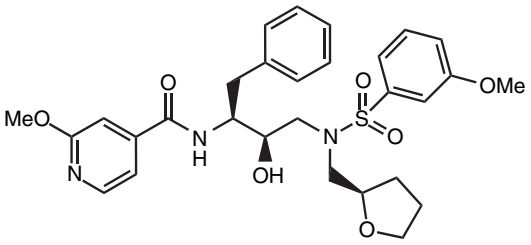
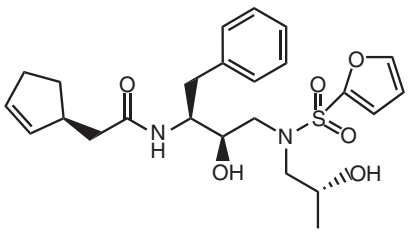
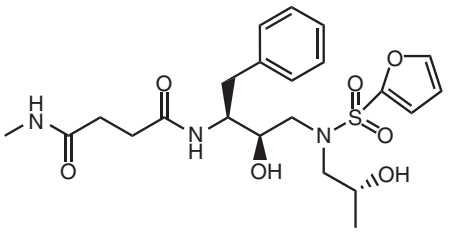
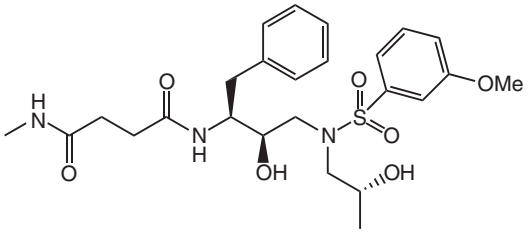
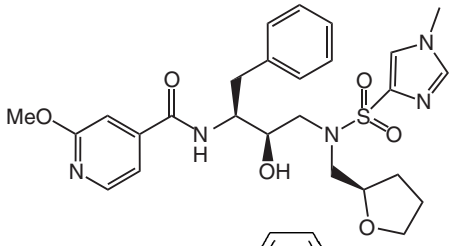
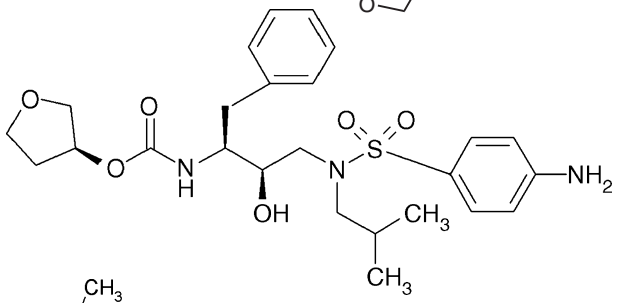
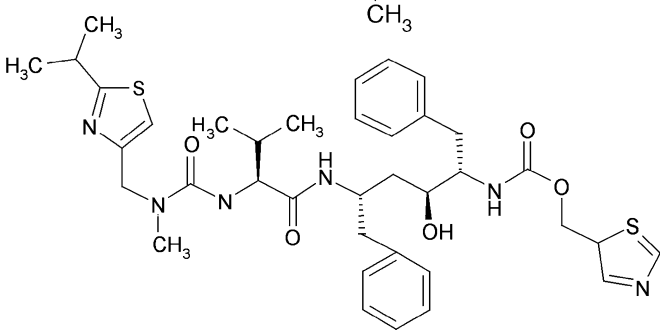
Name	Structure	K_i (nM)
CARB-AD37		24
CARB-KB45		58
CARB-AD08		110
CARB-KB51		260
CARB-KB49		540
CARB-KB32		760

Table 2: (Continued)

Name	Structure	K_i (nM)
CARB-AC97		910
APV		0.1
RTV		0.06

The conformation of the scaffold is similar across all four docked conformations of CARB-AD37, so this aspect of the docked conformation is insensitive to sulfonamide geometry (Figure 11). The scaffold of CARB-KB45 is somewhat more sensitive to the sulfonamide geometry (Figure 12). The conformation of the R1 substituents of both CARB-AD37 and CARB-KB45 vary significantly among the docked poses, but no clear correlation is observed between the sulfonamide geometry and the agreement with the crystal structures (Figures 11 and 12).

Table 3: Measured inhibition constants (nM) of designed compounds CARB-AD37 and CARB-KB45, along with six clinical inhibitors, for wild-type HIV protease (WT) and three clinically relevant mutants (M1, M2, M3)

		K_i (nM)			
Compound	V_{out}	WT	M1	M2	M3
CARB-AD37	155	24	360	63	372
CARB-KB45	139	58	1300	130	2900
APV	128	0.1	0.15	0.21	1.4
NFV	166	0.28	15	3.5	19
LPV	170	0.005	6.1	0.040	0.90
IDV	180	0.18	34	0.73	21
SQV	213	0.07	90	1.0	78
RTV	256	0.06	3.0	0.46	2.8

V_{out} computed volume outside the substrate envelope (Å^3).
APV, amprenavir; NFV, nelfinavir; LPV, lopinavir; IDV, indinavir; SQV, saquinavir; RTV, ritonavir.

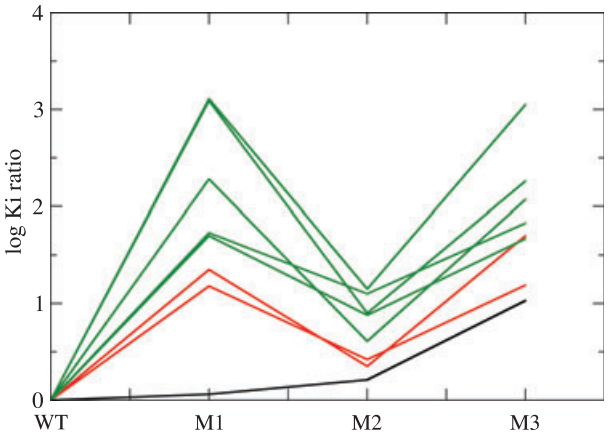


Figure 4: Susceptibilities of HIV protease inhibitors to mutation-sets M1, M2, and M3, compared with wild type (WT), where susceptibility is expressed as $\log\frac{K_i^{mut}}{K_i^{wt}}$. Each graph represents one inhibitor. Green: NFV, LPV, IDV, SQV, RTV. Black: amprenavir. Red: CARB-AD37, CARB-KB45.

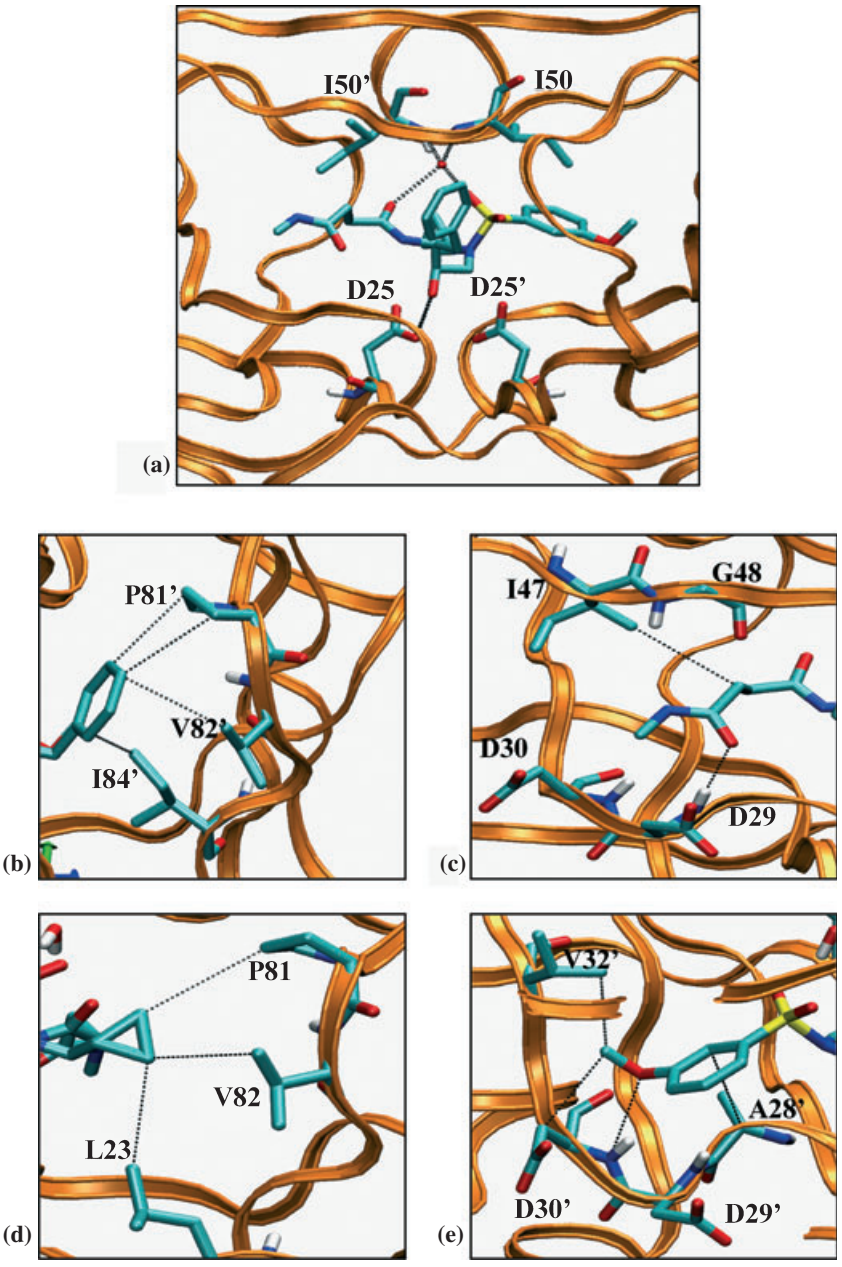


Figure 5: Conformation of CARB-AD37 docked into HIV protease from crystal structure 1HPV. (a) Overview of structure, highlighting ligand–protein interactions of the scaffold. (b–e) Interactions of phenyl, R1, R2, and R3 substituents, respectively. Dotted lines represent non-bonded interactions between ligand and atoms within 4.1 Å. Molecular graphics here and in other figures generated with the program VMD (30).

The cyclopropyl substituent at R2 of both ligands overlays well on the crystal conformation when the R sulfonamide geometry is used for docking (light and dark blue conformations in Figures 11 and 12), but not when the S form is used (red and orange conformations). The sensitivity of the cyclopropyl group's geometry to the sulfonamide's geometry is reasonable because it is linked directly to the nitrogen in question. The R3 substituents are also linked directly to the sulfonamide group and, like the cyclopropyl group, are best positioned when the R-conformation of the sulfonamide is used during docking (Figures 11 and 12). The errors in the predicted position of the R3 substituent of CARB-KB45 are larger than those for the R3 group of CARB-AD37. Perhaps the relative lack of strong interactions of the CARB-KB45 substituent at the S2′ subsite leaves it more sensitive to the geometry of the nitrogen.

Discussion

This paper presents the first application of the substrate envelope hypothesis in the design of HIVP inhibitors with broad specificity against clinically relevant mutants. Incorporation of fit to the substrate envelope as a design criterion led to two new HIVP inhibitors with relatively flat affinity profiles against a panel of clinically relevant mutants. Comparison with first-generation clinical HIVP inhibitors indicates that the new inhibitors are less susceptible to the mutations studied here, except in the case of APV. Interestingly, this exception is consistent with the hypothesis because APV proves to fit the substrate envelope slightly better than the new inhibitors. These results lend support to the hypothesis that compounds which fit the envelope will resist clinically relevant mutations.

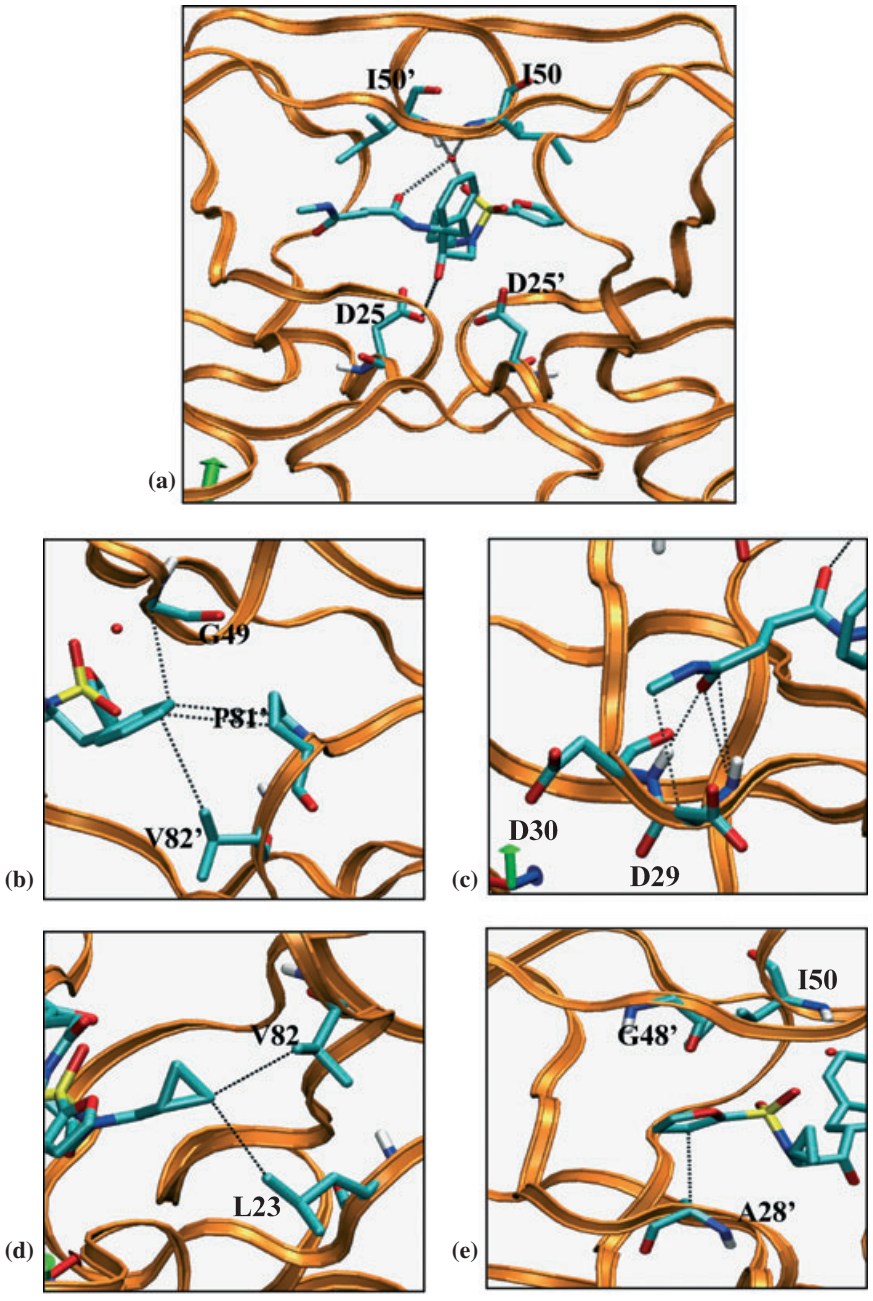


Figure 6: Conformation of CARB-KB45 docked into HIV protease from crystal structure 1HPV. (a) Overview of structure, highlighting ligand–protein interactions of the scaffold. (b–e) Interactions of phenyl, R1, R2, and R3 substituents, respectively. Dotted lines represent non-bonded interactions between ligand and atoms within 4.1 Å. Molecular graphics here and in other figures generated with the program VMD.

Crystallographic studies of compounds CARB-AD37 and CARB-KB45 show generally good agreement between the predicted and crystal structures. However, detailed analysis reveals inversion at the sulfonamide nitrogens relative to the conformation used in the design calculations. Redocking the designed compounds after correcting their geometry indicate that the R2 and R3 substituents would have been more accurately positioned had the correct geometry been known in advance. The variability of the sulfonamide's geometry is consistent with the observation of two different sulfonamide conformations for the chemically similar inhibitor darunavir bound in two different sites of HIV-1 protease (29). The darunavir in the classical binding site was found to exist as the R-entantiomer, consistent with the present crystal structures and not with that seen for APV in 1HPV.

The lowest K_i value achieved by the present compounds is 24 nM, significantly higher than the first-generation inhibitors. This presumably results in large part from the inaccuracy of the energy model, as well perhaps as from the specific pool of substituents to which the virtual library was limited. In addition, repeating the full design procedure with the correct sulfonamide conformation might lead to compounds of greater affinity. More generally, the present results indicate that there might be considerable value in a docking procedure that would automatically sample alternative geometries of

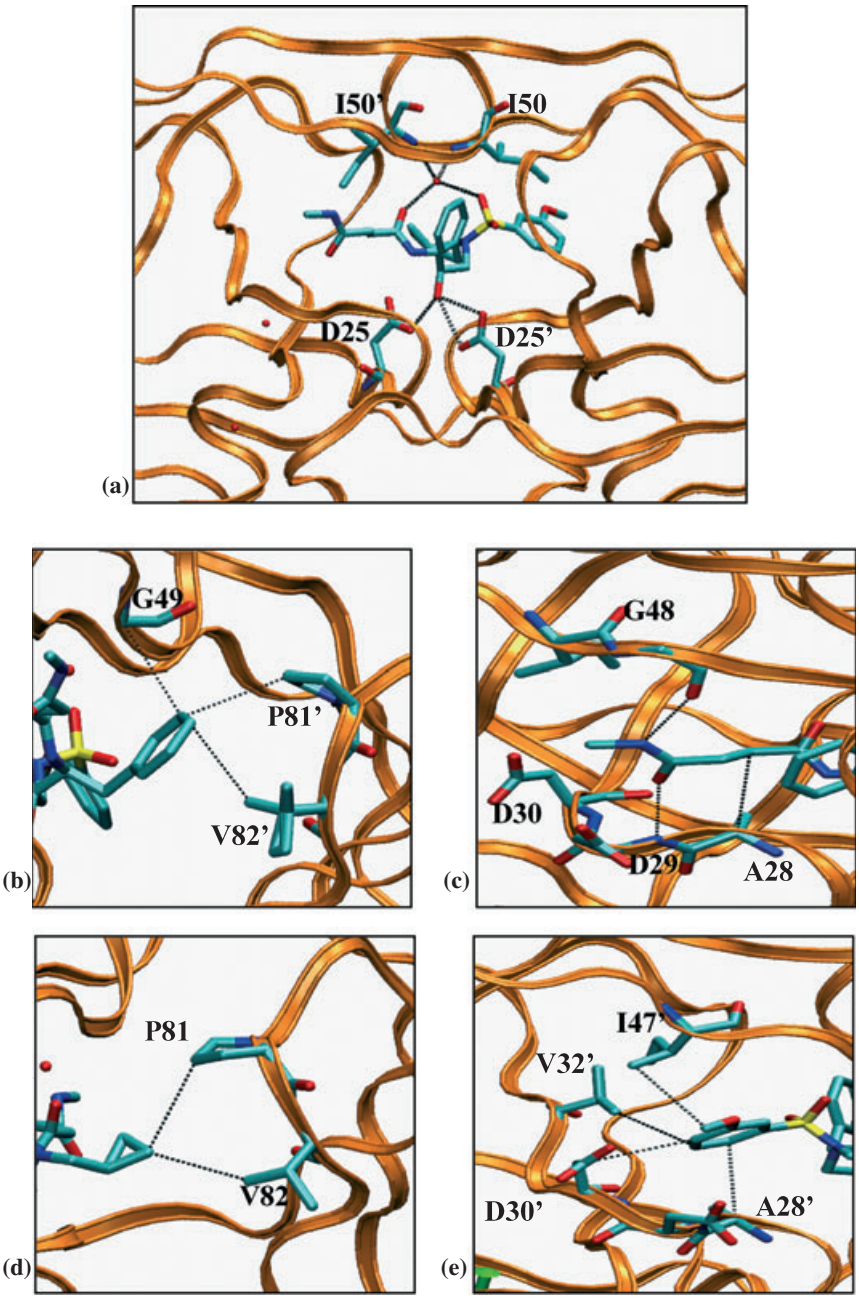


Figure 7: Crystal structure of the complex of CARB-AD37 with HIV protease. (a) Overview of structure, highlighting ligand–protein interactions of the scaffold. (b–e) Interactions of phenyl, R1, R2, and R3 substituents, respectively. Dotted lines represent non-bonded interactions between ligand and atoms within 4.1 Å. Molecular graphics here and in other figures generated with the program VMD.

invertible nitrogen atoms. Finally, it might be possible to discover compounds in this series with higher affinity and still reasonably good fits to the substrate envelope by adjusting the combined score (eqn 2) to place more weight on energy and less on V_{out}.

Overall, the present results are encouraging as they suggest that a simple shape-based criterion can help guide inhibitor design toward compounds that will better resist mutation. The same principle may also be applicable to other drug-targets that are known to be highly mutable. On the other hand, molecular recognition is a complex process involving subtle interactions and conformational adjustments. Therefore, the present broad-brush approach is probably best viewed as a helpful approximation whose applicability and limitations need to be delineated by further study.

Experimental Section

Chemistry

General

^{1}H- and ^{13}C-NMR spectra were recorded on a Varian 400 MHz NMR spectrometer (Palo Alto, CA, USA), operating at 400 MHz for

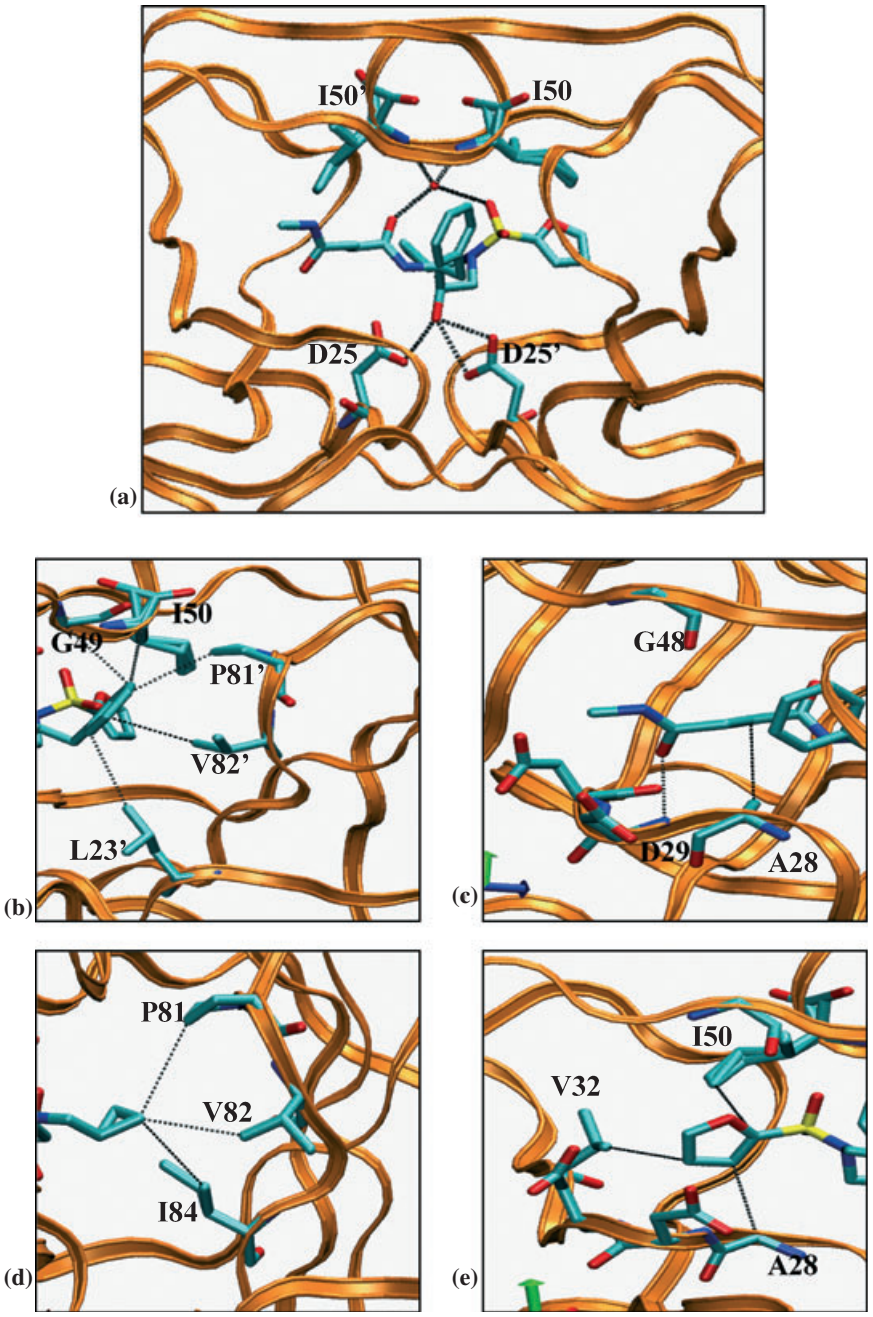


Figure 8: Crystal structure of the complex of CARB-KB45 with HIV protease. (a) Overview of structure, highlighting ligand–protein interactions of the scaffold. (b–e) Interactions of phenyl, R1, R2, and R3 substituents, respectively. Dotted lines represent non-bonded interactions between ligand and atoms within 4.1 Å. Molecular graphics here and in other figures generated with the program VMD.

^{1}H and 100 MHz for ^{13}C. Chemical shifts are reported in p.p.m. relative to the solvent signal, and coupling constant (J) values are reported in Hertz (Hz). Thin layer chromatography (TLC) was performed on E-Merck (Darmstadt, Germany) silica gel 60-F-254 plates and spots were visualized with UV light. Flash column chromatography was performed using 230–400 mesh silica gels (E-Merck). High-resolution mass spectra (HRMS) were recorded on Waters Q-TOF Premier mass spectrometer (Milford, MA, USA) by direct infusion of solutions of each compound using electrospray ionization (ESI) in positive mode. Low-resolution mass spectra were obtained using Waters Alliance HT/Micromass ZQ system (Milford, MA, USA) (ESI). Anhydrous dichloromethane, *N,N*-dimethylformamide (DMF), and toluene were purchased from Sigma-Aldrich (St. Louis, MO, USA) and used as such. All reagents and chemicals were purchased from commercial vendors and used as received.

General procedure for the ring opening of epoxide with amines

A solution of the chiral epoxide 1 (1-oxiranyl-2-phenylethyl)carbamic acid tert-butyl ester) (10 mmol) in EtOH or iPrOH (50 mL) was added to the amine (11 mmol) and the mixture was heated at 80 °C for 3 h. After cooling to room temperature, solvents were removed under reduced pressure. The residue was purified by recrystallization

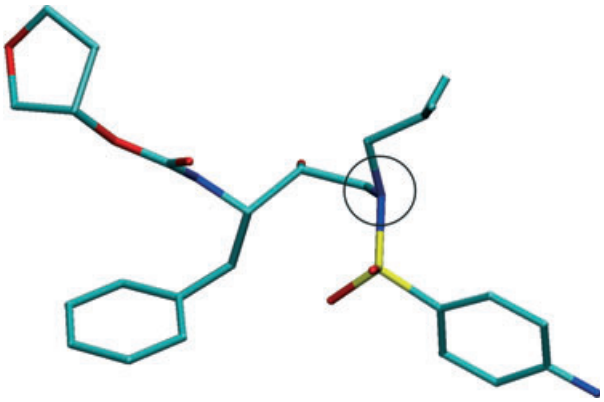


Figure 9: Structure of amprenavir from PDB entry 1HPV. Circle: sulfonamide nitrogen in S conformation.

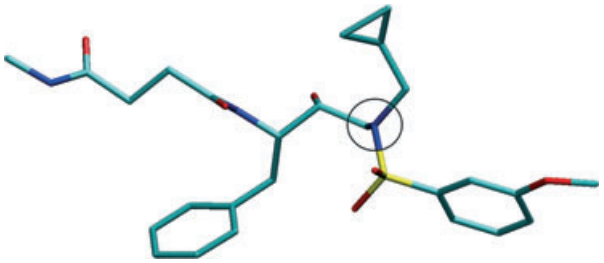


Figure 10: Structure of CARB-AD37 from present crystal structure. Circle: sulfonamide nitrogen in R conformation.

Table 4: Root-mean-square deviations, relative to current crystal structure, of conformations of CARB-AD37 docked into the listed crystal structures of HIVP with the sulfonamide in either the S (Figure 9) or R (Figure 10) conformation of the sulfonamide nitrogen

Crystal structure	Sulfonamide geometry	RMSD (Å)
1HPV	S	1.32
1HPV	R	0.90
KB60	S	0.73
KB60	R	0.65

Table 5: Root-mean-square deviations, relative to current crystal structure, of conformations of CARB-KB45 docked into the listed crystal structures of HIVP with the sulfonamide in either the S (Figure 9) or R (Figure 10) conformation of the sulfonamide nitrogen

Crystal structure	Sulfonamide geometry	RMSD (Å)
1HPV	S	1.08
1HPV	R	0.83
KB60	S	1.31
KB60	R	1.18

from ethyl acetate–hexanes mixture to provide the product as white solid in excellent yield.

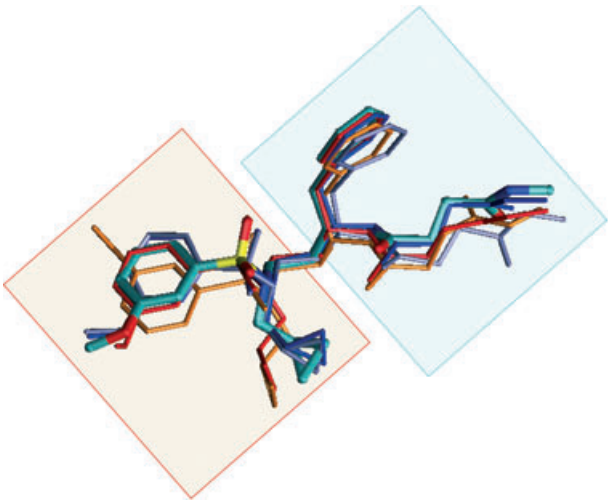


Figure 11: Superposed conformations of bound CARB-AD37. Thick tubes colored by element (C: cyan, N: blue, O: red, S: yellow): crystallographic conformation. Thin single-colored tubes: docked conformations. Orange and red: original sulfonamide geometry docked, respectively, into protease from crystal structures 1HPV and KB60. Pale blue and blue: inverted sulfonamide geometry, docked, respectively, into 1HPV and KB60. Blue box: R1 and phenyl substituents. Red box: R2 and R3 substituents.

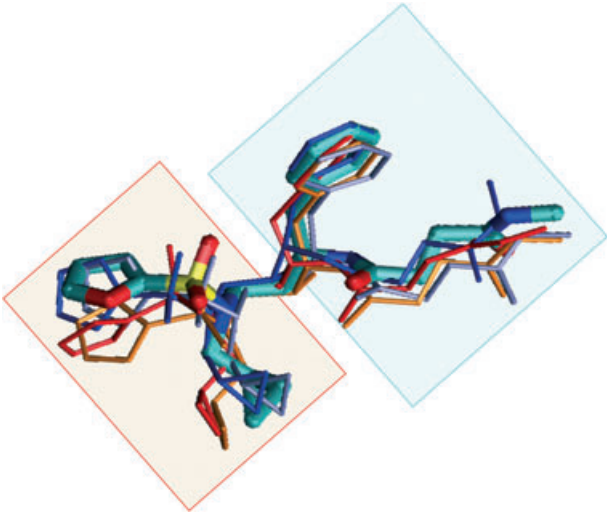


Figure 12: Superposed conformations of bound CARB-KB45. Thick tubes colored by element (C: cyan, N: blue, O: red, S: yellow): crystallographic conformation. Thin single-colored tubes: docked conformations. Orange and red: original sulfonamide geometry docked, respectively, into protease from crystal structures 1HPV and KB60. Pale blue and blue: inverted sulfonamide geometry, docked, respectively, into 1HPV and KB60. Blue box: R1 and phenyl substituents. Red box: R2 and R3 substituents.

General procedure for the synthesis of (R)-(hydroxyethylamino)sulfonamides

To an ice-cooled solution of the secondary amine 3 (5 mmol) in CH_2Cl_2 (20 mL) was added an aqueous solution of Na_2CO_3 (8 mmol in 5 mL H_2O) followed by the slow addition of sulfonyl chloride (5 mmol) solution in CH_2Cl_2 (5 mL). After 15 min the reaction mixture was warmed to ambient temperature and stirred till no starting material was detected by TLC. Reaction mixture was diluted with CH_2Cl_2 and layers were separated. Organic extract was washed with saturated aqueous NaCl solution, dried (Na_2SO_4), filtered and evaporated under reduced pressure. The residue was purified by flash chromatography on silica gel using mixture of ethyl acetate and hexanes as eluent to afford pure product as solid.

General procedure for the removal of Boc protection

To the solution of the sulfonamide 5 (0.5 mmol) in CH_2Cl_2 (15 mL) was added TFA (5 mL) and the mixture was stirred at room temperature for 1 h. Solvents were removed under reduced pressure, and the residue was dissolved in toluene (5 mL) and evaporated under reduced pressure to provide the amine that was used as such in the coupling step.

General procedure for the coupling reaction

To an ice-cooled solution of the carboxylic acid 7 (0.6 mmol) in DMF-CH_2Cl_2 mixture (1:1; 8 mL) were added solid EDCI (0.6 mmol) and HOBt (0.6 mmol) under dry N_2 atmosphere. After stirring 15 min, a solution of the deprotected amine 6 in DMF-CH_2Cl_2 mixture (1:1; 5 mL) was slowly added followed by the addition of DIPEA (1 mmol). The reaction mixture was warmed to room temperature and stirred until reaction was complete (monitored by TLC). Small amount of water and CH_2Cl_2 were added and layers were separated. The organic extract was washed with saturated aqueous NaCl solution, dried (Na_2SO_4), filtered and evaporated under reduced pressure. The residue was purified by flash chromatography on silica gel using mixture of ethyl acetate and hexanes (in some cases, methanol/chloroform mixture) as eluent to provide the target compound as solid.

N-[(1S,2R)-1-benzyl-2-hydroxy-3-[[(3-methoxyphenyl)sulfonyl]-[(R)-(tetrahydro-2-furanyl)methyl]amino]propyl]-2-methoxypyridine-4-carboxamide (CARB-AD08)

^{1}H-NMR (400 MHz, $CDCl_3$) 8.20 (d, J = 5.6 Hz, 1H), 7.31 (q, J = 8.0 Hz, 1H), 7.28–7.18 (m, 6H), 7.13 (t, J = 2.4 Hz, 1H), 7.07–7.03 (m, 2H), 6.96 (s, 1H), 6.37 (d, J = 9.2 Hz, 1H), 5.37 (d, J = 2.8 Hz, 1H), 4.34–4.21 (m, 2H), 3.96 (m, 1H), 3.93 (s, 3H), 3.77 (dt, J = 5.6, 1.2 Hz, 2H), 3.72 (s, 3H), 3.63–3.57 (m, 2H), 3.12 (dd, J = 14.0, 4.8 Hz, 1H), 3.06 (dd, J = 13.6, 6.8 Hz, 1H), 2.86 (d, J = 8.4 Hz, 1H), 2.82 (d, J = 8.4 Hz, 1H), 2.02 (m, 1H), 1.93–1.85 (m, 2H), 1.47 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) 165.54, 165.05, 160.26, 148.06, 144.71, 139.39, 137.42, 130.52, 129.99 (2C), 128.74 (2C), 126.89, 119.38, 119.16, 114.0, 112.43, 108.94, 80.64, 73.97, 68.39, 57.18, 56.64, 55.77, 54.02, 53.04, 35.70, 29.12, 25.50. HRMS (ESI) *m/z*: $C_{29}H_{36}N_3O_7S$ $(M + H)^+$; calcd, 570.2274; found, 570.2191.

N1-[(1S,2R)-1-benzyl-3-[(cyclopropylmethyl)[(3-methoxyphenyl)-sulfonyl]amino]-2-hydroxypropyl]-N4-methylsuccinamide (CARB-AD37)

^{1}H-NMR (400 MHz, $CDCl_3$) 7.43–7.35 (m, 2H), 7.31 (t, J = 2.0 Hz, 1H), 7.28–7.17 (m, 5H), 7.09 (m, 1H), 6.48 (d, J = 6.0 Hz, 1H), 5.85 (br. s, 1H), 4.19 (m, 1H), 4.16 (s, 1H), 3.98 (m, 1H), 3.85 (s, 3H), 3.37 (dd, J = 14.8, 4.4 Hz, 1H), 3.20 (dd, J = 14.8, 7.6 Hz, 1H), 3.14 (dd, J = 14.4, 6.4 Hz, 1H), 3.08–2.99 (m, 2H), 2.95–2.87 (m, 1H), 2.71 (d, J = 4.8 Hz, 3H), 2.43–2.26 (m, 4H), 0.85 (m, 1H), 0.48 (m, 2H), 0.16 (m, 2H); ^{13}C-NMR (100 MHz, $CDCl_3$) 173.28, 172.81, 160.17, 140.53, 138.22, 130.47, 129.53 (2C), 128.69 (2C), 126.70, 119.55, 118.98, 112.49, 72.46, 55.93, 54.80, 54.51, 51.64, 35.23, 31.93, 31.66, 26.56, 9.99, 4.74, 4.35. HRMS (ESI) *m/z*: $C_{26}H_{36}N_3O_6S$ $(M + H)^+$; calcd, 518.2325; found, 518.2269.

N1-[(1S,2R)-1-benzyl-2-hydroxy-3-[(2-hydroxypropyl)[(3-methoxyphenyl)-sulfonyl]amino]propyl]-N4-methylsuccinamide (CARB-KB32)

^{1}H-NMR (400 MHz, $CDCl_3$) 7.42 (t, J = 7.6 Hz, 1H), 7.34 (d, J = 7.6 Hz, 1H), 7.28 (t, J = 2.4 Hz, 1H), 7.27–7.16 (m, 5H), 7.11 (m, 1H), 6.59 (d, J = 8.8 Hz, 1H), 6.09 (d, J = 4.8 Hz, 1H), 4.22–4.15 (m, 2H), 4.08 (m, 1H), 3.84 (s, 3H), 3.53 (dd, J = 14.8, 3.2 Hz, 1H), 3.37 (dd, J = 14.4, 2.0 Hz, 1H), 3.03 (dd, J = 14.0, 4.4 Hz, 1H), 2.91 (dd, J = 15.2, 8.8 Hz, 1H), 2.84–2.76 (m, 2H), 2.70 (d, J = 4.4 Hz, 3H), 2.46–2.40 (m, 2H), 2.28 (m, 2H), 1.14 (d, J = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$) 173.27, 172.88, 160.25, 139.51, 138.15, 130.59, 129.57 (2C), 128.61 (2C), 126.64, 119.58, 119.09, 112.70, 74.02, 68.33, 59.66, 55.94, 55.69, 53.61, 35.22, 31.81, 31.63, 26.62, 20.64. HRMS (ESI) *m/z*: $C_{25}H_{36}N_3O_7S$ $(M + H)^+$; calcd, 522.2274; found, 522.2227.

N1-[(1S,2R)-1-benzyl-3-[(cyclopropylmethyl)[(furan-2-sulfonyl]amino]-2-hydroxypropyl]-N4-methylsuccinamide (CARB-KB45)

^{1}H-NMR (400 MHz, $CDCl_3$) 7.52 (q, J = 0.8 Hz, 1H), 7.28–7.15 (m, 5H), 7.01 (dd, J = 3.6, 0.8 Hz, 1H), 6.60 (d, J = 8.0 Hz, 1H), 6.48 (dd, J = 3.6, 2.0 Hz, 1H), 6.03 (d, J = 4.4 Hz, 1H), 4.20–4.13 (m, 2H), 3.97 (m, 1H), 3.52 (dd, J = 15.2, 4.4 Hz, 1H), 3.31 (dd, J = 14.8, 8.0 Hz, 1H), 3.25–3.12 (m, 2H), 3.02 (dd, J = 14.4, 4.8 Hz, 1H), 2.94–2.84 (m, 1H), 2.71 (d, J = 5.2 Hz, 3H), 2.45–2.26 (m, 4H), 0.88 (m, 1H), 0.48 (m, 2H), 0.17 (m, 2H); ^{13}C-NMR (100 MHz, $CDCl_3$) 173.34, 172.94, 148.64, 146.28, 138.21, 129.52 (2C), 128.67 (2C), 126.69, 116.67, 111.40, 72.57, 54.60 (2C), 51.28, 35.12, 31.91, 31.66, 26.56, 9.71, 4.54, 4.16. HRMS (ESI) *m/z*: $C_{23}H_{32}N_3O_6S$ $(M + H)^+$; calcd, 478.2012; found, 478.1995.

N1-[(1S,2R)-1-benzyl-2-hydroxy-3-[(2-hydroxypropyl)[(furan-2-sulfonyl]amino]-propyl]-N4-methylsuccinamide (CARB-KB49)

^{1}H-NMR (400 MHz, $CDCl_3$) 7.52 (q, J = 0.8 Hz, 1H), 7.24–7.12 (m, 5H), 6.99 (m, 2H), 6.79 (br. s, 1H), 6.48 (q, J = 1.6 Hz, 1H), 4.12 (m, 2H), 4.0 (m, 1H), 3.63 (dd, J = 14.8, 3.6 Hz, 1H), 3.52 (dd,

J = 14.8, 2.0 Hz, 1H), 2.98 (dd, J = 14.4, 4.8 Hz, 1H), 2.87 (dd, J = 15.2, 9.2 Hz, 1H), 2.75–2.69 (m, 2H), 2.67 (d, J = 4.4 Hz, 3H), 2.42–2.31 (m, 4H), 2.30–2.20 (m, 2H), 1.10 (d, J = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$ + 1 drop CD3OD) 173.54, 173.01, 147.62, 146.67, 138.15, 129.46 (2C), 128.52 (2C), 126.55, 117.23, 111.33, 73.69, 68.36, 59.91, 55.54, 53.38, 34.94, 31.77, 31.62, 26.35, 20.38. HRMS (ESI) *m/z*: $C_{22}H_{32}N_3O_7S$ $(M + H)^+$; calcd, 482.1961; found, 482.1921.

N-[(1S,2R)-1-benzyl-2-hydroxy-3-[(2-hydroxypropyl)[(furan-2-sulfonyl]amino]propyl]-2-[(S)-cyclopent-2-enyl]acetamide (CARB-KB51)

^{1}H-NMR (400 MHz, $CDCl_3$) 7.54 (q, J = 0.8 Hz, 1H), 7.24–7.12 (m, 5H), 6.94 (m, 2H), 6.62 (d, J = 6.4 Hz, 1H), 6.44 (m, 1H), 5.75 (m, 1H), 5.67 (m, 1H), 4.27–4.16 (m, 3H), 4.40 (br. s, 1H), 3.75 (dd, J = 15.2, 2.4 Hz, 1H), 3.55 (dd, J = 14.8, 2.0 Hz, 1H), 3.12 (dd, J = 14.4, 4.8 Hz, 1H), 3.05–2.97 (m, 2H), 2.82 (dd, J = 14.8, 8.8 Hz, 1H), 2.38–2.10 (m, 5H), 1.13 (d, J = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$) 173.23, 147.86, 146.34, 138.13, 134.29 (134.23), 131.44 (131.33), 129.87 (2C), 128.55 (2C), 126.54, 117.03, 110.97, 72.76, 68.32, 56.24, 55.27, 52.39, 43.21 (43.12), 42.73 (42.67), 34.28, 32.71 (32.64), 32.04, 29.77 (29.72). HRMS (ESI) *m/z*: $C_{24}H_{33}N_2O_6S$ $(M + H)^+$; calcd, 477.2059; found, 477.2032.

N-[(1S,2R)-1-benzyl-2-hydroxy-3-[(1-methyl-1H-imidazole-4-sulfonyl)[(R)- (tetrahydro-2-furanyl)methyl]amino]propyl]-2-methoxypyridine-4-carboxamide (CARB-AC97)

^{1}H-NMR (400 MHz, $CDCl_3$) 8.18 (dd, J = 4.8, 0.8 Hz, 1H), 7.35 (d, J = 1.6 Hz, 1H), 7.27–7.16 (m, 5H), 7.05 (dd, J = 5.2, 1.6 Hz, 1H), 6.96 (q, J = 0.8 Hz, 1H), 6.60 (d, J = 8.8 Hz, 1H), 5.80 (d, J = 2.4 Hz, 1H), 4.31 (m, 1H), 4.23 (m, 1H), 3.96 (t, J = 8.0 Hz, 1H), 3.92 (s, 3H), 3.78–3.66 (m, 4H), 3.65 (s, 3H), 3.16 (dd, J = 15.2, 8.4 Hz, 1H), 3.10 (dd, J = 14.0, 4.8 Hz, 1H), 3.03 (dd, J = 13.6, 7.2 Hz, 1H), 2.91 (dd, J = 15.2, 8.8 Hz, 1H), 2.0 (m, 1H), 1.90–1.83 (m, 3H), 1.53–1.45 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) 165.44, 164.94, 147.93, 145.09, 139.53, 139.23, 137.73, 130.03 (2C), 128.57 (2C), 126.67, 124.57, 114.24, 108.93, 80.64, 72.42, 68.30, 56.37, 55.29, 54.01, 53.07, 35.41, 34.20, 29.12, 25.55. HRMS (ESI) *m/z*: $C_{26}H_{34}N_5O_6S$ $(M + H)^+$; calcd, 544.2230; found, 544.2197.

Acknowledgments

We thank Ellen A. Nalivaika for providing protease specimens. Some protease inhibitors were obtained through the AIDS Research and Reference Reagent Program, NIAID, NIH. We also thank Kaneka, USA for the generous gifts of chiral epoxides. This publication was made possible by Grant GM066524 from the National Institute of General Medical Sciences of the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of General Medical Sciences.

References

1. Anonymous (2001) HIV and AIDS-United States, 1981–2000. CDC: MMWR Morb Mortal Wkly Rep;50:430–434.
2. Ohtaka H., Freire E. (2005) Adaptive inhibitors of the HIV-1 protease. Prog Biophys Mol Biol;88:193–208.
3. Prabu-Jeyabalan M., Nalivaika E.A., Schiffer C.A. (2000) How does a symmetric dimer recognize an asymmetric substrate. J Mol Biol;301:1207–1220.
4. Prabu-Jeyabalan M., Nalivaika E.A., King N.M., Schiffer C.A. (2003) Viability of a drug resistant HIV-1 protease variant: structural insights for better anti-viral therapy. J Virol;77:1306–1315.
5. King N.M., Prabu-Jeyabalan M., Nalivaika E.A., Schiffer C.A. (2004) Combating susceptibility to drug resistance: lesions from HIV-1 protease. Chem & Biol;11:1333–1338.
6. Chellappan S., Kairys V., Fernandes M.X., Schiffer C., Gilson M.K. (2007) Evaluation of the substrate envelope hypothesis for inhibitors of HIV-1 protease. Prot Struct Funct Bioinf Published Online May 1, 2007. doi: 10.1002/Proof.21431.
7. Irwin J.J., Shoichet B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. J Chem Inf Model;45:177–182.
8. Leach A.R. (2001) Molecular Modeling: Principles and Applications, 2nd edn. England: Prentice-Hall.
9. Kim E.E., Baker C.T., Dwyer M.D., Murcko M.A., Rao B.G., Tung R.D., Navia M.A. (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. J Am Chem Soc;117:1181.
10. David L., Luo R., Gilson M.K. (2001) Ligand-receptor docking with the Mining Minima optimizer. J Comput Aided Mol Des;15: 157–171.
11. Kairys V., Gilson M.K. (2002) Enhanced docking with the mining minima optimizer: acceleration and side-chain flexibility. J Comput Chem;23:1656–1670.
12. Gilson M.K., Gilson H.S.R., Potter M.J. (2003) Fast assignment of accurate partial atomic charges: an electronegativity equalization method that accounts for alternate resonance forms. J Chem Inf Comput Sci;43:1982–1997.
13. Polar Hydrogen Parameter Set for CHARMM. Waltham, MA: Molecular Simulations Inc.
14. Mayo S.L., Olafson B.D., Goddard W.A. III (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem;94:8897–8909.
15. Ali A., Reddy G.S.K.K., Cao H., Anjum S.G., Nalam M.N.L., Schiffer C.A., Rana T.M. (2006) Discovery of HIV-1 protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel P2 ligands. J Med Chem;49:7342–7356.
16. King N.M., Prabu-Jeyabalan M., Nalivaika E.A., Wigerinck P., de Bethune M.-P., Schiffer C.A. (2004) Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. J Virol;78:12012–12021.
17. Shafer R.W., Stevenson D., Chan B. (1999) Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res;27:348–352.
18. King N.M., Melnick L., Prabu-Jeyabalan M., Nalivaika E.A., Yang S.S., Gao Y., Nie X., Zepp C., Heefner D.L., Schiffer C.A. (2002)

Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. Prot Sci;11:418–429.
19. Minor W. (1993) XDISPLAYF. West Lafayette, IN: Purdue University.
20. Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol;276: 307–326.
21. Collaborative Computational Project, Numbery (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst D;50:760–763.
22. Navaza J. (1994) AMoRe: an automated package for molecular replacement. Acta Cryst A;50:157–163.
23. Morris R.J., Perrakis A., Lamzin V.S. (2002) ARP/wARP's model-building algorithms: I. The main chain. Acta Cryst D;58:968–975.
24. Jones T.A., Bergdoll M., Kjeldgaard M. (1990) O: a macromolecular modeling environment. In: Bugg C., Ealich S., editor. Crystallographic and Modeling Methods in Molecular Design. Berlin, Germany: Springer-Verlag Press;p. 189–195.
25. Murshudov G.N., Vagin A.A., Dodson E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst D;53:240–255.
26. Kuriyan J., Weis W.I. (1991) Rigid protein motion as a model for crystallographic temperature factors. Proc Natl Acad Sci USA; 88:2773–2777.
27. Schomaker V., Trueblood K.N. (1968) On the rigid-body motion of molecules in crystals. Acta Cryst B;24:63–76.
28. Tickle I.J., Moss D.S. (1999) Modelling Rigid-body Thermal Motion in Macromolecular Crystal Structure Refinement, IUCr99 Computing School. Available at: http://people.cryst.bbk.ac.uk/~tickle/iucr99/iucrcs99.html
29. Kovalevsky A.Y., Liu F., Leshchenko S., Ghosh A., Louis J.M., Harrison R.W., Weber I.T. (2006) Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. J Mol Biol;363:161–173.
30. Humphrey W., Dalke A., Schulten K. (1996) VMD – visual molecular dynamics. J Mol Graph;14:33–38.