

# Testing the Substrate-Envelope Hypothesis with Designed Pairs of Compounds

Yang Shen,<sup>†,○,||,#</sup> Michael D. Altman,<sup>‡,⊥,#</sup> Akbar Ali,<sup>§,||,#</sup> Madhavi N. L. Nalam,<sup>§</sup> Hong Cao,<sup>§,||</sup>  
Tariq M. Rana,<sup>\*,||,∇</sup> Celia A. Schiffer,<sup>\*,§</sup> and Bruce Tidor<sup>\*,†,○</sup>

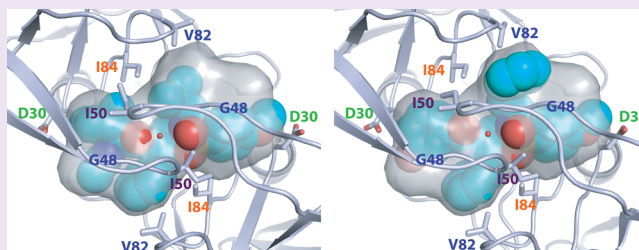
<sup>†</sup>Department of Biological Engineering, <sup>‡</sup>Department of Chemistry, and <sup>○</sup>Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

<sup>§</sup>Department of Biochemistry and Molecular Pharmacology and <sup>||</sup>Chemical Biology Program, University of Massachusetts Medical School, Worcester, Massachusetts 01605, United States

<sup>∇</sup>Sanford-Burnham Medical Research Institute, La Jolla, California 92037, United States

## S Supporting Information

**ABSTRACT:** Acquired resistance to therapeutic agents is a significant barrier to the development of clinically effective treatments for diseases in which evolution occurs on clinical time scales, frequently arising from target mutations. We previously reported a general strategy to design effective inhibitors for rapidly mutating enzyme targets, which we demonstrated for HIV-1 protease inhibition [Altman et al. *J. Am. Chem. Soc.* 2008, 130, 6099–6113]. Specifically, we developed a computational inverse design procedure with the added constraint that designed inhibitors bind entirely inside the substrate envelope, a consensus volume occupied by natural substrates. The rationale for the substrate-envelope constraint is that it prevents designed inhibitors from making interactions beyond those required by substrates and thus limits the availability of mutations tolerated by substrates but not by designed inhibitors. The strategy resulted in subnanomolar inhibitors that bind robustly across a clinically derived panel of drug-resistant variants. To further test the substrate-envelope hypothesis, here we have designed, synthesized, and assayed derivatives of our original compounds that are larger and extend outside the substrate envelope. Our designs resulted in pairs of compounds that are very similar to one another, but one respects and one violates the substrate envelope. The envelope-respecting inhibitor demonstrates robust binding across a panel of drug-resistant protease variants, whereas the envelope-violating one binds tightly to wild type but loses affinity to at least one variant. This study provides strong support for the substrate-envelope hypothesis as a design strategy for inhibitors that reduce susceptibility to resistance mutations.



Many chemical and biological agents have been developed to deliver quality-of-life improvements, especially in health care and agriculture. Advances include herbicides and pesticides, as well as medical tests, treatments, and interventions. However, therapeutic and environmental agents impose selective pressure that can lead to acquired resistance. Although resistance can be attained through many mechanisms,<sup>1–6</sup> for infectious diseases and cancer, in particular, mutation of the direct target is often the source of resistance and is associated with accelerated mutation rates. Such resistance limits the effectiveness of therapies that are time-consuming and expensive to develop and approve through regulatory processes. Thus, it is crucial to devise general strategies to incorporate into standard drug discovery paradigms that lead to elimination of resistance or at least to a large reduction in its incidence. Often the targets are enzymes essential to disease maintenance or progression. For such cases, the development of inhibitors that are as similar as possible to the disease-related substrates has been hypothesized as an effective strategy for reducing resistance.<sup>7</sup> The notion behind

such substrate mimicry is that enzyme mutations that reduce inhibitor binding would then, by similarity arguments, be very likely to also reduce substrate binding and possibly also turnover, and thus not be well tolerated. Inhibitors that are similar to substrates thus restrict resistance by introducing significantly higher genetic barriers that must be surmounted in their presence.

HIV-1 protease has proven an extremely useful case for studying the development and avoidance of resistance. HIV-1 protease was the first therapeutic target for which structure-based drug design was broadly applied.<sup>8</sup> Ten HIV-1 protease inhibitors have been approved by the U.S. FDA since the 1990s. All but tipranavir<sup>9</sup> are peptidomimetics that are similar to transition-state analogues of substrates, including saquinavir,<sup>10</sup> ritonavir,<sup>11</sup> indinavir,<sup>12</sup> nelfinavir,<sup>13</sup> amprenavir<sup>14</sup>/fosam-

Received: June 26, 2013

Accepted: August 16, 2013

Published: August 16, 2013

Table 1. Parent Compounds (AD-93 and KB-83) and Designed New Compounds

Compound Name	Chemical Structure	Compound Name	Chemical Structure
AD-93 <sup>25</sup>		KB-83 <sup>25</sup>	
AG-23 (14)		AF-68 (18)	
AF-72 (15)		AF-77 (19)	
AF-69 (16)		AF-78 (20)	
AF-71 (17)		AF-80 (21)	

prenavir,<sup>15</sup> lopinavir,<sup>16</sup> atazanavir,<sup>17</sup> and darunavir.<sup>18</sup> These drugs allow longer life expectancy for AIDS patients but still fall short of a cure. Therapeutic failure is due to rapid evolution of viral strains and accumulation of resistance-endowing mutations under selective drug pressure. In an effort to understand how drug-resistant HIV-1 protease maintains the ability to recognize and cleave its substrates, the Schiffer group compared the crystal structures of an inactive variant of HIV-1 protease with six peptides that correspond to the natural substrate cleavage sites.<sup>19</sup> It was discovered that the protease recognizes the substrates of diverse sequences through an asymmetric shape commonly adopted by the substrates, whose surface has been referred to as “the substrate envelope”. The same group further showed that many inhibitors protrude beyond the substrate envelope and make contacts with enzyme residues not contacted by substrates and that are the sites of primary resistance mutations.<sup>7</sup> It was thus hypothesized that “an inhibitor contained within the substrate envelope, interacting only with the same residues that are necessary to recognize substrate, may be less susceptible to drug resistance.”<sup>7</sup> This so-

called “substrate-envelope hypothesis” was assessed in a computational framework for inhibitors of HIV-1 protease<sup>20,21</sup> as well as Abl kinase, chitinase, thymidylate synthase, DHFR, and neuraminidase.<sup>22</sup> It was shown that the volume of an inhibitor molecule that protrudes outside the substrate envelope correlates with average mutation sensitivity. Recent crystallography data and analyses suggested that the substrate-envelope hypothesis also applies to the inhibitors of the NS3/4A protease, a target for hepatitis C virus (HCV) infection.<sup>23,24</sup>

Implementing a general strategy to design robustly effective inhibitors for mutable enzyme targets, we previously performed computational inverse design of HIV-1 protease inhibitors that reside inside the substrate envelope.<sup>25</sup> Promising candidates were synthesized, assayed, and characterized, which led to the identification of subnanomolar inhibitors to drug-resistant variants. Thirty-six compounds (comprising the MIT-2 library) were potent inhibitors of wild-type HIV-1 protease, with  $K_i$  values ranging from 14 pM to 4 nM. Ten of the most potent inhibitors were further assayed in antiviral activity against a panel of four drug-resistant HIV-1 protease

Table 2. Measured Binding Affinity Values against Wild-Type and Mutant Panel

	wild type <sup>a</sup>	M1 (L10I, <b>G48V</b> , <b>I54V</b> , <b>V82A</b> ) <sup>a</sup>	M2 (D30N, <b>N88D</b> ) <sup>a</sup>	M3 (L10I, A71V, G73S, <b>I84V</b> , <b>L90M</b> ) <sup>a</sup>	M4 ( <b>I50V</b> , <b>A71V</b> ) <sup>a</sup>	worst-fold affinity loss <sup>b</sup>
AD-93	0.046 ± 0.023	0.724 ± 0.377	0.905 ± 0.124	1.713 ± 1.042	<b>1.920 ± 0.834</b>	41
AG-23	0.021 ± 0.009	<b>0.376 ± 0.168</b>	0.028 ± 0.025	0.048 ± 0.022	0.092 ± 0.025	18
AF-72	0.129 ± 0.028	1.328 ± 0.851	0.383 ± 0.097	2.753 ± 0.609	<b>4.823 ± 3.116</b>	37
AF-69	0.070 ± 0.005	<b>19.18 ± 3.84</b>	2.004 ± 0.298	9.965 ± 0.530	2.265 ± 0.643	274
AF-71	3.329 ± 0.663	18.20 ± 8.12	2.951 ± 0.128	<b>44.82 ± 22.42</b>	30.04 ± 3.79	13
KB-83	0.194 ± 0.007	1.560 ± 0.124	0.283 ± 0.097	<b>2.021 ± 0.033</b>	1.101 ± 0.241	10
AF-68	0.146 ± 0.071	0.819 ± 0.255	0.049 ± 0.034	<b>1.029 ± 0.134</b>	0.144 ± 0.115	7
AF-77	0.546 ± 0.028	2.045 ± 0.502	0.490 ± 0.100	2.469 ± 0.860	<b>3.404 ± 0.593</b>	6
AF-78	0.437 ± 0.132	<b>42.93 ± 0.559</b>	2.543 ± 0.462	23.80 ± 4.17	5.887 ± 2.755	98
AF-80	2.493 ± 0.251	<b>28.13 ± 1.37</b>	1.261 ± 0.176	10.53 ± 0.16	13.53 ± 4.10	11

<sup>a</sup>The measured  $K_i$  of each inhibitor to each enzyme is given as the mean ± standard deviation, in units of nM. Measurements are reported for wild-type HIV-1 protease as well as for four drug-resistant variants (M1, M2, M3, and M4) for which the specific mutations are given in parentheses, with the key resistance mutations in bold face (relative to the PDB ID 1T3R wild type). <sup>b</sup>The worst-fold loss is given as the weakest  $K_i$  to one of the drug-resistant variants (bold face) divided by the  $K_i$  to wild type for each inhibitor.

variants derived from clinically isolated, drug-resistant strains. Two compounds remained potent against the panel with less than 15-fold affinity loss, and four had moderate shifts of less than 80-fold loss.<sup>25</sup>

In this study, we further investigate the implications of the substrate-envelope hypothesis in designing robust inhibitors that are of low susceptibility to drug-resistance mutations. The scaffold forming the basis of the MIT-2 library was adopted from amprenavir and darunavir, the latter of which has an especially flat resistance profile. Thus, we sought to investigate whether the relatively high incidence of these compounds with flat binding profiles in the MIT-2 library was due to their being designed to bind within the substrate envelope or was instead a result of their being based on an advantageous scaffold. Here, we computationally designed two series of compounds that progressively protrude outward toward and through the substrate envelope but are otherwise identical, in order to isolate the effect of the substrate-envelope constraint and study its impact on binding specificity profiles; all compounds in the current study share the amprenavir/darunavir scaffold. Eight compounds were designed based on two MIT-2 parent compounds, with protrusions beyond various subpockets of the binding site (including two levels of protrusions at the P1' site). Two compounds that were predicted and shown experimentally to violate significantly the substrate envelope were shown to bind tightly to wild type but poorly to members of the mutant panel; in contrast, the parent compounds respect the substrate envelope and are flat binders against the panel, with less than 50-fold affinity loss. Four newly designed compounds that essentially respect the substrate envelope were also flat binders. Taken together these results indicate that the six broadly binding compounds along with earlier designs<sup>25</sup> benefit from residing fully within the substrate envelope and thus decreasing the chance of losing interactions with a drug-resistant variant, rather than benefitting from an advantaged scaffold. Moreover, the substrate-envelope violators lost the most binding affinity to protease variants with mutations nearby the site of the violation. Thus, these results greatly strengthen the notion that the substrate-envelope hypothesis provides a valuable framework for developing enzyme inhibitors that are robust against both wild type and resistant variants.

## RESULTS AND DISCUSSION

Drug-resistant mutations in HIV-1 protease occur in various positions and combinations. Principles for the design of therapeutics that avoid resistance mutations for HIV-1 protease and other targets would be valuable. The substrate-envelope hypothesis is a general principle for avoiding resistant variants for enzyme targets. Here we further tested the substrate-envelope hypothesis by designing paired inhibitors of HIV-1 protease that differ only in whether they respect the substrate envelope, chemically synthesizing them, and performing inhibitory and antiviral activity assessments.

**Computational Design of Inhibitors.** Computational molecular design was applied to target the HIV-1 protease active site using a design library consisting of a scaffold with three variable positions. The goal was to probe changes in binding specificity as increasing functional group size approached and eventually pierced the substrate envelope. Two parent compounds reported previously were selected for this study, MIT-2-AD-93 and MIT-2-KB-83, chosen for their high affinities and relatively flat binding profiles.<sup>25</sup> Designs were carried out in the wild-type active site, using larger substituents that are allowed to pierce the substrate envelope. From the sets of predicted high-affinity designs, we selected eight molecules for synthesis and testing that had functional groups larger than those in the corresponding parent compound and thus probed the substrate envelope with the potential to extend beyond it (Table 1).

The selected compounds involve the same single substitutions to both parent compounds, with an *N*-acetyl-isoleucine substitution at R1, a smaller cyclohexylmethyl and a larger 3-phenylpropyl hydrophobic substitution at R2, and 5-(isoxazol-5-yl)-2-thiophene at R3. The numbers of non-hydrogen atoms in the substitutions (original function group) were 9 (7), 7 (4 or 5), 9 (4 or 5), and 10 (8 or 9), respectively, which corresponds to a net increase of 1–5 non-hydrogen atoms in each substitution.

**Inhibitor Synthesis and Activity.** The compounds were synthesized (see Methods for synthetic methods and compound characterization). They were assayed for inhibitory activity against wild type and a panel of drug-resistant protease variants using modified high-throughput enzyme inhibition assays compared to the original characterization of the parent compounds.<sup>25,26</sup> The mutant panel included two multi-drug-resistant variants M1 (L10I, **G48V**, **I54V**, **V82A**) and M3

(L10I, A71V, G73S, I84V, L90M), a signature variant of nelfinavir resistance M2 (D30N, N88D), and a signature variant of amprenavir and darunavir resistance M4 (I50V, A71V). (In each case the boldface mutations are the primary drug-resistant ones.) The inhibition data for the parent and variant compounds against the panel are given in Table 2. Because all inhibitors appear to act competitively, we interpret the measured inhibition constants ( $K_i$ ) as inhibitor affinities (disassociation constants  $K_d$ ).

The parent compounds are potent, subnanomolar binders to wild-type protease and show flat resistance profiles across the panel of resistant variants. The binding profile is characterized using two metrics, the worst-fold affinity loss and the weakest binding affinity. The worst-fold affinity loss is the ratio between the weakest binding affinity ( $K_i$ ) of a compound for a protease variant and its affinity for wild-type protease. Specifically, AD-93 (KB-83) had a wild-type affinity of 0.046 nM (0.19 nM) and 41-fold (10-fold) loss against the mutant panel, where the weakest binding affinity was 1.92 nM (2.02 nM) measured with the M4 (M3) variant. Several observations were made for the new compounds. (1) Relative to the parent compounds, six of the eight new compounds had comparable or improved binding affinity with wild-type protease. The two exceptions, AF-71 and AF-80, both involved the same substitution at R3, remained nanomolar binders to wild-type protease, but suffered 72-fold and 13-fold loss, respectively, in affinity compared to their respective parent compounds. (2) The binding specificity profiles of the new compounds fall into two classes: six that remained potent or had a moderate shift against the mutant panel (<50-fold affinity loss) and two that were more susceptible to resistance mutations (>80-fold affinity loss). The latter two, AF-69 and AF-78, both involve the large 3-phenylpropyl hydrophobic substitution at R2 and in the crystal structures showed a large expected violation of the substrate envelope (see Table 3). Both suffered their largest affinity loss

**Table 3. Measured Substrate Envelope Violations (in Å<sup>3</sup>) and Their Decompositions**

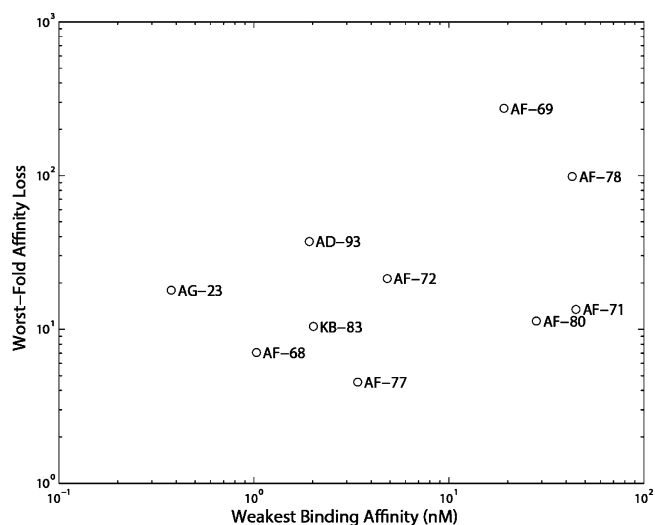
compound	total	R1 (near P2)	R2 (near P1')	R3 (near P2')
AD-93	7.8	0.5	0.0	5.1
AG-23	11.0	2.1	0.8	4.8
AF-72	13.6	0.3	6.2	5.5
AF-69-A/B <sup>a</sup>	14.5/20.0	1.1/0.1	6.0/13.4	5.8/5.7
AF-71	19.9	0.3	0.0	17.4
KB-83-A/B <sup>a</sup>	7.4/10.1	0.4/3.0	0.0	5.2
AF-68	10.9	2.3	0.0	6.3
AF-77	15.2	0.2	6.7	5.9
AF-78-A/B <sup>a</sup>	29.0/29.5	0.1/0.6	21.5	5.7
AF-80	20.3	0.5	0.0	18.2

<sup>a</sup>“A/B” indicates that there exist two conformations (A and B) that often have equal occupancies. The atomic volumes of substrate-envelope violations for these compounds were provided as in the two conformations separated by “/”(where applicable).

when binding variant M1. Interestingly, AF-71 and AF-80, the two relatively weak binders to wild-type protease, had rather flat binding specificity profiles with worst-fold affinity losses of 13 and 11, respectively. This is consistent with a common perception that it is less difficult to identify broadly binding compounds with low affinity than it is to find broadly binding ones with high affinity.<sup>27</sup> (3) Of the 10 compounds, six maintained subnanomolar to nanomolar affinity in their

weakest complexes. They include parent compounds and new compounds involving the *N*-acetyl-isoleucine substitution at R1 or the smaller cyclohexylmethyl substitution at R2. In both the AD-93 and the KB-83 series, the larger 3-phenylpropyl hydrophobic substitutions at R2 and the 5-(isoxazol-5-yl)-2-thiophene substitutions at R3 resulted in weakest binding affinities, with  $K_i$ 's greater than 20 nM.

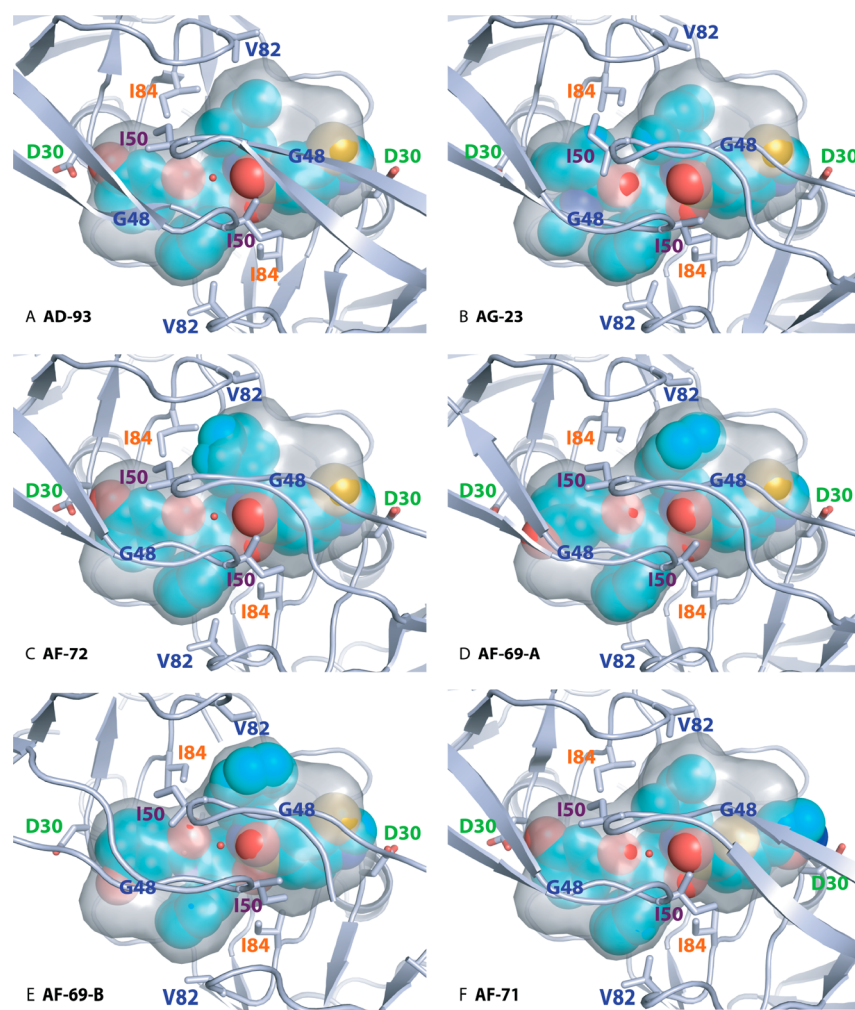
In designing potent inhibitors that are of low susceptibility to resistance mutations, it is not clear which metric, weakest binding affinity or worst-fold affinity loss, is more important. Therefore, we show in Figure 1 a two-dimensional plot of both



**Figure 1.** Pareto plot of affinity and specificity. Plot shows two metrics for each of 10 inhibitors studied, weakest binding affinity and worst fold loss, against the wild-type and mutant panel.

metrics for all 10 compounds against the wild-type and mutant panel. The lower-left corner of the plot indicates where ideal low-susceptibility inhibitors would lie. Inhibitors located here have both tight and flat binding specificity profiles. AG-23, AF-68, and AF-77 are the best compounds judged by this type of Pareto optimality; that is, no other compound achieves lower values in both weakest binding affinity and worst-fold affinity loss. One would shift priority from AG-23 to AF-68 and then to AF-77 as increasing priority was put on worst-fold affinity loss relative to worst binding affinity.

**Crystal Structures of Bound Complexes.** To understand further the resistance patterns described above, crystal structures were solved for the complexes of wild-type HIV-1 protease and all eight new compounds. The AD-93 and KB-83 complexes were reported previously<sup>25</sup> (PDB accession codes 2QI4 and 3SA8, respectively). These crystal structures are shown in Figure 2 (AD-93 series) and Figure 3 (KB-83 series). Violations of the substrate envelope were computed from the crystal structures (total and decomposition by functional groups are given in Table 3; crystallographic details are in Methods and Supplementary Table 1). For the 10 compounds, observed envelope violation volumes correlated well with corresponding substituent sizes in the designed structures. Parent compounds AD-93 and KB-83, both flat binders, had negligible violation of the substrate envelope. New compounds with small violations all displayed flat specificity profiles. In particular, AG-23, AF-72, AF-68, and AF-77 did not violate the substrate envelope by more than 15.2 Å<sup>3</sup>, which is smaller than the volume of a carbon atom (20.6 Å<sup>3</sup> in the model used here),

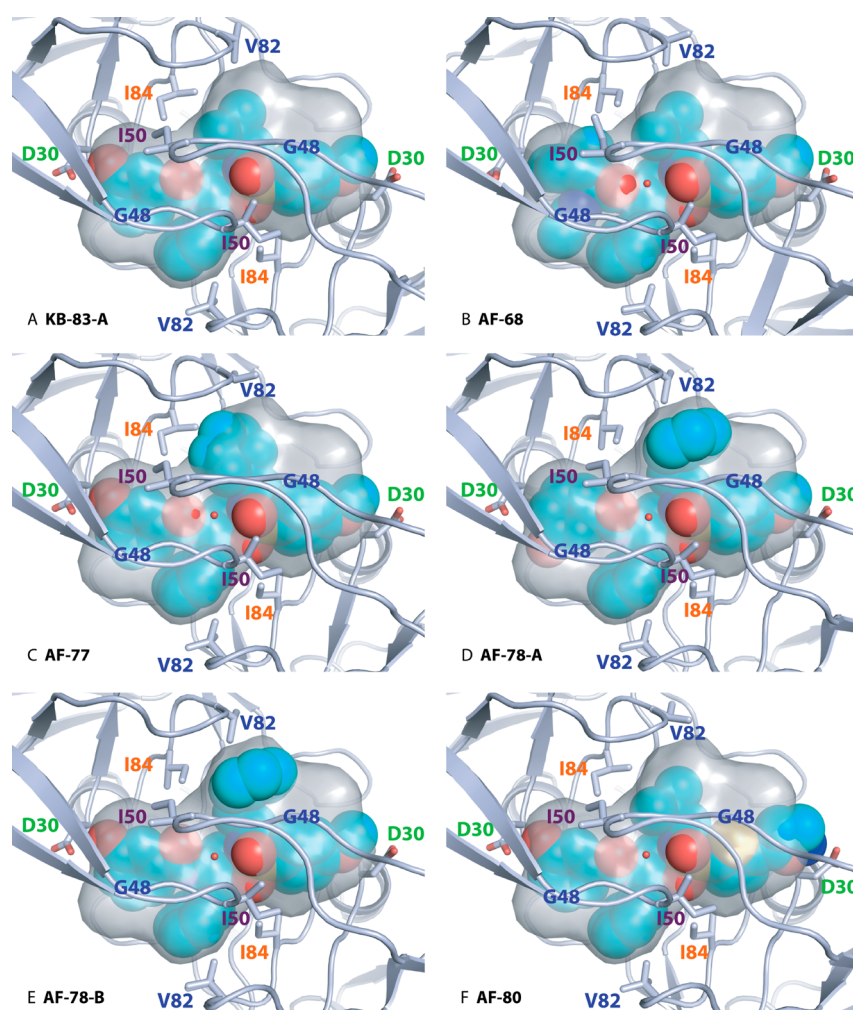


**Figure 2.** Crystal structures of AD-93 series. Proteases are shown in gray cartoon with sticks representing side chains that are mutated in the mutant panel. One-letter amino acid code and residue index are used to label these residues, color-coded by various drug-resistant variants (blue for M1, green for M2, orange for M3, and purple for M4). These structures are aligned and viewed from the top (flap region), and this view was applied to all panels in Figures 2 and 3. All compounds are represented as van der Waals spheres. Carbon atoms are gray for proteases and cyan for inhibitors. The remaining atoms follow the same color-coding rule: red for oxygen, blue for nitrogen, and yellow for sulfur (hydrogen atoms are not shown). A gray, half-transparent surface around each compound represents the substrate envelope. Thus parts of spheres that are not veiled by the gray surface correspond to atomic volumes that pierce the substrate envelope. (A) AD-93, (B) AG-23, (C) AF-72, (D) AF-69, conformation A, (E) AF-69, conformation B, (F) AF-71.

and the maximum worst-fold affinity loss across the set is 37. These substrate-envelope violation volumes are calculated to be a little higher than those for amprenavir ( $3.73 \text{ \AA}^3$ ) or darunavir ( $4.10 \text{ \AA}^3$ ) that have low susceptibility to drug-resistant variations but are much lower than those for drugs with high susceptibility, such as saquinavir ( $80.40$  or  $64.73 \text{ \AA}^3$ ; PDB accession code 3OXC<sup>28</sup> [A or B]). In contrast, compounds with substantial substrate-envelope violations had poor specificity profiles. In particular, AF-69 (in conformation B with 50% occupancy) and AF-78 (in both conformations A and B with equal occupancies) had violations of about 20 and  $29 \text{ \AA}^3$ , respectively, which are comparable to 1–1.5 carbon atoms, and their worst-fold affinity loss was 274 and 98, respectively. There seems to be a relationship between the location at which an inhibitor protrudes outside the substrate envelope and the mutant sustaining the greatest loss in affinity. Namely, binding is most sensitive to mutations near that envelope violation, which might be most likely to disrupt interactions with the protruding group. For example, the greatest structural violations of both AF-69 and AF-78 were due to the R2

group, where the same bulky hydrophobic substitution (3-phenylpropyl) was introduced, and their weakest binding affinities were measured with M1, which includes the P1' site mutations (G48V and V82A) near where R2 packs in the bound complex. Taken together, the data presented here show that zero or small violations of the substrate envelope can lead to compounds with relatively flat binding profiles and that larger violations can produce inhibitors that are susceptible to resistant variants. This is by no means a highly quantitative relationship; for example, compounds with larger envelope violations can have smaller worst-fold affinity losses (AF-78 vs AF-69), particularly against a relatively small panel of mutants.

Note that two compounds violate the envelope but bind robustly across the mutant panel. AF-71 and AF-80, involving the same R3 substitution with a double-ring aromatic system, had violation amounts of about  $20 \text{ \AA}^3$  but worst-fold affinity losses of only about 10. In the P2' site surrounding the R3 group, only one residue (Asp 30) is involved in the mutations included in the mutant panel (D30N in M2). However, AF-71 and AF-80 have improved affinity with M2 compared to that of



**Figure 3.** Crystal structures of KB-83 series. See the caption of Figure 2 for details. (A) KB-83, conformation A (conformation B of compound KB-83 only differs slightly in R1 compared to conformation A and is thus not shown here), (B) AF-68, (C) AF-77, (D) AF-78, conformation A, (E) AF-78, conformation B, (F) AF-80.

**Table 4. Antiviral Activity ( $EC_{50}$  Values in nM) against Selected Wild-Type and Drug-Resistant HIV Clones**

	wild-type (CNDO control strain) <sup>a</sup>	wild-type clade A (I13V, E35D, M36I, R41K, H69K, L89M) <sup>a</sup>	wild-type clade B (N37H, R41K, V77I, I93L) <sup>a</sup>	wild-type clade C (I15V, M36I, R41K, H69K, L89M, I93L) <sup>a</sup>	MDRC4 (L10I, G16E, R41K, M46I, I54V, L63T, T74S, V82A, I93L) <sup>b</sup>	MDR1 (L10I, T12K, N37T, R41K, M46I, I54V, D60E, I62V, L63P, I64V, V77I, V82A, L90M)	worst-fold activity loss <sup>c</sup>
amprenavir	3.0	2.6	1.8	3.7	<b>17.5</b>	15.6	5.5
darunavir	0.25	0.3	0.15	0.18	<b>0.6</b>	0.23	2.4
AG-23	2.7	2.3	1.9	3.5	<b>8.7</b>	7.2	3.2
AF-72	63.7	48.6	37.9	94.6	<b>165.2</b>	99.1	2.6
AF-68	3.5	2.4	1.6	3.0	<b>15.4</b>	15.1	4.4
AF-77	47.3	38.4	22.4	65.7	<b>105.5</b>	86.9	2.2

<sup>a</sup>Wild-type control strain CNDO and patient-derived strains of wild-type HIV-1 from clades A, B, and C. <sup>b</sup>MDRC4, multi-drug-resistant control strain R268. (Genbank accession numbers for all strains are given in the Supporting Information; mutant residues listed here are relative to the CNDO wild-type control strain.). <sup>c</sup>The worst-fold loss is given as the weakest  $EC_{50}$  value to one of the variants (bold face) divided by that to wild-type for each inhibitor.

wild type. This is consistent with the notion that the panel does not have mutations nearby that can challenge these violations, but that these inhibitors could bind poorly to other functional HIV-1 protease variants with more disruptive mutations in the vicinity. Interestingly, the solved X-ray crystal structure had a second conformation for AF-69 (conformation A), which had a smaller R2 violation. If these are both equally accessible, one might imagine that binding to mutants could occur in at least conformation A, and so the loss in affinity would be minor

( $k_b T \ln 2$ ). That this is not the case could be due to unequal occupancy of these two conformations.

**Antiviral Activity Profiles of Inhibitors.** Cell-based antiviral assays were carried out by Monogram Biosciences (South San Francisco, CA) using a panel of various strains of wild-type and variant proteases (Table 4). Four designed compounds that all respected the substrate envelope were tested (the Pareto optimal AG-23 and AF-68 and the somewhat bulkier AF-72 and AF-77) together with two FDA-approved

drugs (amprenavir and darunavir, whose inhibition and antiviral activities were previously characterized<sup>25,26</sup>). All four designed compounds had flat antiviral activity profiles across the viral panel with worst-fold activity losses of between 2.2 and 4.4, which were comparable to those of amprenavir (5.5) and darunavir (2.4). In terms of absolute antiviral activity against wild-type virus, the EC<sub>50</sub>'s of AG-23 (2.7 nM) and AF-72 (3.5 nM) were similar to that of amprenavir (5.5 nM) but an order of magnitude weaker than that of darunavir (0.25 nM). The EC<sub>50</sub>'s of AF-68 (63.7 nM) and AF-77 (47.4 nM) were another order of magnitude higher. Taken together, these results further strengthen the substrate-envelope hypothesis, by showing that the flat binding profiles of envelope-respecting compounds extends to flat activity profiles.

**Conclusion.** In this study, we provide a detailed test of the substrate-envelope hypothesis as a means of designing robust inhibitors against drug-resistant variants. While the use of substrate and transition-state analogues as a starting point for inhibitor design is common, the rationale there is one of achieving affinity: because substrates and transition-states bind to the target, their analogues could bind as well. The substrate-envelope hypothesis relies on more detailed substrate mimicry so the inhibitor does not inadvertently rely on binding interactions nonessential to substrates and thus susceptible to resistance mutations. In other words, the goal of substrate-envelope mimicry is to borrow specificity rather than just affinity from substrates.

To this end, two series of related HIV-1 protease inhibitors were computationally designed. Compounds within each series probed single functional group changes that expanded toward and eventually breached the substrate envelope. Affinity, specificity, and antiviral activity assays were performed on the synthesized compounds using drug-resistant panels of protease and virus, and complexes with the wild-type protease were structurally characterized with X-ray crystallography.

Although all compounds share the amprenavir/darunavir scaffold, which is associated with relatively flat binding profiles, we found pairs of compounds that differed by only a single functional group that bound to the wild-type enzyme in a similar manner with similar affinity but differed greatly in their ability to inhibit drug-resistant proteases variants (AD-93 vs AF-69 and KB-83 vs AF-78). The single distinguishing feature separating flat from nonflat binders in these pairs was that flat binders respected (AD-93 and KB-83) and nonflat binders violated (AF-69 and AF-78) the substrate envelope. This suggests that the envelope is more important than the shared scaffold in the development of inhibitors of low susceptibility to drug-resistant mutations.

Interestingly, mutations in the protease occur near different locations on the envelope. There is excellent correspondence between losses in affinity and the location of envelope-violating inhibitor excursions. AF-69 and AF-78 both have R2 extensions that violate the envelope at P1/P1', and both inhibitors lose their greatest affinity to the M1 variant with characteristic mutations G48V and V82A, which are close to the envelope at P1/P1', consistent with the notion that the inhibitors violate the envelope to make non-substrate-like P1/P1' interactions that are disturbed in the M1 variant. In contrast, two of the designed compounds, AF-71 and AF-80, violate the envelope substantially but exhibit flat binding profiles against the mutant panel. As they involve the same R3 substitution and the panel lacks mutations in the P2/P2' site nearby, we predict that they would become susceptible to other drug-resistant variants that

are not present in the panel studied. In other words, these two cases should not be regarded as counterexamples of the substrate-envelope hypothesis but rather examples showcasing potential predictive powers of the hypothesis.

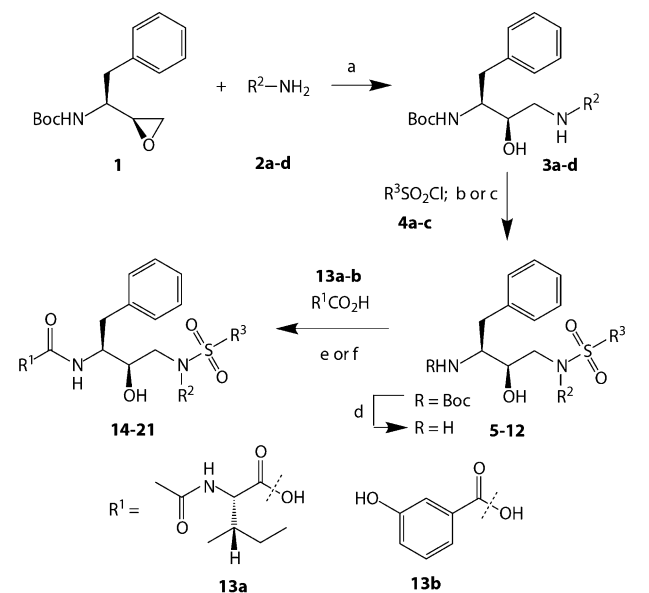
Although the substrate-envelope hypothesis, in a narrow sense, applies only to enzyme targets, one can generalize the envelope hypothesis for non-enzyme targets through extending the definition of the envelope to include other necessary functions of a target besides binding substrates. Consequentially, designed compounds that hide inside some "envelope of function" could be immune to target mutations that maintain native function.

## METHODS

**Computational Design.** In our previous study,<sup>25</sup> 36 MIT-2 compounds were computationally designed, and the binding specificity profiles for 10 of them against the wild-type and mutant panel were experimentally measured. Two compounds that exhibited flat binding profiles, namely, MIT-2-AD-93 and MIT-2-KB-83, were chosen as parent compounds. A 1.2-Å crystal structure of darunavir-bound wild-type HIV-1 protease (PDB accession code 1T3R) was used as the fixed target into which inhibitor compounds were computationally designed. All inhibitors share the (*R*)-(hydroxyethylamino)-sulfonamide scaffold derived from the clinically approved inhibitors amprenavir<sup>14</sup> and darunavir<sup>18</sup> but possess various functional groups (that attach to the scaffold at different positions) derived from the ZINC database<sup>29</sup> and commercial catalogs. The same design framework and parallel synthetic components were used for the previously reported MIT-2 library.<sup>25</sup> The special design strategy here was to compare the lists of top candidates designed with the substrate-envelope constraint and the maximal-envelope constraint, identify functional group choices at each position that violate the substrate-envelope constraint and are otherwise energetically favored against the background of other positions, and generate the list of derivative compounds that are one functional group away from each parent compound covering various positions and molecular properties.

**Synthesis.** The designed protease inhibitors were synthesized using a stepwise synthetic approach. The Boc-protected intermediates (*R*)-(hydroxyethylamino)sulfonamides **5–12** (Scheme 1) were prepared from the commercially available chiral epoxide **1**, (1*S*,2*S*)-(1-oxiran-2-yl)phenylethyl carbamic acid *tert*-butyl ester, using methods described previously.<sup>25,30</sup> Briefly, regioselective ring-opening of epoxide **1** with selected primary amines **2a–d** provided the  $\beta$ -amino alcohols **3a–d**, respectively. Selective reactions of secondary amino group in **3a–d** with 4-methoxybenzene sulfonyl chloride using aqueous sodium carbonate (Na<sub>2</sub>CO<sub>3</sub>) and with benzothiazole-6-sulfonyl chloride and 5-(isoxazol-5-yl)thiophene-2-sulfonyl chloride using diisopropylethylamine in anhydrous CH<sub>2</sub>Cl<sub>2</sub> provided the (*R*)-(hydroxyethylamino)sulfonamide intermediates **5–12**. Removal of the Boc protection using trifluoroacetic acid in CH<sub>2</sub>Cl<sub>2</sub> followed by the reactions of the resulting free amines with carboxylic acids **13a,b** using either EDCI/HOBt/DIEA in a DMF–CH<sub>2</sub>Cl<sub>2</sub> (1:1) mixture (Method A) or EDCI/HOBt in a H<sub>2</sub>O–CH<sub>2</sub>Cl<sub>2</sub> (1:1) mixture (Method B)<sup>25</sup> provided the designed inhibitors **14–21**.

**Assays for Inhibition and Antiviral Activities.** HIV protease inhibitor activities were determined by the fluorescence resonance energy transfer (FRET) method.<sup>31,32</sup> Protease substrate, Arg-Glu-(EDANS)-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-Lys(DABCYL)-Arg, was labeled with the energy transfer donor (EDANS) and acceptor (DABCYL) dyes at its two ends to perform FRET. Fluorescence measurements were carried out on an EnVision plate reader (PerkinElmer). Drug susceptibility assays were carried out by Monogram Biosciences against wild-type HIV-1 control, three patient-derived strains of wild-type HIV-1 from clades A, B, and C, and two multi-drug-resistant HIV-1 variants including MDRC4 as control strain and MDR1 strain with protease mutations M46I, I54V, V82A, and L90M.

Scheme 1. Synthesis of Paired Compound Library<sup>a</sup>

14 $R^1 = 13a$ ;	5 $R^2 = (S)$ -2-methylbutyl; $R^3 = 6$ -benzo[ <i>d</i> ]thiazole
15 $R^1 = 13b$ ;	6 $R^2 =$ cyclohexylmethyl; $R^3 = 6$ -benzo[ <i>d</i> ]thiazole
16 $R^1 = 13b$ ;	7 $R^2 = 3$ -phenylpropyl; $R^3 = 6$ -benzo[ <i>d</i> ]thiazole
17 $R^1 = 13b$ ;	8 $R^2 = (S)$ -2-methylbutyl; $R^3 = 5$ -(isoxazol-5-yl)-2-thiophene
18 $R^1 = 13a$ ;	9 $R^2 =$ isobutyl; $R^3 = 4$ -methoxybenzene
19 $R^1 = 13b$ ;	10 $R^2 =$ cyclohexylmethyl; $R^3 = 4$ -methoxybenzene
20 $R^1 = 13b$ ;	11 $R^2 = 3$ -phenylpropyl; $R^3 = 4$ -methoxybenzene
21 $R^1 = 13b$ ;	12 $R^2 =$ isobutyl; $R^3 = 5$ -(isoxazol-5-yl)-2-thiophene

<sup>a</sup>Reagents and conditions: (a) EtOH or *i*PrOH, 80 °C, 3 h; (b) aq Na<sub>2</sub>CO<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C to rt, overnight; (c) Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C to rt, overnight; (d) TFA, CH<sub>2</sub>Cl<sub>2</sub>, rt, 1 h; (e) EDCl, HOBT, DIEA, DMF-CH<sub>2</sub>Cl<sub>2</sub> (1:1), 0 °C to rt, overnight; (f) EDCl, HOBT, H<sub>2</sub>O-CH<sub>2</sub>Cl<sub>2</sub> (1:1), 0 °C, 24 h.

**Crystallography.** The expression, isolation, and purification of wild-type and drug-resistant HIV-1 protease variants used for crystallization and binding experiments were carried out as previously described.<sup>33</sup> More details are provided in Supporting Information. The data collection and refinement statistics are shown in Supplementary Table 1.

**Structural Analysis.** The wild-type protease-inhibitor crystallographic complexes were evaluated in terms of substrate-envelope hypothesis. The extent to which an inhibitor violates the substrate envelope was measured by the volume of its non-hydrogen atoms that protrude outside (calculated in total or decomposed at various sites). The protrusion volume was calculated based on a grid with 0.2-Å spacing.

## ■ ASSOCIATED CONTENT

### Supporting Information

Preparation of the design-target structure of HIV-1 protease, preparation of scaffold and functional group library, calculation of the substrate and maximal envelope, computational design for paired compounds, synthesis conditions, crystallography protocols, FRET assays, drug susceptibility assays, and measurement of substrate-envelope hypothesis violation; crystallography and refinement statistics. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### Accession Codes

Coordinates of the 8 new compounds co-crystallized with wild-type HIV-1 protease in the PDB, accession codes 3SA3, 3SA4, 3SA5, 3SA6, 3SA7, 3SA8, 3SA9, 3SAA, and 3SAB.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: tidor@mit.edu.

\*E-mail: celia.schiffer@umassmed.edu.

\*E-mail: trana@sanfordburnham.org.

### Present Addresses

<sup>¶</sup>Toyota Technological Institute at Chicago, 6045 S. Kenwood Avenue, Chicago, IL 60637, USA.

<sup>†</sup>Merck Research Laboratories-Boston, 33 Avenue Louis Pasteur, Boston, MA 02115, USA.

### Author Contributions

<sup>#</sup>These authors contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was partially supported by the National Institute of General Medical Sciences of the NIH (GM066524, GM065418, GM082209, AI41404, and AI43198). Crystallographic data were measured at beamline X29A of the NSLS through the mail-in crystallography program and at BioCARS Sector 14 of Advanced Photo Source (DE-AC02-06CH11357 and RR007707). The authors thank A. Héroux at NSLS, W. Royer, and S. Shandilya for crystallographic data collection, and N. Kurt Yilmaz for proofreading the manuscript. Computations were carried out on hardware partially provided by a gift from Intel Corporation and through a grant from the US National Science Foundation (0821391).

## ■ REFERENCES

- Hayes, J. D., and Wolf, C. R. (1990) Molecular mechanisms of drug resistance. *Biochem. J.* 272, 281–295.
- McManus, M. C. (1997) Mechanisms of bacterial resistance to antimicrobial agents. *Am. J. Health-Syst. Pharm.* 54, 1420–1433.
- Walsh, C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature* 406, 775–781.
- Krishna, R., and Mayer, L. D. (2000) Multidrug resistance (MDR) in cancer: Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs. *Eur. J. Pharm. Sci.* 11, 265–283.
- Gottesman, M. M. (2002) Mechanisms of cancer drug resistance. *Annu. Rev. Med.* 53, 615–627.
- Tenover, F. C. (2006) Mechanisms of antimicrobial resistance in bacteria. *Am. J. Med.* 119, S3–S10.
- King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem. Biol.* 11, 1333–1338.
- Wlodawer, A., and Erickson, J. W. (1993) Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* 62, 543–585.
- Turner, S. R., Strohbach, J. W., Tommasi, R. A., Aristoff, P. A., Johnson, P. D., Skulnick, H. I., Dolak, L. A., Seest, E. P., Tomich, P. K., Bohanon, M. J., Horng, M.-M., Lynn, J. C., Chong, K.-T., Hinshaw, R. R., Watenpugh, K. D., Janakiraman, M. N., and Thaisrivongs, S. (1998) Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J. Med. Chem.* 41, 3467–3476.
- Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Krohn, A., and Al, E. (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* 248, 358–361.
- Kempf, D. J., Marsh, K. C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentge, C. A., Green, B. E., Fino, L., Park, C. H., and Kong, X. P. (1995) ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2484–2488.



- (12) Vacca, J. P., Dorsey, B. D., Schleif, W. A., Levin, R. B., McDaniel, S. L., Darke, P. L., Zugay, J., Quintero, J. C., Blahy, O. M., and Roth, E. (1994) L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 91, 4096–4100.
- (13) Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997) Viracept (nelfinavir mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40, 3979–3985.
- (14) Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., and 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, 1181–1182.
- (15) Baker, C., Chaturvedi, P. R., Hale, M. R., Bridson, G., Heiser, A., Furfine, E. S., Spaltenstein, A., and Tung, R. D. (1999) Discovery of VX-175/GW433908, a novel, water-soluble prodrug of amprenavir, in 39th Interscience Conference on Anti-microbial Agents and Chemotherapy, San Francisco, 26–29 Sep 1999, abstract 313.
- (16) Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C.-M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G. R., Sun, E., Japour, A. J., Leonard, J. M., Plattner, J. J., and Norbeck, D. W. (1998) ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* 42, 3218–3224.
- (17) Robinson, B. S., Riccardi, K. A., Gong, Y., Guo, Q., Stock, D. A., Blair, W. S., Terry, B. J., Deminie, C. A., Djang, F., Colonno, R. J., and Lin, P. (2000) BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.* 44, 2093–2099.
- (18) Surleraux, D. L. N. G., Tahri, A., Verschuere, W. G., Pille, G. M. E., de Kock, H. A., Jonckers, T. H. M., Peeters, A., De Meyer, S., Azijn, H., Pauwels, R., de Bethune, M.-P., King, N. M., Prabu-Jeyabalan, M., Schiffer, C. A., and Wigerinck, P. B. T. P. (2005) Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* 48, 1813–1822.
- (19) Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* 10, 369–381.
- (20) Chellappan, S., Kairys, V., Fernandes, M. X., Schiffer, C., and Gilson, M. K. (2007) Evaluation of the substrate envelope hypothesis for inhibitors of HIV-1 protease. *Proteins Struct. Funct. Bioinf.* 68, 561–567.
- (21) Nalam, M. N. L., Ali, A., Altman, M. D., Reddy, G. S. K. K., Chellappan, S., Kairys, V., Ozen, A., Cao, H., Gilson, M. K., Tidor, B., Rana, T. M., and Schiffer, C. A. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.* 84, 5368–5378.
- (22) Kairys, V., Gilson, M. K., Lather, V., Schiffer, C. A., and Fernandes, M. X. (2009) Toward the design of mutation-resistant enzyme inhibitors: Further evaluation of the substrate envelope hypothesis. *Chem. Biol. Drug Des.* 74, 234–245.
- (23) Romano, K. P., Ali, A., Royer, W. E., and Schiffer, C. A. (2010) Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20986–20991.
- (24) Romano, K. P., Laine, J. M., Deveau, L. M., Cao, H., Massi, F., and Schiffer, C. A. (2011) Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease. *J. Virol.* 85, 6106–6116.
- (25) Altman, M. D., Ali, A., Kumar Reddy, G. S. K., Nalam, M. N. L., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M., and Tidor, B. (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *J. Am. Chem. Soc.* 130, 6099–6113.
- (26) Parai, M. K., Huggins, D. J., Cao, H., Nalam, M. N. L., Ali, A., Schiffer, C. A., Tidor, B., and Rana, T. M. (2012) Design, synthesis, and biological and structural evaluations of novel HIV-1 protease inhibitors to combat drug resistance. *J. Med. Chem.* 55, 6328–6341.
- (27) Huggins, D. J., Sherman, W., and Tidor, B. (2012) Rational approaches to improving selectivity in drug design. *J. Med. Chem.* 55, 1424–1444.
- (28) Tie, Y., Kovalevsky, A. Y., Boross, P., Wang, Y.-F., Ghosh, A. K., Tozser, J., Harrison, R. W., and Weber, I. T. (2007) Atomic resolution crystal structures of HIV-1 protease and mutants V82A and I84V with saquinavir. *Proteins Struct. Funct. Bioinforma.* 67, 232–242.
- (29) Irwin, J. J., and Shoichet, B. K. (2005) ZINC—A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182.
- (30) Ali, A., Kiran Kumar Reddy, G. S., Nalam, M. N. L., Anjum, S. G., Cao, H., Schiffer, C. A., and Rana, T. M. (2010) Structure-based design, synthesis and structure-activity relationship studies of HIV-1 protease inhibitors incorporating phenylloxazolidinones. *J. Med. Chem.* 53, 7699–7708.
- (31) Matayoshi, E. D., Wang, G. T., Krafft, G. A., and Erickson, J. (1990) Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* 247, 954–958.
- (32) Ali, A., Reddy, G. S. K. K., Cao, H., Anjum, S. G., Nalam, M. N. L., Schiffer, C. A., and 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.
- (33) King, N. M., Melnick, L., Prabu-Jeyabalan, M., Nalivaika, E. A., Yang, S., Gao, Y., Nie, X., Zepp, C., Heefner, D. L., and Schiffer, C. A. (2002) Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci.* 11, 418–429.