Selection of HIV-1 for Resistance to Fourth Generation Protease Inhibitors Reveals Two Independent Pathways to High-Level Resistance

Ean Spielvogel[1,2], Sook-Kyung Lee[2], Shuntai Zhou[2], Gordon J. Lockbaum[3], Mina Henes[3], Amy Sondgeroth[1,2], Klajdi Kosovrasti[3], Akbar Ali[3], Nese Kurt Yilmaz[3], Celia A. Schiffer[3*], Ronald Swanstrom[2,4*]

[1]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[3]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[4]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

*Corresponding Author

Ronald Swanstrom: Phone: +1 919 966 5710; ron_swanstrom@med.unc.edu

Celia A. Schiffer: Phone: +1 508 856 8008; Celia.Schiffer@umassmed.edu

Summary

Well-designed viral protease inhibitors (PIs) potently inhibit replication as well as create a high genetic barrier for resistance. Through in vivo selective pressure, we have generated high-level resistance against ten HIV-1 PIs and their precursor, the FDA-approved drug darunavir (DRV), achieving 1,000-fold resistance over the starting EC50. The accumulation of mutations revealed two pathways to high-level resistance, resulting in protease variants with up to 14 mutations in and outside of the active site. The two pathways demonstrate the interplay between drug resistance and viral fitness. Replicate selections showed that one inhibitor could select for resistance through either pathway, although subtle changes in chemical structure of the inhibitors led to preferential use of one pathway over the other. Viral variants from the two pathways showed differential selection of compensatory mutations in Gag cleavage sites. These results reveal the high-level of selective pressure that is attainable with these fourth-generation protease inhibitors, and the interplay between selection of mutations to confer resistance while maintaining viral fitness.

38 **Introduction**

39 Highly active antiretroviral therapy (HAART) against HIV-1 with combinations of three
40 or more drugs effectively block viral replication and preclude the evolution of drug resistance.
41 Each drug by itself can select for resistance, and successive addition of the same three drugs that
42 together are suppressive would lead to multi-drug resistance. Thus, only fully suppressing viral
43 replication allows successful therapy, while sub-optimal inhibition leads to selection of
44 resistance. Accordingly, the population size of the replicating virus is an important determinant
45 of the evolution of resistance. In an early clinical study of monotherapy with a protease
46 inhibitor, the time to the appearance of resistance was correlated with the nadir of viral load
47 before rebound (Kempf et al., 1998), i.e. the lower the nadir viral load, the longer the time to the
48 appearance of resistant rebound virus. Three major factors interplay to define the emergence of
49 resistance *in vivo*: i) the active drug concentration relative to its inhibitory activity; ii) the level
50 of resistance conferred by one or more mutations; and iii) the fitness cost of the resistance
51 mutations.

52 The early identification of the retroviral protease as a member of the aspartyl proteinase
53 family and the determination of a number of cleavage site sequences led to the development of
54 first generation inhibitors that validated the HIV-1 protease as a drug target (Katoh et al., 1987,
55 Richards et al., 1989, Seelmeier et al., 1988). A second generation of inhibitors was quickly
56 developed for use in humans, becoming the third drug in a three-drug regimen that achieved
57 sustained suppression of viral load with no evolution of resistance (Gulick et al., 1997). The
58 third generation of inhibitors had improved properties with regard to side effects and efficacy.
59 In addition, the strategy of "boosting" protease inhibitor levels with ritonavir (RTV), which at
60 low doses inhibits cytochrome P450-3A4 metabolizing HIV-1 PIs, allowed for increased drug

61 levels needed to inhibit replication. These properties have been further enhanced with a fourth-
62 generation PI, darunavir (DRV), which achieves drug levels in plasma (>1 μM) that is 1,000
63 fold greater than its inhibitory activity in cell culture (Ali et al., 2010, Nalam et al., 2013,
64 Yilmaz et al., 2009). The high efficacy of a fourth-generation inhibitor such as DRV can be
65 inferred from an attempt to use this drug in monotherapy (Katlama et al., 2010, Pulido et al.,
66 2011). In the cases of virologic failure there was no significant resistance to DRV in the
67 rebound virus (Katlama et al., 2010, Pulido et al., 2011). Thus the observed rebound is most
68 easily attributed to issues with adherence or possibly poor drug penetration in some tissues.
69 Under selective pressure, such as inhibition with small molecules, for survival the virus
70 has to maintain a balance between mutations that confer inhibitor resistance while maintaining
71 the enzyme's necessary catalytic function to allow viral replication. Typically, at low inhibitor
72 concentrations a less resistant but more fit virus will be selected, while at higher inhibitor
73 concentrations a more resistant but less fit virus may have to be selected. A clear example of
74 this is with HIV-1 PI nelfinavir (NFV) where in patient isolates the resistance mutation D30N
75 was typically observed, while in cell culture I84V was readily selected and provides greater
76 resistance and cross-resistance (Grossman et al., 2004, De Meyer et al., 2005, Ntemgwa et al.,
77 2007) but lower fitness. When resistance can only be achieved by one or more mutations that
78 are deleterious to enzyme function, requiring the selection of additional compensatory
79 mutations to restore fitness, such inhibitors are considered to have a high genetic barrier to
80 resistance. Thus, it becomes increasingly difficult for virus in a small population size to survive
81 the fitness loss long enough to accumulate the additional needed mutations, either as the
82 population size is rapidly declining during therapy initiation or in sites where there might be
83 low level replication on therapy.

84 We have previously designed a series of highly potent protease inhibitors, UMASS1-10,
85 that fit within the substrate envelope, which is the shared volume occupied by natural protease
86 substrates when bound to the active site (Nalam et al., 2013). These inhibitors are less
87 susceptible to resistance because a mutation affecting such inhibitors will simultaneously affect
88 substrate processing. The designed inhibitors share a common chemical scaffold with DRV but
89 have modified chemical moieties that further fill the substrate envelope, and all bind tighter than
90 <5 pM to purified wildtype HIV-1 protease. These inhibitors retained robust binding to many
91 multi-drug resistant protease variants and viral strains. Thus, the substrate envelope proved to
92 be a powerful tool to guide the design of potent and robust inhibitors, by minimizing
93 susceptibility to resistance mutations.
94 In this study we have examined the evolutionary path that HIV-1 follows to attain high
95 level resistance to a panel of fourth-generation PIs by selecting for resistance under conditions
96 of escalating inhibitor concentration during viral replication in cell culture. While DRV and
97 UMASS1-10 potently inhibit wild-type and single mutant variants, under persistent pressure of
98 sub-optimal inhibition the virus evolves to accumulate mutations and escape inhibition. In most
99 cases, selection was carried out until the inhibitor concentration was over 1,000 times the
100 starting EC50, with the final concentration approximating that achieved by DRV *in vivo*. While
101 it is possible to select for high level resistance to second and third generation protease
102 inhibitors, these high levels of resistance are not relevant given that these drugs do not reach
103 comparably high concentrations *in vivo* (Watkins et al., 2003). Selections against the UMASS
104 series of protease inhibitors were performed twice, in the presence and absence of an initial pool
105 of common single-site resistance mutations, which had a long-term impact on the sequence
106 diversity in the culture. Resistance overall followed one of two pathways, one defined by

107 higher drug resistance but lower viral fitness and the other defined by higher viral fitness but
108 lower drug resistance. Relatively minor modifications in inhibitor structure influenced selection
109 of one or the other pathway to resistance, although both pathways eventually led to high levels
110 of cross-resistance between inhibitors. The viral passaging experiments resulted in proteases
111 with up to 14 resistance-associated mutations, and deep sequencing analysis showed persistent
112 heterogeneity in the viral population within the culture. These results reveal the extremely high
113 genetic barrier to resistance for fourth-generation protease inhibitors at inhibitor concentrations
114 that can be achieved *in vivo*, and the complex evolutionary pathways required to achieve
115 resistance.
116
117 **Results**
118 **Panel of highly potent and analogous HIV-1 protease inhibitors**
119 HIV-1 protease inhibitors were designed by modifications to DRV to increase favorable
120 interactions within the substrate envelope thereby increasing potency while minimizing
121 evolution of resistance (Nalam et al., 2013). A panel of ten DRV analogues were chosen with
122 enzymatic inhibition constants (K_i) in the single or double-digit picomolar range to wild-type
123 NL4-3 protease and the I84V and I50V/A71V drug resistant variants, respectively [**Table 1**]
124 (Mittal et al., 2013, Lockbaum et al., 2019). These PIs contained modified P1' positions with
125 (S)-2-methylbutyl or 2-ethyl-n-butyl groups (R1-1 and R1-2, respectively) in combination with
126 five diverse P2′ phenyl-sulfonamides (R2-1 to R2-5), with the inhibitors named UMASS-1
127 through -10 [**Table 1**]. These inhibitors and DRV were tested in a cell culture-based viral
128 inhibition assay. The EC50 constants (the amount of inhibitor needed to inhibit 50% of the
129 infectivity of the virus when the drug was present during virus production) for DRV and

130 UMASS analogues ranged from 2.4 to 9.1 nM, significantly more potent than the second and

131 third generation protease inhibitors (**Figure S1**).

132

133 **Table 1**. Inhibition constants (K_i) in enzymatic assays and half maximal effective

134 concentrations (EC50) in viral inhibition assays of DRV and analogous protease inhibitors.

Inhibitor	Structure	K_i (pM)			EC50 (nM)
		WT	I84V	I50V/A71V	WT
DRV		< 5.0	25.6 ± 5.6	74.5 ± 5.6	7.7
UMass 1		< 5.0	26.1 ± 3.7	110.3 ± 8.8	5.9
UMass 2		< 5.0	< 5.0	15.0 ± 2.7	2.4
UMass 3		< 5.0	9.9 ± 2.7	79.9 ± 5.9	9.1
UMass 4		< 5.0	10.5 ± 1.8	32.9 ± 3.0	3.2
UMass 5		< 5.0	7.0 ± 1.7	7.8 ± 0.9	4.0
UMass 6		< 5.0	12.8 ± 3.1	100.0 ± 9.9	5.2
UMass 7		< 5.0	12.1 ± 4.5	18.2 ± 3.0	3.1
UMass 8		< 5.0	< 5.0	55.4 ± 4.0	4.2
UMass 9		< 5.0	7.6 ± 1.6	42.3 ± 2.6	6.4
UMass 10		< 5.0	14.3 ± 9.3	5.8 ± 1.1	4.1

135

136 **Selection for high-level resistance *in vitro***

137 To evaluate the potential of each inhibitor to select for mutations that would confer high-
138 level resistance and to compare these mutations across analogous inhibitors, we performed viral
139 passaging under conditions of escalating inhibitor concentration in cell culture. Virus in the
140 cultures was periodically sequenced after selection to specific inhibitor concentrations. The
141 selection experiments were performed under two separate starting conditions, a mixture of 26
142 viruses with known single-site mutations associated with drug resistance in an NL4-3
143 background, or with virus generated from only the wild-type NL4-3 clone (which closely
144 approximates the clade B consensus sequence for the protease amino acid sequence). Notably,
145 only about one-half of the selected mutations were present in the initial mixture, indicating that
146 even in the selection that was seeded with the pool of resistance mutations there was sufficient
147 evolutionary capacity to explore additional mutational space. Inhibitor/Selective pressure
148 started at low nanomolar concentrations and increased by a factor of 1.5 with each subsequent
149 viral passage. All of the selections starting with wild type virus reached at least 5 μM of
150 inhibitor concentration. For technical reasons, only 5 of the selections starting with the mixture
151 of mutants reached an inhibitor concentration of 400 nM and are included in this report (**Figure**
152 **S2**). To assess variability in the selection scheme, selection against DRV was replicated four
153 separate times starting with the same mutant mixture.

154

155 **Two major mutational pathways to resistance determined by next generation sequencing**
156 **(NGS) of viral culture during *in vitro* selection**

157 Resistance mutations selected in the protease coding domain during the escalating
158 selective pressure of increased protease inhibitor concentration were examined at various time-
159 points using a next generation sequencing (NGS) protocol that included Primer ID with the
160 MiSeq platform (Zhou et al., 2015). In this approach, individual cDNA molecules are tagged
161 with 11 random degenerate bases in the cDNA primer to give a unique molecular identifier to
162 each cDNA/RNA template before the PCR step, allowing quantification of the number of
163 templates sequenced (by the number of different Primer ID identifiers recovered). A Template
164 Consensus Sequences (TCS) was generated using the multiple reads associated with each
165 Primer ID identifier/template, which greatly lowers the error rate. The abundance of viral RNA
166 templates recovered from the culture supernatants made it possible to sequence thousands of
167 templates, which validated the sampling sensitivity by detecting several copies of minor
168 variants representing less than 0.1% of the population.
169 Each of the viral cultures showed an accumulation of protease mutations with increasing
170 selective pressure. NGS analysis revealed very few fixed variants until the inhibitor
171 concentration reached 3 nM, with some exceptions occurring at sub-EC50 concentrations.
172 Multiple resistance variants were observed in relatively high abundance after the drug
173 concentration surpassed the EC50 values above 3 nM, highlighting the high genetic diversity in
174 the culture. Additional compensatory mutations became linked at higher drug concentrations,
175 which was followed by a fairly stable population through the rest of the time points. An
176 average of 6 mutations were present by the time the drug concentration reached 100 nM, while
177 an average of 10 mutations (and up to 14) were seen for the selections that reached greater than
178 1 μM inhibitor concentration (**Figure S3, S4**).

179 The mutations observed in the most abundant protease variant present at the highest
180 inhibitor concentration achieved in each selection are shown in **Figure 1**. These end-point
181 protease variants illustrate two largely independent pathways to resistance, centered around the
182 active site mutations I84V or I50V, although in some cases both mutations were observed.
183 Also, the most abundant and the second most abundant genotypes in each culture typically
184 differ by a single compensatory mutation, indicative of a necessary "backbone" of resistance
185 mutations shared by a majority of successful variants (not shown). Finally, certain mutations are
186 linked to one or the other pathway while others are shared (see below).
187
188 **Both resistance pathways confer high levels of cross-resistance to all PIs**
189 To quantify the resistance associated with each selection pathway, a subset of viruses
190 that reached the 5 μM inhibitor concentration in the selection cultures were chosen to be tested
191 in an EC50 infectivity assay. The EC50 values were obtained using pools of viruses that
192 contained mostly homogenous populations, which aimed to minimize any confounding
193 variables in the dose-response curve, although there was some sequence heterogeneity in the
194 cultures. The pools were sequenced and all viral pools used in the EC50 experiments had a
195 single variant representing at least 80% of their population.
196 The viruses tested revealed EC50 values 100 to over 10,000-fold higher than WT virus
197 (ND) across the different inhibitors [**Figure 2**]. Cross-resistance against all inhibitors was
198 observed at high levels. The virus pool that contained both 50V and 84V mutations showed the
199 highest levels of resistance across the panel of inhibitors. Thus, the selections were successful
200 in generating highly resistant variants to these fourth-generation inhibitors after the
201 accumulation of mutations in over 10% of the sequence of the protease or more.

202

203 **Sequence diversity (entropy) varied over the course of selection**

204 As previously mentioned, on average 6 mutations were observed at 100 nM inhibitor
205 concentration, and on average 10 mutations were observed when the selective pressure was
206 above 1 μM, indicating the increasing number of mutations necessary for viability under
207 increasing selective pressure. However, deep sequencing at selected inhibitor concentrations
208 revealed mutations accumulated in complex patterns. We assessed the sequence complexity of
209 each culture by calculating the Shannon Entropy to allow comparison of changes in diversity in
210 the cultures. Entropy profiles are shown for all of the selections in **Figure 3**. When we
211 examined the entropy values for all selections that reached 1 μM in inhibitor concentration we
212 found that cultures starting with the mixture of resistant viruses averaged a nearly two-fold
213 higher entropy value compared to the cultures where the selection started with just the virus
214 generated from the NL4-3 clone (3.0 vs 1.6, $P<0.0001$ Mann-Whitney test). This was
215 unexpected, as both sets of selections passed through many genetic bottlenecks. This result is
216 most easily explained if the rates of recombination were fairly high throughout the culture
217 period.

218 The additional entropy plots of each individual selection with the times when mutations
219 appeared (**Figure 3**) in the culture show the early appearance of the I84V mutation was
220 associated with peaks in entropy, reflecting high genetic diversity, followed by a decrease in
221 entropy when the mutation became fixed. The I50V mutation was not associated with drops in
222 entropy when it entered the population, rather these populations maintained high genetic
223 diversity even at higher drug concentrations. We interpret these patterns as indicative of I84V
224 conferring some level of resistance without a dramatic loss in fitness, allowing a more

225 homogeneous culture (i.e. less entropy). In contrast, I50V confers a higher level of resistance
226 but at a greater fitness cost, thus requiring greater diversity in the culture either as compensatory
227 mutations or as other combinations of mutations with lesser resistance but higher fitness. We
228 previously showed I50V significantly reduces the fitness of the virus relative to the fitness loss
229 of a virus with I84V (Henderson et al., 2012). In contrast, I50V (with A71V) was on average
230 significantly less sensitive to inhibition by this series of inhibitors (Table 1).
231
232 **Inhibitor structure influences the resistance pathway**
233 Selections were performed with 11 analogous inhibitors derived from a common
234 scaffold (**Table 1**) (Nalam et al., 2013, Paulsen et al., 2017). We were interested to see if subtle
235 chemical differences between the inhibitors could impact selection for different resistance
236 pathways. We found the P1' group, either (S)-2-methylbutyl (R1-1) or 2-ethyl-n-butyl (R1-2),
237 influenced the resistance pathway. With the UMASS 1-5 series (the smaller R1-1 group), the
238 I84V pathway was favored. In contrast, the UMASS 6-10 series with the larger R1-2 group,
239 favored the I50V mutation. Overall, 7 of the 8 cultures with an R1-1 inhibitor first had I84V,
240 while 8 of 9 cultures with R1-2 had I50V (P=0.003, Fisher's Exact Test). We considered the
241 possibility that the mixture of mutant viruses in the first selection might skew the pathway
242 selected. However, in only 1 of the 8 cultures with sufficient data from both selections was
243 there a switch from the I84V pathway to the I50V pathway between the first and second
244 selections (cultures of UMASS 6 with an R1-2 group). Thus, we conclude that the P1' group of
245 the inhibitor is a strong determinant of the pathway selected.
246 While the cultures of inhibitors with R1-1 groups favored I84V there was also some
247 selection of I50V; this is in contrast to cultures with the R1-2 group inhibitors which strongly

248 favored the I50V pathway and excluded the I84V pathway. This preference is explained with
249 analysis of the protease-inhibitor cocrystal structures. The R1-1 has one more methyl group
250 than DRV which packs against residue 82, but this group loses significant vdW contacts with
251 residue I84 due to I84V mutation. The R1-2 group has one more methyl than the R1-1 group
252 which packs against residue 84 and thus better maintains vdW contacts [**Figure 4**] (Lockbaum
253 et al., 2019). Similar to I84V, the I50V mutation causes a steric reduction of a residue side
254 chain in the hydrophobic S1' pocket. Like I84V, the I50V mutation causes loss of vdW contacts
255 with the R1-1 group, but unlike with the I84V mutation, the R1-2 is unable to accommodate the
256 I50V mutation due to the flaps adopting a subtly different conformation in the presence of the
257 mutation.

258 To examine the broader pattern of mutations selected based on the R1 group, the
259 abundance data from UMASS1-5 and UMASS6-10 was pooled and examined sequentially at
260 different levels of drug concentration. In this analysis (**Figure 5**), the mutations selected against
261 UMASS1-5 are depicted pointing upwards, while the mutations resulting from UMASS6-10
262 point downwards. These data show the strong preference for the I50V pathway for the R1-2-
263 containing inhibitors, and suggest there may be specific and shared mutations in the two
264 pathways (see below). Since the R1 group was a strong determinant of resistance pathway
265 chosen, we did not analyze the data for the larger set of R2 groups.

266 We further explored how deterministic pathway choice was by analyzing the four
267 replicates of DRV selection starting from a pool of 26 single resistance-associated mutation
268 variants (**Figure S5**). In one of the cultures the virus was lost during the escalation of inhibitor
269 concentration, suggesting the selection protocol provides strong selection pressure at or near
270 levels that can extinguish the virus. The sequence analysis for the other three cultures showed

271 that HIV-1 can evolve DRV resistance using both pathways. Of the replicate selections, two
272 out of the three selections followed the I84V pathway, with I50V in the remaining selection.
273 These results show there is a stochastic element in which resistance pathway is used under these
274 conditions of escalating selective pressure. These results are also consistent with the smallest
275 R1 group (even smaller than in the R1-1 series) in DRV being able to use both resistance
276 pathways, while the larger R1-2 is more selective for the I50V pathway.
277
278 **Linked versus shared mutations in evolution of high-level resistance through the two**
279 **pathways**
280 To examine the order in which the 8 to 14 mutations accumulated in the protease gene to
281 confer high level resistance, and determine if there was specificity between the two pathways
282 (I50V and I84V), the abundance data from multiple selections that ended in one or the other
283 pathway were pooled and examined sequentially at different levels of drug concentration. In
284 this analysis (**Figure** 6), the selections resulting in the I84V pathway point up, with I84V
285 reaching 100% penetrance, by definition. Similarly, those selections that fixed I50V are shown
286 pointing downward, with I50V reaching 100% penetrance.
287 The mutational data grouped by I50V and I84V penetrance show that the mutations are
288 often close in three-dimensional space. The I84V pathway shows a strong correlation with
289 V32I (specifically) and V82I, two hydrophobic residues that have a direct steric relationship
290 with residue 84 and most likely participate in hydrophobic sliding [**Figure 7A**] (Foulkes-
291 Murzycki et al., 2007, Ragland et al., 2014, Ragland et al., 2017, Mittal et al., 2012).
292 Hydrophobic repacking has been observed when I84V mutates and adopts an alternate rotamer
293 in the B chain which also affects the rotamer of residue 32 (Lockbaum et al., 2019). The V32I

294 mutation has been shown to work cooperatively with the L33F mutation to achieve higher levels
295 of resistance than either mutation on their own (Ragland et al., 2014), although L33F was
296 observed in both pathways. Active site hydrophobic packing is also altered with the I47V
297 mutation which is selected mostly in the I50V pathway, while L76V is unique to the I84V
298 pathway, although it appears at a low frequency. In addition to being in close proximity to
299 I84V, the V82I mutation is also near the L10F mutation, observed in both pathways [**Figure**
300 **7B**]. Lastly the I54L mutation has been shown to be critical in conferring very high levels of
301 drug resistance at the expense of catalytic efficiency, which is probably why that mutation is
302 only observed at high inhibitor concentrations (Henes et al., 2019).

303 While both pathways have an L33F mutation which adds steric bulk to the hydrophobic
304 core of the protease, the I50V mutation uniquely utilizes the I13V mutation to relieve that steric
305 pressure [**Figure 7B**]. Both pathways also have an M46I mutation a critical site for resistance
306 (Ragland et al., 2014) which modulates flap dynamics. Only the I50V pathway has the F53L
307 mutation that directly interacts with residue 46 on the outer surface of the flap, likely providing
308 flap stability. L10I, G16E, I47A, L76S, I85V, and L89T/I mutations appear at lower frequency,
309 making it challenging to assess if they are specific or critical to either pathway.

310 To further evaluate the loss in potency, HIV-1 protease variants with high levels of drug
311 resistance from viral selection were chosen to span the diversity in sequence, and represent the
312 I50V, I84V and I50V/I84V pathways. The protease variants were expressed/purified, and
313 enzymatic activity and inhibition assayed against DRV and the ten inhibitor analogs. Although
314 the enzymatic activity of some of the proteases is ~10-fold compromised relative to wildtype
315 (~17 $(s*\mu M)^{-1}$) some have retained near-WT activity (Sel-U5s-7Mut). The chosen set of 9
316 protease variants (**Table 2**) includes 4 that contain I50V (red), 3 that contain I84V (blue), and 2

317 that have both I84V/I50V (purple). These proteases contain 6–14 mutations relative to the
318 wildtype enzyme, and the potency of the inhibitors has dropped from pM range to 1–100 nM
319 (**Figure 8**). U5 and U10 retain potency against some of the variants, but all 11 inhibitors are
320 compromised by two highly resistant proteases. Very high levels of resistance occur with
321 proteases that the contain I50V pathway variants (red), the I84V pathway variants (blue) or the
322 variants with both I50V/I84V (purple).
323
324 **Table 2**: Highly mutated inhibitor resistant end-point protease variants chosen to quantify
325 inhibitor resistance and catalytic efficiency.

	L	V	I	I	V	L	N	K	K	M	I	I	F	I	L	I	A	I	G	T	L	V	I	I	L	T	Q	Avg. Inhib. Res.	Catalytic Eff.
	10	11	13	15	32	33	37	43	45	46	47	50	53	54	63	66	71	72	73	74	76	82	84	85	89	91	92	Ki (nM)	Kcat/Km (s*uM)^-1
Sel-U9s-6Mut	F									I		V	L		P				S									4.62 ± 0.51	3.53 ± 0.28
Sel-U3-12Mut	F					F				I	V	V	L		P		V				S	I		V	I			24.90 ± 1.18	0.56 ± 0.27
Sel-U8-10Mut	F									I	V	V	L		P			V	S			I		V				39.61 ± 1.98	1.91 ± 0.40
Sel-U9e-14Mut	F		V			F			R	I	V	V	L	L		F	V			A	S	I						40.36 ± 1.82	0.60 ± 0.18
Sel-U5s-7Mut	F				I	F			I								V					I	V					3.61 ± 0.23	17.94 ± 0.23
Sel-DRV3-10Mut	F				I	F	D			I					P					S	V		V		M			9.80 ± 0.82	9.14 ± 0.30
Sel-U4-14Mut	F	I	V		I	F		T		L				L			V					I	V		M	S	R	35.38 ± 1.84	1.04 ± 0.40
Sel-U2-9Mut	I				I	F			I	I		V					V					I	V					14.51 ± 0.57	10.67 ± 0.45
Sel-U5e-10Mut	F			V						I	V	V	L				V					I	V		T			85.85 ± 4.47	1.12 ± 0.29

326
327
328 **Analysis of Gag cleavage-site mutations**
329 HIV-1 protease is responsible for cleaving 10 different substrates during viral
330 maturation. Although these substrates do not have high sequence identity, the amino acids
331 corresponding to each cleavage site have a similar/conserved size and shape when bound to
332 protease active site (Prabu-Jeyabalan et al., 2002). Under inhibitor selective pressure, the
333 protease accumulates mutations that alter the active site, which may perturb the binding affinity
334 and processing of substrates. As the protease mutates to confer drug resistance, certain cleavage
335 sites are known to co-evolve to maintain protease binding affinity and the relative rates at which
336 the substrates are cleaved (Doyon et al., 1996, Zhang et al., 1997, Mammano et al., 1998, Kolli
337 et al., 2009a, Ozen et al., 2012b). Protease-substrate coevolution particularly occurs at the

338 cleavage sites flanking the spacer peptide SP2 in Gag (NC/SP2 and SP2/p6) (Prabu-Jeyabalan et
339 al., 2004, Kolli et al., 2006, Kolli et al., 2014, Kolli et al., 2009b, Lee et al., 2012, Ozen et al.,
340 2011, Ozen et al., 2012a, Ozen et al., 2014a). We sequenced the protease cleavage sites
341 encoded in the viral *gag* gene in the pools of selected viruses where the inhibitor concentration
342 had reached a level of greater than 1 μM [**Figure 9**].
343 An analysis of four cultures that had I84V as the major resistance mutation showed they all had
344 a mutation at the NC/SP2 cleavage site at position P2, with a change from the wild type alanine
345 amino acid to either of the larger hydrophobic amino acids valine or isoleucine. In addition,
346 three of the four I84V cultures had a mutation at the adjacent SP2/p6 cleavage site, either at P1'
347 (leucine to phenylalanine) or P5' (proline to leucine). Conversely, all seven cultures where the
348 I50V mutation was the major resistance mutation there was a mutation in the SP2/p6 cleavage
349 site, but not in the NC/SP2 site. Four of the seven cultures had leucine to phenylalanine
350 mutations at the P1' position, two had proline to leucine mutations at the P5' position, and one
351 had both of these mutations together. Finally, in the three cultures where the protease evolved
352 both the I50V and I84V mutations, Gag mutations were observed only at the SP2/p6 cleavage
353 site. One culture had the proline to leucine mutation at the P5' position, while the other two had
354 both the P1' (leucine to phenylalanine) and P5' (proline to leucine) mutations.
355 The patterns of protease-substrate coevolution from these cultures suggest two
356 phenomena are at work. First, P2 mutations in the NC/SP2 cleavage site are compensatory for
357 the I84V resistance mutation but are likely to be antagonistic for the I50V mutation, since they
358 do not appear in the cultures with I50V either alone or in combination with I84V. Second, the
359 effects of the SP2/p6 mutations at the P1' and P5' positions may be mechanistically related. The
360 P1' leucine to phenylalanine mutation increases the size of the P1' side chain and thus occupies

361 more space in the S1' subsite of the protease. Combining the I84V and I50V selections, nine of
362 ten cases have either the SP2/p6 P1' or the P5' position is mutated, with only one I50V culture
363 where they appear together. This suggests the proline to leucine P5' mutation may indirectly
364 increase wild type P1' leucine interactions with the S1' subsite. Consistent with this, we
365 previously found by solving crystal structures that the proline to leucine mutation at the P5′
366 position causes a distal conformational change in the protease flap and alters substrate–protease
367 interactions (Ozen et al., 2014b). When both I50V and I84V were present in the protease
368 together, the P1' and P5' mutations appeared together (in 2 of 3 cultures), which we would
369 predict further increased S1' subsite interactions to compensate for the smaller amino acids at
370 both protease residues, 50 and 84.

371 Previous work has shown that the I84V variant will more rapidly cleave NC/SP2 when
372 alanine is mutated to valine at position P2 (Kolli et al., 2009a). I50V was also shown to cleave
373 the SP2/p6 site more rapidly when leucine is mutated to phenylalanine at position P1' and when
374 proline is mutated to leucine at position P5' (Ozen et al., 2014b). To understand the molecular
375 basis of HIV-1 protease coevolution with SP2/p6 cleavage site mutations, crystal structures
376 were examined. In our model, when alanine was substituted for valine, we observed that the P2
377 residue in the NC/SP2 site occupies the vdW space in the substrate pocket previously filled with
378 a methyl group on isoleucine at position 84 of protease. This mutation is not observed in the
379 I50V pathway, possibly due to the fact that the isoleucine does not occupy the same space.
380 However, the SP2/p6 cleavage site mutations in P1' and P5' the I50V pathway result in
381 increased cleavage of SP2/p6, which would provide more starting product required for cleavage
382 of the NC/SP2 site, the least active site in all of Gag. Analyses of the protease–substrate
383 interactions indicated that restoration of active site dynamics is an additional constraint in the

384 selection of coevolved mutations. Additionally, compensatory coevolved mutations such as
385 ProP5'Leu in the substrate do not directly restore interactions lost due to protease mutations but
386 induce distal changes. Hence, protease–substrate coevolution permits mutational, structural,
387 and dynamic changes via molecular mechanisms that involve distal effects contributing to drug
388 resistance.
389
390 **Patterns of mutation beyond protease and cleavage sites that occur during resistance**
391 **selection**.
392 When HIV-1 evolves drug resistance, resistance is being evolved by the whole viral
393 system in the environment where the virus is replicating. Thus although HIV-1 protease
394 inhibitors target the viral protease, for viruses to attain high level resistance the entire virus
395 likely adapts to this selective pressure. Over recent years there have been a number of reports
396 of site mutations in the cleavage sites as well as other locations within the Gag polyprotein and
397 potentially Env gp41 (Doyon et al., 1996, Cote et al., 2001, Prabu-Jeyabalan et al., 2004, Kolli
398 et al., 2006, Banke et al., 2009, Dam et al., 2009, Parry et al., 2011). However the mechanism
399 by which these changes contribute to protease inhibitor drug resistance (Rabi et al., 2013) or
400 how co-evolution may otherwise compensate as the virus acquires high levels of drug resistance
401 is unknown.
402 Our viral selection experiments provide a unique opportunity to examine both the
403 mutations that occur both within the protease gene, and throughout the viral genome and
404 potential alterations in host response. Given the large number of selections performed we can
405 begin to elucidate the role of compensatory changes outside protease. Mutations that were
406 observed in more than 2 selections and/or involved a change in charge (shaded yellow) are

407 shown in **Figure 10**, with reversions to consensus subtype B or mutations observed in the no-
408 drug control (i.e. simply adaptation to tissue culture passage) not included. Overall this involved
409 changes at 99 different sites: 31 sites in Gag, 29 sites in RT/integrase, 30 sites in Envelope
410 (Env) and 9 in Vif. Only 5 changes were in cleavage sites, one in Nucleocapsid and 4 in p6
411 (**Figure 9**). Approximately 35% of the selection-associated mutations are consistent with
412 APOBEC3G/F (A3G/F) driven mutations. In total, 48 sites involved changes in charge, most
413 often making the resistant selected virus more positively charged. A total of 14 mutations were
414 observed in five or more drug selection experiments; these include: Capsid (V27I, Q67H,
415 P207S); Nucleocapsid (*R32K*); p6 (L1F, P5L, F17S); RT (*E194K*, *D237N*, *E297K*); Integrase
416 (D6H, *D41N*, M154I) and Vif (I31N) – (italics indicate likely APOBEC mutations, underlined
417 previously reported, and bold cleavage site mutation).

418 The Capsid mutation H87Q in the cyclophilin A (CypA) loop is well-known and allows
419 HIV-1 to escape the Trim5α restriction factor (Kootstra et al., 2007, Bosco et al., 2010). We
420 observed this mutation in our passaging experiments, including in the no-drug control indicating
421 an adaptation to cell culture (data not shown). Three other mutations were observed in 5 or more
422 of the viral drug selections in Capsid (V27I, Q67H, P207S) as described above. Using the
423 available crystal and cryoEM structures (Bhattacharya et al., 2014, Zhao et al., 2013), we analyzed
424 where these three positions are physically located within the Capsid structure (Fig **11**). All three
425 sites appear to be at pivotal locations: V27I is located in a region between the N and C-terminal
426 domains, facing a hydrophobic region that is not optimally packed (Fig **11**). The V27I mutation
427 may improve this packing and was previously observed to rescue infectivity (Rong et al., 2001).
428 This pocket is also targeted by a number of antiretroviral inhibitors which also elicit resistance as
429 V27A/I (Lemke et al., 2012). Q67H is located on a capsid-capsid interface within the capsid

430 hexamer. Modeling suggests that Q67H may act by improving inter-capsid monomer interactions
431 with Y169' and by forming intramolecular hydrogen bonds with Q63. Q67H was previously
432 observed to both confer resistance and enhance infectivity (Shi et al., 2015) to PF-3450074 (PF74)
433 which targets capsid assembly. P207S is located prominently at the pentameric interface between
434 capsid hexamers of the viral structure. Structurally, P207S can potentially form either direct or
435 water-mediated hydrogen bonds with the other subunits. The P207S mutation has been identified
436 as critical for evading the host restriction factors MxB (Busnadiego et al., 2014) and possibly
437 SUN2 (Donahue et al., 2016). Thus all three mutations we observed frequently within resistant
438 viruses have been previously associated with enhanced infectivity often by evading host factors.
439

440 **Discussion**

441 The development of anti-HIV-1 therapeutics has been a successful endeavor to control
442 viral replication and restore long term health to those living with HIV-1. To prevent the
443 emergence of resistance, combination therapies targeting multiple viral targets (RT, PR, IN) are
444 successfully used in the clinic. However, rather than the number of targets, the combined
445 potency of the drugs is important to effectively suppress viral replication. The initial
446 demonstration of suppressive therapy was accomplished with three drugs directed at two
447 targets. As the potency of the individual inhibitors has increased there has been an interest in
448 exploring reducing the number of drugs in a regimen. This includes initial suppression with a
449 combination of three drugs then maintenance therapy with fewer drugs. To date attempts at
450 maintenance with a single potent drug, an HIV-1 protease inhibitor, have been partially
451 successful ,with some people maintaining suppression while others experience virologic

452 rebound (Katlama et al., 2010, Pulido et al., 2011). Since incomplete suppression leads to
453 resistance, any strategy that can cause virologic failure/rebound is not tenable.
454 Virologic failure can result from several causes. In one case the virus is able to replicate
455 in the presence of subinhibitory concentrations of drug and evolve resistance, which leads to
456 higher levels of replication. This situation is easily recognized by the presence of resistance
457 mutations in the target gene. In another case there can be failure due to poor adherence leading
458 to uncontrolled virus growth and viral rebound without the presence of resistance mutations. A
459 more confusing situation is rebound without resistance mutations but under circumstances
460 where there is reason to believe adherence was high. The first two cases can be distinguished
461 by the presence or absence of resistance mutations, while the last case is a challenge to account
462 for. It is worth knowing that when DRV was clinically tested in monotherapy 85% of
463 participants maintained virologic suppression and those who did experience virologic failure
464 had no evidence of significant DRV resistance (Katlama et al., 2010, Pulido et al., 2011).
465 DRV and the UMASS series of inhibitors have EC50 values in the range of 1-10 nM in
466 cell culture, and DRV reaches a level of >1μM as the maximum plasma concentration *in vivo*,
467 in the range of 1000-fold over the EC50. In this manuscript we have selected for resistance to
468 DRV and to 10 analogues of DRV with similar or increased potency to the drug levels that can
469 be achieved *in vivo*. To select for viral replication at this level of drug the viral protease
470 incorporated between 8 and 14 mutations, remodeling over 10% of its entire sequence. In
471 culture this was achieved over 50-60 passages of the virus under conditions of escalating drug
472 concentration to allow the sequential addition of mutations. This is not how the virus
473 experiences drug selective pressure *in vivo*. There exposure to high levels of drug (relative to
474 the EC50) is achieved quickly and largely sustained. Under these circumstances there is no

475 opportunity for the virus to undergo the significant evolution required to fix the large number of
476 mutations needed for resistance to DRV. The near absence of resistance mutations in the
477 virologic failures in the DRV monotherapy trial (Katlama et al., 2010, Pulido et al., 2011)
478 suggests that there was selective problems with adherence in that arm, that the drug was
479 differentially metabolized in a subset of people such that there was virtually no systemic drug
480 exposure to the virus, or that there were compartments within the body that had very low drug
481 exposure and allowed the production of enough virus to appear in the blood as virologic failure.
482 Given the extremely large differential of EC50 and blood drug concentration it will be
483 important to distinguish among these reasons for virologic failure as dual therapy combinations
484 are entertained.

485 We tested 10 DRV analogues for pathways to resistance, in addition to DRV. Mutations
486 accumulated over most of the course of the increasing selective pressure and revealed two
487 distinct pathways to high-level resistance, i.e. the major resistance mutation I50V or I84V.
488 Replicate selections showed that HIV-1 can evolve PI resistance to these inhibitors using both
489 pathways, confirming that selections with the same inhibitor can produce different outcomes
490 and that variants maintain a dynamic behavior over the course of a selection. However, with the
491 largest P1'-equivalent moiety we found a strong preference for the I50V pathway while the
492 smaller P1'-equivalent moieties were able to utilize either the I50V or the I84V pathway to high
493 level resistance. In the first set of selections the I50V mutation was not included in the starting
494 mixture, while the second selection was started with a homogeneous unmutated population.
495 Thus these cultures were not limited in their ability to select among the familiar resistance
496 mutation pathways even with different starting points. Selection to >1 μM inhibitor
497 concentration resulted in broad cross resistance across the entire panel of inhibitors.

498 Although these pathways generally developed independently, they are not mutually
499 exclusive. Several highly-selected cultures assembled linked "hybrid" variants with both I50V
500 and I84V at the higher drug concentrations. Although there were only a few examples of this,
501 they appeared to add I84V into the I50V pathway rather than the reverse.
502 In all of the cultures with the early appearance of the I84V mutation, this was coincident
503 with or quickly followed by the addition of a mutation at position 32. Mutation V32I surfaced
504 early in each culture, simultaneously or after the mutation at I84V. Mutations selected at the
505 highest drug concentrations show I84V linked to mutations at positions 10, 33, 46, 71, and 82.
506 Early appearance of mutations at I84V are associated with peaks in entropy, reflecting high
507 diversity at that time point. This was followed by a decrease in entropy when the mutation
508 becomes fixed as other populations died off. There appears to be a temporal order in one
509 pathway where I50V is added first followed by I47V then F53L and I13V (or I85V). The I50V
510 mutation was not associated with drops in entropy like the 84V mutation, and was followed by
511 an increase entropy with higher drug concentrations. The selection for specific mutations could
512 be interpreted based on structural studies.
513 Protease cleavage site mutations are known to coevolve with protease inhibitor
514 resistance within the protease itself REF. The two resistance pathways were associated with
515 distinctive patterns of evolution within the NC/SP2 and SP2/p6 cleavage sites (Doyon et al.,
516 1996, Zhang et al., 1997, Mammano et al., 1998). The NC/SP2 site has a suboptimal alanine at
517 the P2 site in the cleavage site sequence, with the resistance associated mutation placing a more
518 favorable valine or isoleucine in that position to make the cleavage site sequence more
519 favorable (Potempa et al., 2018). This interaction can be accounted for in the structure of
520 protease bound to substrate. However, selection at this position does not occur in the I50V

521 pathway, even in those viruses where both I50V and I84V are present, suggesting an
522 antagonistic effect on I50V or an absence of any benefit in cleavage site rate with this P2
523 change. In contrast, the mutations in the P1' site and the P5' site seem to have similar effects
524 even though the P5' change is outside of the cleavage recognition sequence. The P1' change in
525 the SP2/p6 site is leucine to phenylalanine, with larger hydrophobic amino acids in the P1'
526 position being preferred (Pettit et al., 2002). The P5' change in this site from proline to leucine
527 must effect a similar change by allowing the P1' leucine to move further into the S1' subsite to
528 improve the rate of cleavage. In the most resistant viruses, with both I50V and I84V, both the
529 P1' and P5' mutations appeared together.

530 We found an array of mutations across the genome in the resistant viruses. The presence
531 of mutations that appear to be the result of APOBEC3G modifications suggest the cultures went
532 through significant bottlenecks to fix these likely deleterious mutations. In contrast, some
533 sequence changes represented reversion to the consensus subtype B mutations, suggesting
534 selection for improved fitness. A more interesting set of mutations were in the capsid sequence
535 frequently appearing at subunit interfaces. This raises the possibility that one of the
536 compensatory mechanisms of a less active protease may be in the subunit recognition/assembly
537 of the capsid.

538 In this manuscript we have examined the evolutionary pathways that confer high level
539 resistance to DRV and a series of DRV variants. We showed that resistance to the level of
540 DRV that is achieved in plasma *in vivo* requires extensive mutagenesis both within the outside
541 of the protease. These levels of evolution are not attainable during the rapid decline of the viral
542 population size with the initiation of multidrug therapy nor likely to occur during sporadic,
543 isolated viral replication in tissue. While monotherapy even with potent drugs has not been as

544 robust in achieving viral suppression as triple drug therapy, there is potential for two drug
545 therapies. To date dual therapy often includes the nucleoside analog 3TC which has a very low
546 genetic barrier (a single mutation M184V in RT). A more reasonable strategy would be to pair
547 two drugs where both drugs have a high genetic barrier, with a protease inhibitor such as DRV
548 as an obvious choice. Finally, given the high level of drug that can be obtained in the blood
549 there is the potential to use DRV as a platform for a fifth generation of protease inhibitors that
550 have additional properties such as reduced protein binding or enhance blood-brain-barrier
551 penetration.
552

553 **Materials and Methods**

554 **Cell lines and viruses.**

555 CEMx174 cells were maintained in RPMI 1640 medium with 10% fetal calf serum and
556 penicillin-streptomycin. TZM-bl and 293T cells were maintained in Dulbecco's modified
557 Eagle-H medium supplemented with 10% fetal calf serum and penicillin-streptomycin.
558 CEMx174 cell line was obtained from the National Institutes of Health AIDS Research and
559 Reference Reagent Program. A wild-type virus stock NL4-3 was prepared by transfection of
560 the pNL4-3 plasmid (purified using the Qiagen plasmid Maxikit) into HeLa cells.

561

562 **Selections.**

563 An aliquot of 3 x 10^6 CEMx174 cells was incubated at 37°C for 2 to 3 h with 250 µl of a virus
564 stock generated from the HIV-1 infectious molecular clone pNL4-3. The culture volume was
565 then brought to 10 ml with RPMI medium. Each flask received one of the following inhibitors
566 at escalating concentrations: UMass1, UMass2, UMass3, UMass4, UMass5, UMass6, UMass7,
567 UMass8, UMass9, UMass10, DRV and no drug (ND). After 48 h and every 48 h after, the cells
568 were pelleted by centrifugation and 10ml of fresh medium and inhibitors were added. When the
569 culture had undergone extensive cytopathic effect (CPE) indicative of viral replication, the
570 supernatant medium and the cells were harvested separately and stored at -80°C. The virus-
571 containing supernatant was used to start the next round of infection, and after several rounds at
572 the initial concentration, the inhibitor concentration was increased 1.5-fold at each subsequent
573 round of virus passage. The level of resistance (50% inhibitory concentration [EC50]) of the
574 single inhibitor-selected virus pools was determined by a TZM infection assay in which the

575 protease inhibitor is added to productively infected cells and the titers of supernatant virus made
576 in the presence of the inhibitor are determined.
577
578 **TZM Infection Assay**
579 Protease inhibitor dilutions were prepared by taking 10 µM stocked and performing a 5-fold
580 serial dilution using RPMI media (final drug concentration is 100 µM). One dilution of drug
581 was added to each well of a 24 well plate and repeated so each virus would have a full set of
582 dilutions. Viruses for the assay were made by seeding 3 x 10^6 CEM cells in a 24 well plate and
583 incubating with 250 µl of virus at 37°C for 2 to 3 h before bringing the culture to 10ml with
584 RPMI media. After 48 h the medium was changed and repeated every 48 h after until the
585 culture had undergone CPE. Infected CEM cells were collected and diluted so that 1ml of cells
586 could be plated in each well containing a unique drug dilution. Then 24 hours later the virus
587 supernatant was collected from each well followed by filtering through a 0.45 µM filter then
588 placed in -80°C. Viruses were thawed and added to 96 well plates in triplicate. TZM-bl cells
589 were collected and diluted to a concentration of $2x10^5$ cells/ml, 100 µl were added on top of the
590 pre-plated viruses. Plates were kept in 37°C, 5% CO_2 in an incubator for 48 hours. After 48
591 hours, the cells in the plates were lysed by removing the medium, washing two times with 100
592 µl PBS and then lysed with 1x lysis buffer (made from 5x Promega Firefly Lysis Buffer).
593 Plates were frozen for at least 24 hours and then thawed for 2 hours before analyzing with
594 Promega Firefly Luciferase Kit on a luminometer.
595
596 **DNA preparation and amplification of the protease-coding region.**

597 Total cellular DNA was isolated from infected cell pellets by using the QIAamp blood kit
598 (Qiagen). The protease-coding domain of viral DNA was amplified by nested PCR. The PCR
599 conditions are available on request. PCR products were purified by using QIAquick PCR
600 purification kit (Qiagen) and directly sequenced or cloned into the pT7Blue vector (Novagen)
601 and sequenced.
602
603 **Primer-ID Deep Sequencing of viral RNA**
604 We used the PID protocol to prepare MiSeq PID libraries with multiplexed primers. Viral RNA
605 was extracted from plasma samples using the QIAamp viral RNA mini kit (Qiagen, Hilden,
606 Germany). Complementary DNA (cDNA) was synthesized using a cDNA primer mixture
607 targeting protease (PR) with a block of random nucleotides in each cDNA primer serving as the
608 PID, and SuperScript III RT (ThermoFisher). After 2 rounds of bead purification of the cDNA,
609 we amplified the cDNA using a mixture of a forward primer that targeted the upstream coding
610 region, followed by a second round of PCR to incorporate the Illumina adaptor sequences. Gel-
611 purified libraries were pooled and sequenced using the MiSeq 300 base paired-end sequencing
612 protocol (Illumina). The sequencing covered the HIV-1 PR region (HXB2 2648–2914, 3001–
613 3257).
614 We used the Illumina bcl2fastq pipeline for the initial processing and constructed
615 template consensus sequences (TCSs) with TCS pipeline version 1.33
616 (https://github.com/SwanstromLab/PID). We then aligned TCSs to an HXB2 reference to
617 remove sequences not at the targeted region or that had large deletions.
618
619 **Protease expression and purification**

620 The highly mutated, resistant, protease variant genes were purchased on a pET11a plasmid with
621 codon optimization for protein expression in *E. coli* (Genewiz). A Q7K mutation was included
622 to prevent autoproteolysis (Rose et al., 1993). The expression, isolation, and purification of WT
623 and mutant HIV-1 proteases used for enzymatic assays were carried out as previously described
624 (Ozen et al., 2014b, King et al., 2002, Henes et al., 2019). Briefly, the gene encoding the
625 desired HIV-1 protease was subcloned into the heat-inducible pXC35 expression vector
626 (ATCC) and transformed into *E. coli* TAP-106 cells. Cells grown in 6 L of Terrific Broth were
627 lysed with a cell disruptor twice, and the protein was purified from inclusion bodies (Hui et al.,
628 1993). Inclusion bodies, isolated as a pellet after centrifugation, were dissolved in 50% acetic
629 acid followed by another round of centrifugation at 19,000 rpm for 30 minutes to remove
630 insoluble impurities. Size exclusion chromatography was carried out on a 2.1-L Sephadex G-75
631 Superfine (Sigma Chemical) column equilibrated with 50% acetic acid to separate high
632 molecular weight proteins from the desired protease. Pure fractions of HIV-1 protease were
633 refolded using a 10-fold dilution of refolding buffer (0.05 M sodium acetate at pH 5.5, 5%
634 ethylene glycol, 10% glycerol, and 5 mM DTT). Folded protein was concentrated to 0.5–3
635 mg/ml and stored. The stored protease was used in K_M and K_i assays.
636

637 **Enzymatic Assays**

638 ***K_m Assay.*** K_m values were determined as previously described (Lockbaum et al., 2019, Henes et
639 al., 2019, Windsor and Raines, 2015, Matayoshi et al., 1990). Briefly, a 10-amino acid
640 substrate containing the natural MA/CA cleavage site with an EDANS/DABCYL FRET pair
641 was dissolved in 8% DMSO at 40 nM and 6% DMSO at 30 nM. The 30 nM substrate was 4/5
642 serially diluted from 30 nM to 6 nM. HIV-1 protease was diluted to 120 nM and, and 5 µl were

643 added to the 96-well plate to obtain a final concentration of 10 nM. Fluorescence was observed
644 using a PerkinElmer Envision plate reader with an excitation at 340 nm and emission at 492 nm,
645 and monitored for 200 counts. A FRET inner filter effect correction was applied as previously
646 described (Liu et al., 1999). Data corrected for the inner filter effect was analyzed with Prism7.
647
648 ***K_i Assay.*** Enzyme inhibition constants (K_i values) were determined as previously described
649 (Lockbaum et al., 2019, Henes et al., 2019, Windsor and Raines, 2015, Matayoshi et al., 1990).
650 Briefly, in a 96-well plate, inhibitors were serially diluted down from 2000-10,000 nM
651 depending on protease resistance. All samples were incubated with 5 nM protein for 1 hour. A
652 10-amino acid substrate containing an optimized protease cleavage site (Windsor and Raines,
653 2015), purchased from Bachem, with an EDANS/DABCYL FRET pair was dissolved in 4%
654 DMSO at 120 mM. Using a PerkinElmer Envision plate reader, 5 µl of the 120 mM substrate
655 were added to the 96-well plate to a final concentration of 10 mM. Fluorescence was observed
656 with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts. Data was
657 analyzed with Prism7.
658

659 **References**

660

661 ALI, A., BANDARANAYAKE, R. M., CAI, Y., KING, N. M., KOLLI, M., MITTAL, S.,
662 MURZYCKI, J. F., NALAM, M. N., NALIVAIKA, E. A., OZEN, A., PRABU-
663 JEYABALAN, M. M., THAYER, K. & SCHIFFER, C. A. 2010. Molecular Basis for
664 Drug Resistance in HIV-1 Protease. *Viruses,* 2**,** 2509-35.
665 BANKE, S., LILLEMARK, M. R., GERSTOFT, J., OBEL, N. & JORGENSEN, L. B. 2009.
666 Positive selection pressure introduces secondary mutations at Gag cleavage sites in
667 human immunodeficiency virus type 1 harboring major protease resistance mutations. *J*
668 *Virol,* 83**,** 8916-24.
669 BHATTACHARYA, A., ALAM, S. L., FRICKE, T., ZADROZNY, K., SEDZICKI, J.,
670 TAYLOR, A. B., DEMELER, B., PORNILLOS, O., GANSER-PORNILLOS, B. K.,
671 DIAZ-GRIFFERO, F., IVANOV, D. N. & YEAGER, M. 2014. Structural basis of HIV-
672 1 capsid recognition by PF74 and CPSF6. *Proc Natl Acad Sci U S A,* 111**,** 18625-30.
673 BOSCO, D. A., EISENMESSER, E. Z., CLARKSON, M. W., WOLF-WATZ, M.,
674 LABEIKOVSKY, W., MILLET, O. & KERN, D. 2010. Dissecting the microscopic
675 steps of the cyclophilin A enzymatic cycle on the biological HIV-1 capsid substrate by
676 NMR. *J Mol Biol,* 403**,** 723-38.
677 BUSNADIEGO, I., KANE, M., RIHN, S. J., PREUGSCHAS, H. F., HUGHES, J., BLANCO-
678 MELO, D., STROUVELLE, V. P., ZANG, T. M., WILLETT, B. J., BOUTELL, C.,
679 BIENIASZ, P. D. & WILSON, S. J. 2014. Host and viral determinants of Mx2
680 antiretroviral activity. *J Virol,* 88**,** 7738-52.
681 COTE, H. C., BRUMME, Z. L. & HARRIGAN, P. R. 2001. Human immunodeficiency virus
682 type 1 protease cleavage site mutations associated with protease inhibitor cross-
683 resistance selected by indinavir, ritonavir, and/or saquinavir. *J Virol,* 75**,** 589-94.
684 DAM, E., QUERCIA, R., GLASS, B., DESCAMPS, D., LAUNAY, O., DUVAL, X.,
685 KRAUSSLICH, H. G., HANCE, A. J., CLAVEL, F. & GROUP, A. S. 2009. Gag
686 mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-
687 experienced patients besides compensating for fitness loss. *PLoS Pathog,* 5**,** e1000345.
688 DE MEYER, S., AZIJN, H., SURLERAUX, D., JOCHMANS, D., TAHRI, A., PAUWELS, R.,
689 WIGERINCK, P. & DE BETHUNE, M. P. 2005. TMC114, a novel human
690 immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-
691 resistant viruses, including a broad range of clinical isolates. *Antimicrob Agents*
692 *Chemother,* 49**,** 2314-21.
693 DONAHUE, D. A., AMRAOUI, S., DI NUNZIO, F., KIEFFER, C., PORROT, F., OPP, S.,
694 DIAZ-GRIFFERO, F., CASARTELLI, N. & SCHWARTZ, O. 2016. SUN2
695 Overexpression Deforms Nuclear Shape and Inhibits HIV. *J Virol,* 90**,** 4199-4214.
696 DOYON, L., CROTEAU, G., THIBEAULT, D., POULIN, F., PILOTE, L. & LAMARRE, D.
697 1996. Second locus involved in human immunodeficiency virus type 1 resistance to
698 protease inhibitors. *J Virol,* 70**,** 3763-9.
699 FOULKES-MURZYCKI, J. E., SCOTT, W. R. & SCHIFFER, C. A. 2007. Hydrophobic
700 sliding: a possible mechanism for drug resistance in human immunodeficiency virus
701 type 1 protease. *Structure,* 15**,** 225-33.

702 GROSSMAN, Z., PAXINOS, E. E., AVERBUCH, D., MAAYAN, S., PARKIN, N. T.,
703 ENGELHARD, D., LORBER, M., ISTOMIN, V., SHAKED, Y., MENDELSON, E.,
704 RAM, D., PETROPOULOS, C. J. & SCHAPIRO, J. M. 2004. Mutation D30N is not
705 preferentially selected by human immunodeficiency virus type 1 subtype C in the
706 development of resistance to nelfinavir. *Antimicrob Agents Chemother,* 48**,** 2159-65.
707 GULICK, R. M., MELLORS, J. W., HAVLIR, D., ERON, J. J., GONZALEZ, C.,
708 MCMAHON, D., RICHMAN, D. D., VALENTINE, F. T., JONAS, L., MEIBOHM, A.,
709 EMINI, E. A. & CHODAKEWITZ, J. A. 1997. Treatment with indinavir, zidovudine,
710 and lamivudine in adults with human immunodeficiency virus infection and prior
711 antiretroviral therapy. *N Engl J Med,* 337**,** 734-9.
712 HENDERSON, G. J., LEE, S. K., IRLBECK, D. M., HARRIS, J., KLINE, M., POLLOM, E.,
713 PARKIN, N. & SWANSTROM, R. 2012. Interplay between single resistance-associated
714 mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor
715 sensitivity. *Antimicrob Agents Chemother,* 56**,** 623-33.
716 HENES, M., LOCKBAUM, G. J., KOSOVRASTI, K., LEIDNER, F., NACHUM, G. S.,
717 NALIVAIKA, E. A., LEE, S. K., SPIELVOGEL, E., ZHOU, S., SWANSTROM, R.,
718 BOLON, D. N. A., KURT YILMAZ, N. & SCHIFFER, C. A. 2019. Picomolar to
719 Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring
720 Drug Resistance. *ACS Chem Biol*.
721 HUI, J. O., TOMASSELLI, A. G., REARDON, I. M., LULL, J. M., BRUNNER, D. P.,
722 TOMICH, C. S. & HEINRIKSON, R. L. 1993. Large scale purification and refolding of
723 HIV-1 protease from Escherichia coli inclusion bodies. *J Protein Chem,* 12**,** 323-7.
724 KATLAMA, C., VALANTIN, M. A., ALGARTE-GENIN, M., DUVIVIER, C., LAMBERT-
725 NICLOT, S., GIRARD, P. M., MOLINA, J. M., HOEN, B., PAKIANATHER, S.,
726 PEYTAVIN, G., MARCELIN, A. G. & FLANDRE, P. 2010. Efficacy of
727 darunavir/ritonavir maintenance monotherapy in patients with HIV-1 viral suppression:
728 a randomized open-label, noninferiority trial, MONOI-ANRS 136. *AIDS,* 24**,** 2365-74.
729 KATOH, I., YASUNAGA, T., IKAWA, Y. & YOSHINAKA, Y. 1987. Inhibition of retroviral
730 protease activity by an aspartyl proteinase inhibitor. *Nature,* 329**,** 654-6.
731 KEMPF, D. J., RODE, R. A., XU, Y., SUN, E., HEATH-CHIOZZI, M. E., VALDES, J.,
732 JAPOUR, A. J., DANNER, S., BOUCHER, C., MOLLA, A. & LEONARD, J. M. 1998.
733 The duration of viral suppression during protease inhibitor therapy for HIV-1 infection
734 is predicted by plasma HIV-1 RNA at the nadir. *AIDS,* 12**,** F9-14.
735 KING, N. M., MELNICK, L., PRABU-JEYABALAN, M., NALIVAIKA, E. A., YANG, S. S.,
736 GAO, Y., NIE, X., ZEPP, C., HEEFNER, D. L. & SCHIFFER, C. A. 2002. Lack of
737 synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci,* 11**,**
738 418-29.
739 KOLLI, M., LASTERE, S. & SCHIFFER, C. A. 2006. Co-evolution of nelfinavir-resistant
740 HIV-1 protease and the p1-p6 substrate. *Virology,* 347**,** 405–409.
741 KOLLI, M., OZEN, A., KURT-YILMAZ, N. & SCHIFFER, C. A. 2014. HIV-1 protease-
742 substrate coevolution in nelfinavir resistance. *J. Virol.,* 88**,** 7145–7154.
743 KOLLI, M., STAWISKI, E., CHAPPEY, C. & SCHIFFER, C. A. 2009a. Human
744 immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance
745 inhibitor resistance. *J Virol,* 83**,** 11027-42.

746 KOLLI, M., STAWISKI, E., CHAPPEY, C. & SCHIFFER, C. A. 2009b. Human
747 immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance
748 inhibitor resistance. *J. Virol.,* 83**,** 11027–11042.
749 KOOTSTRA, N. A., NAVIS, M., BEUGELING, C., VAN DORT, K. A. & SCHUITEMAKER,
750 H. 2007. The presence of the Trim5alpha escape mutation H87Q in the capsid of late
751 stage HIV-1 variants is preceded by a prolonged asymptomatic infection phase. *AIDS,*
752 21**,** 2015-23.
753 LEE, S. K., POTEMPA, M., KOLLI, M., OZEN, A., SCHIFFER, C. A. & SWANSTROM, R.
754 2012. Context surrounding processing sites is crucial in determining cleavage rate of a
755 subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral
756 protease. *J. Biol. Chem.,* 287**,** 13279–13290.
757 LEMKE, C. T., TITOLO, S., VON SCHWEDLER, U., GOUDREAU, N., MERCIER, J. F.,
758 WARDROP, E., FAUCHER, A. M., COULOMBE, R., BANIK, S. S., FADER, L.,
759 GAGNON, A., KAWAI, S. H., RANCOURT, J., TREMBLAY, M., YOAKIM, C.,
760 SIMONEAU, B., ARCHAMBAULT, J., SUNDQUIST, W. I. & MASON, S. W. 2012.
761 Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site
762 within the N-terminal domain of the viral CA protein. *J Virol,* 86**,** 6643-55.
763 LIU, Y., KATI, W., CHEN, C. M., TRIPATHI, R., MOLLA, A. & KOHLBRENNER, W.
764 1999. Use of a fluorescence plate reader for measuring kinetic parameters with inner
765 filter effect correction. *Anal Biochem,* 267**,** 331-5.
766 LOCKBAUM, G. J., LEIDNER, F., RUSERE, L. N., HENES, M., KOSOVRASTI, K.,
767 NACHUM, G. S., NALIVAIKA, E. A., ALI, A., YILMAZ, N. K. & SCHIFFER, C. A.
768 2019. Structural Adaptation of Darunavir Analogues against Primary Mutations in HIV-
769 1 Protease. *ACS Infect Dis,* 5**,** 316-325.
770 MAMMANO, F., PETIT, C. & CLAVEL, F. 1998. Resistance-associated loss of viral fitness in
771 human immunodeficiency virus type 1: phenotypic analysis of protease and gag
772 coevolution in protease inhibitor-treated patients. *J Virol,* 72**,** 7632-7.
773 MATAYOSHI, E. D., WANG, G. T., KRAFFT, G. A. & ERICKSON, J. 1990. Novel
774 fluorogenic substrates for assaying retroviral proteases by resonance energy transfer.
775 *Science,* 247**,** 954-8.
776 MITTAL, S., BANDARANAYAKE, R. M., KING, N. M., PRABU-JEYABALAN, M.,
777 NALAM, M. N., NALIVAIKA, E. A., YILMAZ, N. K. & SCHIFFER, C. A. 2013.
778 Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance
779 in HIV-1 protease with mutations at residue 50. *J Virol,* 87**,** 4176-84.
780 MITTAL, S., CAI, Y., NALAM, M. N., BOLON, D. N. & SCHIFFER, C. A. 2012.
781 Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. *J Am Chem*
782 *Soc,* 134**,** 4163-8.
783 NALAM, M. N., ALI, A., REDDY, G. S., CAO, H., ANJUM, S. G., ALTMAN, M. D.,
784 YILMAZ, N. K., TIDOR, B., RANA, T. M. & SCHIFFER, C. A. 2013. Substrate
785 envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chem Biol,*
786 20**,** 1116-24.
787 NTEMGWA, M., BRENNER, B. G., OLIVEIRA, M., MOISI, D. & WAINBERG, M. A. 2007.
788 Natural polymorphisms in the human immunodeficiency virus type 2 protease can
789 accelerate time to development of resistance to protease inhibitors. *Antimicrob Agents*
790 *Chemother,* 51**,** 604-10.

791 OZEN, A., HALILOGLU, T. & SCHIFFER, C. A. 2011. Dynamics of preferential substrate
792 recognition in HIV-1 protease: redefining the substrate envelope. *J. Mol. Biol.,* 410**,**
793 726–744.
794 OZEN, A., HALILOGLU, T. & SCHIFFER, C. A. 2012a. HIV-1 Protease and Substrate
795 Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern. *J.*
796 *Chem. Theory Comput.,* 8**,** 703–714.
797 OZEN, A., HALILOGLU, T. & SCHIFFER, C. A. 2012b. HIV-1 Protease and Substrate
798 Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern. *J*
799 *Chem Theory Comput,* 8.
800 OZEN, A., LIN, K. H., KURT YILMAZ, N. & SCHIFFER, C. A. 2014a. Structural basis and
801 distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc.*
802 *Natl. Acad. Sci. U. S. A.,* 111**,** 15993–15998.
803 OZEN, A., LIN, K. H., KURT YILMAZ, N. & SCHIFFER, C. A. 2014b. Structural basis and
804 distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc*
805 *Natl Acad Sci U S A,* 111**,** 15993-8.
806 PARRY, C. M., KOLLI, M., MYERS, R. E., CANE, P. A., SCHIFFER, C. & PILLAY, D.
807 2011. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and
808 replication capacity. *Antimicrob. Agents Chemother.,* 55**,** 1106–1113.
809 PAULSEN, J. L., LEIDNER, F., RAGLAND, D. A., KURT YILMAZ, N. & SCHIFFER, C. A.
810 2017. Interdependence of Inhibitor Recognition in HIV-1 Protease. *J Chem Theory*
811 *Comput,* 13**,** 2300-2309.
812 PETTIT, S. C., HENDERSON, G. J., SCHIFFER, C. A. & SWANSTROM, R. 2002.
813 Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag
814 processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol,*
815 76**,** 10226-33.
816 POTEMPA, M., LEE, S. K., KURT YILMAZ, N., NALIVAIKA, E. A., ROGERS, A.,
817 SPIELVOGEL, E., CARTER, C. W., JR., SCHIFFER, C. A. & SWANSTROM, R.
818 2018. HIV-1 Protease Uses Bi-Specific S2/S2' Subsites to Optimize Cleavage of Two
819 Classes of Target Sites. *J Mol Biol,* 430**,** 5182-5195.
820 PRABU-JEYABALAN, M., NALIVAIKA, E. & SCHIFFER, C. A. 2002. Substrate shape
821 determines specificity of recognition for HIV-1 protease: analysis of crystal structures of
822 six substrate complexes. *Structure,* 10**,** 369-81.
823 PRABU-JEYABALAN, M., NALIVAIKA, E. A., KING, N. M. & SCHIFFER, C. A. 2004.
824 Structural basis for coevolution of a human immunodeficiency virus type 1
825 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J.*
826 *Virol.,* 78**,** 12446–12454.
827 PULIDO, F., ARRIBAS, J. R., HILL, A., VAN DELFT, Y. & MOECKLINGHOFF, C. 2011.
828 Analysis of drug resistance during HIV RNA viraemia in the MONET trial of
829 darunavir/ritonavir monotherapy. *Antivir Ther,* 16**,** 59-65.
830 RABI, S. A., LAIRD, G. M., DURAND, C. M., LASKEY, S., SHAN, L., BAILEY, J. R.,
831 CHIOMA, S., MOORE, R. D. & SILICIANO, R. F. 2013. Multi-step inhibition explains
832 HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest,* 123**,** 3848-60.
833 RAGLAND, D. A., NALIVAIKA, E. A., NALAM, M. N., PRACHANRONARONG, K. L.,
834 CAO, H., BANDARANAYAKE, R. M., CAI, Y., KURT-YILMAZ, N. & SCHIFFER,
835 C. A. 2014. Drug resistance conferred by mutations outside the active site through

836 alterations in the dynamic and structural ensemble of HIV-1 protease. *J Am Chem Soc,*
837 136**,** 11956-63.
838 RAGLAND, D. A., WHITFIELD, T. W., LEE, S. K., SWANSTROM, R., ZELDOVICH, K. B.,
839 KURT-YILMAZ, N. & SCHIFFER, C. A. 2017. Elucidating the Interdependence of
840 Drug Resistance from Combinations of Mutations. *J Chem Theory Comput,* 13**,** 5671-
841 5682.
842 RICHARDS, A. D., ROBERTS, R., DUNN, B. M., GRAVES, M. C. & KAY, J. 1989.
843 Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic
844 proteinases. *FEBS Lett,* 247**,** 113-7.
845 RONG, L., RUSSELL, R. S., HU, J., GUAN, Y., KLEIMAN, L., LIANG, C. & WAINBERG,
846 M. A. 2001. Hydrophobic amino acids in the human immunodeficiency virus type 1 p2
847 and nucleocapsid proteins can contribute to the rescue of deleted viral RNA packaging
848 signals. *J Virol,* 75**,** 7230-43.
849 ROSE, J. R., SALTO, R. & CRAIK, C. S. 1993. Regulation of autoproteolysis of the HIV-1 and
850 HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem,* 268**,** 11939-45.
851 SEELMEIER, S., SCHMIDT, H., TURK, V. & VON DER HELM, K. 1988. Human
852 immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin
853 A. *Proc Natl Acad Sci U S A,* 85**,** 6612-6.
854 SHI, J., ZHOU, J., HALAMBAGE, U. D., SHAH, V. B., BURSE, M. J., WU, H., BLAIR, W.
855 S., BUTLER, S. L. & AIKEN, C. 2015. Compensatory substitutions in the HIV-1 capsid
856 reduce the fitness cost associated with resistance to a capsid-targeting small-molecule
857 inhibitor. *J Virol,* 89**,** 208-19.
858 WATKINS, T., RESCH, W., IRLBECK, D. & SWANSTROM, R. 2003. Selection of high-level
859 resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob*
860 *Agents Chemother,* 47**,** 759-69.
861 WINDSOR, I. W. & RAINES, R. T. 2015. Fluorogenic Assay for Inhibitors of HIV-1 Protease
862 with Sub-picomolar Affinity. *Sci Rep,* 5**,** 11286.
863 YILMAZ, A., IZADKHASHTI, A., PRICE, R. W., MALLON, P. W., DE MEULDER, M.,
864 TIMMERMAN, P. & GISSLEN, M. 2009. Darunavir concentrations in cerebrospinal
865 fluid and blood in HIV-1-infected individuals. *AIDS Res Hum Retroviruses,* 25**,** 457-61.
866 ZHANG, Y. M., IMAMICHI, H., IMAMICHI, T., LANE, H. C., FALLOON, J.,
867 VASUDEVACHARI, M. B. & SALZMAN, N. P. 1997. Drug resistance during
868 indinavir therapy is caused by mutations in the protease gene and in its Gag substrate
869 cleavage sites. *J Virol,* 71**,** 6662-70.
870 ZHAO, G., PERILLA, J. R., YUFENYUY, E. L., MENG, X., CHEN, B., NING, J., AHN, J.,
871 GRONENBORN, A. M., SCHULTEN, K., AIKEN, C. & ZHANG, P. 2013. Mature
872 HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics.
873 *Nature,* 497**,** 643-6.
874 ZHOU, S., JONES, C., MIECZKOWSKI, P. & SWANSTROM, R. 2015. Primer ID Validates
875 Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation
876 Sequencing of HIV-1 Genomic RNA Populations. *J Virol,* 89**,** 8540-55.
877

878

879

Figures

1st Selection: Most Abundant Variants Detected at the Last Time Point

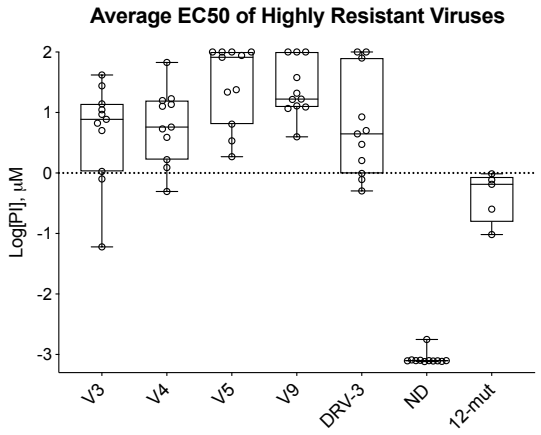
Passage #	Drug Conc. (nM)	R1	R2	PIs	Most Abundant Variants Detected at the Last Time Point	Abundance
53	4000	1	2	UMASS-2	10I 28S 32I 33F 46I 71V 82I 84V	31.0%
45	1275	1	4	UMASS-4	10F 28S 46I 84V	9.4%
53	5000	2	1	UMASS-6	13V 16E 32I 33F 45I 46I 54L 71V 76V 82F 84V	57.4%
58	400	2	2	UMASS-7	10F 46I 50V 63P 85V	21.0%
61	1300	2	3	UMASS-8	10F 46I 47V 50V 53L 63P 72V 73S 82I 85V	63.9%
39	3000	-	1	DRV	13V 16E 32I 33F 45I 46I 82F 84V	45.3%
61				ND	93L	36.7%

2nd Selection: Most Abundant Variants Detected at the Last Time Point

Passage #	Drug Conc. (nM)	R1	R2	PIs	Most Abundant Variants Detected at the Last Time Point	Abundance
75	5000	1	1	UMASS-1	10I 16E 32I 33F 46I 54L 71V 76V 82I 84V	55.4%
82	5000	1	2	UMASS-2	10F 33F 46I 47V 50V 71V 82I 84V	37.5%
93	5000	1	3	UMASS-3	10F 33F 46I 47V 50V 53L 63P 71V 76S 82I 85V 89I	85.7%
81	5000	1	4	UMASS-4	10F 11I 13V 32I 33F 43T 46L 54L 71V 82I 84V 89M 91S 92R	87.3%
77	5000	1	5	UMASS-5	10F 15V 46I 47V 50V 53L 71V 82I 84V 89T	67.9%
46	5000	2	1	UMASS-6	10F 13V 33F 46I 47A 50V 53L 71V	45.1%
76	5000	2	2	UMASS-7	10F 13V 33F 46I 47V 50V 53L 54L 71V 82I	94.5%
67	5000	2	3	UMASS-8	10F 12K 33F 46I 47V 50V 53L 54L 63P 71V 82I 85V	55.4%
78	5000	2	4	UMASS-9	10F 13V 33F 45R 46I 47V 50V 53L 54L 66F 71V 74A 76S 82I	92.3%
77	5000	2	5	UMASS-10	10F 13V 33F 43T 46I 47V 50V 54L 71V 82L 88S	91.2%
43	1350	-	1	DRV	32I 41I 47V 82I 84V 85V	62.8%
75	0			ND		WT 82.9%

Figure 1. Point mutations of the most abundant variant observed at the highest drug concentration from each selection pool. The green shaded box and the blue shaded box indicate R1-1 and R1-2 side chains, respectively. The red lettering is for mutations and abundance in the first selection started with the 26 mutants; the blue lettering is for the mutations and abundance in the second selection started with just the wild type virus.

890



891

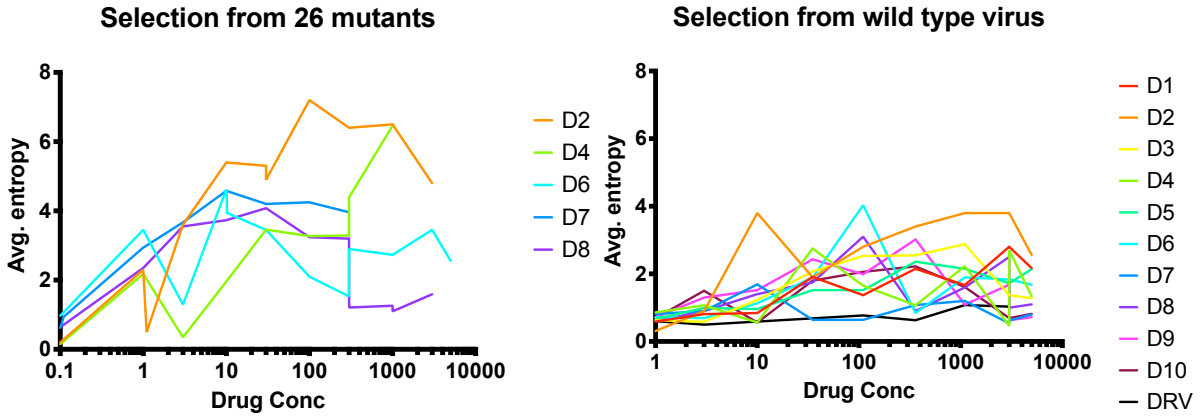
Drug Conc. (nM)	R1	R2	PIs	Most Abundant Variants Detected at the Last Time Point	Abundance
5000	1	3	UMASS-3	10F 33F 46I 47V 50V 53L 63P 71V 76S 82I 85V 89I	85.7%
5000	1	4	UMASS-4	10F 11I 13V 32I 33F 43T 46L 54L 71V 82I 84V 89M 91S 92R	87.3%
5000	1	5	UMASS-5	10F 15V 46I 47V 50V 53L 71V 82I 84V 89T	67.9%
5000	2	4	UMASS-9	10F 13V 33F 45R 46I 47V 50V 53L 54L 66F 71V 74A 76S 82I	92.3%

892

893 **Figure 2**. EC50 values of select highly resistant end-point variants vs all 11 inhibitors. V3, V4,
894 etc. refer to the virus pool for the specific UMASS inhibitor culture (e.g. UMASS-3 is V3, etc.).
895 Each virus pool, with the most abundant sequence in the pool shown, was tested against all 10
896 UMASS inhibitors with the EC50 values shown as points along with the mean, inner quartile in
897 the box, and the total range as whiskers. ND is the culture passaged with no drug selection. 12-
898 mut is a recombinant virus with just 12 mutations in the protease.

899

900



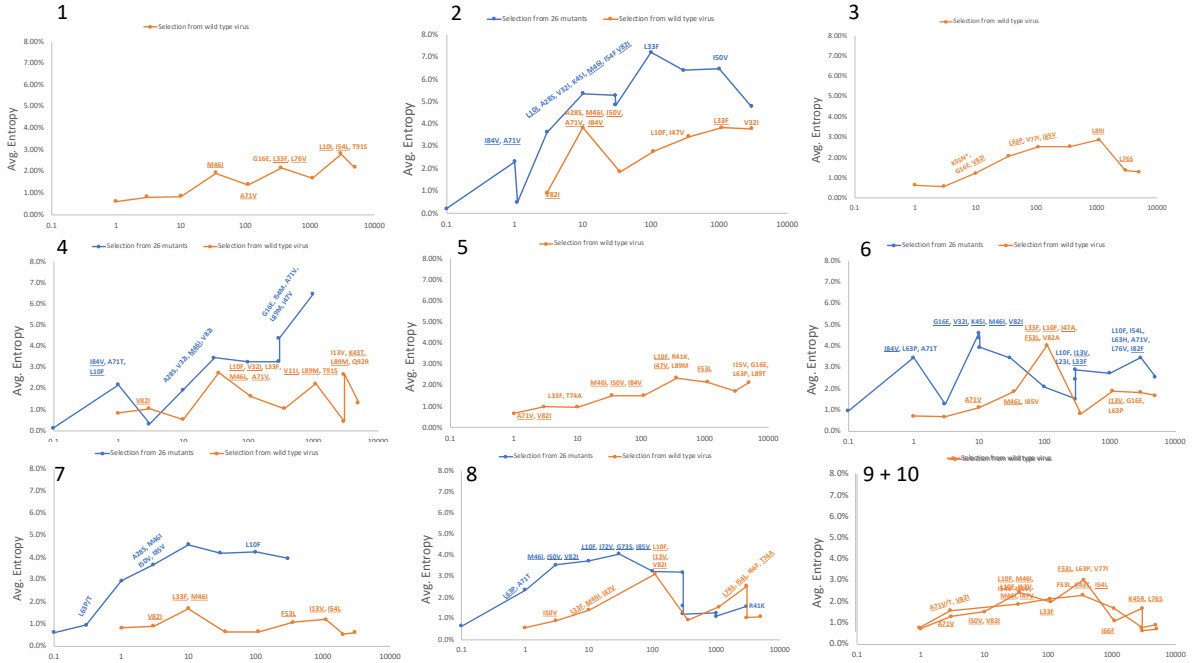


Figure 3. Shannon entropy of the viral pools for 15 selections of U1-10 – each trajectory was passaged between 40-80 times to attain high resistance.

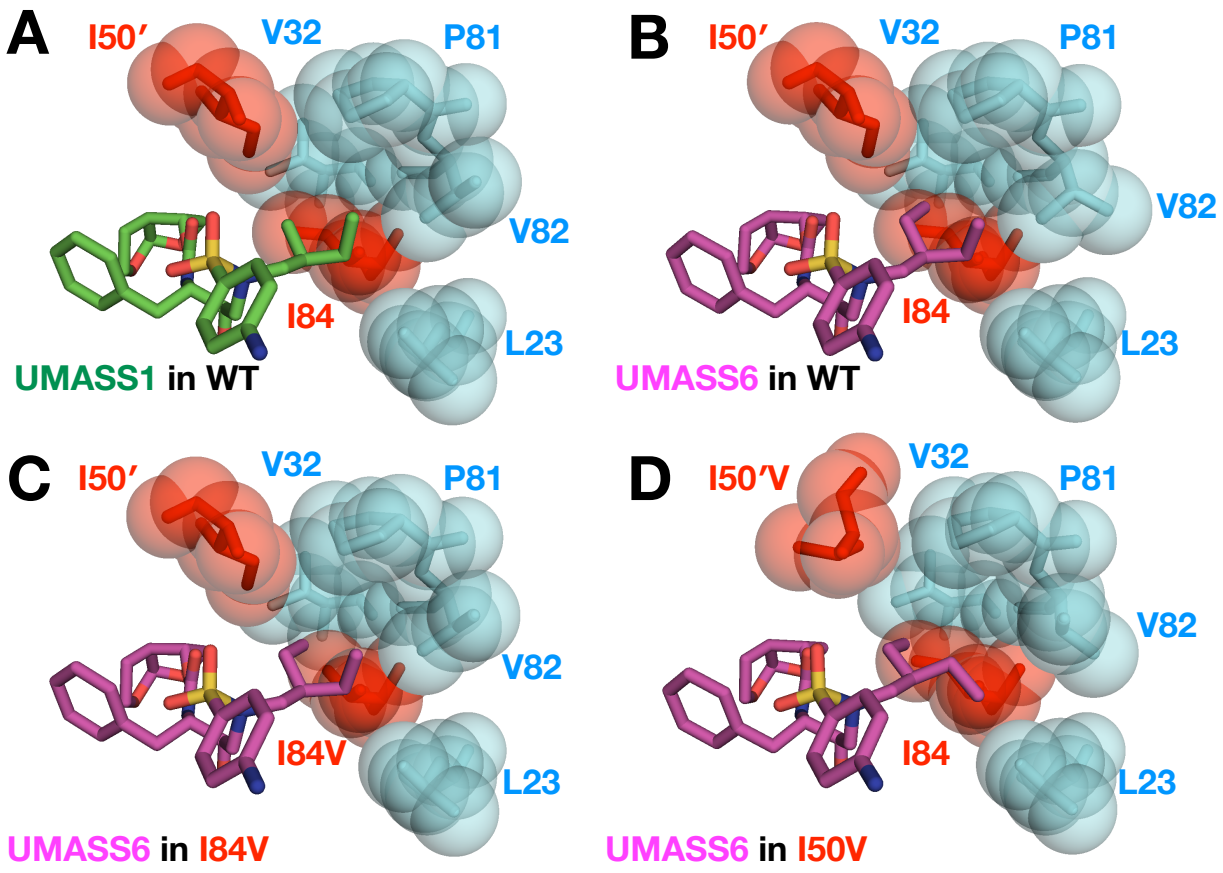


Figure 4. Hydrophobic packing in the S1' subsite. A) DRV in WT protease (PDB: 6DGX). B) UMass6 in WT protease (PDB: 6DGZ). C) UMass6 in I84V variant (PDB: 6DH2). Inhibitor and surrounding residues are identical to UM6-WT. D) UMass6 in I50V variant (PDB: 6DH8). Inhibitor and I50V residue both adopt alternate conformations, ultimately reducing protease-inhibitor vdW contacts.

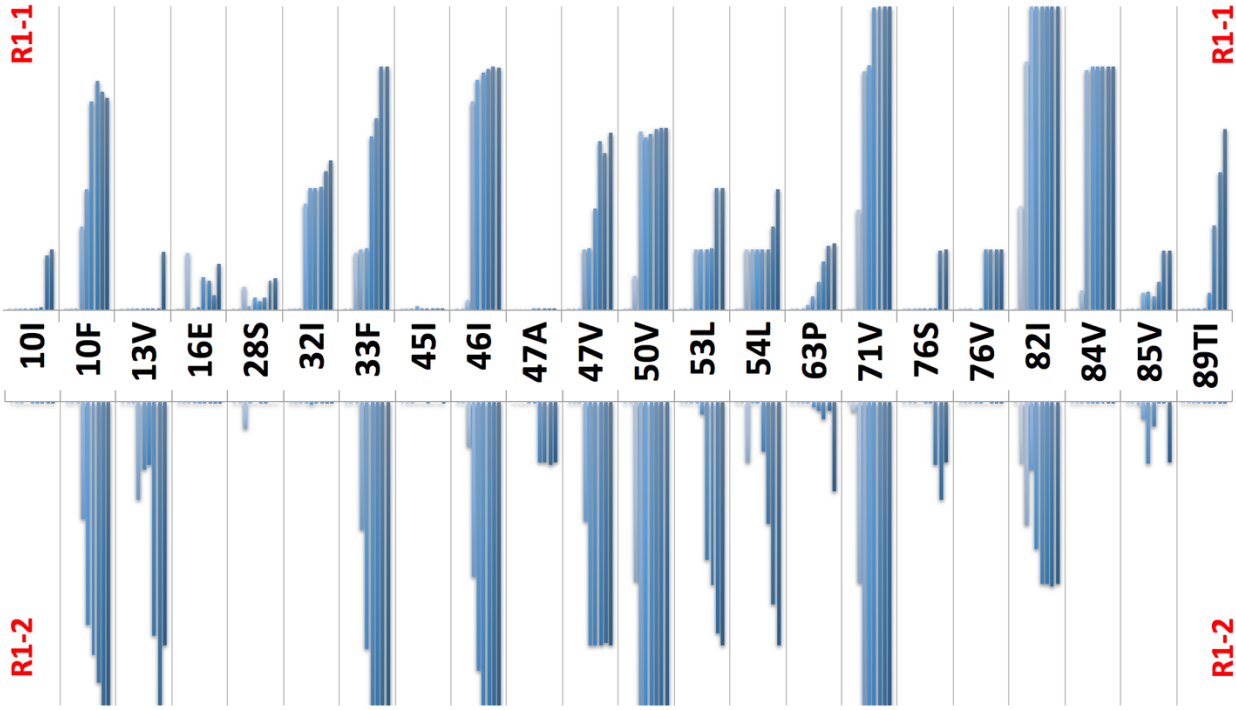


Figure 5. The abundance data from multiple selections pooled and examined sequentially at different levels of drug concentration/selective pressure. Mutations resulting from selections that were challenged with an inhibitor containing a R1 moiety (UMASS1-5) point upwards, while mutations derived from selections that were challenged with an inhibitor containing a R2 moiety (UMASS6-10) point downwards.

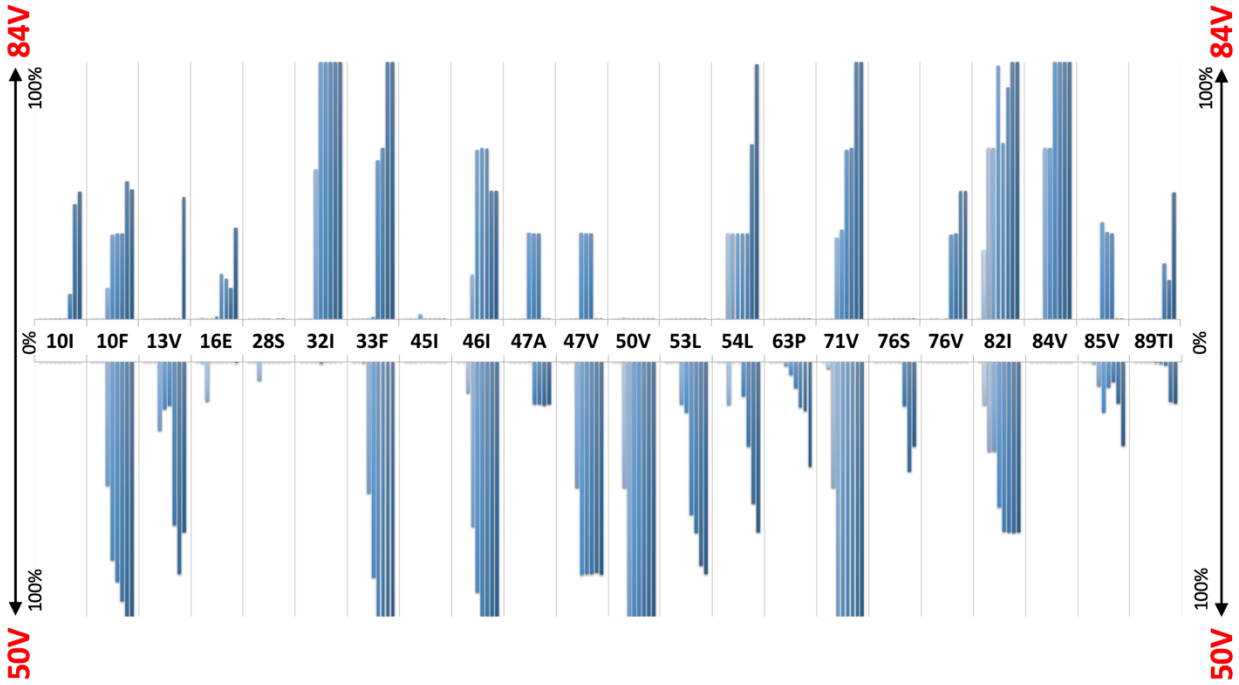


Figure 6. The abundance data from multiple selections that ended in one or the other pathway was pooled and examined sequentially at different levels of drug concentration/selective pressure. The selections resulting in the I84V pathway point up, with I84V reaching 100% penetrance by definition. Similarly, those selections that fixed I50V are shown pointing downward, with I50V reaching 100% penetrance.

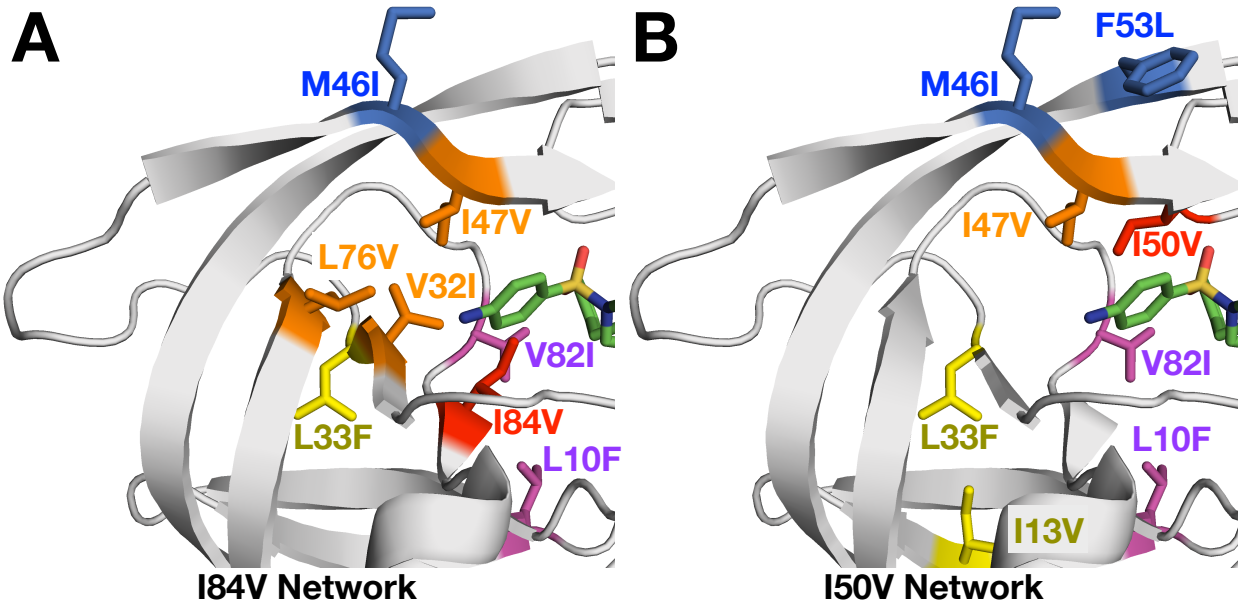


Figure 7. **A)** Mutations associated with I84V penetrance. **B)** Mutations associated with I50V penetrance. Key mutation shown in red (I84V or I50V), outer flap residues in blue (M46I and F53L), inner flap residues in orange (V32I, I47V, L76V), Hydrophobic Core in yellow (L33F and I13V), V82I in purple, and Darunavir shown as green sticks.

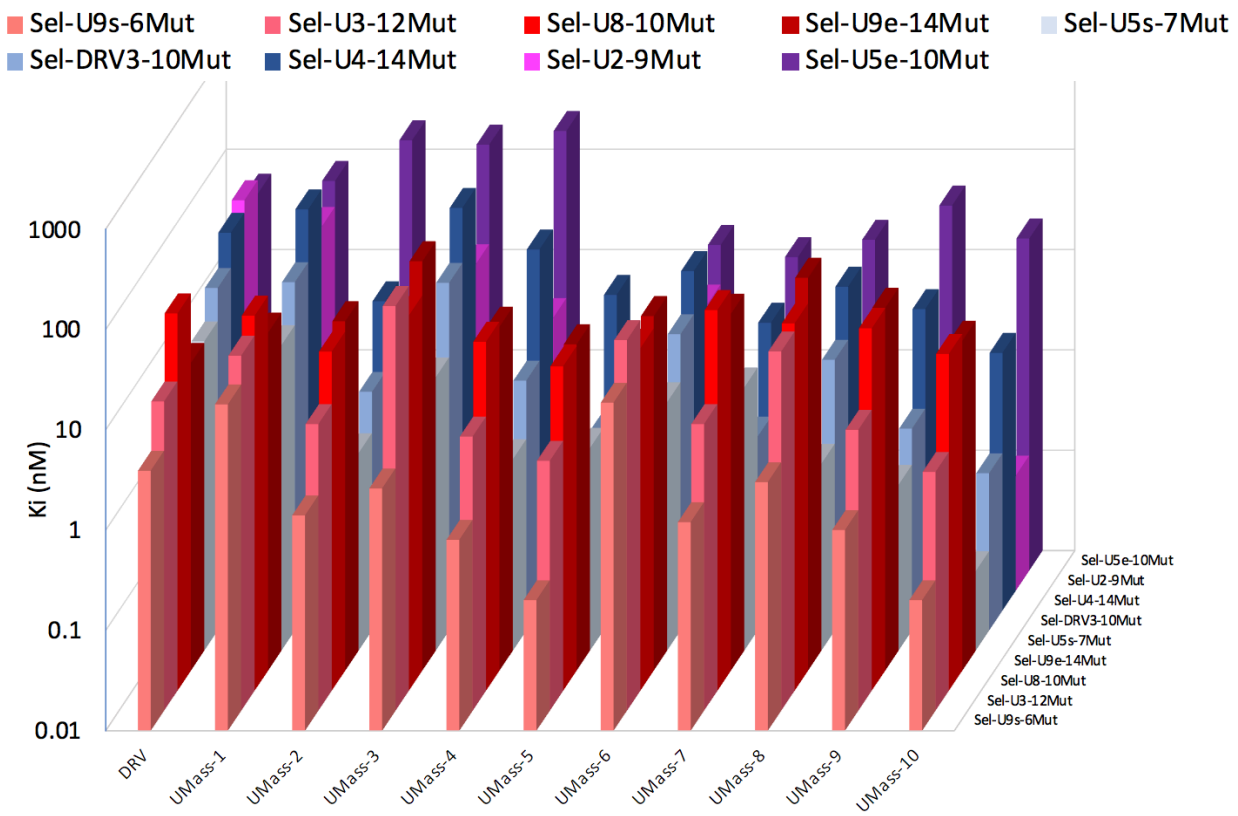


Figure 8. Ki of chosen end-point protease variants vs all 11 inhibitors. Variants with and I50V mutation are colored red, variants with I84V are colored blue, and variants with both I50V/I84V are colored purple.

			NC	SP2	P6	
			P2	-	P1'	P5'
D1 *wt*	5000	84V	V			L
D4 *mut*	1275	84V	V		F	
D4 *wt*	5000	84V	I			I
D6 *mut*	5000	84V	V			
D3 *wt*	5000	50V			F	L
D6 *wt*	5000	50V				L
D7 *wt*	5000	50V			F	
D8 *mut*	1300	50V				L
D8 *wt*	5000	50V			F	
D9 *wt*	5000	50V			F	
D10 *wt*	5000	50V			F	
D2 *mut*	4000	50V+84V				L
D2 *wt*	5000	50V+84V			F	L
D5 *wt*	5000	50V+84V			F	L

Figure 9. Mutations observed near protease cleavage sites for different inhibitors at the noted inhibitor concentrations. Inhibitors are grouped by possession of one or both of the I50V and I84V mutations.

| | MA | | | | | | | | | CA | | | | | | | NC | | | | | p6 | | | | | | | | | | | RT - RH | | | | | | | | | | |

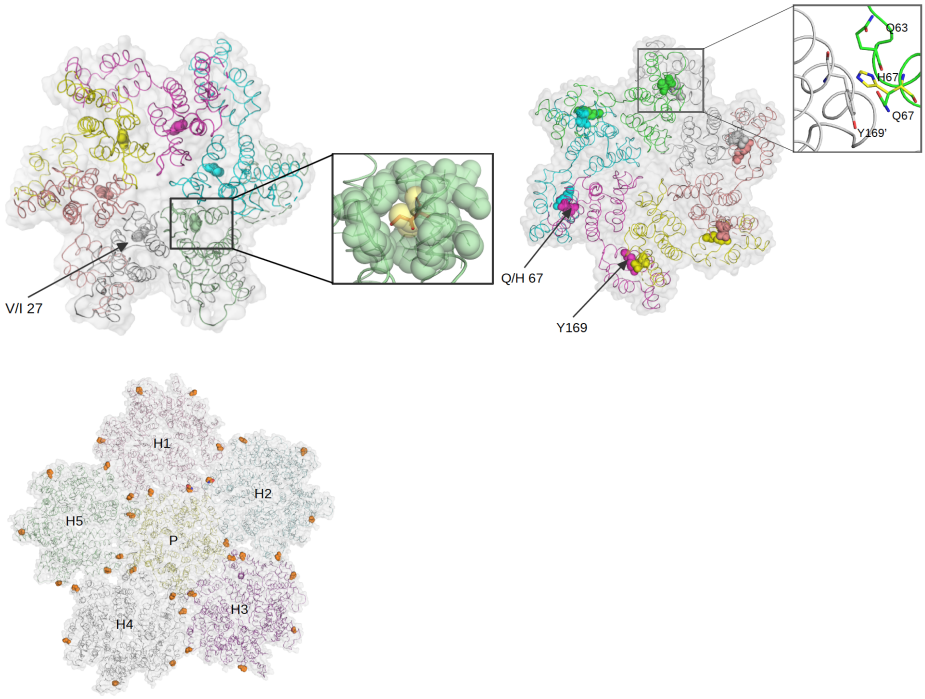


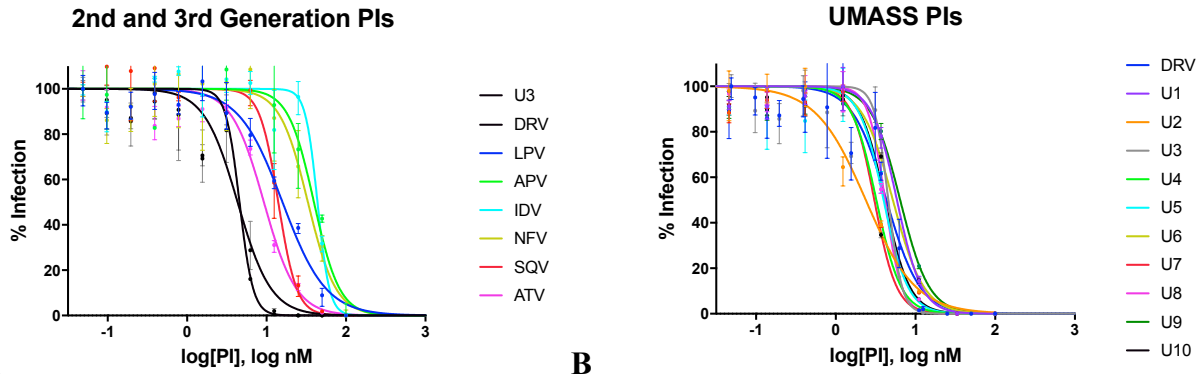
Figure 11. Observed changes in Capsid appear likely to stabilize the capsid structure. Specifically V27I packs (Val in yellow vdW) more tightly filling a pocket (left), Q67H may form additional hydrogen bonding between monomers in the capsid hexamer (middle)(4YWM) and P207S occurs at the interface of hexamers in the viral pentamer (shown in orange spheres) (3J3Q) (right).

964

965

966 **Supplemental Figures**

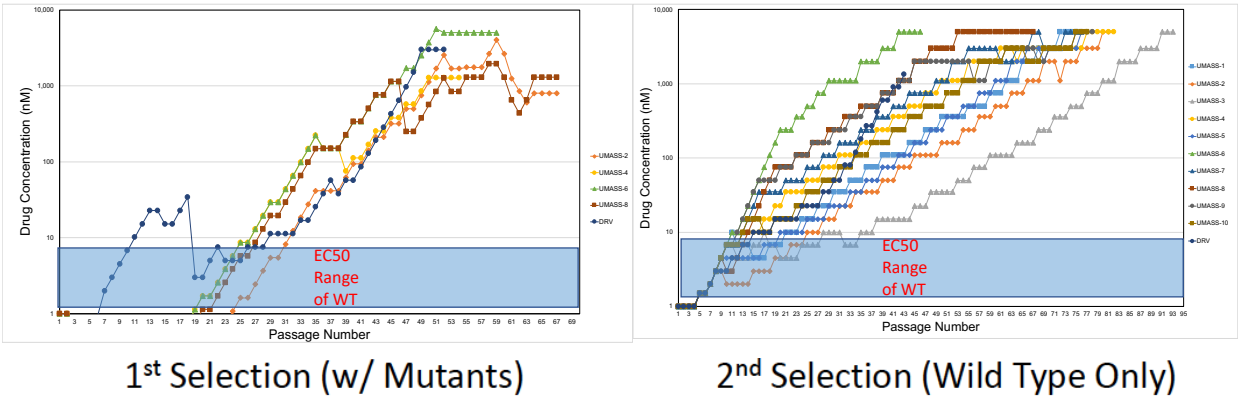
967



968 **A** **B**

969 **Figure S1.** EC50 inhibition curves for **A)** old and new protease inhibitors and **B)** analogs of

970 DRV.



971

972

973 **Figure S2**. Viral selections with DRV and UMASS1-10 have increasing inhibitor

974 concentrations during passaging. Passages are increased when extensive CPE is observed

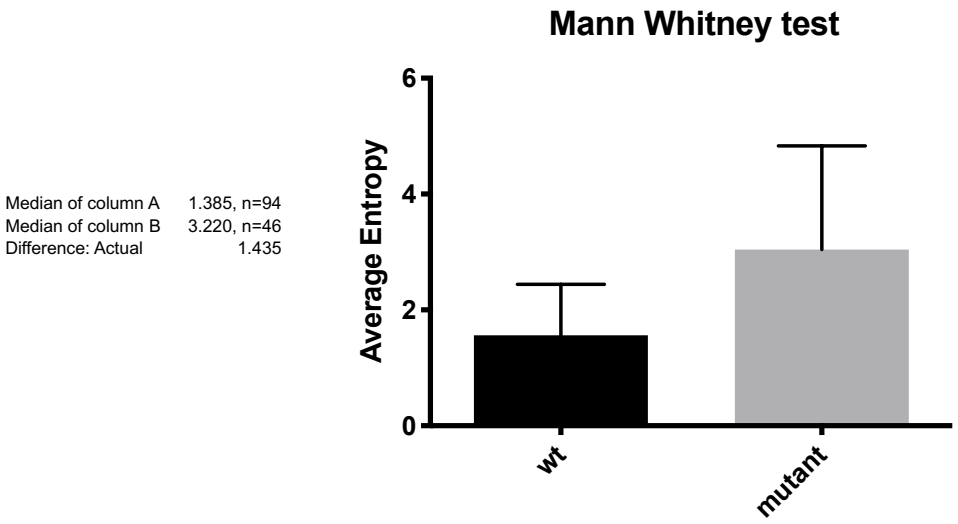
975 during PI selection. Virus generated from the 26 single mutants or NL4-3 molecular clone was

976 passaged in the presence of increasing inhibitor concentration as described in the text. The

977 period of time (in days) until the virus-infected culture displayed maximal CPE (massive

978 syncytia) was 4-7 days on average.

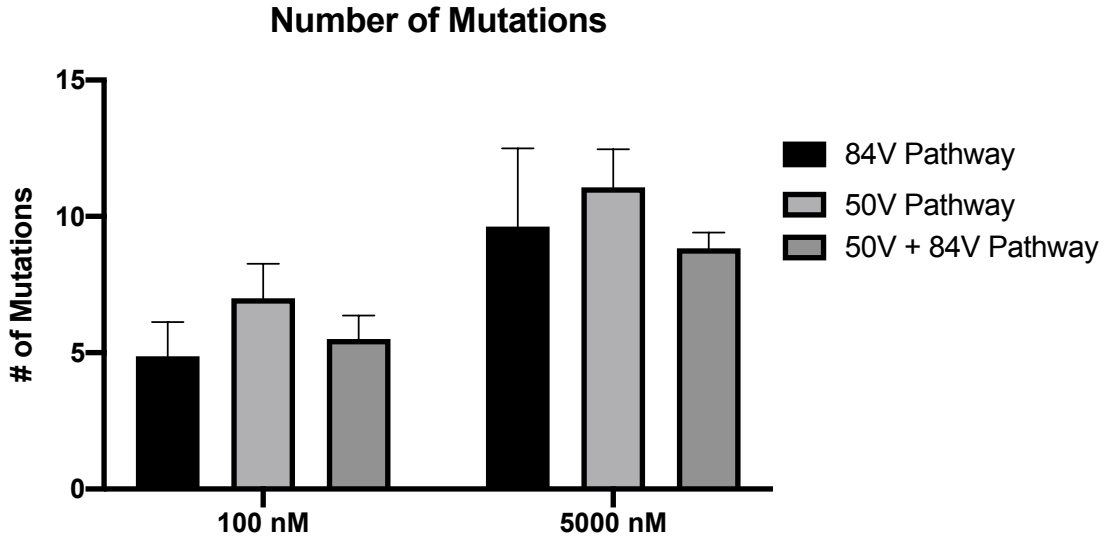
979



980

981 **Figure S3.** Mann-Whitney rank sum test for Shannon's entropy of the viral pools for 4 26-

982 mutant selections and 10 wild-type selections.



983

984

985 **Figure S4**. Average number of mutations for each resistance pathway. Mutation numbers at

986 drug concentrations of 100 nM and 5000 nM are shown.

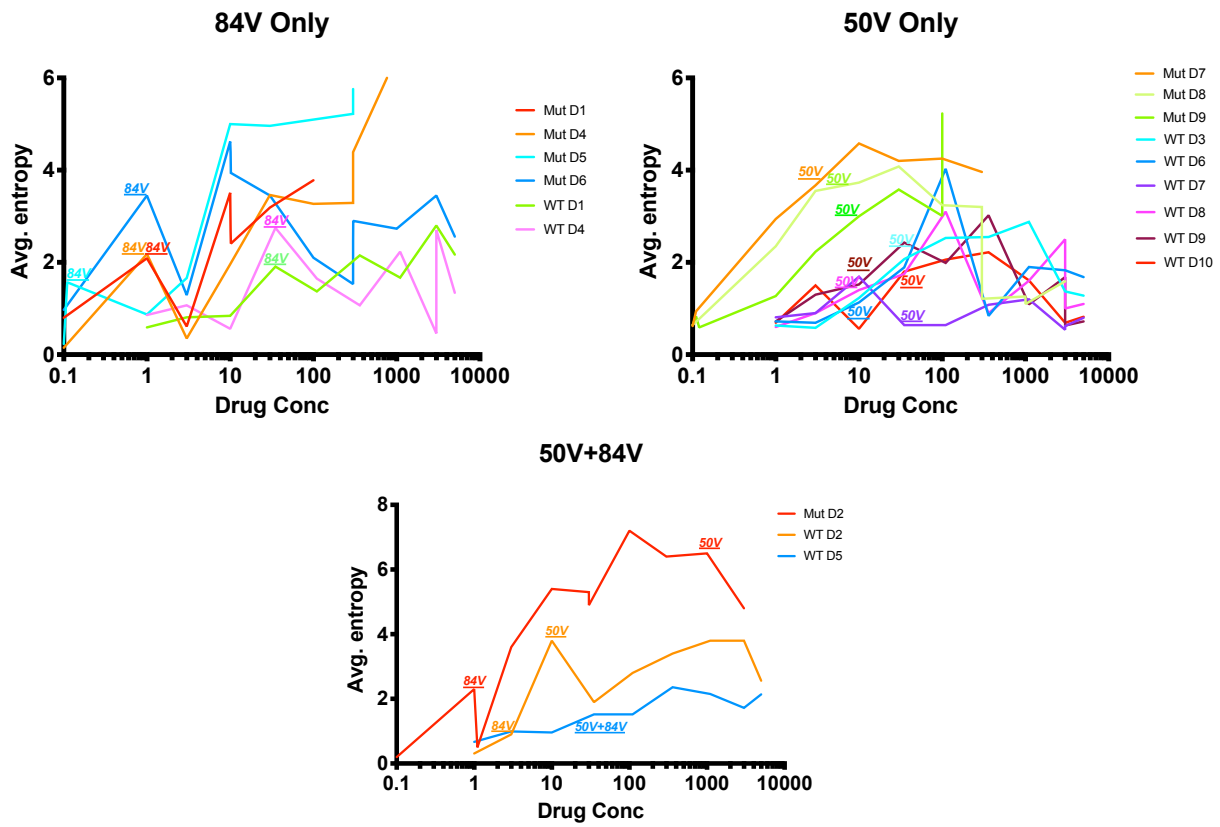


Figure S5. Shannon entropy of the viral pools for 20 selections of U1-10 – each trajectory was passaged between 40-80 times to attain high resistance

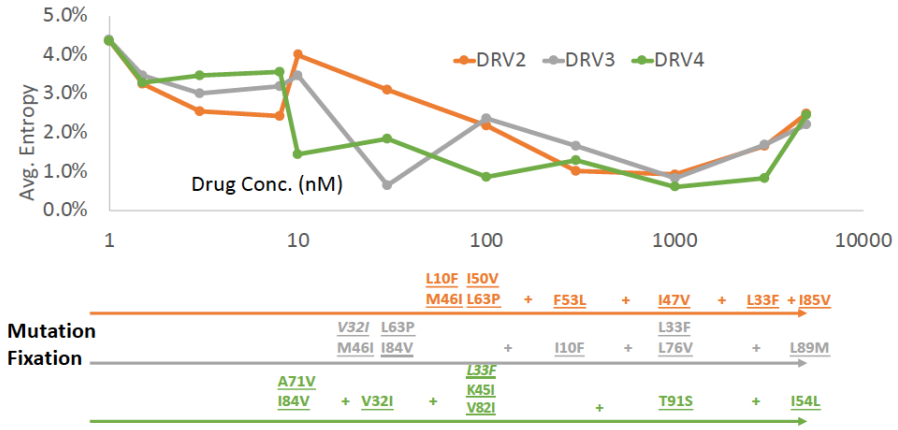


Fig. S6 Changes of average entropy within the protease region through three DRV selections. When mutations are fixed they are labeled in the bottom part of the figure. All the underlined mutations were eventually fixed in the viral population.