

## Supplemental Material

Table S4. Parameters describing LPV binding to PR1 and PR2 in crystal structures.

	<b>PR1</b>	<b>PR2</b>
Total number of LPV VdW contacts (<4.5Å)	104	102
contacts with A monomer	60	55
contacts with B monomer	44	46
contacts with solvent	17	17
Number of direct hydrogen bonds	8	8
Number of water-mediated hydrogen bonds <sup>a</sup>	3	3
Inhibitor buried surface area (Å <sup>2</sup> ) <sup>a</sup>	733	741
Inhibitor buried surface area (%) <sup>a</sup>	87.4	88.2
Gap volume index <sup>a, b</sup>	0.33	0.27

<sup>a</sup> values calculated using the Protein-protein interaction server (15)

<sup>b</sup> Gap Volume Index = GV /ASA, where GV is the volume of the gaps between the two interacting subunits calculated using the program SURFNET (20) and ASA is the accessible surface area buried upon complex formation.

Table S5

Frequency of protease amino acid substitutions in sequences with an insertion between positions 32 and 41<sup>1</sup>

Position	Percent of insertions (n=278)	OR (95% CI) <sup>2</sup>	P <sup>3</sup>
11	10%	2.9 (2.0 - 4.1)	3.1E-09
13	50%	1.6 (1.3 - 1.8)	2.1E-07
16	14%	2.4 (1.8 - 3.3)	9.3E-09
30	2.2%	0.16 (0.08 - 0.32)	2.0E-07
36	74%	1.8 (1.6 - 2.1)	4.3E-18
47	17%	2.5 (1.9 - 3.3)	6.9E-11
57	26%	2.0 (1.6 - 2.5)	4.0E-09
77	19%	0.55 (0.42 - 0.72)	1.2E-05
88	1.4%	0.10 (0.05 - 0.23)	2.7E-08
89	22%	2.4 ( 1.9 - 3.1)	9.4E-13
91	6.8%	2.8 (1.8 - 4.3)	2.3E-06

<sup>1</sup> The frequency of amino acid substitutions 1-99 in 278 viruses submitted for drug resistance determination between 2002 and 5/1/2007 were compared to the substitution frequencies in viruses with at least one genotypically predicted PI resistance (n = 34,522). Positions with wild type and mutant mixed amino acid substitutions were excluded

<sup>2</sup> Odds Ratio and 95% Confidence Interval

<sup>3</sup> P value was estimated for the chi squared statistic. A Benjamini-Hochberg false discovery rate of 0.01 was used to account for multiple comparisons