Supporting Information

Molecular determinants of epistasis in HIV-1 protease: Elucidating the interdependence of L89V and L90M mutations in resistance

Mina Henes[1], Klajdi Kosovrasti[1], Gordon J. Lockbaum[1], Florian Leidner[1], Gily S. Nachum[1], Ellen A. Nalivaika[1], Daniel N.A. Bolon[1], Nese Kurt Yilmaz[1], Celia A. Schiffer*[,1], Troy W. Whitfield*[,2,3]

[1]Department of Biochemistry and Molecular Pharmacology, [2]Department of Medicine, and [3]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

*Corresponding Author

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

Troy W. Whitfield: Phone: +1 508 856 4401; Troy.Whitfield@umassmed.edu

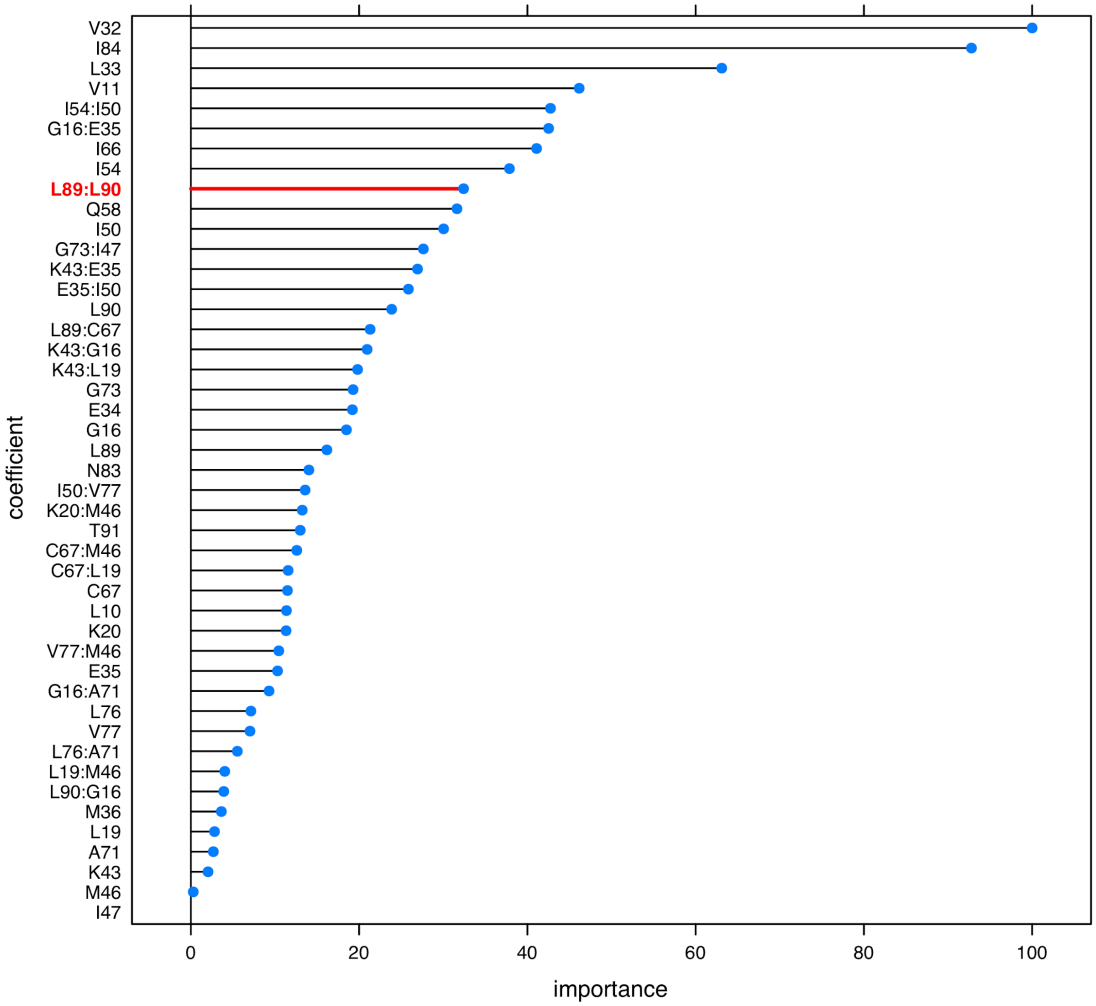

Figure S1. Feature importance for a sequence-based linear model of darunavir binding by HIV-1 protease. Curated in vitro susceptibility data (IC_{50}) for darunavir[1] were analyzed for detection of pairwise non-additive effects of mutations on susceptibility. In total, 605 IC_{50} measurements were included, each corresponding to a complete HIV-1 protease sequence from patient isolates. The fitted model was of the form $\log_{10}(IC_{50}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{1,2} x_1 x_2 + \ldots + \beta_{p-1,p} x_{p-1} x_p$, where there are p "main effects" (i.e. mutations occurring at a single residue) and at most p(p-1)/2 interaction terms. The features x_i are defined as indicator variables for a mutation at site i, such that $x_i=0$ for wild-type amino acid at site i and $x_i=1$ for any amino acid substitution at that site. Feature selection for this model was done in a two-stage "main effects first" procedure[2], where a model was first fitted for the main effects (i.e. no interaction terms) by multiple regression under the elastic net penalty[3], $\lambda\Sigma_i(1-\alpha)\beta_i^2/2 + \alpha|\beta|$, with sparsity enforced by favoring the l_1-norm penalty, choosing $\alpha=0.95$. Next, the interaction terms were selected by fitting the residuals of the main effects model under the elastic net penalty. At each of these stages, feature selection was carried out under 5-fold cross validation. After feature selection, the coefficients, β_i, of the final model were fitted under a relaxed ($\lambda=0$) penalty. The "importance" of these features above is reported as $|t|/|t|_{max}$. The co-occurrence of mutations at residues 89 and 90, highlighted in red above, is an important feature in determining susceptibility to darunavir inhibition.

```
                    5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
NL4-3          1  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMNLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
KY             1  PQITLWQRPVVVVKVGGQLMEALLDTGADDTIFEEMNLPGRWTPKIVGGIGGFMKVRQYENVPIEIYGKKILSTVLIGPTPANIIGRNVMTQIGCTLNF
KY (V89L)      1  PQITLWQRPVVVVKVGGQLMEALLDTGADDTIFEEMNLPGRWTPKIVGGIGGFMKVRQYENVPIEIYGKKILSTVLIGPTPANIIGRNLMTQIGCTLNF
KY (M90L)      1  PQITLWQRPVVVVKVGGQLMEALLDTGADDTIFEEMNLPGRWTPKIVGGIGGFMKVRQYENVPIEIYGKKILSTVLIGPTPANIIGRNVLTQIGCTLNF
KY (DM)        1  PQITLWQRPVVVVKVGGQLMEALLDTGADDTIFEEMNLPGRWTPKIVGGIGGFMKVRQYENVPIEIYGKKILSTVLIGPTPANIIGRNLLTQIGCTLNF
```

Figure S2. Sequence alignment of NL4-3 wild-type HIV-1 protease and KY variants. Amino acid substitutions at residues 89 and 90 are highlighted in red.

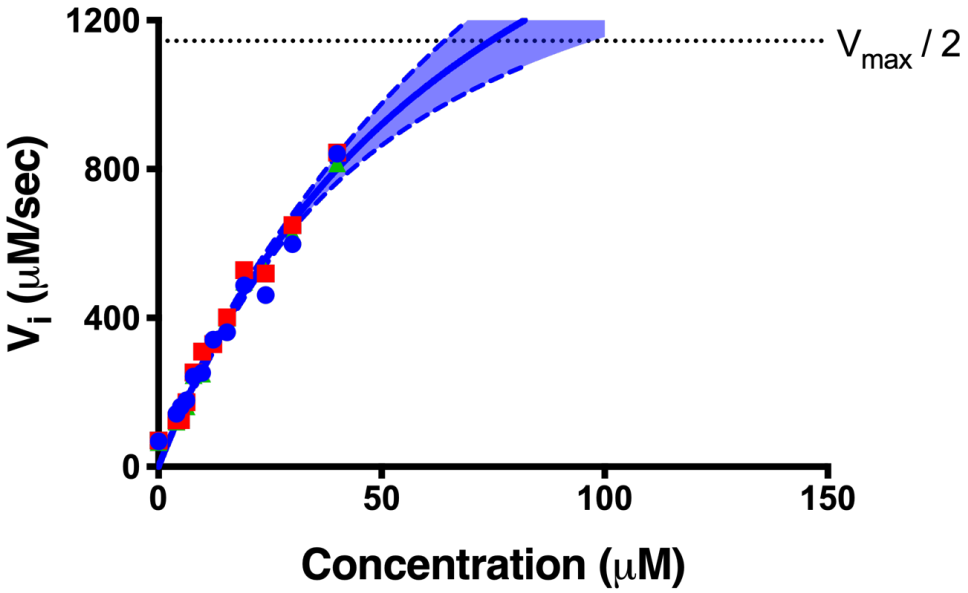

Figure S3. Michaelis-Menten plot for the KY variant. Points for three independent replicates, each corrected for the inner-filter effect, are plotted using different colors. After performing a global non-linear fitting to the Michaelis-Menten equation, the resulting best fit is given (solid blue curve), along with 95% confidence intervals (dashed blue curves). The estimated K_M is 74.4 ± 13.4μM.

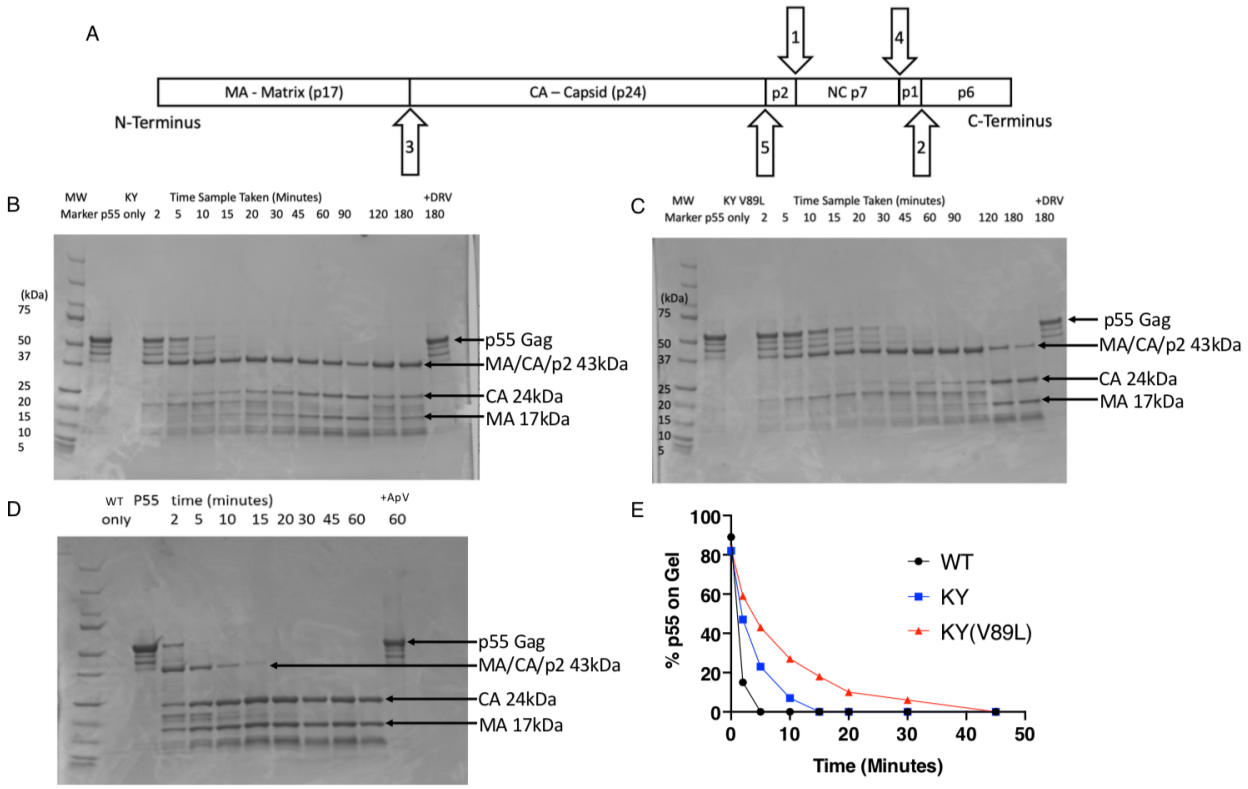

Figure S4. (A) Schematic of the p55 Gag polyprotein used in the gel cleavage assays with product sizes noted in kDa. Sites and order of cleavage by WT protease are denoted by arrows[4]. p55 Gag polyprotein cleavage by (B) KY, (C) KY(V89L), and (D) NL4-3 WT. (D) Percent p55 left on the gel reveals that the enzymatic rate of the KY(V89L) variant is approximately 3 times slower compared to KY and approximately 9 times slower than NL4-3 WT. Only the upper band in the p55 lane was used for quantifying the starting amount of p55. Minor bands are impurities.

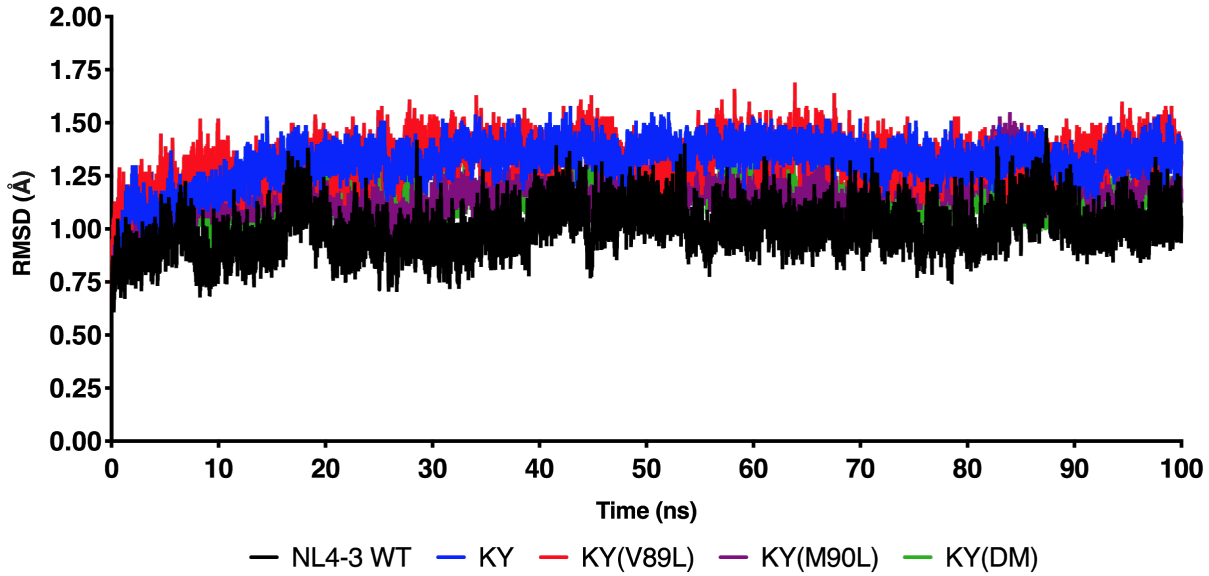

Figure S5. Root-mean-square deviation (RMSD) for WT, KY, KY(V89L), KY(M90L), and KY(DM) calculated from 100 ns molecular dynamics simulations.

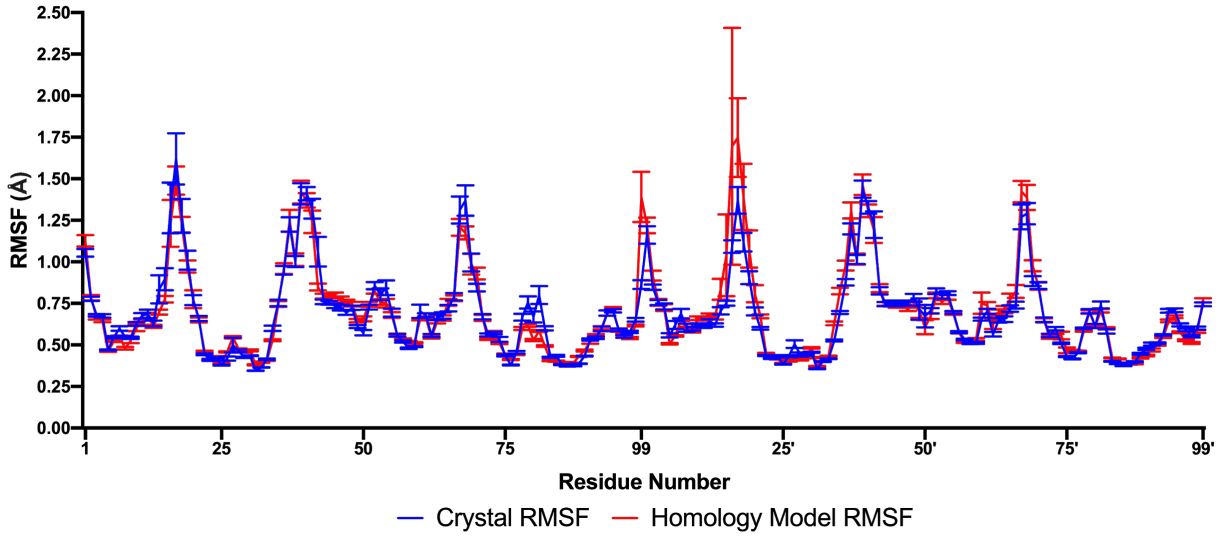

Figure S6. Comparing the Root-mean-square fluctuation (RMSF) profile of C_α atoms utilizing a protease variant with 8 mutations relative to NL4-3 WT: I13V, G16E, V32I, L33F, K45I, M46I, V82F, I84V. Three independent 100 ns simulations were carried out starting from a crystal structure (PDB: 6OPV) and a homology model generated using NL4-3 (PDB: 6DGX) as a starting point. The RMSF profiles generated starting from the homology model and from the crystal structure exhibit no significant differences.

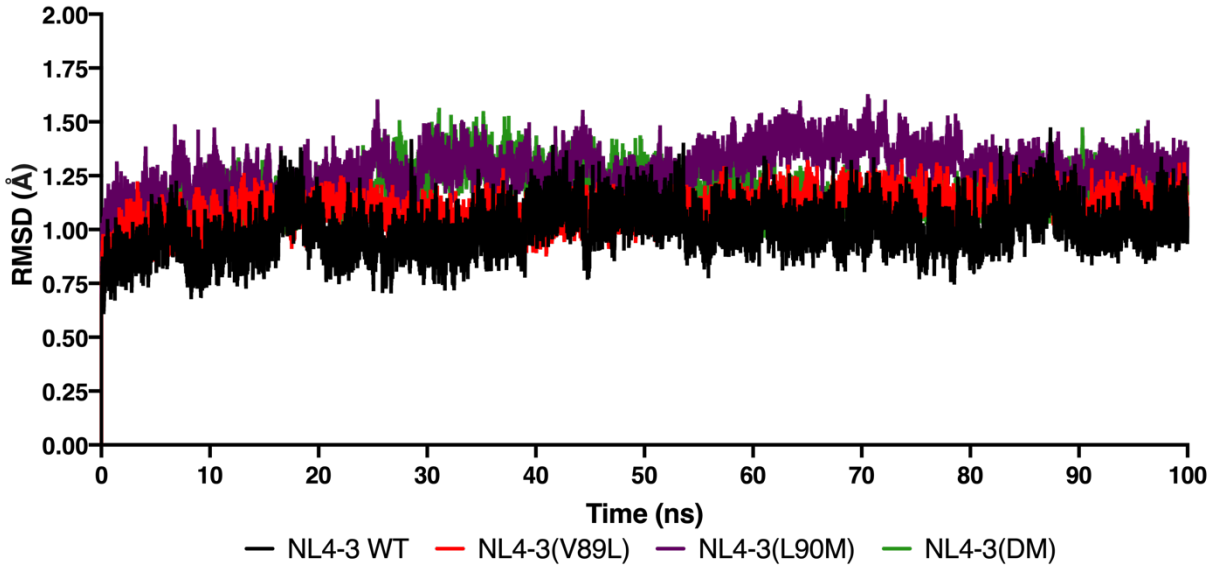

Figure S7. Root-mean-square deviation (RMSD) for WT, NL4-3(L89V), NL4-3(L90M), and NL4-3(DM) calculated from 100 ns molecular dynamics simulations

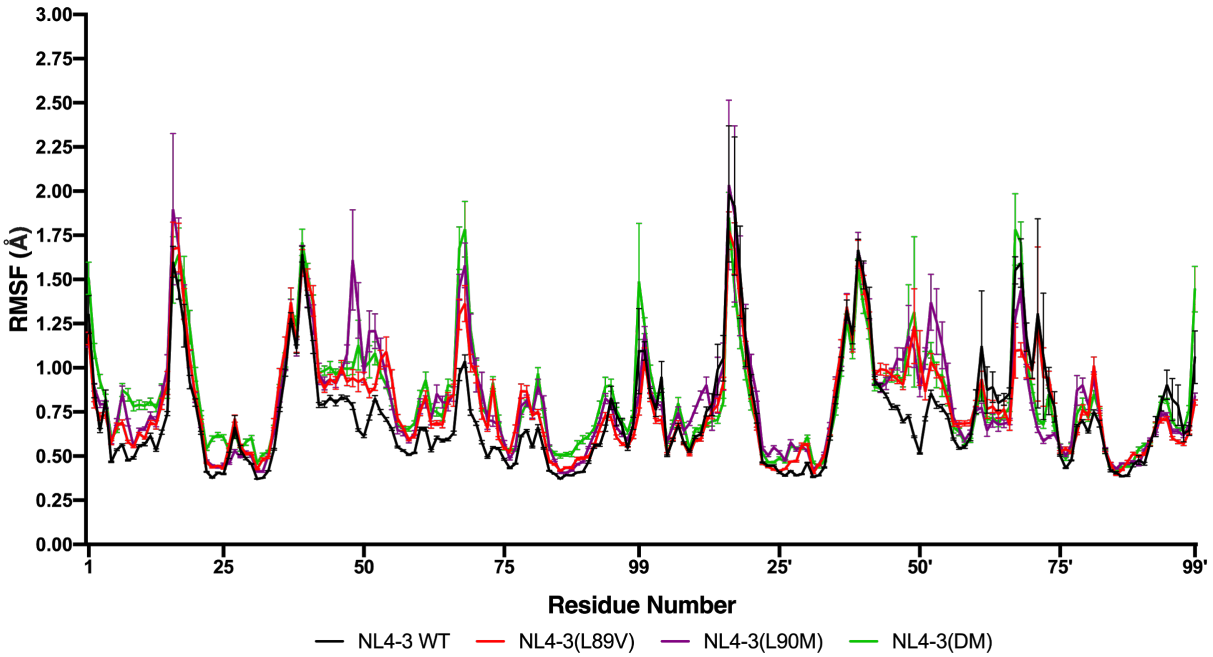

Figure S8. Root-mean-square fluctuation (RMSF) of C_α atoms for the NL4-3, NL4-3(L89V), NL4-3(L90M), and NL4-3(DM) mutations on the WT background. The L90M mutation on the NL4-3 background alters chain B flap dynamics whereas it alters chain A flap dynamics in the KY background.

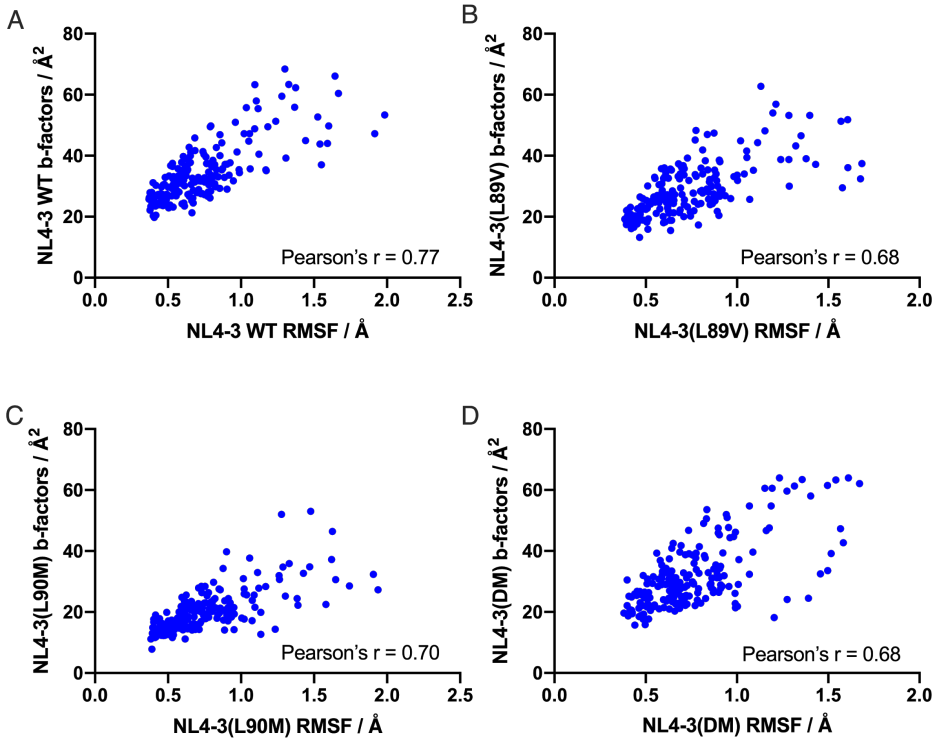

Figure S9. Crystal structure b-factors and root-mean-square fluctuation (RMSF) of C_α atoms from MD simulations of (A) NL4-3 WT, (B) NL4-3(L89V), (C) NL4-3(L90M), and (D) NL4-3(DM) are moderately correlated with Pearson's r between 0.68 to 0.77.

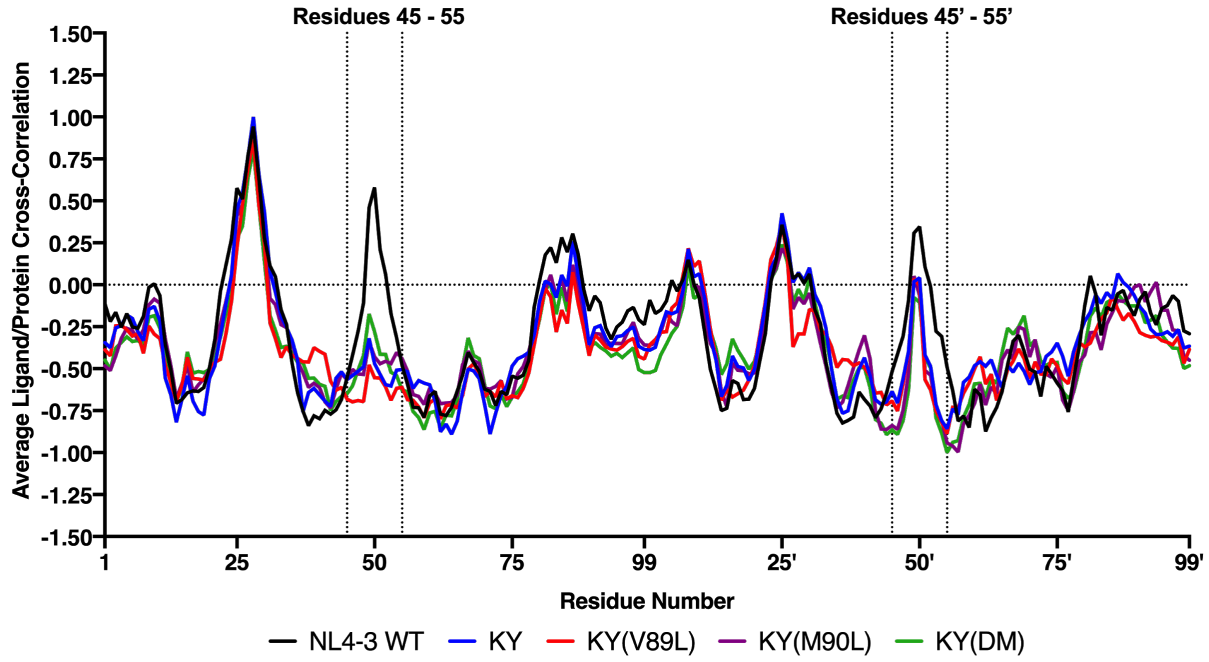

Figure S10. Average per-residue cross-correlation for all DRV heavy atoms with the HIV-1 protease C_α atoms.

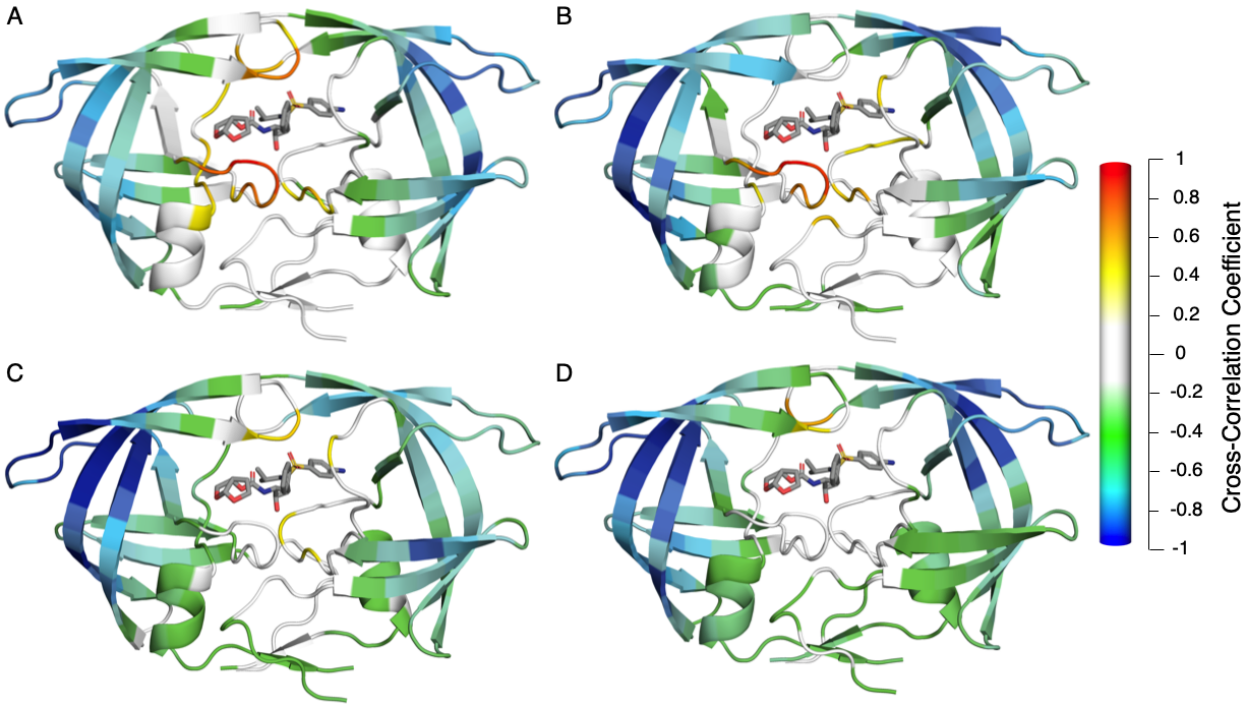

Figure S11. Average cross-correlation between fluctuations of C_α atoms and all DRV heavy atoms during the MD simulations. (A) NL4-3, (B) NL4-3(L89V), (C) NL4-3(L90M), and (D) KY(DM). C_α atoms with a Pearson correlation coefficient between -0.2 and 0.2 are colored white.

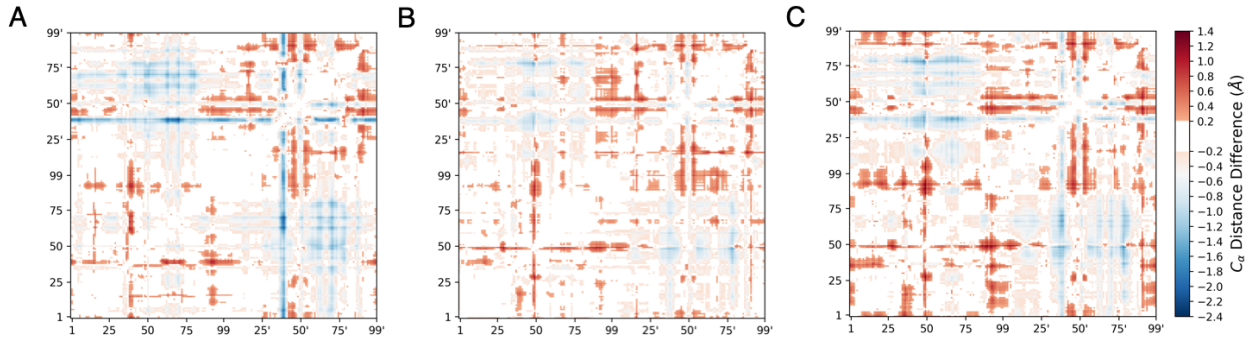

Figure S12. Differences in mean intra-protease C_α-C_α distance comparing (A) KY minus KY(V89L), (B) KY minus KY(M90L), and (C) KY minus KY(DM). Positive differences indicate that the C_α-C_α distance is greater in KY, while negative differences indicate the opposite. In KY(V89L) and KY(DM), bearing the larger leucine side chain, the B chain 70's β-sheet is displaced compared to KY. Unique to the KY(V89L) variant, the B chain flap elbow, 40's loop, is displaced away from the core of the protease and into the solvent.

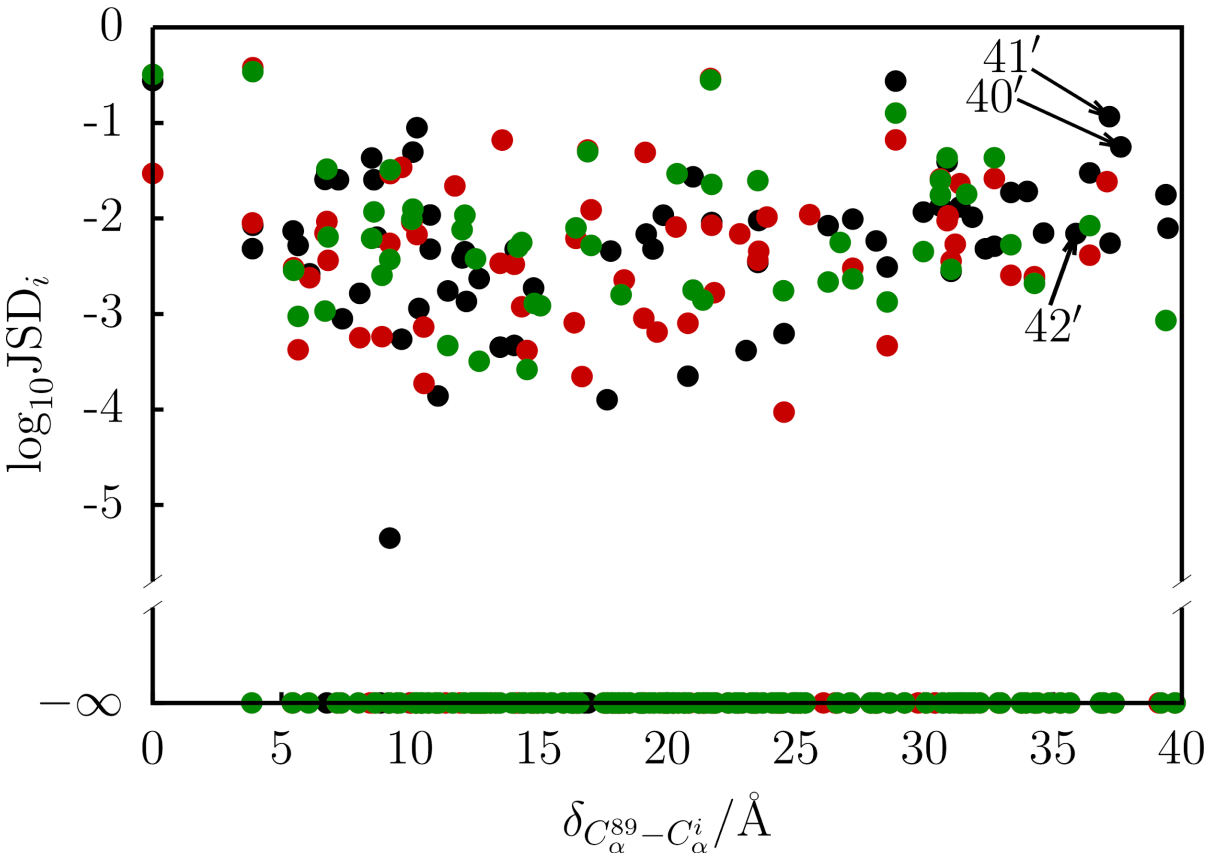

Figure S13. Scatter plot showing Jensen-Shannon divergence for each dihedral angle, i, in the HIV-1 homodimer as a function of approximate distance from the site of mutation (at residues 89, 90 or both). Distances are computed between C_α atoms. The colors of the points correspond to the comparisons: KY versus KY(V89L) (black), KY versus KY(M90L) (red) and KY versus KY(DM) (green). Points along the abscissa are dihedral angles where the Jensen-Shannon divergence signal was not distinguishable from noise. Highlighted resides 40′-42′ are in the chain B flap elbow and are different in the KY(V89L) variant compared with the other KY variants.

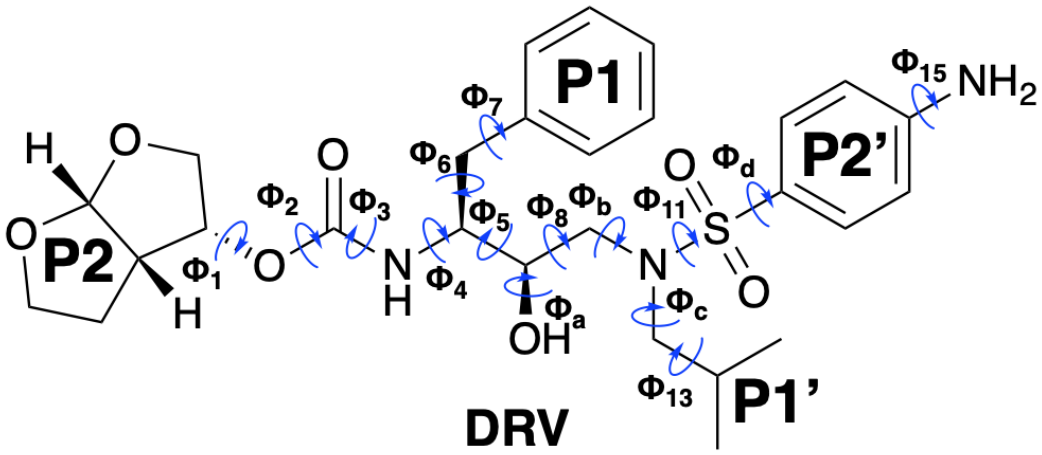

Figure S14. 2D structure of DRV with rotatable bonds labeled.

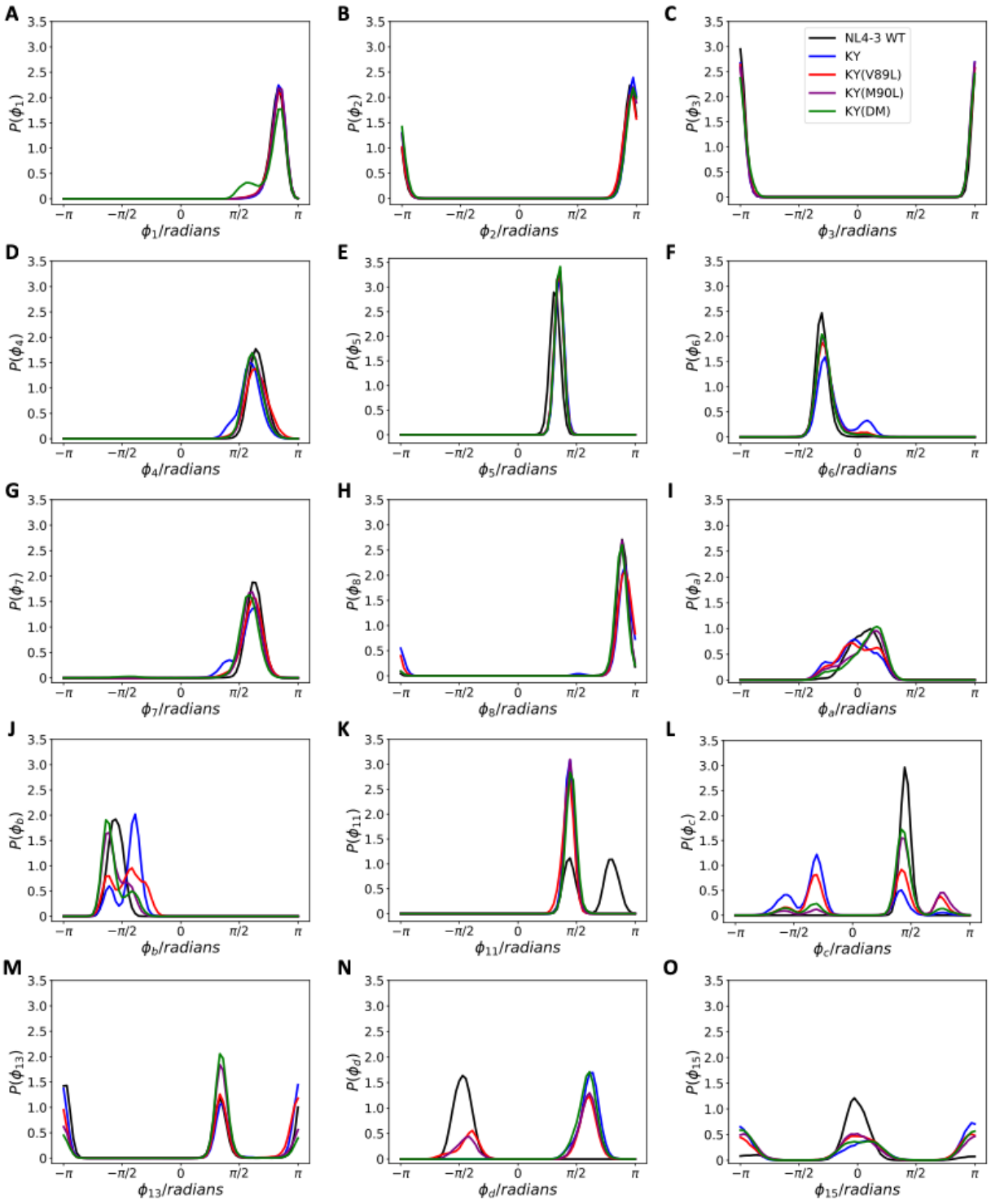

Figure S15. Probability distributions collected from molecular dynamics simulations of the 15 DRV rotatable bonds (see **Figure S14** for a definition of dihedral angles).

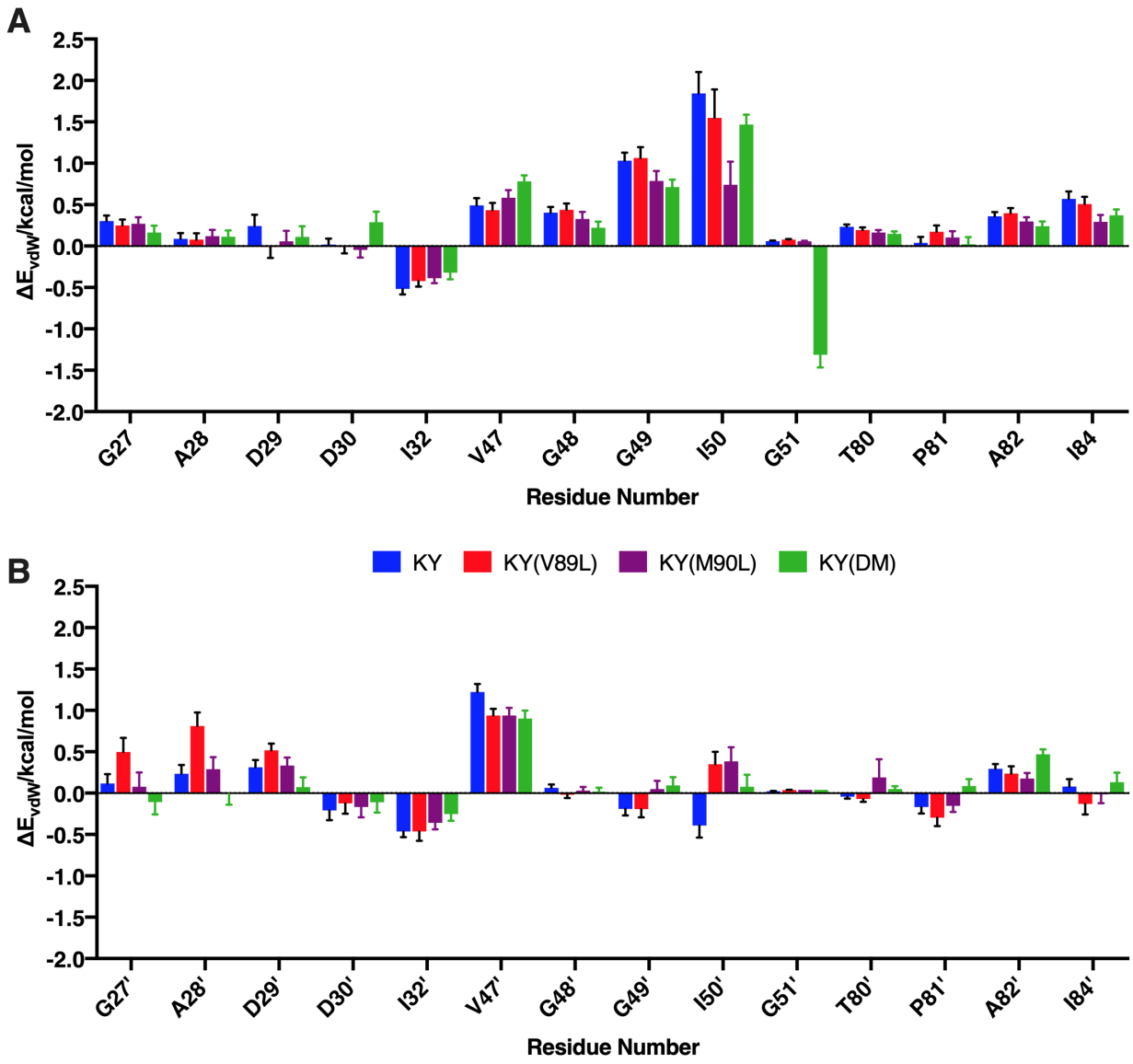

Figure S16. Change in darunavir per-residue van der Waals energy (ΔE_{vdW}) relative to NL4-3. (A) Changes in chain A. (B) Changes in chain B. Positive values indicate loss of favorable interactions relative to NL4-3.

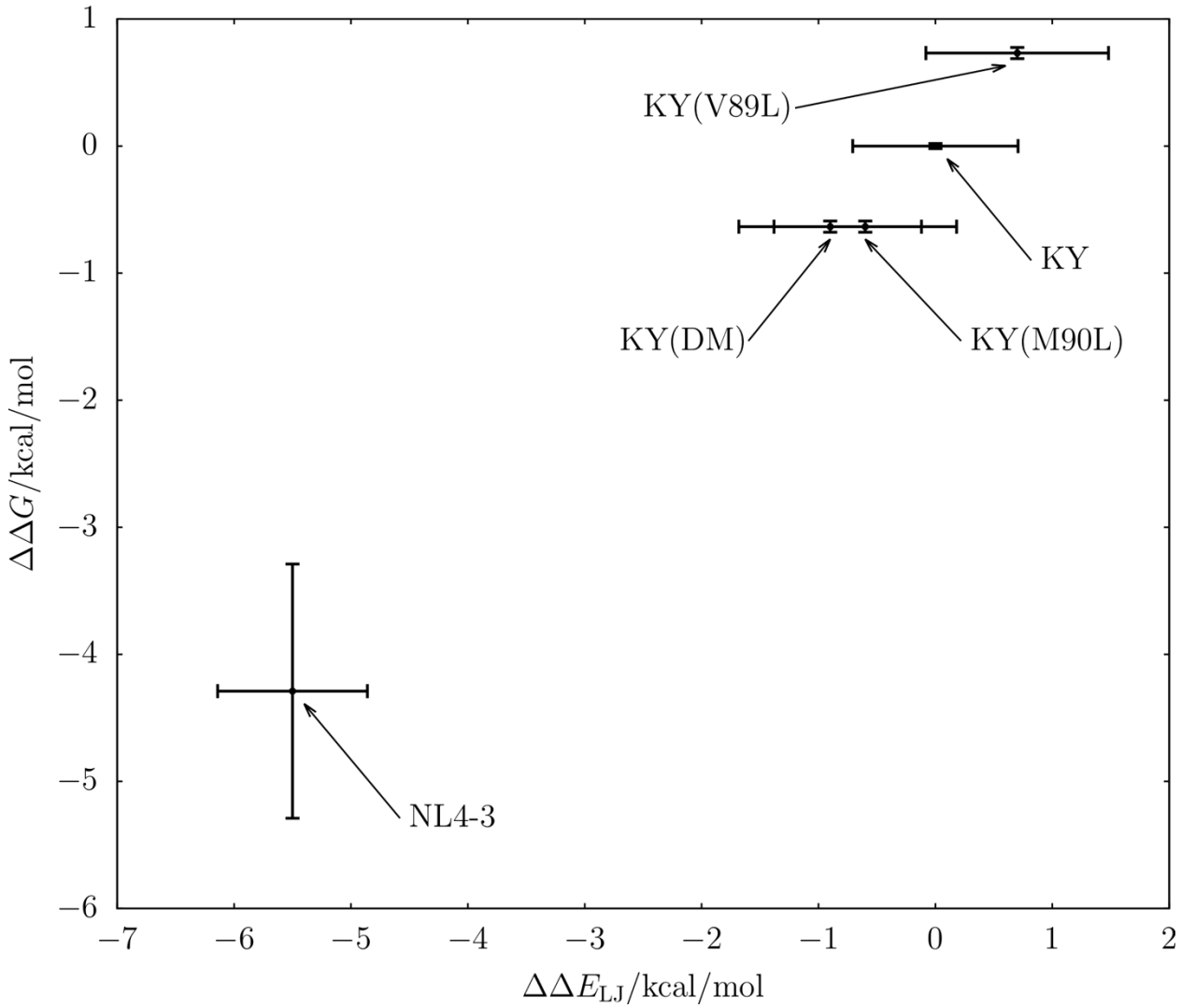

Figure S17. Measured changes in DRV binding with different protease variants correspond to changes in computed van der Waals interaction energy (Pearson's r=0.99). The reference protease variant above is taken to be KY: for a protease variant k, $\Delta\Delta E_{LJ}=\Delta E_{LJ}^{(k)}-\Delta E_{LJ}^{(KY)}$.

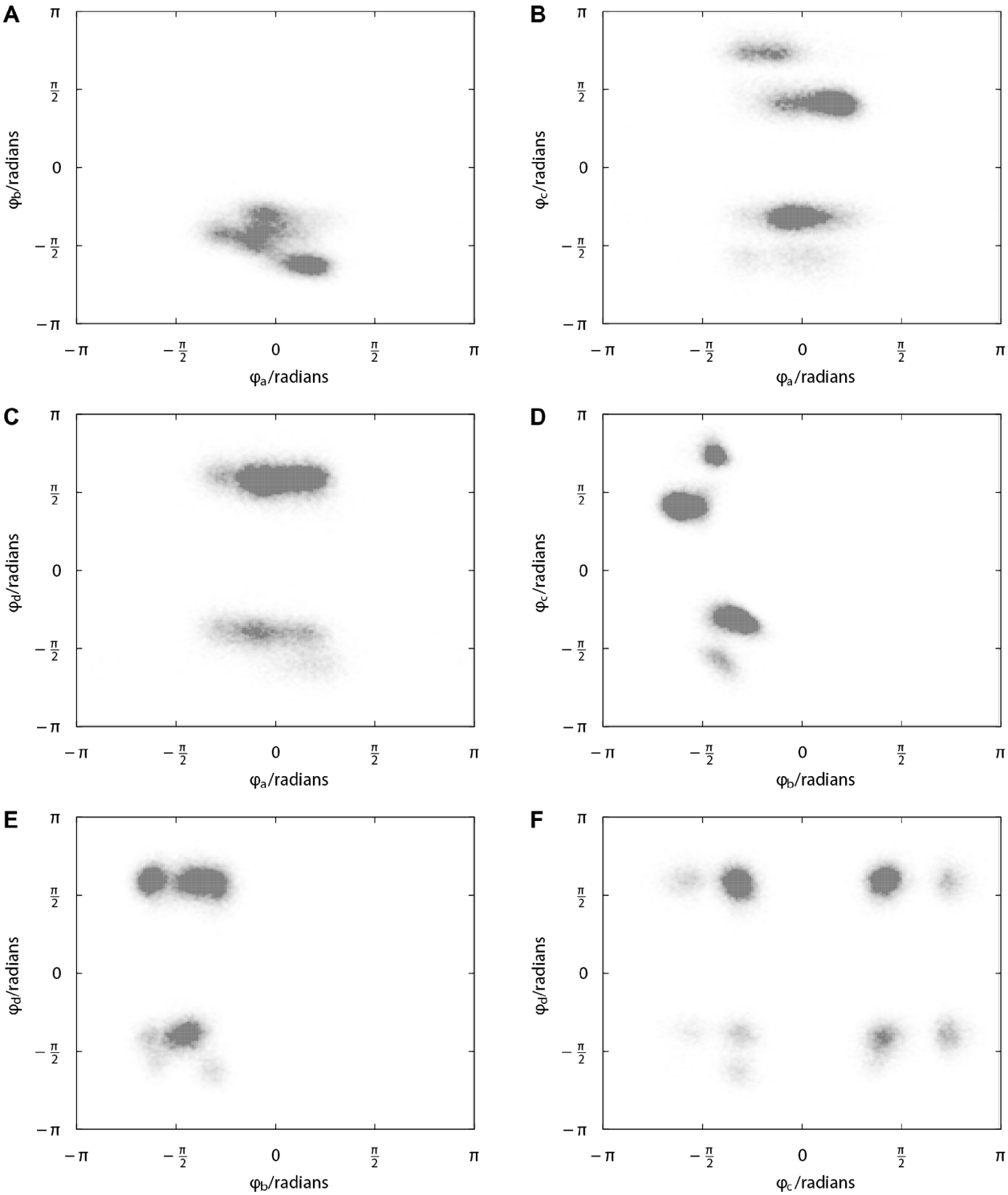

Figure S18. In general, sampling among different DRV dihedral angles is coupled. As an example of coupling between the ϕ_a- ϕ_d dihedral angles, joint probability distributions $P(\phi_i, \phi_j)$ are plotted for the DRV-bound KY(V89L) protease variant. The joint probability distributions plotted are (A) $P(\phi_a, \phi_b)$, (B) $P(\phi_a, \phi_c)$, (C) $P(\phi_a, \phi_d)$, (D) $P(\phi_b, \phi_c)$, (E) $P(\phi_b, \phi_d)$, (F) $P(\phi_c, \phi_d)$.

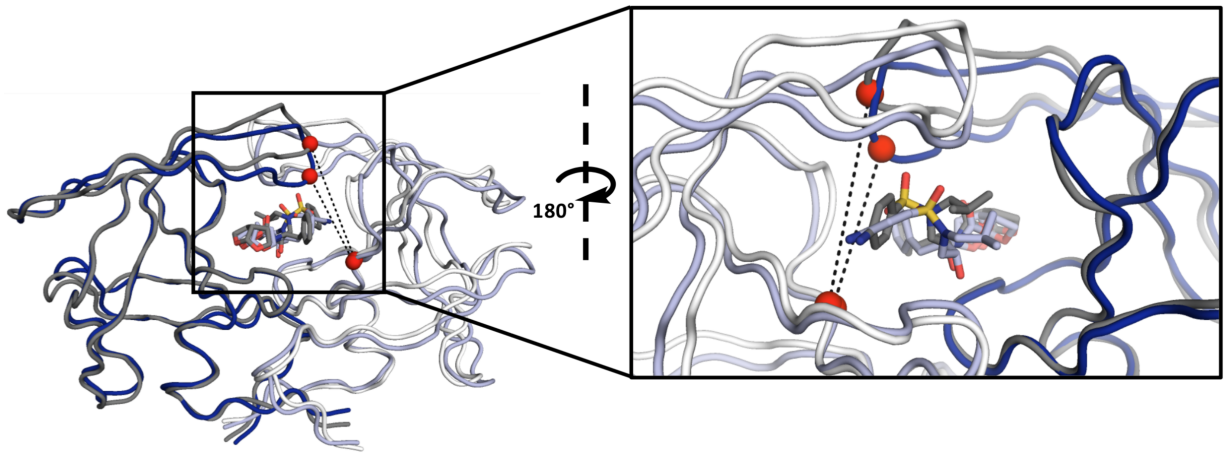

Figure S19. The ϕ_b dihedral of DRV is associated with widening of the active site in KYV89L. Two frames from the MD simulations showing the ϕ_b distribution associated with a closer C_α^{I50} - $C_\alpha^{I84'}$ distance (shaded blue - chain A is dark blue and chain B is light blue) and the ϕ_b distribution associated with a more open C_α^{I50} - $C_\alpha^{I84'}$ distance (shaded gray - chain A is dark gray and chain B is white). The C_α atoms of I50 and I84' are shown as red spheres. DRV associated with a closer C_α^{I50} - $C_\alpha^{I84'}$ distance is shown in light blue while DRV associated with a more open C_α^{I50} - $C_\alpha^{I84'}$ distance is shown in gray (B) Opening of chain A flap is associated with DRV sampling a difference conformation compared to a closed state.

Table S1. X-Ray Crystallography Statistics

PDB ID	6DGX	6OOU	6OOS	6OOT
Protein	NL4-3	NL4-3(V89L)	NL4-3(L90M)	NL4-3(DM)
Inhibitor	DRV	DRV	DRV	DRV
Resolution (A)	2.00	2.13	1.90	1.82
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P6_1$
a (A)	51.0	51.3	51.0	62.2
b (A)	58.3	58.3	58.0	62.2
c (A)	61.9	62.4	62.4	82.9
Completeness	98.1	95.5	95.6	97.4
Tot. Reflections	84332	29727	104194	257556
Uniq. Reflect.	12758	10309	27535	15901
Avg I/Sig	31.1	13.4	26.1	33.6
Redundancy	6.6	2.9	3.8	16.2
R-Merge (%)	5.7	6.9	5.1	10.0
RMSD Bonds*	0.003	0.005	0.005	0.002
RMSD Angles*	0.590	0.649	0.836	0.603
R-Free*	22.8	25.5	19.8	23.5
R-Work*	19.2	20.4	16.4	19.2

*Based on Phenix Program

Table S2. Kinetics and binding measurements for the NL4-3 wild-type and variants with the introduction of the following mutations: L89V, L90M, and the L89V, L90M double mutation. K_M and the turnover number, k_{cat}, were measured using a natural substrate sequence. The enzyme catalytic efficiency is k_{cat}/K_M. Mean protease-DRV van der Waals energy, ΔE_{vdW}, is reported.

	NL4-3	NL4-3(L89V)	NL4-3(L90M)	NL4-3(DM)
K_M (μM)	**71.4 ± 6.8**	**46.9 ± 4.2**	**139.2 ± 18.9**	**180.3 ± 56.2**
k_{cat} (s^{-1})	**1282.7 ± 0.06**	**13.3 ± 0.6**	**5.9 ± 1.1**	**108.6 ± 0.03**
k_{cat} / K_M ($\mu M^{-1}s^{-1}$)	**17.1 ± 0.1**	**0.3 ± 0.1**	**0.04 ± 0.2**	**0.6 ± 0.3**
K_i (nM)	**< 0.005**	**< 0.005**	**< 0.005**	**< 0.005**
ΔE_{vdW} (kcal/mol)	**-58.5 ± 0.4**	**-58.1 ± 0.6**	**-57.5 ± 0.5**	**-55.8 ± 0.6**

Table S3. Configurational entropy, $S_j = -k_B \sum_i P_i(\phi_j) \ln P_i(\phi_j)$, for each of the selected DRV dihedral angles $1 \leq j \leq 4$. ϕ_a, ϕ_c, and ϕ_d entropy are moderately correlated with DRV binding (Pearson's r=0.70, 0.60, 0.64, respectively). ϕ_b entropy is strongly correlated with DRV binding (Pearson's r=0.87).

	NL4-3 WT	**KY**	**KY(V89L)**	**KY(M90L)**	**KY(DM)**
S_a/k_B	3.8	4.2	4.3	4.1	4.0
S_b/k_B	2.7	3.2	3.8	3.3	3.3
S_c/k_B	2.2	4.4	4.6	4.1	4.0
S_d/k_B	3.1	3.2	4.2	4.0	3.2

References

1. Rhee, S.-Y.; Taylor, J.; Fessel, W. J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; Shafer, R. W., HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob Agents Chemother* **2010,** *54* (10), 4253-4261.
2. Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R., Least angle regression. *Ann. Statist.* **2004,** *32* (2), 407-499.
3. Zou, H.; Hastie, T., Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **2005,** *67* (2), 301-320.
4. Pettit, S. C.; Henderson, G. J.; Schiffer, C. A.; Swanstrom, R., Replacement of the P1 Amino Acid of Human Immunodeficiency Virus Type 1 Gag Processing Sites Can Inhibit or Enhance the Rate of Cleavage by the Viral Protease. *Journal of Virology* **2002,** *76* (20), 10226-10233.