Supplementary information:

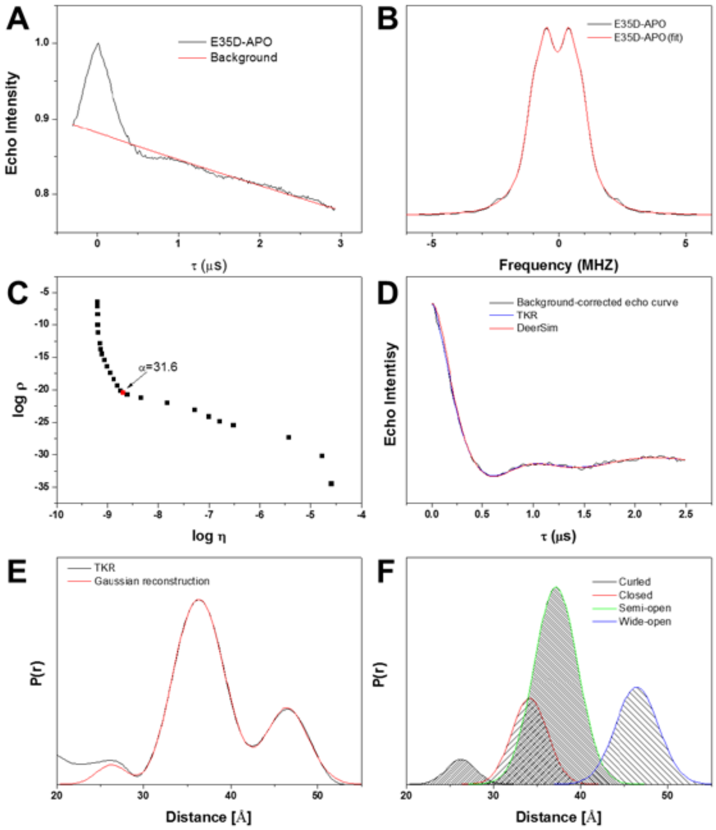
Effects of Hinge Region Natural Polymorphisms on HIV-1 Protease Structure, Dynamics and Drug-Resistance Evolution

Zhanglong Liu[1], Xi Huang[1a], Lingna Hu[1b], Linh Pham[1c], Katye M. Poole[2], Yan Tang[2], Brian P. Mahon[2], Wenxing Tang[2], Kunhua Li[1], Nathan E. Goldfarb[2d], Ben M. Dunn[2], Robert McKenna[2], Gail E. Fanucci[1*]

[1] Department of Chemistry, University of Florida, Gainesville, 32611
[2] Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610
[a] Current address: Gilead Science Inc., Foster City, CA, 94404, USA.
[b] Current address: Baxter International Inc., B3 I-park, SIP, Suzhou 215021, China.
[c] Current address: Department of Science and Mathematics, Texas A&M University - Central Texas, Killeen, TX 76549, USA
[d] Current address: California Health Sciences University, 120 North Clovis Avenue, Clovis, California 93612

Outline:

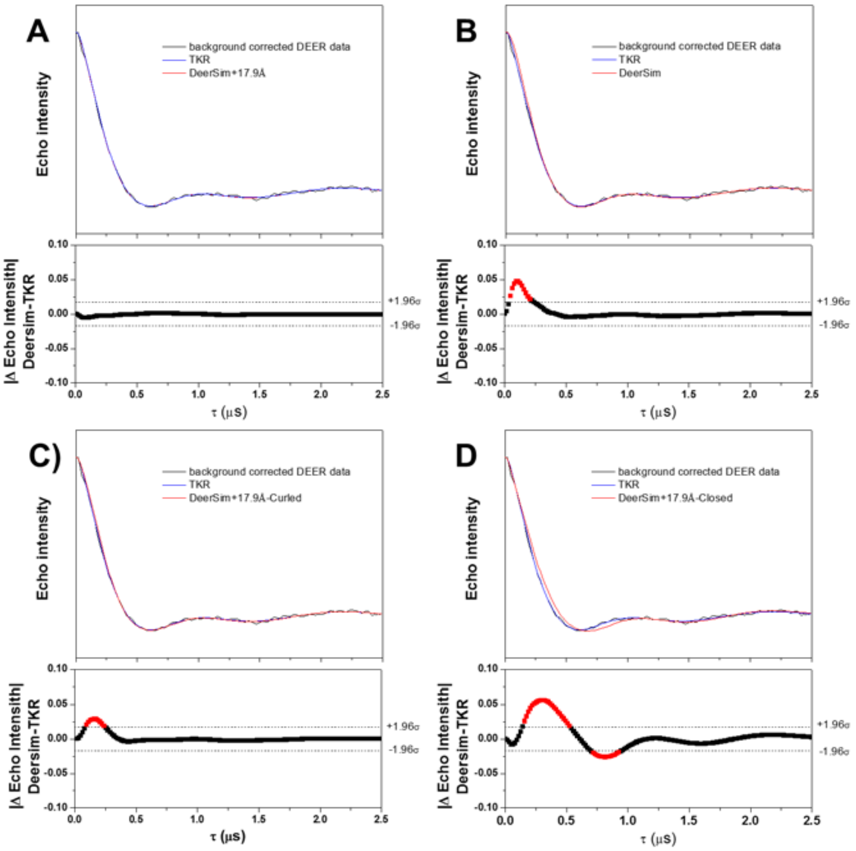
I. DEER error analyses are shown in Figure S1-S8.
II. NMR relaxation data, Model-free NMR analysis statistics and order parameters of subtype B, E35D and PR5 constructs are shown in Figure S9-S10 and Table S1.
III. The structural comparisons are shown in Figure S11.



G

E35D-APO	Distance (Å)	FWHM (Å)	Population
Curled	26.3	3.5	5%
Closed	34.2	4.1	18%
Semi-open	37.2	5.0	52%
Wide-open	46.5	4.9	25%

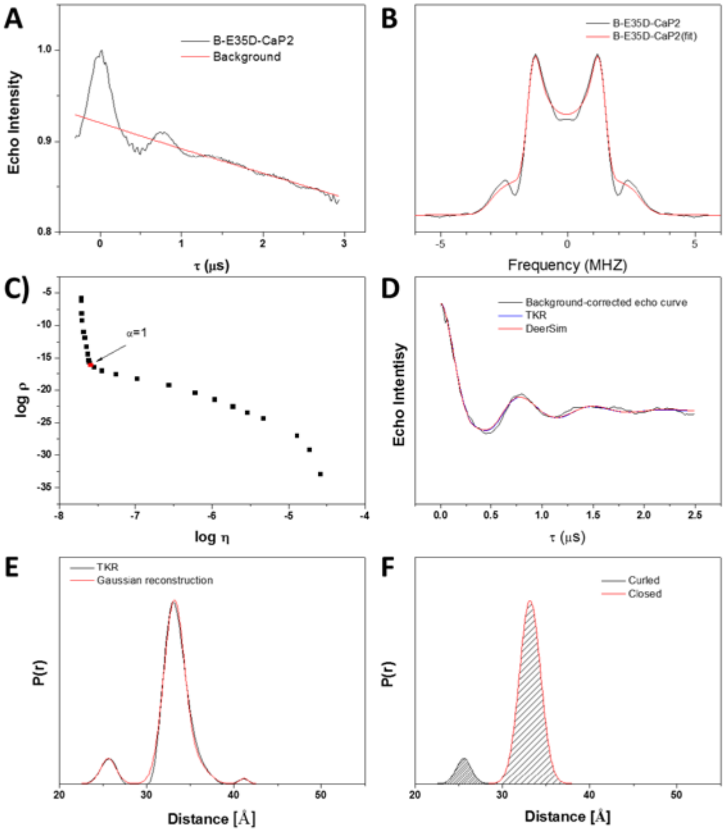
Figure S1. DEER data analysis for HIV-1PR apo-E35D, A) Raw dipolar evolution curve and background; B) Frequency domain spectrum; C) L-curve derived from TKR fit to obtain the optimal regulation parameter; D) Background corrected dipolar evolution curve (black) overlaid with the fit from TKR (blue) and the theoretical curve generated from the Gaussian reconstruction (red). E) The corresponding distance profile generated via analysis with TKR (black) and the sum of Gaussian functions used in the reconstruction (red). F) The individual Gaussian functions used in the reconstruction labeled according to the corresponding conformation of HIV-1 PR; G) Table that summarize the conformational population parameters. A relative population error of 5% was estimated. [1]



E

E35D-APO	Distance (Å)	FWHM (Å)	%	Final %
17.9Å	17.9	8.8	15.6%	0
Curled	26.3	3.5	3.8%	5%
Closed	34.2	4.1	15.6%	18%
Semi-open	37.2	5.0	43.7%	52%
Wide-open	46.5	4.9	21.3%	25%

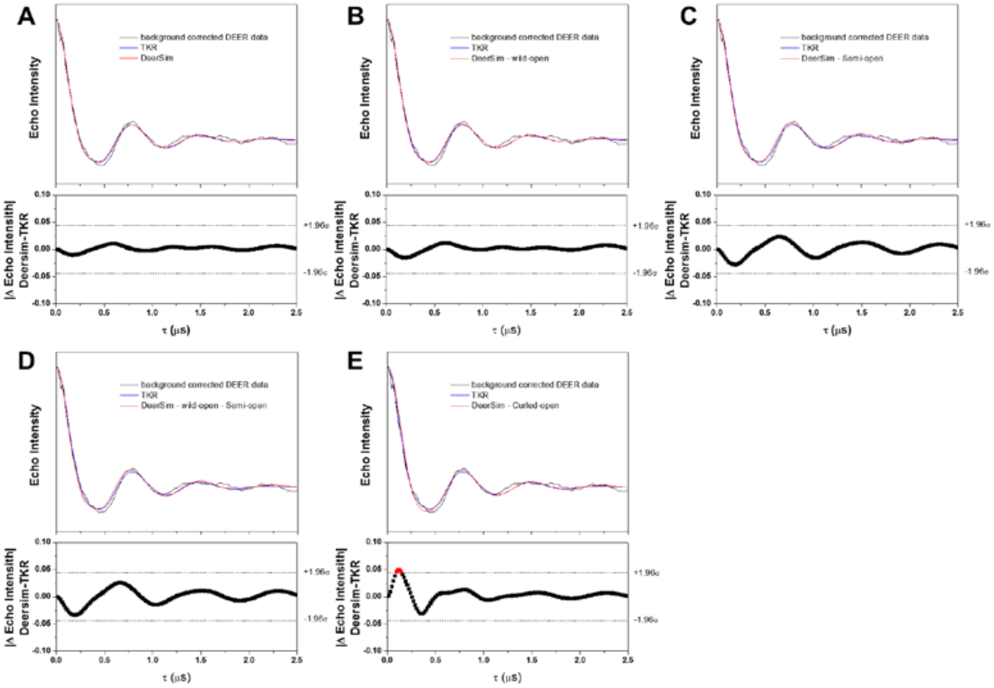
Figure S2. Error Analysis for populations <20% of HIV-1PR apo- E35D, A-D) (upper)Background corrected dipolar echo curve (black) overlaid with the TKR fit (blue) and the modified echo curve for the distance profile with one or more populations suppressed (red); (lower) The differences between TKR fit and Gaussian reconstruction echo curve. 95% confidence limit is marked as dashed line. Points marked as red represent there is significant difference between TKR fit and Gaussian reconstruction curve. E) Table of values summarizing the populations from TKR analysis and Gaussian reconstruction procedure, peak at 1.79 nm (in red) was suppressed although it was above the suppression limit because it was far below the lower distance limit of the protease conformation and might be caused by the protein aggregation. A relative population error of 5% was estimated. [1]



G

E35D-CaP2	Distance (Å)	FWHM (Å)	Population
Curled	25.6	1.8	9%
Closed	33.2	2.4	91%

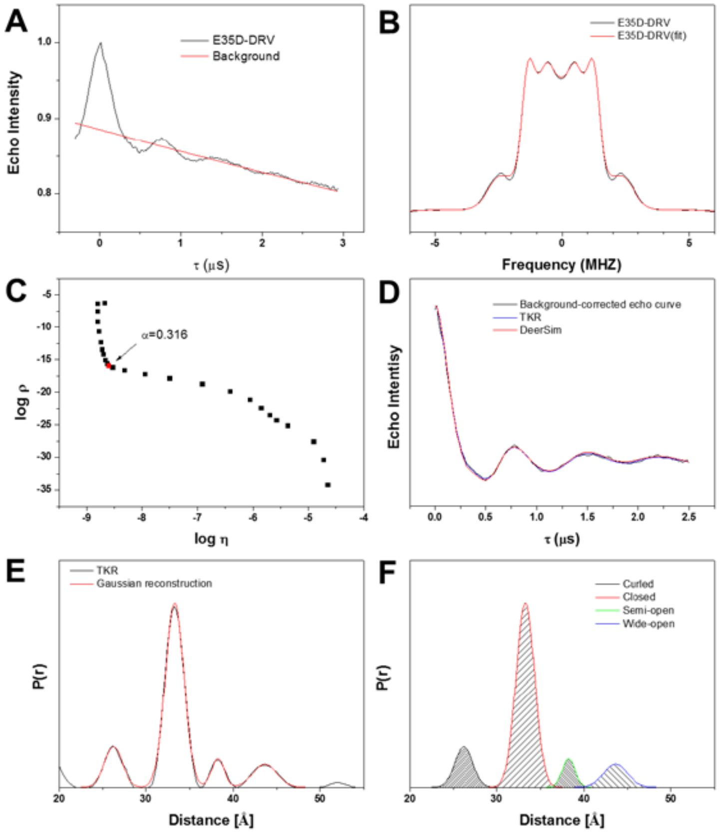
Figure S3. DEER data for HIV-1PR CaP2-bound E35D, A) Raw dipolar evolution curve and background; B) Frequency domain spectrum; C) L-curve derived from TKR fit to obtain the optimal regulation parameter; D) Background corrected dipolar evolution curve (black) overlaid with the fit from TKR (blue) and the theoretical curve generated from the Gaussian reconstruction (red). E) The corresponding distance profile generated via analysis with TKR (black) and the sum of Gaussian functions used in the reconstruction (red). F) The individual Gaussian functions used in the reconstruction labeled according to the corresponding conformation of HIV-1 PR; G) Table that summarize the conformational population parameters. A relative population error of 5% was estimated. [1]



F

E35D-CaP2	Distance (Å)	FWHM (Å)	%	Final %
Curled	25.6	1.8	8.7%	9%
Closed	33.2	2.4	83.7%	91%
Semi-open	36.2	2.2	6.6%	0%
Wide-open	41.1	0.9	1.0%	0%

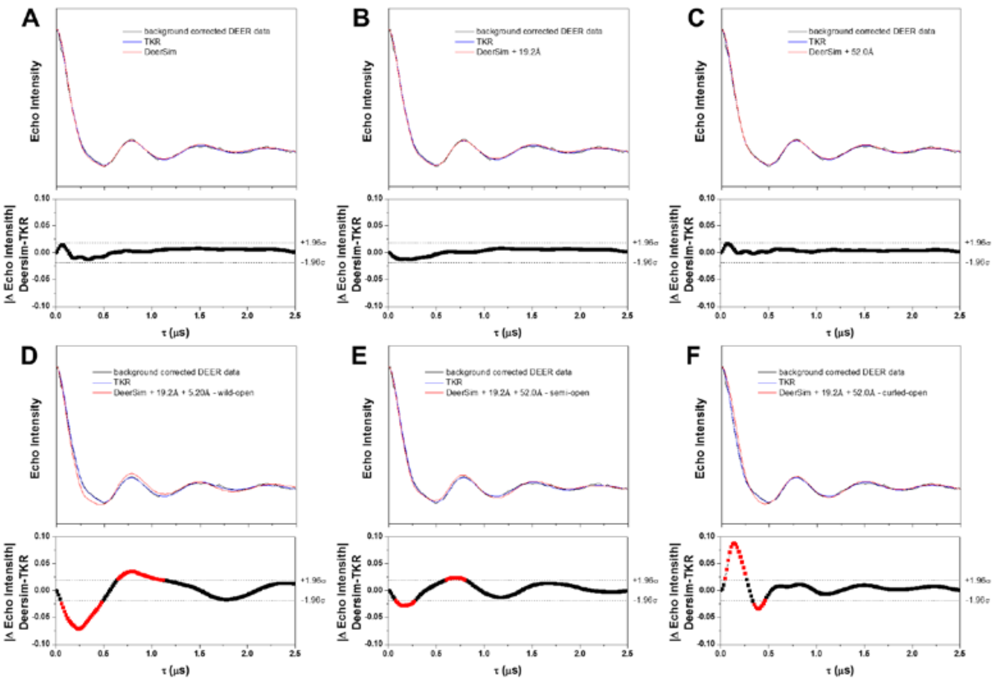
Figure S4. Error Analysis for populations <20% of HIV-1PR CaP2-bound E35D, A-E) (upper) Background corrected dipolar echo curve (black) overlaid with the TKR fit (blue) and the modified echo curve for the distance profile with one or more populations suppressed (red); (lower) The differences between TKR fit and Gaussian reconstruction echo curve. 95% confidence limit is marked as dashed line. Points marked as red represent there is significant difference between TKR fit and Gaussian reconstruction curve. F) Table of values summarizing the populations from TKR analysis and Gaussian reconstruction procedure, Semi-open (in red) and wide-open (in red) were suppressed. A relative population error of 5% was estimated. [1]



G

E35D-DRV	Distance (Å)	FWHM (Å)	Population
Curled	26.2	2.1	15%
Closed	33.3	2.1	67%
Semi-open	38.2	1.5	7%
Wide-open	43.6	2.8	11%

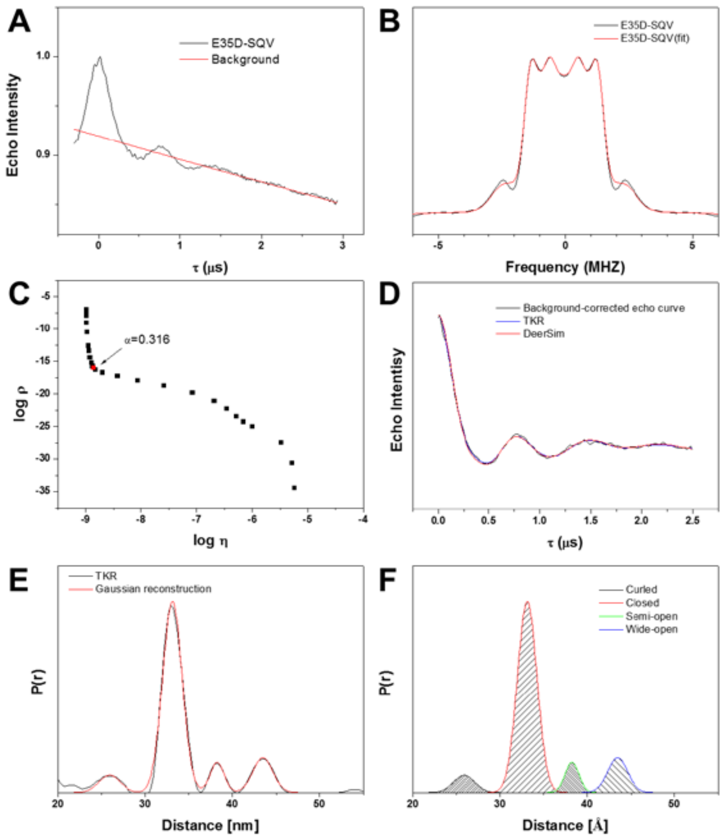
Figure S5. DEER data for HIV-1PR DRV-bound E35D, A) Raw dipolar evolution curve and background; B) Frequency domain spectrum; C) L-curve derived from TKR fit to obtain the optimal regulation parameter; D) Background corrected dipolar evolution curve (black) overlaid with the fit from TKR (blue) and the theoretical curve generated from the Gaussian reconstruction (red). E) The corresponding distance profile generated via analysis with TKR (black) and the sum of Gaussian functions used in the reconstruction (red). F) The individual Gaussian functions used in the reconstruction labeled according to the corresponding conformation of HIV-1 PR; G) Table that summarize the conformational population parameters. A relative population error of 5% was estimated. [1]



G

E35D-DRV	Distance (Å)	FWHM (Å)	%	Final %
19.2Å	19.2	2.0	9.1%	0%
Curled	26.2	2.1	13.0%	15%
Closed	33.3	2.1	60.3%	67%
Semi-open	38.2	1.5	6.5%	7%
Wide-open	43.6	2.8	9.8%	11%
52.0Å	52.0	1.7	1.2%	0%

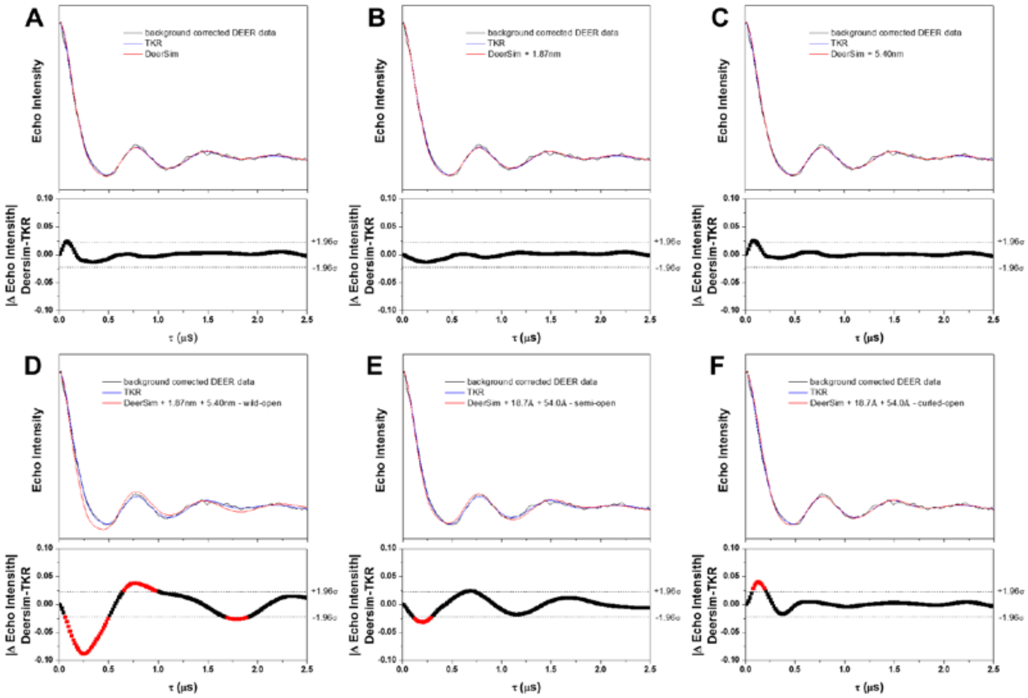
Figure S6. Error Analysis for populations <20% of HIV-1PR DRV-bound E35D, A-F) (upper)Background corrected dipolar echo curve (black) overlaid with the TKR fit (blue) and the modified echo curve for the distance profile with one or more populations suppressed (red); (lower) The differences between TKR fit and Gaussian reconstruction echo curve. 95% confidence limit is marked as dashed line. Points marked as red represent there is significant difference between TKR fit and Gaussian reconstruction curve. G) Table of values summarizing the populations from TKR analysis and Gaussian reconstruction procedure, peaks at 1.92nm (in red) and 5.2nm (in red) were suppressed. A relative population error of 5% was estimated. [1]



G

E35D-SQV	Distance (Å)	FWHM (Å)	Population
Curled	25.9	2.4	7%
Closed	33.1	2.2	72%
Semi-open	38.2	1.6	8%
Wide-open	43.5	2.3	13%

Figure S7. DEER data for HIV-1PR SQV-bound E35D, A) Raw dipolar evolution curve and background; B) Frequency domain spectrum; C) L-curve derived from TKR fit to obtain the optimal regulation parameter; D) Background corrected dipolar evolution curve (black) overlaid with the fit from TKR (blue) and the theoretical curve generated from the Gaussian reconstruction (red). E) The corresponding distance profile generated via analysis with TKR (black) and the sum of Gaussian functions used in the reconstruction (red). F) The individual Gaussian functions used in the reconstruction labeled according to the corresponding conformation of HIV-1 PR; G) Table that summarize the conformational population parameters. A relative population error of 5% was estimated. [1]



G

E35D-SQV	Distance (Å)	FWHM (nmÅ)	%	Final %
18.7Å	18.7	5.9	8.8%	0%
Curled	26.2	2.4	6.4%	7%
Closed	33.3	2.2	64.7%	72%
Semi-open	38.2	1.6	7.3%	8%
Wide-open	43.6	2.3	12.0%	13%
54.0Å	54.0	1.5	0.8%	0%

Figure S8. Error Analysis for populations <20% of HIV-1PR SQV-E35D, A-F) (upper)Background corrected dipolar echo curve (black) overlaid with the TKR fit (blue) and the modified echo curve for the distance profile with one or more populations suppressed (red); (lower) The differences between TKR fit and Gaussian reconstruction echo curve. 95% confidence limit is marked as dashed line. Points marked as red represent there is significant difference between TKR fit and Gaussian reconstruction curve. G) Table of values summarizing the populations from TKR analysis and Gaussian reconstruction procedure, peaks at 1.87nm (in red) and 5.4nm (in red) were suppressed. A relative population error of 5% was estimated. [1]

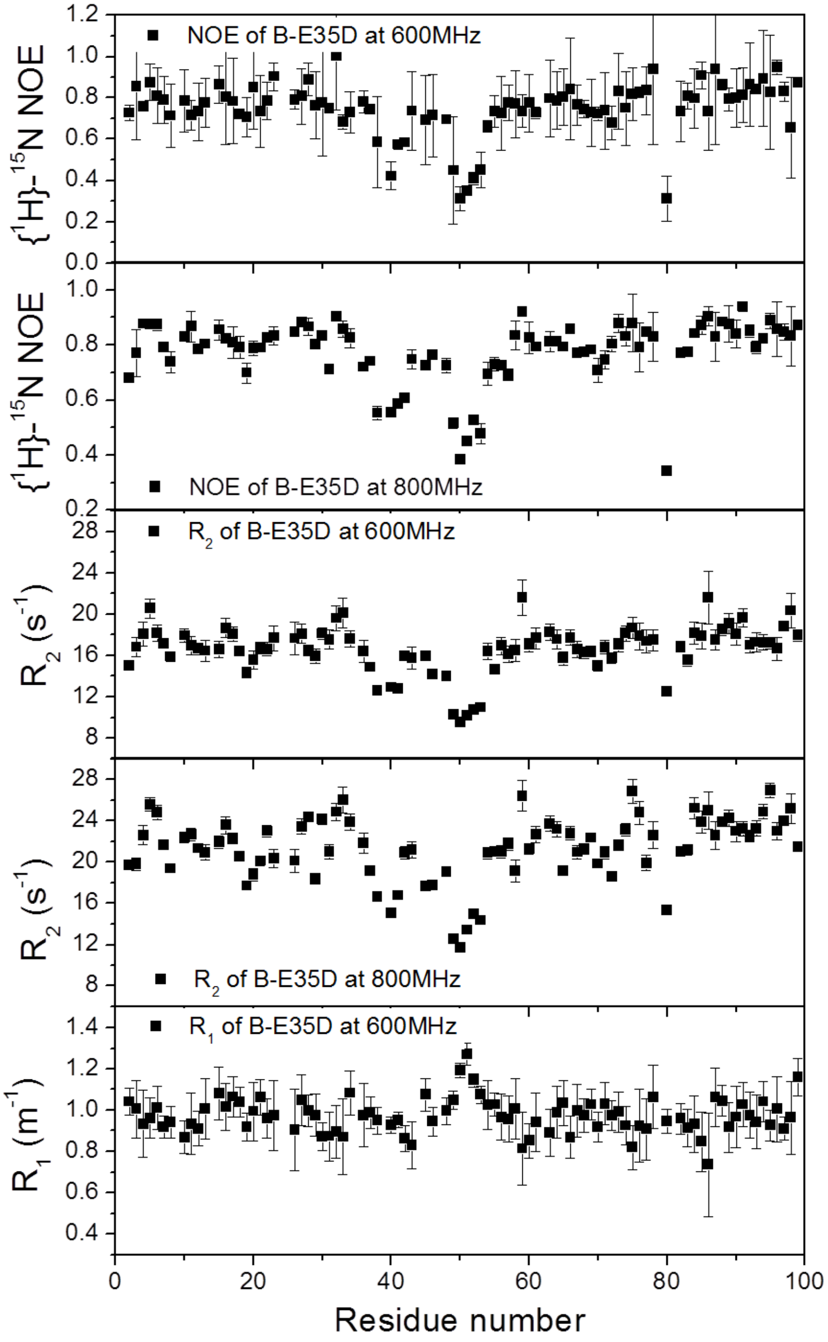


Figure S9. NMR relaxation data for E35D HIV-1PR at 600 MHz and 800 MHz.

Table S1. The number of residues was used in Model-free analysis of the NMR data for the backbone dynamics, where only the models of tm0-tm5 were applied for calculation.

Sample:	tm1	tm2	tm3	tm4	tm 5	tm 0	Total residues
B	16	13	25	23	13	2	92
E35D	6	13	23	32	7	0	81
PR5	10	12	15	41	11	0	89

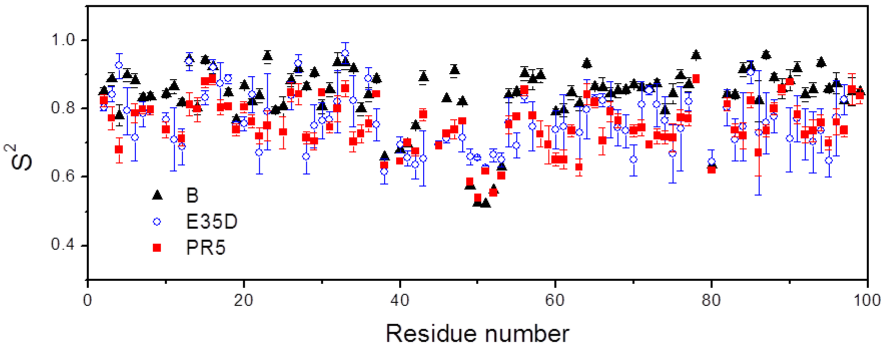


Figure S10. The backbone order parameter (S^2) of E35D (blue circle) compared to values obtained previously for subtype B (black upper triangle) and PR5(red lower square)[2] .

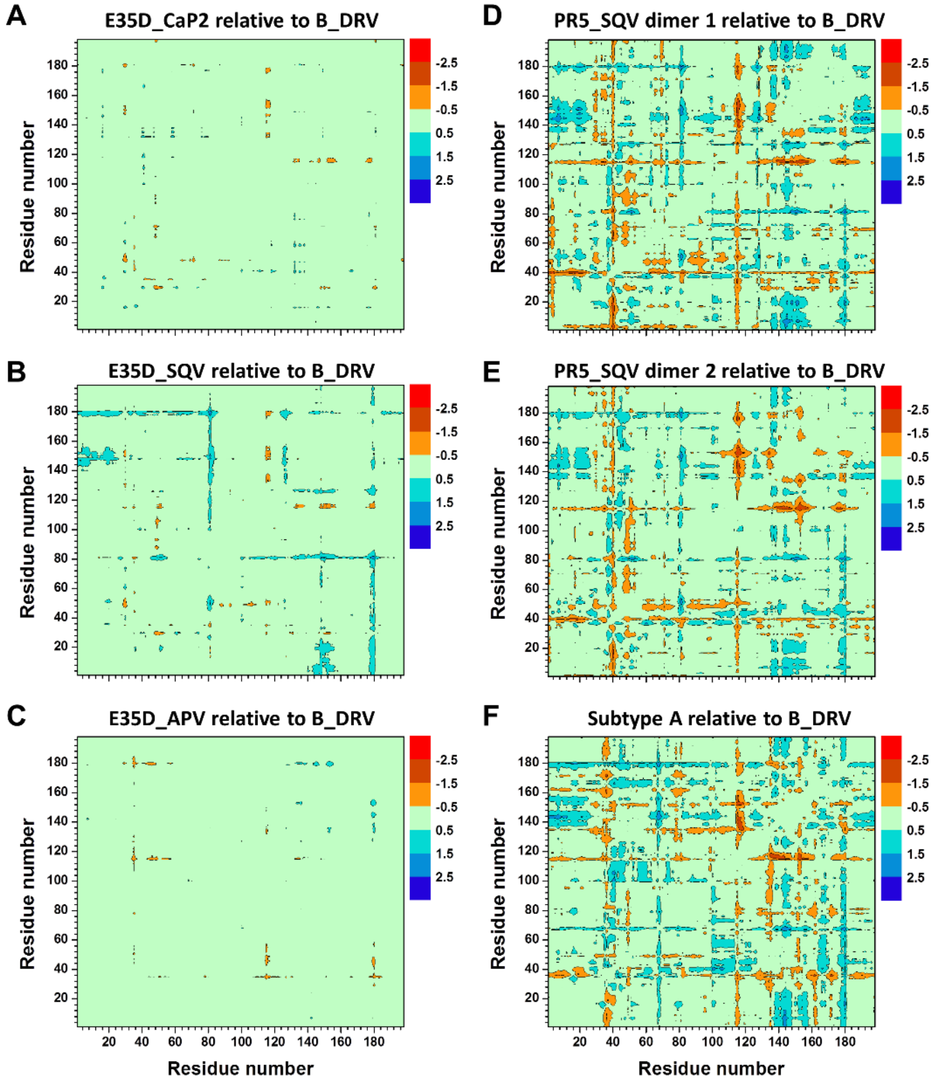


Figure S11. Double difference plots of the backbone α-carbon position of CaP2-bound E35D in A), SQV-bound E35D in B), APV-bound E35D in C), SQV-bound PR5 in D and E) and apo subtype A construct in F) comparing with DRV-bound B. The red color indicates that backbone distances of the HIV-1 PR constructs between the corresponding residues are shorter than that of the subtype B construct.

Reference:

1. de Vera, I. M. S., Smith, A. N., Dancel, M. C. A., Huang, X., Dunn, B. M., and Fanucci, G. E. (2013) Elucidating a Relationship between Conformational Sampling and Drug Resistance in HIV-1 Protease, *Biochemistry-Us 52*, 3278-3288.
2. Huang, X., Britto, M. D., Kear, J. L., Christopher, B. D., Rocca, J. R., Simmerling, C., McKenna, R., Bieri, M., Gooley, P. R., Dunn, B. M., and Fanucci, G. E. (2014) The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics, *The Journal of biological chemistry 289*, 17203 - 17214.