Supporting Information

Prediction of drug-target binding kinetics by Comparative Binding Energy analysis.

Gaurav K. Ganotra[a,c] and Rebecca C. Wade[a,b,c*]

[a]Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

[b]Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

[c]Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, Heidelberg, Germany

*Corresponding author

Email Id: Gaurav.ganotra@h-its.org, Rebecca.wade@h-its.org

Table of Content

Supplementary Methods

Structures of protein-inhibitor complexes for HSP90 and HIV-1 protease were either obtained from crystal structures downloaded from the PDB database or modeled by analogy by introducing small substitutions into crystal structures of similar compounds complexed with the proteins. The Protein Preparation wizard of the Schrodinger suite (Release 2015-4) was used to pre-process the structures of the bound complexes, to add missing side chains, to add disulphide bonds, and for optimizing the H-bond network to assign hydrogen atom positions. The protonation states were assigned at pH 7.0. Crystallographic waters were deleted after optimizing the H-bond network. To get rid of any bad contacts or steric clashes, all of the structures were subjected to the default energy minimization procedure in Schrodinger using the Impref module[1] and the OPLS3 force field. The Impref minimization involves a two-step relaxation in which first the rotatable hydrogen atoms are minimized with all the torsional potentials removed, and then an all-atom minimization is performed that is terminated either when the system is fully converged or when it reaches a heavy-atom RMSD from the initial structure of 0.30 Å.

The force field parameters and topology files for all the systems were constructed using the LEap program of the Amber14 software[2]. ff14SB[3] was used for the proteins and the inhibitor parameters were generated based on the general Amber force field (GAFF). The partial atomic charges of the inhibitors were calculated using the RESP[4] program to fit the atom-centered charges to the molecular electrostatic potential grid computed using the GAMESS program[5]. Then, all the structures were energy minimized using the PMEMD module of the Amber14 software[2]. The energy minimization protocol involved 4 separate minimization procedures with gradually decreasing harmonic restraints on heavy atoms in the first 3 minimization procedures with a restraint force constant of 500 kcal mol^{-1} $Å^{-2}$, 100 kcal mol^{-1} $Å^{-2}$ and 5 kcal mol^{-1} $Å^{-2}$, respectively. In the final minimization procedure, no positional restraints were used. For each minimization procedure, 500 steps of steepest-descent minimization followed by 500 steps of conjugate-gradient minimization were applied. Minimization was performed using implicit solvent and a distance dependent dielectric constant ($\varepsilon=4r$) was used. The gCOMBINE program[6] was used for the calculation of the Coulombic and LJ interaction energies. The gCOMBINE program decomposes the total Coulombic and LJ energies into an array of energy terms with each energy term corresponding to the interaction energy between one of the amino acid residues of the protein and the bound inhibitor. The subsequent chemometric PLS analysis was also performed using gCOMBINE.

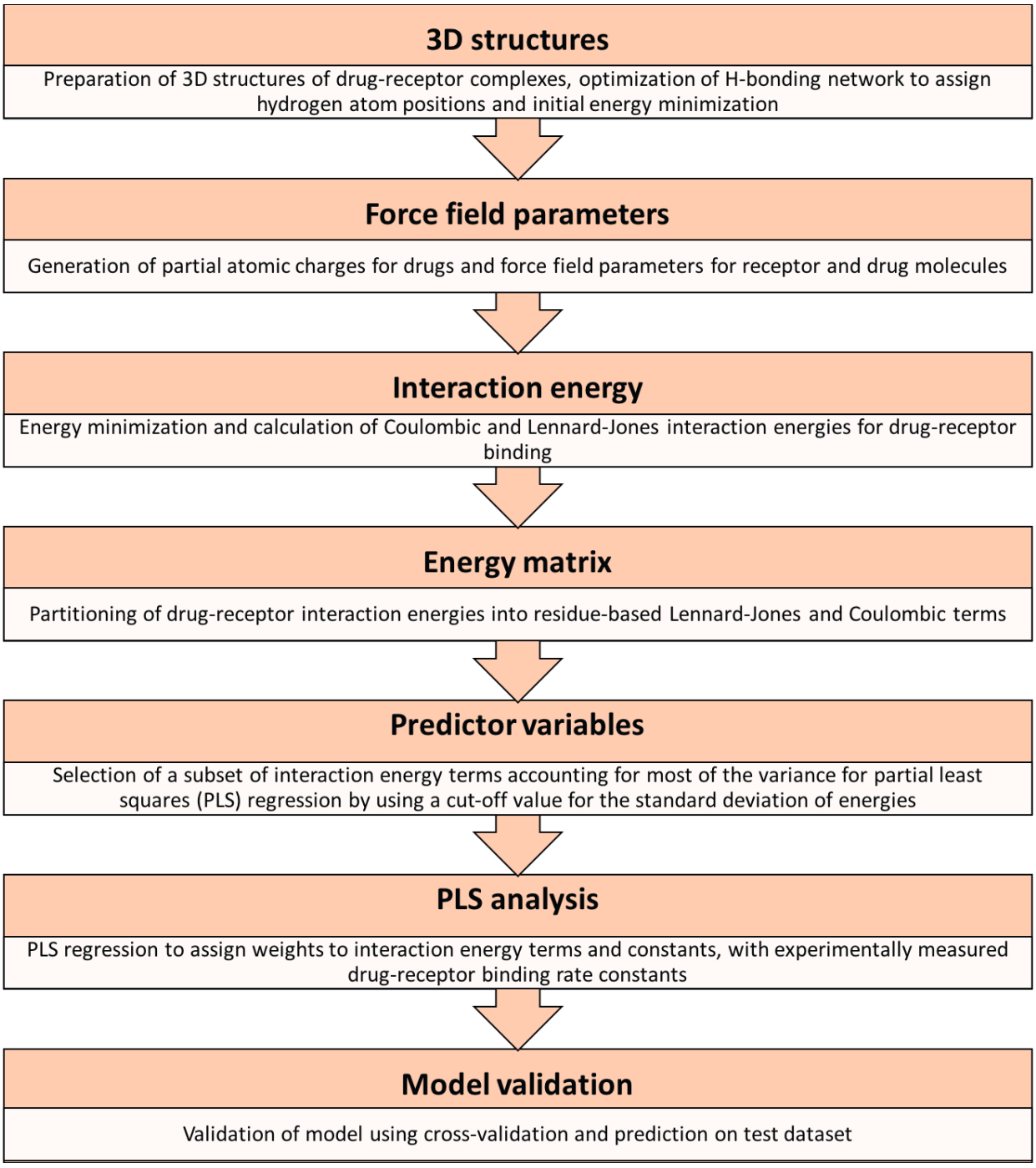

Figure S1: Schematic outline of the different steps involved in applying COMBINE analysis to derive a QSKR to predict drug-binding kinetics.

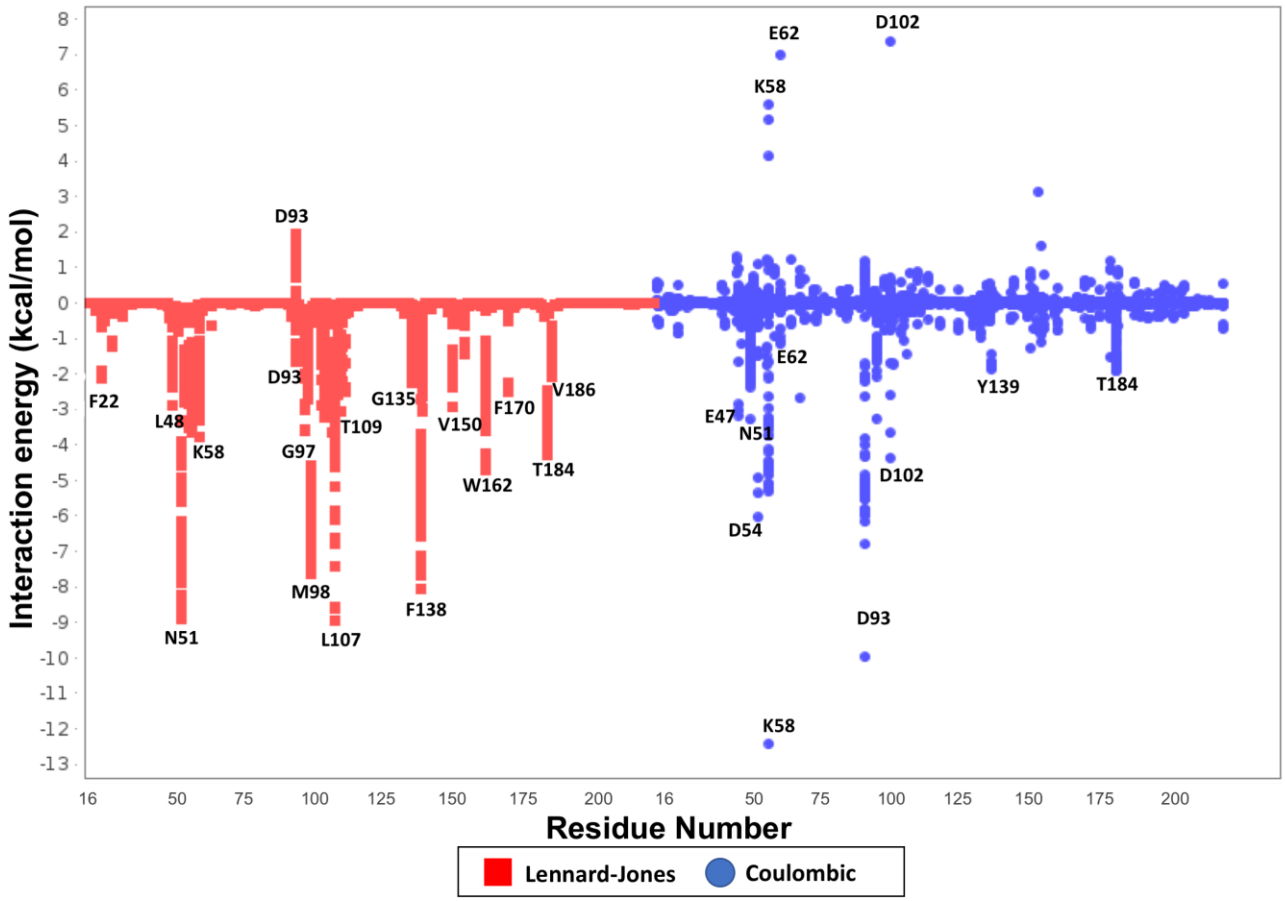

Figure S2. Interaction energies between N-HSP90 amino acid residues and inhibitors. Lennard-Jones and Coulombic interaction energies were computed between inhibitors and 207 residues in the N-terminal domain of HSP90 using the gCOMBINE program. The first 207 columns on the x-axis correspond to Lennard-Jones energies (kcal/mol) for each residue and the last 207 columns correspond to Coulombic energies (kcal/mol) between each inhibitor and different residues. Each column has 66 data points corresponding to the 66 inhibitors used for the COMBINE analysis. Selected amino acid residues, whose interaction energies with the bound inhibitor show a high variance in the dataset, are labelled.

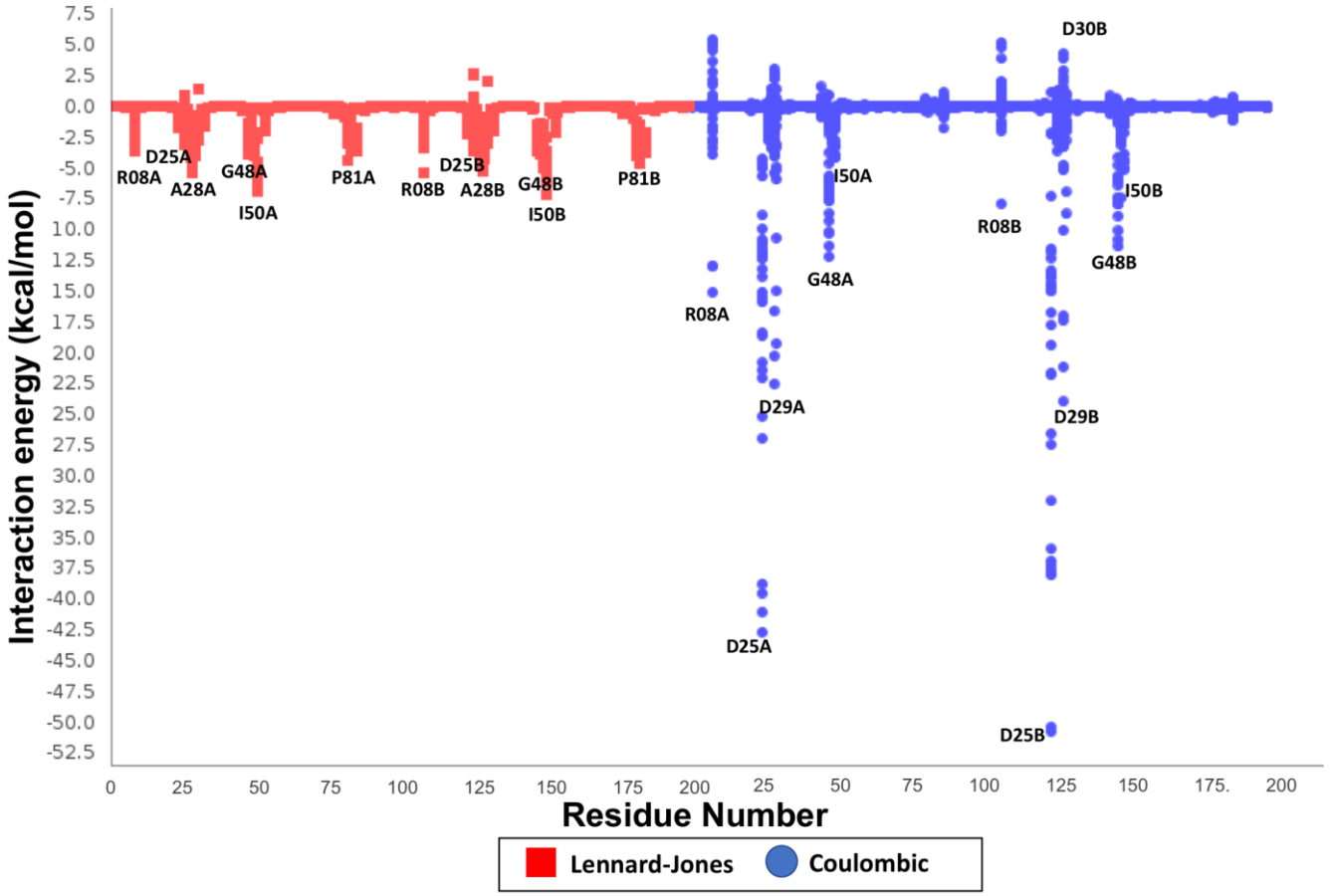

Figure S3. Interaction energies between HIV-1 protease residues and inhibitors. Lennard-Jones and Coulombic interaction energies were computed between the inhibitors and 198 amino acid residues of the protease dimer using the gCOMBINE program. The first 198 columns on the x-axis correspond to Lennard-Jones energies (kcal/mol) for each residue and the last 198 columns correspond to Coulombic energies (kcal/mol) between each inhibitor and different residues. Each column has 33 data points corresponding to the 33 inhibitors used for the COMBINE analysis. Selected amino acid residues, whose interaction energies with the bound inhibitor show a high variance in the dataset, are labelled.

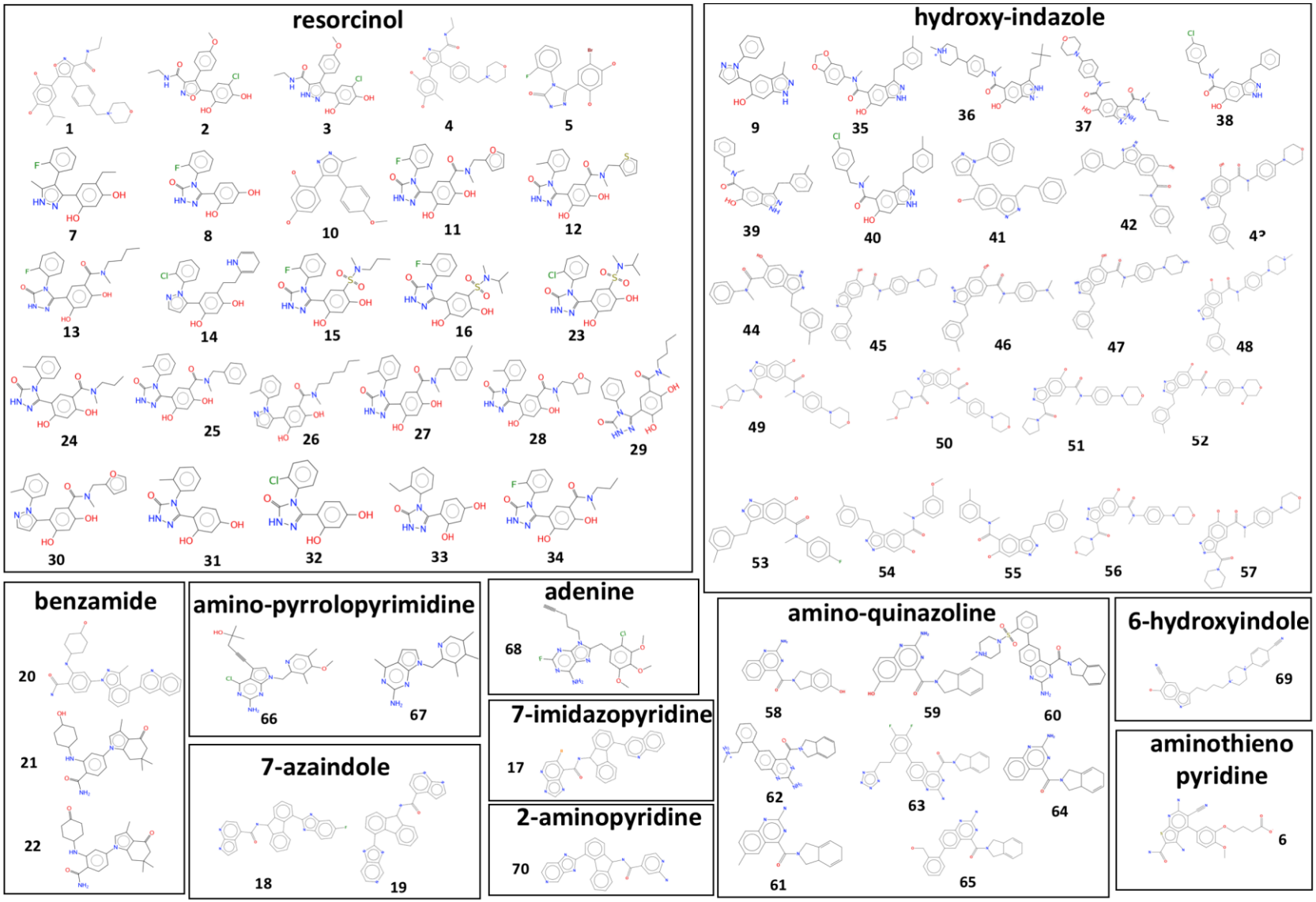

Figure S4: 2D chemical structures of inhibitors of HSP90 used for the COMBINE analysis. These 70 inhibitors belong to 11 different chemical classes: resorcinol, hydroxyl-indazole, aminoquinazoline, benzamide, aminopyrrolopyrimidine, 7-imidazopyridine, 7-azaindole, aminothienopyridine, 6-hydroxyindole, adenine and 2-aminopyridine

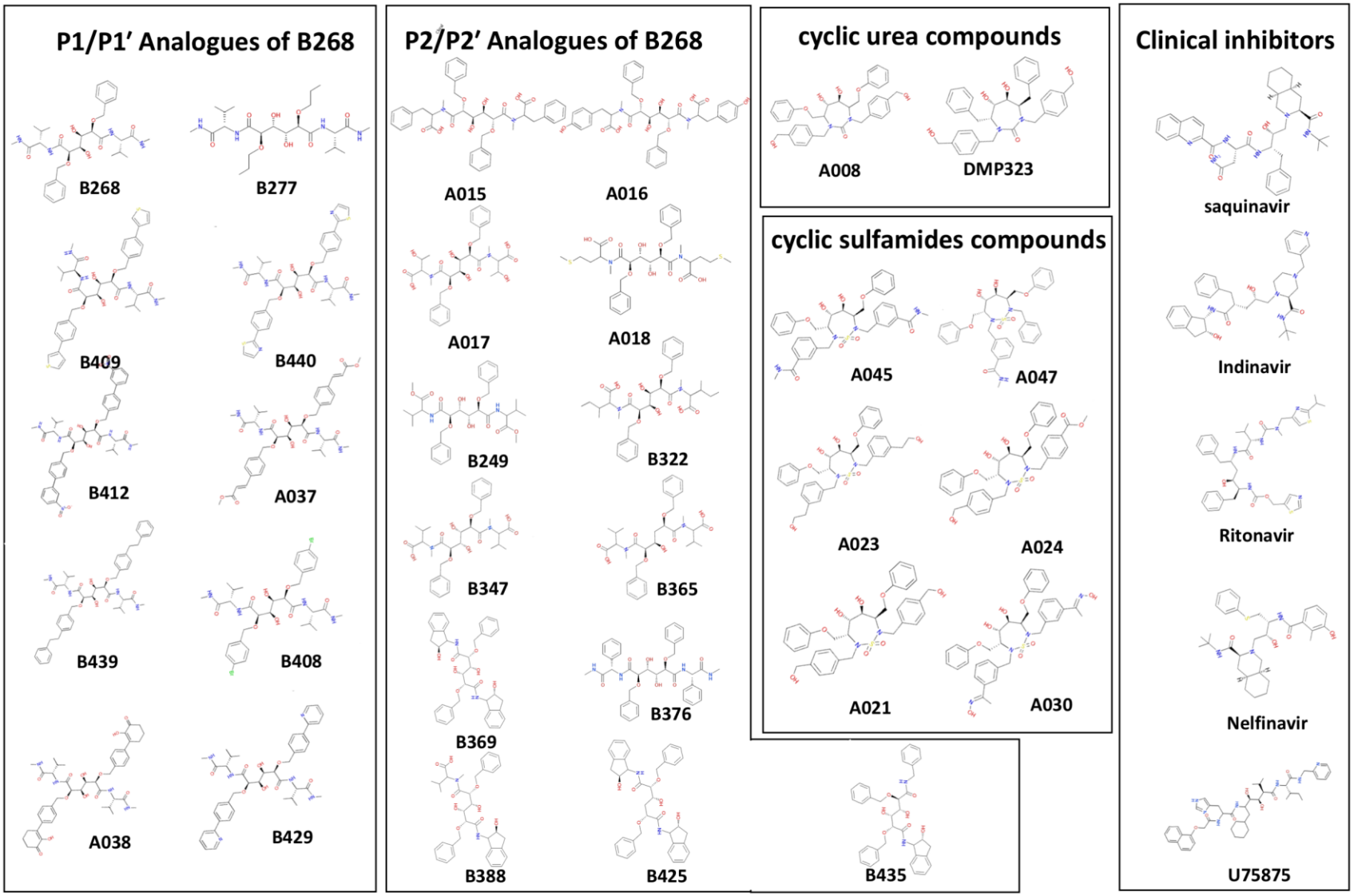

Figure S5: 2D chemical structures of inhibitors of HIV-1 protease used for the COMBINE analysis

Table S1. Summary of the models derived for different numbers of latent variables (LVs) for the COMBINE analysis for k_{off} rate constants of N-HSP90 inhibitors. The models were derived using $\log_{10}(k_{off})$ values (unit of k_{off} rates in s^{-1}) as response variable in the PLS analysis. The table lists the regression coefficient (R^2) for the training set, the correlation coefficient for leave-one-out cross validation sets (Q^2_{LOO}), average absolute errors (AAE_T and AAE_V) and root mean squared errors (RME_T and RME_V) for the training set and leave-one-out validation sets, respectively, the correlation coefficient (R^2_{PRED}) for the test-set (prediction-set), average absolute error (AAE_P) and root mean squared error (RME_P) for the test-set. The model with 3 LVs displayed the best balance between predictive performance and number of LVs.

No. of Latent Variables	R^2	Q^2_{LOO}	AAE_T	AAE_V	RME_T	RME_V	R^2_{PRED}	AAE_P	RME_P
1	0.43	0.31	0.61	0.67	0.78	0.85	0.33	0.60	0.79
2	0.75	0.66	0.39	0.45	0.51	0.60	0.71	0.44	0.52
3	0.80	0.69	0.37	0.45	0.46	0.57	0.86	0.33	0.37
4	0.82	0.72	0.35	0.44	0.44	0.55	0.87	0.32	0.35
5	0.85	0.71	0.32	0.45	0.40	0.55	0.86	0.31	0.36
6	0.87	0.71	0.30	0.45	0.38	0.56	0.89	0.26	0.33
7	0.88	0.69	0.27	0.45	0.36	0.58	0.89	0.27	0.33
8	0.89	0.66	0.26	0.47	0.34	0.60	0.86	0.32	0.36
9	0.90	0.64	0.25	0.46	0.33	0.62	0.85	0.33	0.37
10	0.91	0.55	0.23	0.51	0.31	0.70	0.69	0.48	0.54

Table S2. Comparison of $\log_{10}(k_{off})$ values calculated by the COMBINE analysis model in PLS regression (column 5) and the experimental $\log_{10}(k_{off})$ values from Kokh et al.[7] (column 4) for different N-HSP90 inhibitors used for training the COMBINE analysis model. The $\log_{10}(k_{off})$ values predicted from leave-one-out (LOO) cross-validation for the training set of inhibitors are given in the last column.

Compound Id	N-HSP90 Binding site conformation: loop or helix	PDB ID	Experimental $\log_{10}(k_{off}\ (s^{-1}))$[7]	Fitted $\log_{10}(k_{off}\ (s^{-1}))$ (PLS regression)	Predicted $\log_{10}(k_{off}(s^{-1}))$ (LOO cross-validation)
1	loop	2VCI	-4.00 ± 0.00	-3.45	-2.57
3	loop	2BSM	-2.00 ± 0.04	-2.54	-2.47
5	loop	5J2X	-1.85 ± 0.07	-1.21	-1.16
7	loop	6ELO	-1.20 ± 0.02	-0.77	-0.65
8	loop	5J64	-0.68 ± 0.07	-0.60	-0.65
9	loop	n.a.	-0.08 ± 0.03	-0.86	-1.13
10	loop	6ELN	-0.60 ± 0.03	-1.12	-1.25
11	helix	5J20	-3.48 ± 0.03	-2.47	-2.37
12	helix	5J86	-2.75 ± 0.09	-2.80	-2.79
13	helix	5J9X	-2.77 ± 0.12	-2.43	-2.41
14	helix	6ELP	-0.76 ± 0.06	-1.91	-2.16
15	helix	5J27	-2.19 ± 0.03	-2.02	-2.00
16	helix	5J86	-1.85 ± 0.05	-2.09	-2.13
17	helix	5LRZ	-3.56 ± 0.01	-3.99	-3.89
18	helix	5LR7	-3.72 ± 0.16	-3.18	-2.59
20	helix	5LQ9	-3.87 ± 0.01	-4.18	-4.11
21	helix	5LS1	-3.31 ± 0.12	-2.89	-2.80
22	helix	5T21	-3.12 ± 0.03	-2.61	-2.52
23	helix	n.a.	-2.02 ± 0.02	-1.80	-1.76
24	helix	n.a.	-2.33 ± 0.07	-2.06	-2.04
27	helix	n.a.	-2.92 ± 0.04	-2.53	-2.52
28	helix	n.a.	-2.34 ± 0.08	-2.59	-2.66
29	helix	n.a.	-2.52 ± 0.04	-2.43	-2.44
31	loop	n.a.	-0.96 ± 0.17	-1.26	-1.34

33	loop	n.a.	-1.15 ± 0.10	-0.64	-0.58
36	helix	5LO6	-2.86 ± 0.12	-3.21	-3.32
37	helix	5LNZ	-2.70 ± 0.04	-3.13	-3.14
38	helix	6EY8	-1.54 ± 0.02	-2.09	-2.08
39	helix	6EFU	-1.65 ± 0.02	-1.95	-1.93
40	helix	6EY9	-1.76 ± 0.01	-2.16	-2.13
41	helix	6EY8	-0.63 ± 0.04	-1.72	-1.87
42	helix	n.a.	-2.30 ± 0.08	-2.19	-2.17
43	helix	5OCI	-3.17 ± 0.00	-2.81	-2.77
44	helix	n.a.	-2.04 ± 0.05	-1.93	-1.91
46	helix	n.a.	-2.63 ± 0.07	-2.76	-2.78
47	helix	n.a.	-2.91 ± 0.03	-2.76	-2.75
48	helix	n.a.	-3.12 ± 0.07	-2.98	-2.94
50	helix	5ODX	-3.53 ± 0.02	-3.34	-3.28
51	helix	5NYH	-2.62 ± 0.01	-2.60	-2.56
52	helix	n.a.	-2.86 ± 0.14	-2.78	-2.73
55	helix	n.a.	-2.11 ± 0.27	-2.10	-2.08
56	helix	n.a.	-1.88 ± 0.02	-2.62	-2.67
57	helix	n.a.	-3.04 ± 0.12	-2.78	-2.72
58	helix	n.a.	-0.26 ± 0.12	-0.45	-0.63
59	helix	n.a.	-0.24 ± 0.02	-0.45	-0.57
60	helix	5OD7	-3.62 ± 0.11	-3.85	-3.64
62	helix	6EI5	-2.34 ± 0.04	-2.11	-2.06
63	helix	n.a.	-2.82 ± 0.05	-2.83	-3.25
64	helix	n.a.	-0.26 ± 0.04	-0.21	-0.29
66	helix	n.a.	-2.90 ± 0.08	-2.26	-2.17
67	helix	5LR1	-1.59 ± 0.02	-0.73	-0.68
68	helix	6EL5	-1.48 ± 0.02	-2.04	-2.19
70	helix	2YKJ	-3.00 ± 0.06	-2.65	-2.27

Table S3. Comparison of $\log_{10}(k_{off})$ values predicted by the COMBINE analysis model and the experimental $\log_{10}(k_{off})$ values[7] for different N-HSP90 inhibitors used in the test set for validation of the COMBINE analysis model.

Compound Id	N-HSP90 binding site confirmation: loop or helix	PDB ID	Experimental $\log_{10}(k_{off}(s^{-1}))$[7]	Predicted $\log_{10}(k_{off}(s^{-1}))$ (COMBINE)
2	loop	2UWD	-2.70 ± 0.03	-2.45
4	loop	5NYI	-4.00 ± 0.00	-3.77
19	helix	2YKI	-3.55 ± 0.07	-3.23
25	helix	n.a.	-2.96 ± 0.21	-2.45
26	helix	n.a.	-2.00 ± 0.07	-2.44
32	loop	n.a.	-0.92 ± 0.07	-1.32
34	helix	n.a.	-2.38 ± 0.05	-2.90
35	helix	6EYA	-2.27 ± 0.03	-2.31
45	helix	n.a.	-3.13 ± 0.05	-2.67
49	helix	n.a.	-2.86 ± 0.06	-2.91
53	helix	n.a.	-1.50 ± 0.22	-1.96
54	helix	n.a.	-1.79 ± 0.10	-2.24
61	helix	n.a.	-0.58 ± 0.12	-0.42

Table S4. Summary of the models derived for different numbers of latent variables (LVs) for the COMBINE analysis for k_{off} rate constants of HIV-1 protease inhibitors. The models were derived using $\log_{10}(k_{off})$ (unit of k_{off} rates in s^{-1}) as the response variable in the PLS analysis. The table lists the regression coefficient (R^2) for the training set, the correlation coefficient for leave-one-out cross validation sets (Q^2_{LOO}), average absolute errors (AAE_T and AAE_V) and root mean squared errors (RME_T and RME_V) for the training set and leave-one-out validation sets, respectively. The model with 6 LVs displayed the best predictive performance and least sensitivity in different cross-validation methods used.

No. of Latent Variables	R^2	Q^2_{LOO}	AAE_T	AAE_V	RME_T	RME_V
1	0.20	0.01	1.03	1.14	1.23	1.36
2	0.38	-0.01	0.86	1.09	1.09	1.38
3	0.56	0.06	0.70	1.02	0.91	1.33
4	0.74	0.34	0.55	0.89	0.70	1.11
5	0.83	0.38	0.46	0.83	0.57	1.08
6	0.94	0.70	0.26	0.58	0.34	0.75
7	0.96	0.77	0.23	0.52	0.27	0.66
8	0.97	0.83	0.20	0.47	0.23	0.57
9	0.98	0.83	0.18	0.47	0.20	0.56
10	0.98	0.83	0.16	0.48	0.19	0.57

Table S5. Comparison of $\log_{10}(k_{off})$ values calculated by the COMBINE analysis model in PLS regression (column 4) and the experimental $\log_{10}(k_{off})$ values from Markgren et al.[8] (column 3) for different HIV-1 protease inhibitors used for training the COMBINE analysis model. The $\log_{10}(k_{off})$ values predicted from leave-one-out (LOO) cross-validation for the training set of inhibitors are given in the last column.

Compound Id	PDB ID	Experimental $\log_{10}(k_{off}(s^{-1}))$	Fitted $\log_{10}(k_{off}(s^{-1}))$ (PLS regression)	Predicted $\log_{10}(k_{off}(s^{-1}))$ (LOO cross-validation)
B435	1D4H	-2.19 ± 0.09	-1.48	-1.36
A047	1G2K	-1.16 ± 0.10	-0.81	-0.74
A023	n.a.	-0.86 ± 0.11	-0.66	-0.05
A024	1G35	-1.16 ± 0.10	-1.30	-1.44
B429	n.a.	-3.43 ± 0.09	-3.71	-3.69
B409	1EC1	-3.37 ± 0.13	-3.21	-3.16
B268	n.a.	-2.44 ± 0.05	-2.43	-2.49
A045	n.a.	-0.58 ± 0.09	-0.90	-1.00
B425	1D4I	-0.63 ± 0.00	-1.62	-2.06
A021	n.a.	-1.56 ± 0.04	-1.16	-0.94
saquinavir	3OXC	-3.64 ± 0.06	-4.00	-3.72
indinavir	2BPX	-2.80 ± 0.04	-2.35	-0.77
ritonavir	1HXW	-2.67 ± 0.06	-2.39	-1.69
DMP323	1QBS	1.92 ± 0.12	1.91	1.41
nelfinavir	1OHR	-3.18 ± 0.04	-3.07	-2.08
B369	1EBY	-1.88 ± 0.19	-2.09	-2.38
B388	n.a.	-1.64 ± 0.15	-1.66	-1.67
A038	n.a.	-3.31 ± 0.02	-3.01	-1.89
A037	n.a.	-3.44 ± 0.04	-3.57	-3.05
B440	n.a.	-3.52 ± 0.02	-3.54	-3.51
B439	n.a.	-2.79 ± 0.06	-2.87	-2.89
B408	n.a.	-2.77 ± 0.02	-2.52	-2.37
B412	n.a.	-3.09 ± 0.19	-3.28	-3.29
A008	n.a.	1.64 ± 0.15	1.45	0.51
B277	n.a.	-2.31 ± 0.17	-2.07	-1.96
A030	n.a.	-1.38 ± 0.13	-1.00	-0.71
A015	n.a.	-0.03 ± 0.37	-0.78	-1.35

A016	n.a.	-1.22 ± 0.22	-1.37	-1.50
A017	n.a.	-0.75 ± 0.09	-0.61	-0.71
B322	n.a.	-1.17 ± 0.29	-1.43	-1.60
B365	n.a.	-1.51 ± 0.06	-1.52	-1.60
B347	n.a.	-1.57 ± 0.05	-1.31	-0.88
A018	n.a.	-0.32 ± 0.20	-0.43	-1.28

Table S6. Summary of the models derived for different numbers of latent variables (LVs) for the COMBINE analysis for the K_D of the N-HSP90 inhibitors. The models were derived using $\log_{10}(K_D)$ (unit of K_D is M) as the response variable in the PLS analysis. For deriving the model for K_D, the full dataset of 66 compounds was initially used for training. 3 outliers (compounds 17, 50 and 67) were later removed from the PLS analysis to improve the quality of the model. Therefore, the final model for K_D was trained with 63 compounds. The table lists the regression coefficient (R^2) for the training set, the correlation coefficient for leave-one-out cross validation sets (Q^2_{LOO}), average absolute errors (AAE_T and AAE_V) and root mean squared errors (RME_T and RME_V) for the training set and leave-one-out validation sets, respectively. The model with 3 LVs displayed the best balance between predictive performance and number of LVs used for PLS.

No. of Latent Variables	R^2	Q^2_{LOO}	AAE_T	AAE_V	RME_T	RME_V
1	0.33	0.13	0.66	0.74	0.90	1.02
2	0.55	0.34	0.58	0.68	0.73	0.89
3	0.59	0.41	0.57	0.67	0.70	0.84
4	0.64	0.39	0.52	0.67	0.66	0.85
5	0.69	0.40	0.49	0.67	0.61	0.85
6	0.70	0.32	0.48	0.68	0.60	0.90
7	0.72	0.16	0.47	0.74	0.58	1.00
8	0.74	0.09	0.45	0.76	0.56	1.04
9	0.74	-0.17	0.44	0.82	0.55	1.18
10	0.75	-0.75	0.44	0.91	0.54	1.44

Table S7. Statistical measures of correlation for the COMBINE analysis model derived for K_D of N-HSP90 inhibitors. The table lists the cross-validated correlation coefficient (Q^2), average absolute error (AAE_V) and root mean squared error (RME_V) for different validation methods used. These statistical measures correspond to a model derived with 3 latent variables in PLS analysis.

Validation	R^2	Q^2	AAE_V	RME_V
Leave-one-out (LOO)	0.59	0.41	0.67	0.84
Leave-two-out (L2O)	0.59	0.41	0.68	0.84
Leave-three-out (L3O)	0.59	0.38	0.70	0.86
Random groups of 7 (10 iterations)	0.59	0.37	0.69	0.87

Table S8. Summary of the models derived for different numbers of latent variables (LVs) for the COMBINE analysis for K_D for the HIV-1 protease inhibitors. The models were derived using $\log_{10}(K_D)$ (unit of K_D is M) as the response variable in the PLS analysis. For deriving the model for K_D, the full dataset of 36 compounds was initially used for training. 3 outliers (compounds B435, A037 and B249) were later removed from the PLS analysis to improve the quality of the model. Therefore, the final model for K_D was trained with 33 compounds. The table lists the regression coefficient (R^2) for the training set, the correlation coefficient for leave-one-out cross validation sets (Q^2_{LOO}), average absolute errors (AAE_T and AAE_V) and root mean squared errors (RME_T and RME_V) for the training set and leave-one-out validation sets, respectively. The model with 6 LVs displayed the best balance between predictive performance and number of LVs used for PLS.

No. of Latent Variables	R^2	Q^2_{LOO}	AAE_T	AAE_V	RME_T	RME_V
1	0.30	0.08	1.04	1.22	1.23	1.42
2	0.39	0.16	0.95	1.13	1.15	1.35
3	0.55	0.27	0.78	1.05	0.99	1.26
4	0.64	0.37	0.68	0.96	0.88	1.17
5	0.69	0.45	0.63	0.85	0.82	1.09
6	0.78	0.53	0.50	0.79	0.69	1.02
7	0.80	0.53	0.47	0.78	0.65	1.01
8	0.83	0.53	0.45	0.76	0.60	1.01
9	0.86	0.43	0.41	0.89	0.54	1.11
10	0.89	0.24	0.39	1.01	0.50	1.28

Table S9. Statistical measures of correlation for the COMBINE analysis model derived for K_D of HIV-1 protease inhibitors. The table lists the cross-validated correlation coefficient (Q^2), average absolute error (AAE_V) and root mean squared error (RME_V) for different validation methods used. These statistical measures correspond to a model derived with 6 latent variables in PLS analysis.

Validation	R^2	Q^2	AAE_V	RME_V
Leave-one-out (LOO)	0.78	0.53	0.79	1.02
Leave-two-out (L2O)	0.78	0.46	0.81	1.08
Leave-three-out (L3O)	0.78	0.44	0.81	1.10
Random groups of 5 (10 iterations)	0.78	0.48	0.83	1.06

Table S10. Statistical measures of correlation for the COMBINE analysis model derived for the K_D for the resorcinol series of inhibitors of HSP90. The models were derived using $\log_{10}(K_D)$ (unit of K_D is M) as the response variable in the PLS analysis. For deriving the model for K_D, a smaller dataset of 25 compounds belonging to the resorcinol series was used for training. 3 outliers (compounds 23, 28, 30) were later removed from the PLS analysis to improve the quality of the model. Therefore, the final model for K_D was trained with 22 compounds. The table lists the cross-validated correlation coefficients (Q^2), average absolute error (AAE_V) and root mean squared error (RME_V) for different validation methods used. These statistical measures correspond to a model derived with 4 latent variables in PLS analysis.

Validation	R^2	Q^2	AAE_V	RME_V
Leave-one-out (LOO)	0.79	0.49	0.47	0.57
Leave-two-out (L2O)	0.79	0.45	0.47	0.59
Leave-three-out (L3O)	0.79	0.47	0.46	0.58
Random groups of 5 (10 iterations)	0.79	0.43	0.49	0.61

Table S11. List of the SMILES strings for the HSP90 inhibitors.

Com-pound Id	SMILES
1	CCNC(=O)c1noc(-c2cc(C(C)C)c(O)cc2O)c1-c1ccc(C[NH+]2CCOCC2)cc1
2	CCNC(=O)c1noc(-c2cc(Cl)c(O)cc2O)c1-c1ccc(OC)cc1
3	CCNC(=O)c1[nH]nc(-c2cc(Cl)c(O)cc2O)c1-c1ccc(OC)cc1
4	CCNC(=O)c1noc(c2cc(Cl)c(O)cc2O)c1c3ccc(C[NH+]4CCOCC4)cc3
5	O=c1[nH]nc(-c2cc(Br)c(O)cc2O)n1-c1ccccc1F
6	COc1ccc(-c2c(C#N)c(N)nc3sc(C(N)=O)c(N)c23)cc1OCCCC(=O)O
7	CCc1cc(-c2n[nH]c(C)c2-c2ccccc2F)c(O)cc1O
8	O=c1[nH]nc(-c2ccc(O)cc2O)n1-c1ccccc1F
9	Cc1n[nH]c2cc(O)c(-c3ccnn3-c3ccccc3)cc12
10	COc1ccc(-c2c(-c3ccc(O)cc3O)n[nH]c2C)cc1
11	CN(Cc1ccco1)C(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2F)c(O)cc1O
12	Cc1ccccc1-n1c(-c2cc(C(=O)N(C)Cc3cccs3)c(O)cc2O)n[nH]c1=O
13	CCCCN(C)C(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2F)c(O)cc1O
14	Oc1cc(O)c(-c2ccnn2-c2ccccc2Cl)cc1CCc1ccccn1
15	CCCN(C)S(=O)(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2F)c(O)cc1O
16	CC(C)N(C)S(=O)(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2F)c(O)cc1O
17	Brc1cnc2[nH]cnc2c1C(=O)NC1c2ccccc2-c2c(-c3cnc4ccccc4c3)cccc21
19	O=C(NC1c2ccccc2-c2c(-c3nc4ccncc4[nH]3)cccc21)c1ccnc2[nH]ccc12
20	Cc1nn(-c2ccc(C(N)=O)c(N[C@H]3CC[C@H](O)CC3)c2)c2cccc(-c3cnc4ccccc4c3)c12
21	Cc1cn(-c2ccc(C(N)=O)c(N[C@H]3CC[C@H](O)CC3)c2)c2c1C(=O)CC(C)(C)C2
22	Cc1cn(-c2ccc(C(N)=O)c(NC3CCC(=O)CC3)c2)c2c1C(=O)CC(C)(C)C2
23	CC(C)N(C)S(=O)(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2Cl)c(O)cc1O
24	CCCN(C)C(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2C)c(O)cc1O
25	Cc1ccccc1-n1c(-c2cc(C(=O)N(C)Cc3ccccc3)c(O)cc2O)n[nH]c1=O

26	CCCCCCN(C)C(=O)c1cc(-c2ccnn2-c2ccccc2C)c(O)cc1O
27	Cc1cccc(CN(C)C(=O)c2cc(-c3n[nH]c(=O)n3-c3ccccc3C)c(O)cc2O)c1
28	Cc1ccccc1-n1c(-c2cc(C(=O)N(C)CC3CCCO3)c(O)cc2O)n[nH]c1=O
29	CCCCN(C)C(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2)c(O)cc1O
30	Cc1ccccc1-n1nccc1-c1cc(C(=O)N(C)Cc2ccco2)c(O)cc1O
31	Cc1ccccc1-n1c(-c2ccc(O)cc2O)n[nH]c1=O
32	O=c1[nH]nc(-c2ccc(O)cc2O)n1-c1ccccc1Cl
33	CCc1ccccc1-n1c(-c2ccc(O)cc2O)n[nH]c1=O
34	CCCN(C)C(=O)c1cc(-c2n[nH]c(=O)n2-c2ccccc2F)c(O)cc1O
35	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc5c(c4)OCO5)cc23)c1
36	C[NH+]1CCC(c2ccc(N(C)C(=O)c3cc4c(CCC(C)(C)C)n[nH]c4cc3O)cc2)CC1
37	CCCCN(C)C(=O)c1n[nH]c2cc(O)c(C(=O)N(C)c3ccc(N4CCOCC4)cc3)cc12
38	CN(Cc1ccc(Cl)cc1)C(=O)c2cc3c(Cc4ccccc4)n[nH]c3cc2O
39	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)Cc4ccccc4)cc23)c1
40	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)Cc4ccc(Cl)cc4)cc23)c1
41	Oc1cc2[nH]nc(Cc3ccccc3)c2cc1-c1ccnn1-c1ccccc1
42	Cc1ccc(N(C)C(=O)c2cc3c(Cc4cccc(C)c4)n[nH]c3cc2O)cc1
43	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCOCC5)cc4)cc23)c1
44	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccccc4)cc23)c1
45	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCCCC5)cc4)cc23)c1
46	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N(C)C)cc4)cc23)c1
47	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CC[NH2+]CC5)cc4)cc23)c1
48	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCN(C)CC5)cc4)cc23)c1
49	CO[C@H]1CCN(C(=O)c2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCOCC5)cc4)cc23)C1
50	CO[C@H]1CCCN(C(=O)c2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCOCC5)cc4)cc23)C1
51	CN(C(=O)c1cc2c(C(=O)N3CCCC3)n[nH]c2cc1O)c1ccc(N2CCOCC2)cc1
52	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(N5CCOCC5=O)cc4)cc23)c1

53	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4ccc(F)cc4)cc23)c1
54	COc1cccc(N(C)C(=O)c2cc3c(Cc4cccc(C)c4)n[nH]c3cc2O)c1
55	Cc1cccc(Cc2n[nH]c3cc(O)c(C(=O)N(C)c4cccc(C)c4)cc23)c1
56	CN(C(=O)c1cc2c(cc1O)[nH]nc2C(=O)N1CCOCC1)c1ccc(N2CCOCC2)cc1
57	CN(C(=O)c1cc2c(C(=O)N3CCCCC3)n[nH]c2cc1O)c1ccc(N2CCOCC2)cc1
58	Nc1nc(C(=O)N2Cc3ccc(O)cc3C2)c2ccccc2n1
59	Nc1nc(C(=O)N2Cc3ccccc3C2)c2cc(O)ccc2n1
60	C[NH+]1CCN(S(=O)(=O)c2ccccc2-c2ccc3nc(N)nc(C(=O)N4Cc5ccccc5C4)c3c2)CC1
61	Cc1ccc2nc(N)nc(C(=O)N3Cc4ccccc4C3)c2c1
62	CNCc1ccccc1-c1ccc2nc(N)nc(C(=O)N3Cc4ccccc4C3)c2c1
63	Nc1nc(C(=O)N2Cc3ccccc3C2)c2cc(-c3cc(F)c(F)cc3CCc3nnn[nH]3)ccc2n1
64	Nc1nc(C(=O)N2Cc3ccccc3C2)c2ccccc2n1
65	Nc1nc(C(=O)N2Cc3ccccc3C2)c2cc(-c3ccccc3O)ccc2n1
66	COc1c(C)cnc(Cn2cc(C#CCC(C)(C)O)c3c(Cl)nc(N)nc32)c1C
67	Cc1cnc(Cn2ccc3c(Cl)nc(N)nc32)c(C)c1Cl
68	C#CCCCn1c(Cc2cc(OC)c(OC)c(OC)c2Cl)nc2c(N)nc(F)nc21
69	N#Cc1ccc(N2CCN(CCCc3c[nH]c4cc(O)c(C#N)cc34)CC2)cc1
70	Nc1cc(C(=O)NC2c3ccccc3-c3c(-c4nc5ccncc5[nH]4)cccc32)ccn1

Table S11. List of the SMILES strings for the HIV-1 protease inhibitors.

Compound Id	SMILES
B435	O[C@H]([C@@H](O)[C@@H](OCc1ccccc1)C(=O)NC2[C@H](O)Cc3ccccc23)[C@@H](OCc4ccccc4)C(=O)NCc5ccccc5
A047	CNC(=O)c1cccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5ccccc5)S2(=O)=O)c1
A023	OCCc1cccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5cccc(CCO)c5)S2(=O)=O)c1
A024	COC(=O)c1ccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5ccc(CO)cc5)S2(=O)=O)cc1
B429	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(cc1)c2ccccn2)[C@H](O)[C@@H](O)[C@@H](OCc3ccc(cc3)c4ccccn4)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B409	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(cc1)c2ccsc2)[C@H](O)[C@@H](O)[C@@H](OCc3ccc(cc3)c4ccsc4)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B268	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
A045	CNC(=O)c1cccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5cccc(c5)C(=O)NC)S2(=O)=O)c1
B425	O[C@H](C[C@@H](OCc1ccccc1)C(=O)NC2[C@H](O)Cc3ccccc23)[C@@H](OCc4ccccc4)C(=O)NC5[C@@H](O)Cc6ccccc56
A021	OCc1ccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5ccc(CO)cc5)S2(=O)=O)cc1
saquinavir	CC(C)(C)NC(=O)[C@@H]1C[C@@H]2CCCC[C@@H]2CN1C[C@@H](O)[C@H](Cc3ccccc3)NC(=O)[C@H](CC(=O)N)NC(=O)c4ccc5ccccc5n4
indinavir	CC(C)(C)NC(=O)[C@@H]1CN(Cc2cccnc2)CCN1C[C@@H](O)C[C@@H](Cc3ccccc3)C(=O)N[C@@H]4[C@H](O)Cc5ccccc45
ritonavir	CC(C)[C@H](NC(=O)N(C)Cc1csc(n1)C(C)C)C(=O)N[C@H](C[C@H](O)[C@H](Cc2ccccc2)NC(=O)OCc3cncs3)Cc4ccccc4
DMP323	OCc1ccc(CN2[C@H](Cc3ccccc3)[C@H](O)[C@@H](O)[C@@H](Cc3ccccc3)N(Cc3ccc(CO)cc3)C2=O)cc1
nelfinavir	Cc1c(O)cccc1C(=O)N[C@H](CSc1ccccc1)[C@H](O)CN1C[C@H]2CCCC[C@H]2C[C@H]1C(=O)NC(C)(C)C

B369	O[C@H]([C@@H](O)[C@@H](OCc1ccccc1)C(=O)NC2[C@@H](O)Cc3ccccc23)[C@@H](OCc4ccccc4)C(=O)NC5[C@H](O)Cc6ccccc56
B388	CC(C)C(N(C)C(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)NC3[C@@H](O)Cc4ccccc34)C(=O)O
A038	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(cc1)C2=C(O)C(=O)CCC2)[C@H](O)[C@@H](O)[C@@H](OCc3ccc(cc3)C4=C(O)C(=O)CCC4)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
A037	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(\C=C\C(=O)OC)cc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccc(\C=C\C(=O)OC)cc2)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B440	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(cc1)c2nccs2)[C@H](O)[C@@H](O)[C@@H](OCc3ccc(cc3)c4nccs4)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B439	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(CCc2ccccc2)cc1)[C@H](O)[C@@H](O)[C@@H](OCc3ccc(CCc4ccccc4)cc3)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B408	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(Br)cc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccc(Br)cc2)C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
B412	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccc(cc1)c2cccc(c2)[N+](=O)[O-])[C@H](O)[C@@H](O)[C@@H](OCc3ccc(cc3)c4cccc(c4)[N+](=O)[O-])C(=O)N[C@@H](C(C)C)C(=O)NC)C(C)C
U75875	CC[C@@H](C)[C@H](NC(=O)[C@H](C(C)C)[C@@H](O)[C@H](O)[C@H](CC1CCCCC1)NC(=O)[C@H](Cc2c[nH+]c[nH]2)NC(=O)COc3cccc4ccccc34)C(=O)NCc5ccccn5
A008	OCc1ccc(CN2C(COc3ccccc3)[C@H](O)[C@@H](O)C(COc4ccccc4)N(Cc5ccc(CO)cc5)C2=O)cc1
B277	CCCO[C@H]([C@H](O)[C@@H](O)[C@@H](OCCC)C(=O)N[C@@H](C(C)C)C(=O)NC)C(=O)N[C@@H](C(C)C)C(=O)NC
A030	C\C(=N/O)\c1cccc(CN2[C@H](COc3ccccc3)[C@H](O)[C@@H](O)[C@@H](COc4ccccc4)N(Cc5cccc(c5)\C(=N\O)\C)S2(=O)=O)c1
A015	CN(C(Cc1ccccc1)C(=O)O)C(=O)[C@H](OCc2ccccc2)[C@H](O)[C@@H](O)[C@@H](OCc3ccccc3)C(=O)N(C)C(Cc4ccccc4)C(=O)O
A016	CN(C(Cc1ccc(O)cc1)C(=O)O)C(=O)[C@H](OCc2ccccc2)[C@H](O)[C@@H](O)[C@@H](OCc3ccccc3)C(=O)N(C)C(Cc4ccc(O)cc4)C(=O)O
A017	CC(O)C(N(C)C(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)N(C)C(C(C)O)C(=O)O)C(=O)O

B322	CCC(C)C(N(C)C(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)N(C)C(C(C)CC)C(=O)O)C(=O)O
B365	CC(C)C(N(C)C(=O)[C@@H](C[C@@H](O)[C@@H](OCc1ccccc1)C(=O)N(C)C(C(C)C)C(=O)O)OCc2ccccc2)C(=O)O
B347	CC(C)C(N(C)C(=O)[C@H](OCc1ccccc1)[C@@H](O)[C@H](O)[C@@H](OCc2ccccc2)C(=O)N(C)C(C(C)C)C(=O)O)C(=O)O
A018	CSCCC(N(C)C(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)N(C)C(CCSC)C(=O)O)C(=O)O
B249	COC(=O)C(NC(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)NC(C(C)C)C(=O)OC)C(C)C
B376	CNC(=O)[C@@H](NC(=O)[C@H](OCc1ccccc1)[C@H](O)[C@@H](O)[C@@H](OCc2ccccc2)C(=O)N[C@H](C(=O)NC)c3ccccc3)c4ccccc4

Formulas for statistical measures of correlation

1. Regression coefficient (R^2):

$$R^2 = \frac{[\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \langle\hat{y}\rangle)]^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \langle\hat{y}\rangle)^2}$$

Where,

$\bar{y}$ is the average value of the experimental activities ($y_1,\ y_2,..\ y_N$)

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$\langle\hat{y}\rangle$ is the average value of the calculated activities ($\hat{y}_1, \hat{y}_2, .. \hat{y}_N$)

$$\langle\hat{y}\rangle = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i$$

2. Cross-validated correlation coefficient (Q^2):

$$Q^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

3. Average absolute error (AAE):

$$AAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$$

4. Root mean squared error (RME)

$$RME = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$$

References:

1. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
2. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Wu, X; Kollman, P. A. *AMBER 14*; University of California: San Francisco, 2014.
3. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
4. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.
5. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
6. Gil-Redondo, R.; Klett, J.; Gago, F.; Morreale, A. gCOMBINE: A Graphical User Interface to Perform Structure-Based Comparative Binding Energy (COMBINE) Analysis on a Set of Ligand-Receptor Complexes. *Proteins: Struct., Funct., Bioinform.* **2010**, *78*, 162−172.
7. Kokh, D. B.; Amaral, M.; Bomke, J.; Gradler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; et al. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 3859−3869.
8. Markgren, P.-O.; Schaal, W.; Hämäläinen, M.; Karlén, A.; Hallberg, A.; Samuelsson, B.; Danielson, U. H. Relationships between Structure and Interaction Kinetics for HIV-1 Protease Inhibitors. *J. Med. Chem.* **2002**, *45*, 5430−5439.

Data files for performing COMBINE analysis

Data files for HSP90 COMBINE analysis:

(ml8b00397_si_002.zip, size:21.1 MB)

The following 5 files in this zipped folder were used to generate the COMBINE analysis model for the off rate and K_D of HSP90 inhibitors:

1. hsp90_energy_minimized_pdb_files.zip
   This zipped file contains 70 energy-minimized structures of N-HSP90-inhibitor complexes in PDB format.

2. hsp90_amber_topology_files.zip
   This zipped file contains 70 topology (.top) files for N-HSP90-inhibitor complexes generated by the LEaP program of Amber 14 using ff14SB and GAFF force fields for protein and inhibitors, respectively. These topology files are required for performing the PLS analysis with the gCOMBINE program. The gCOMBINE program is available under a free academic license and can be used for non-profit or non-commercial use. The source files for gCOMBINE can be downloaded from https://ub.cbm.uam.es/drug_design/combine.php

3. hsp90_amber_coordinate_files.zip
   This zipped file contains 70 coordinate (.crd) files for N-HSP90-inhibitor complexes generated by the LEaP program of Amber 14 and they are also required by the gCOMBINE program.

4. hsp90_gCOMBINE_output_models.zip
   This zipped file contains output of 3 different COMBINE models generated in this study using gCOMBINE program. These files can be directly loaded into gCOMBINE program (using the Load model option) to visualize the results of PLS analysis.
   - hsp90_output_model_koff
     COMBINE analysis model generated for k_{off} rates of HSP90 inhibitors.
   - hsp90_output_model_KD
     COMBINE analysis model generated for K_D (binding affinity) of HSP90 inhibitors.
   - hsp90_output_model_KD_resorcinol
     COMBINE analysis model generated for K_D (binding affinity) of resorcinol class of HSP90 inhibitors.

5. Nomenclature_files.xlsx
   The nomenclature of PDB, coordinate and topology files for protein-inhibitor complexes provided in the zipped files above, is different from the nomenclature of inhibitors used in the manuscript. This Excel spreadsheet provides the mapping between inhibitor names mentioned in the manuscript and their corresponding file names in the zipped folders.

Data files for HIV COMBINE analysis:

(ml8b00397_si_003.zip, size:10.5 MB)

The following 5 files in this zipped folder were used to generate the COMBINE analysis model for the off rate and K_D of HIV-1 protease inhibitors:

1. hiv_energy_minimized_pdb_files.zip
   This zipped file contains 36 energy-minimized structures of HIV-1 protease-inhibitor complexes in PDB format.

2. hiv_amber_topology_files.zip
   This zipped file contains 36 topology (.top) files for HIV-1 protease-inhibitor complexes generated by the LEaP program of Amber 14 using ff14SB and GAFF force fields for protein and inhibitors, respectively.

3. hiv_amber_coordinate_files.zip
   This zipped file contains 36 coordinate (.crd) files for HIV-1 protease-inhibitor complexes generated by the LEaP program of Amber 14.

4. hiv_gCOMBINE_output_models.zip
   This zipped file contains output of 2 different COMBINE models generated in this study using gCOMBINE program. These files can be directly loaded into gCOMBINE program (using the Load model option) to visualize the results of PLS analysis.
   - hiv_output_model_koff
     COMBINE analysis model generated for k_{off} rates of HIV-1 protease inhibitors.
   - hiv_output_model_KD
     COMBINE analysis model generated for K_D (binding affinity) of HIV-1 protease inhibitors.

5. Nomenclature_files.xlsx
   The nomenclature of PDB, coordinate and topology files for protein-inhibitor complexes provided in the zipped files above, is different from the nomenclature of inhibitors used in the manuscript. This Excel spreadsheet provides the mapping between inhibitor names mentioned in the manuscript and their corresponding file names in the zipped folders.