**Supplementary Information**

**New Strategy for Identifying Potential Natural HIV-1 Non-nucleoside Reverse Transcriptase Inhibitors Against Drug-Resistance: an *in silico* study**

Yanjing Wang1,2†, Xiangeng Wang1,2†, Yi Xiong1, Aman Chandra Kaushik1, Junaid Muhammad1, Abbas Khan1, Hao Dai3\*, Dong-Qing Wei1\*

**Affiliation:**

1State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, and Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China

2Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nanshan District, Shenzhen, Guangdong, 518055, China

3Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

**†****These authors contributed equally to this work**

Yanjing Wang: wangyanjing@sjtu.edu.cn

Xiangeng Wang: wangxiangeng@sjtu.edu.cn

Yi Xiong: xiongyi@sjtu.edu.cn

Aman Chandra Kaushik: amanbioinfo@sjtu.edu.cn

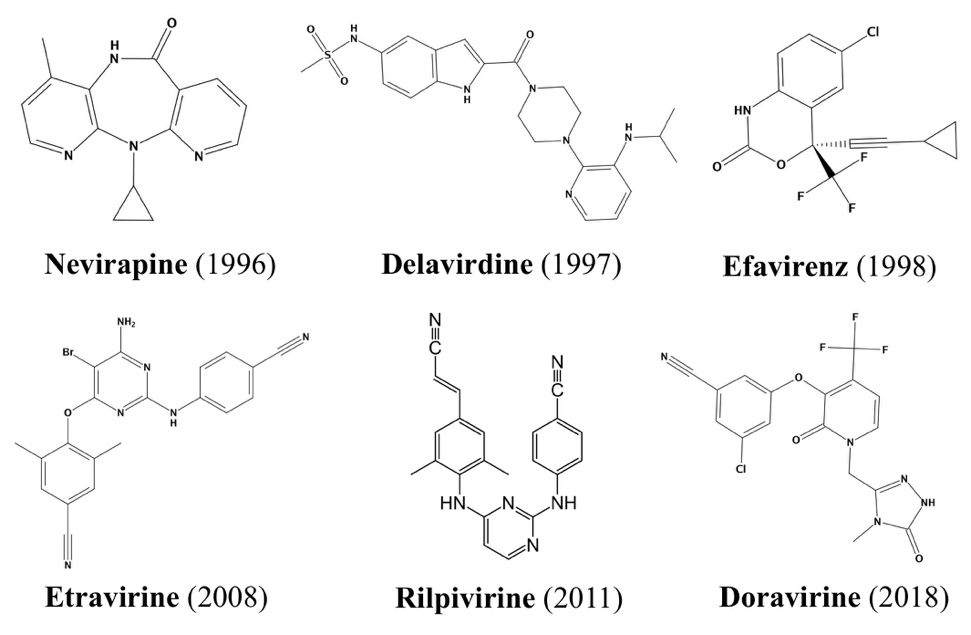
Muhammad Junaid: juni\_sjtu@sjtu.edu.cn

Abbas Khan: biotechuos4@gmail.com

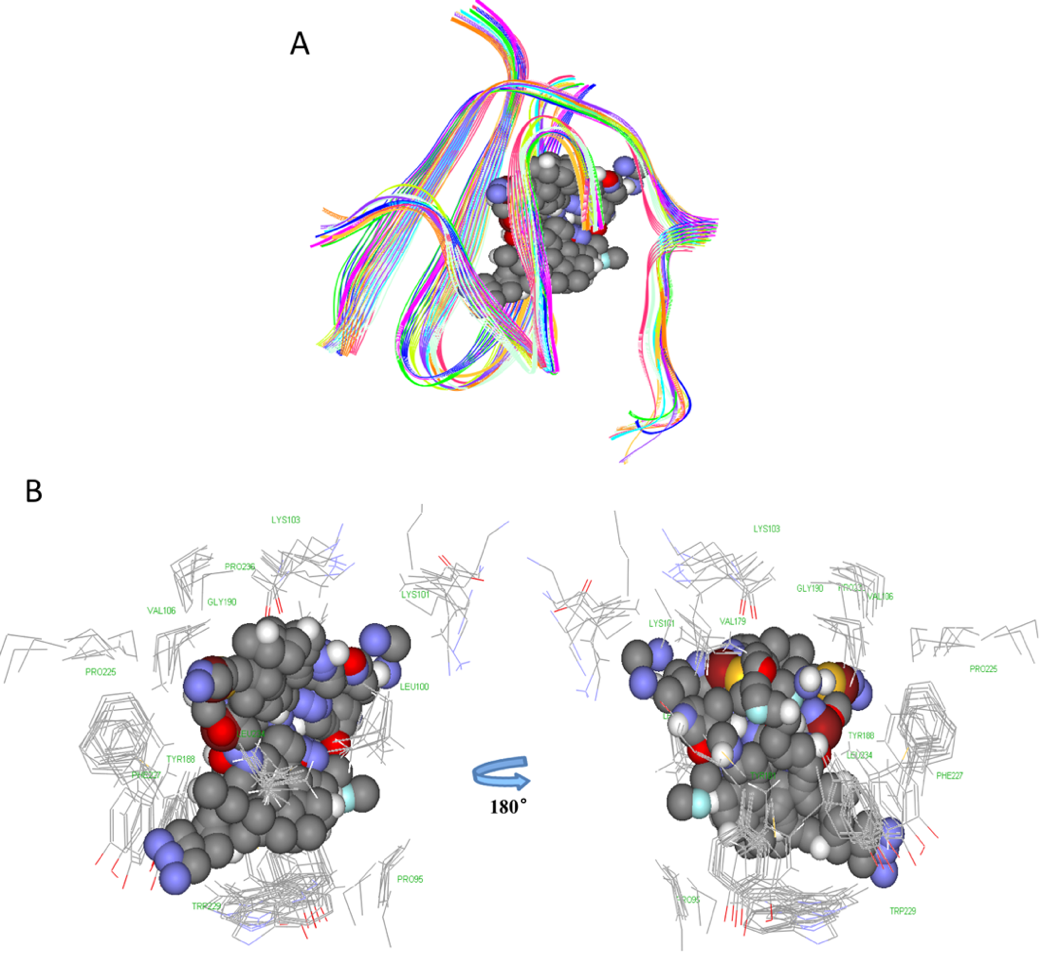
**\*Corresponding author:**

Professor: Dong-Qing Wei, [dqwei@sjtu.edu.cn](mailto:dqwei@sjtu.edu.cn)

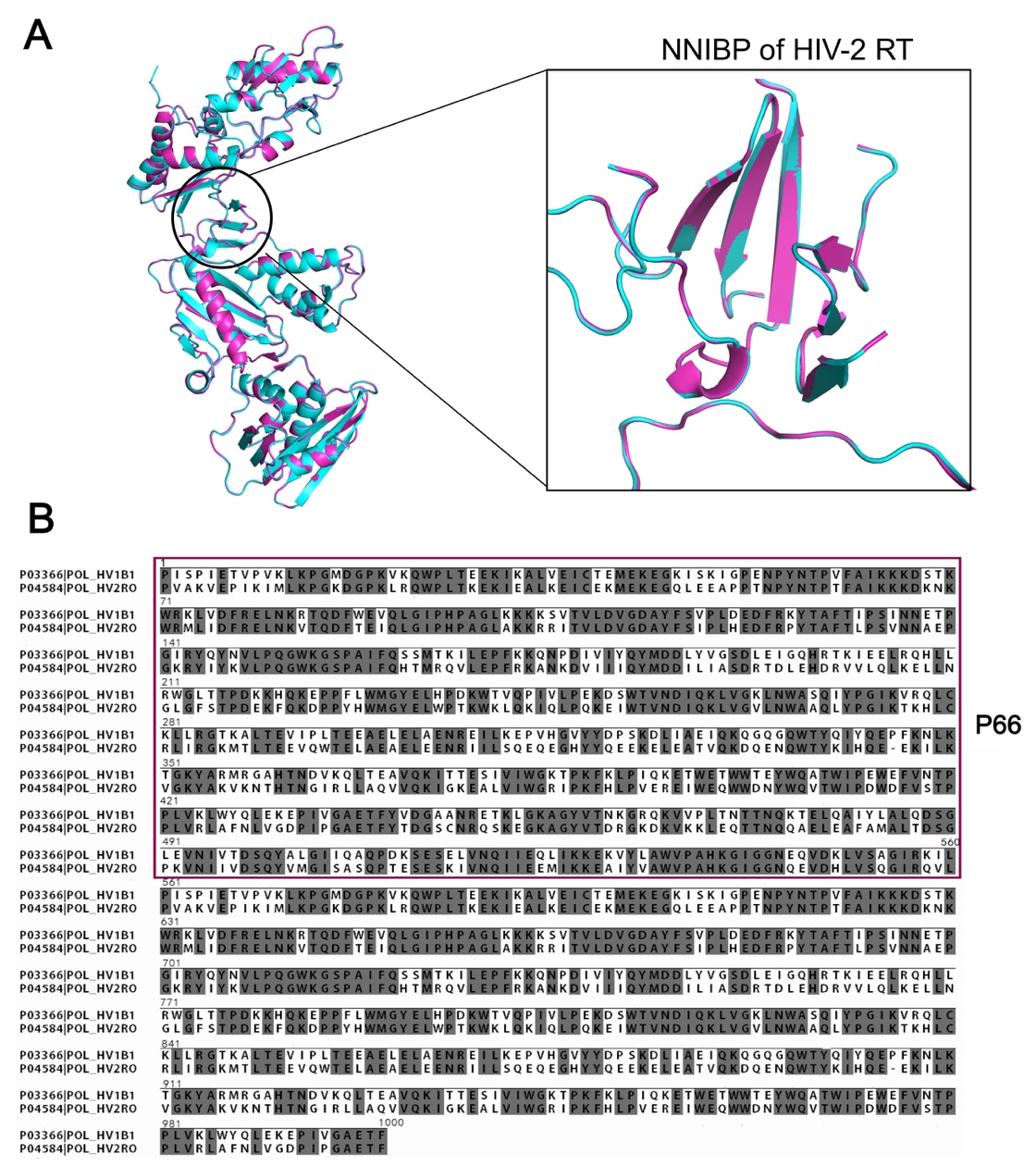
Doctor: Hao Dai, [wys8c764@126.com](mailto:wys8c764@126.com)

****

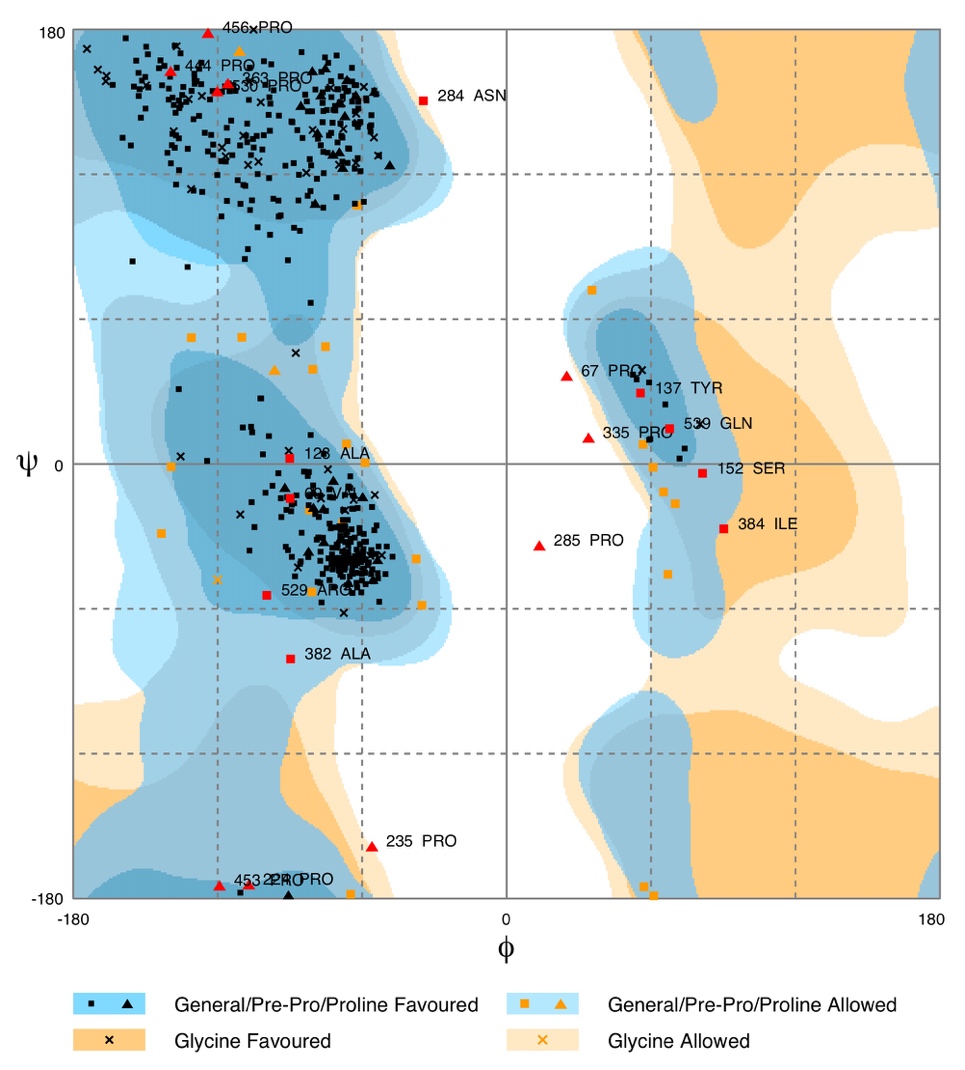
**Figure S1.** Chemical structure of NNRTIs approved by the FDA for HIV-1 treatment



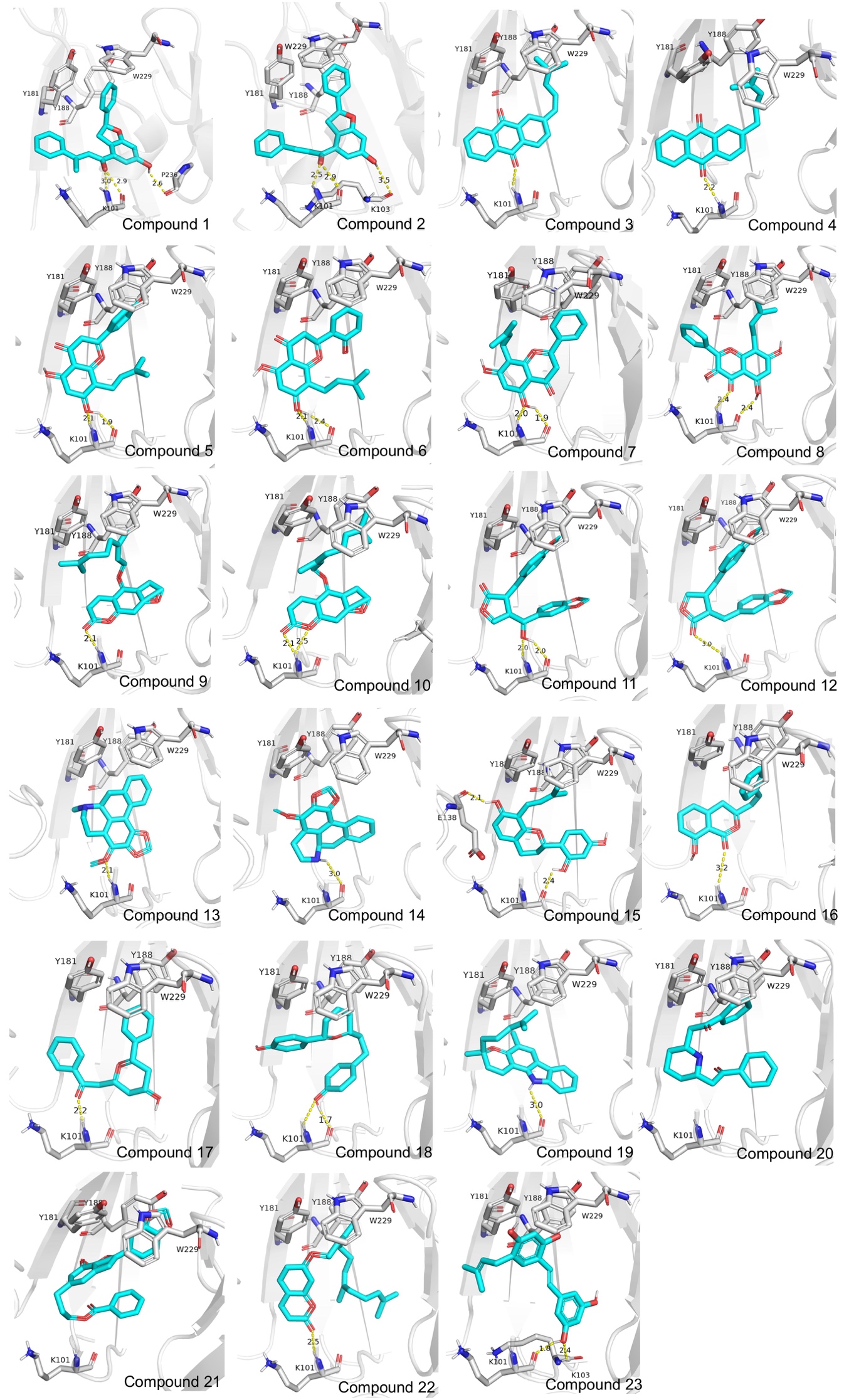
**Figure S2.** Ligand binding landscape(A) Structure alignment in NNIBP of 10 RT crystallographic structures. Grey balls are the area enveloped by the ligands, which represents the NNRTI binding site. (B) Spatial distribution of 14 amino acids near the ligand-binding site.



**Figure S3.** Sequential and Structural comparison between HIV-2 RT and HIV-1 RT. (A) The structure alignment between HIV-2 RT (red) from homology modelling and the template (HIV-1 RT, blue). Only NNRTI was shown. (B) Sequence alignment of template HIV-1 RT sequence with target HIV-2 RT sequence. RT sequences refer to that of HIV-1(PDB:4G1Q, SwissProt: P03366) and HIV-2 (SwissProt: P04584). P66 subunit (1 to 560 residues) has been highlighted in a red box. Sequence identity and similarity were 60.7% and 79.6%, respectively.



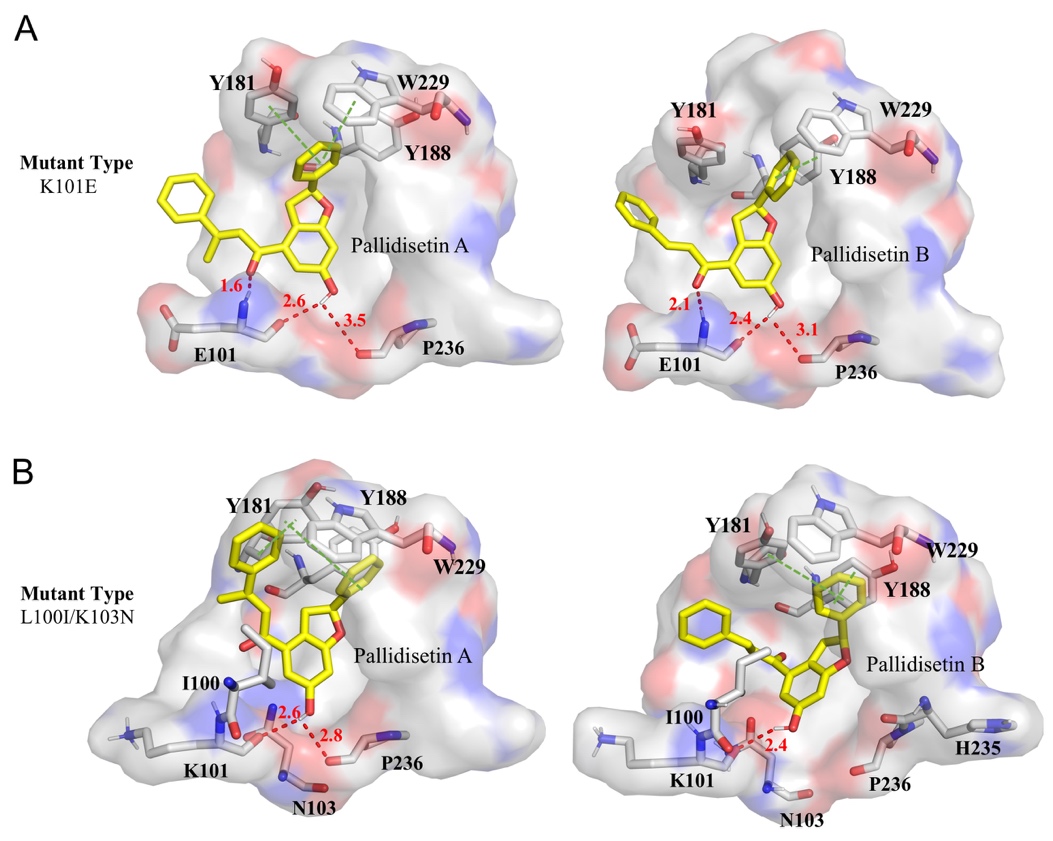
**Figure S4.** Ramachandran plot for the predicted HIV-2 RT model using RAMPAGE server. The model conformation has high stability and rationality (91.4 % residues in the allowed region).

****

**Figure S5.** Representative docking poses of 23 hits.



**Figure S6.** RMSD plot of Backbone atoms RMSD (top), NNIBP (middle) and ligands (bottom) for controls and top 2 candidates in complex with mutant HIV-1 RT protein during 30ns simulations. (A) L100I/K103N variant of HIV-1 RT complexes. (B) K101E variant of HIV-1 RT complexes.



**Figure S7**. The representative binding mode of top 2 hits in NNIBP of mutated HIV-1 RT.

**Table S1.** 10 crystal structures for structure alignment

|  |  |  |  |
| --- | --- | --- | --- |
| PDB ID | Ligand name | Resolution (Å) | Mutation  (chain A) |
| 4G1Q | RPV | 1.5 | - |
| 4I2Q | 51BT | 2.7 | K103N, Y181C |
| 2YKM | NSC700591 | 2.9 | N57S, F227C |
| 2ZE2 | RPV | 2.9 | L100I, K103N, |
| 1EET | MSC204 | 2.7 | P140S, E478Q |
| 1HNI | R95845 | 2.8 | K172R, C280S |
| 1JKH | EFV | 2.5 | Y181C |
| 3DMJ | GWE | 2.6 | V106A, Y181C |
| 2HND | NVP | 2.5 | K101E |
| 2YNF | WHU | 2.3 | Y188L |

**Table S2.** The RMSD values of Cα in the main chain of 10 structures (The unit is Å)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2HND | 0.97 |  | | | | | | | |
| 4I2Q | 1.662 | 1.469 |  | | | | | | |
| 2YKM | 1.224 | 1.042 | 1.729 |  | | | | | |
| 2ZE2 | 1.484 | 1.371 | 0.874 | 1.515 |  | | | | |
| 3DMJ | 1.086 | 1.227 | 1.823 | 1.679 | 1.668 |  | | | |
| 1HNI | 1.413 | 1.401 | 1.846 | 1.03 | 1.603 | 1.826 |  | | |
| 1JKH | 1.448 | 1.345 | 1.881 | 1.08 | 1.696 | 1.751 | 1.182 |  | |
| 1EET | 1.051 | 1.172 | 1.54 | 1.196 | 1.3 | 1.53 | 1.358 | 1.485 |  |
| 4G1Q | 1.553 | 1.271 | 0.742 | 1.569 | 0.698 | 1.702 | 1.658 | 1.735 | 1.313 |
|  | 2YNF | 2HND | 4I2Q | 2YKM | 2ZE2 | 3DMJ | 1HNI | 1JKH | 1EET |

**Table S3.** The 13 drug-resistant mutants located in the NNIBP, and together with the equivalent residues in HIV-2 RT

|  |  |  |
| --- | --- | --- |
| HIV-1 | Drug-resistance mutation | HIV-2 |
| Leu100 | L100I | Leu |
| Lys101 | K101P/E | Ala |
| Lys103 | K103N/S | Lys |
| Val106 | V106A/M/I | Ile |
| Val179 | V179F/D/L/T | Ile |
| Tyr181 | Y181I/C/V | Ile |
| Tyr188 | Y188L/C/H | Leu |
| Gly190 | G190A/S | Ala |
| Pro225 | P225H | Pro |
| Phe227 | F227L/C | Tyr |
| Met230 | M230/L/I | Met |
| Leu234 | L234I | Leu |
| Lys238 | K238T | Lys |

**Table S4.** Docking score of five approved NNRTIs (kcal/mol).

|  |  |  |  |
| --- | --- | --- | --- |
| NNRTI (Year) | HIV-1 | HM | HIV-2 |
| Nevirapine (1996) | -9.5 | -7.3 | -8.9 |
| Delavirdine (1997) | -9.1 | -5.9 | -7.7 |
| Efavirenz (1998) | -10.9 | -8 | -10.2 |
| Etravirine (2008) | -12.7 | -7.2 | -10 |
| Rilpivirine (2011) | -13 | -8 | -9.2 |

**Table S5.** Calculated physicochemical properties of NNRTI Drugs

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | Physicochemical Properties | | | | | |
| Name | PDB | MWb | VSAc | | Vold | cLogPe | HBAf | HBDg | NrotBh |
| Nevirapinea | 3HVT | 266.30 | 278.92 | | 569.13 | 2.05 | 3 | 1 | 1 |
| Delavirdinea | 1KLM | 456.56 | 470.77 | | 594.29 | 1.58 | 4 | 3 | 7 |
| Efavirenza | 1KF9 | 315.67 | 284.05 | | 338.22 | 3.68 | 2 | 1 | 2 |
| Etravirinea | 3MEC | 435.28 | 406.50 | | 510.99 | 6.11 | 5 | 2 | 5 |
| Rilpivirinea | 4G1Q | 366.42 | 409.19 | | 515.92 | 5.64 | 4 | 2 | 7 |
| MSC204 | 1EET | 422.25 | 384.01 | | 467.46 | 2.79 | 4 | 3 | 5 |
| R95845 | 1HNI | 440.13 | 354.21 | | 457.28 | 4.58 | 2 | 2 | 5 |
| NSC700591 | 2YKM | 293.30 | 276.46 | | 333.86 | 3.63 | 3 | 0 | 3 |
| WHU | 2YNF | 499.13 | 422.26 | | 504.20 | 4.58 | 4 | 4 | 6 |
| GWE | 3DMJ | 544.91 | 482.28 | | 592.11 | 5.01 | 5 | 2 | 7 |
| 51BT | 4I2Q | 438.52 | 443.79 | | 404.75 | 3.88 | 6 | 4 | 6 |

aThese are approved NNRTIs.

bMW stands for molecular weight

cVSA represents Area of van der Waals surface

dVol is van der Waals volume

ecLogP means Octanol–water partition coefficient

fHBA stands for Number of H-bond acceptors

gHBD represents Number of H-bond donors

hNrotB means Number of rotatable bonds

**Table S6.** Calculated Physicochemical Properties of 23 hits

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Compound |  | Physicochemical Properties | | | | | | |
| No. | Formula | MW | VSA | Vol | LogP | HBA | HBD | NrotB |
| 1 | C24H20O3 | 356.42 | 365.86 | 505.96 | 5.54 | 3 | 1 | 4 |
| 2 | C23H18O3 | 342.39 | 354.81 | 481.53 | 5.04 | 3 | 1 | 4 |
| 3 | C20H16O2 | 288.35 | 307.07 | 415.94 | 5.93 | 2 | 0 | 2 |
| 4 | C20H20O2 | 290.36 | 319.35 | 421.84 | 6.14 | 2 | 0 | 3 |
| 5 | C20H18O5 | 338.36 | 340.84 | 450.98 | 4.41 | 4 | 3 | 3 |
| 6 | C20H20O5 | 340.38 | 347.44 | 456.89 | 3.70 | 5 | 3 | 4 |
| 7 | C20H20O4 | 324.38 | 342.92 | 448.36 | 4.17 | 4 | 2 | 4 |
| 8 | C20H18O6 | 338.36 | 346.36 | 450.98 | 4.51 | 4 | 3 | 3 |
| 9 | C21H22O4 | 336.39 | 371.52 | 471.17 | 5.97 | 2 | 0 | 6 |
| 10 | C21H20O4 | 338.40 | 375.89 | 477.07 | 6.34 | 2 | 0 | 6 |
| 11 | C20H18O7 | 370.36 | 348.52 | 460.60 | 1.77 | 6 | 1 | 4 |
| 12 | C20H18O6 | 354.36 | 351.69 | 452.07 | 2.33 | 5 | 0 | 4 |
| 13 | C19H19NO3 | 309.36 | 303.57 | 423.05 | 2.80 | 4 | 0 | 1 |
| 14 | C18H17NO3 | 295.34 | 292.72 | 397.00 | 2.52 | 4 | 1 | 1 |
| 15 | C20H22O4 | 326.39 | 353.00 | 455.71 | 3.74 | 4 | 3 | 4 |
| 16 | C17H14O3 | 266.30 | 275.75 | 366.55 | 3.59 | 2 | 1 | 2 |
| 17 | C19H20O3 | 296.37 | 327.88 | 421.31 | 3.19 | 3 | 1 | 6 |
| 18 | C19H20O3 | 296.37 | 338.68 | 422.76 | 3.15 | 3 | 2 | 5 |
| 19 | C23H25NO | 331.46 | 376.60 | 499.92 | 5.72 | 1 | 1 | 3 |
| 20 | C21H23NO2 | 321.42 | 375.56 | 468.46 | 3.84 | 3 | 1 | 8 |
| 21 | C26H22O6 | 430.46 | 433.11 | 575.04 | 5.35 | 4 | 0 | 7 |
| 22 | C24H30O3 | 336.50 | 440.40 | 553.59 | 7.01 | 2 | 0 | 9 |
| 23 | C19H20O3 | 296.37 | 342.72 | 428.61 | 4.51 | 3 | 3 | 4 |

**Table S7.** The predicted HIV-RT IC50 value of 23 hits

|  |  |  |
| --- | --- | --- |
| No | Name | IC50(m) |
| 1 | Pallidisetin A | 51.72 |
| 2 | Pallidisetin B | 35.28 |
| 3 | 2-(4-methylpenta-1,3-dienyl) anthraquinone | 43.51 |
| 4 | 2-(4-methylpent-3-enyl) anthraquinone | 66.8 |
| 5 | Licoflavone C | 58.35 |
| 6 | Kosamol S | 67.67 |
| 7 | Glabranin | 44.28 |
| 8 | Glepidotin A | 56.92 |
| 9 | Anhydronotoptol | 31.81 |
| 10 | Bergamottin | 31.81 |
| 11 | Parabenzlactone | 46.18 |
| 12 | Hinokinin | 23.26 |
| 13 | (-)-Stephalagine | 35.69 |
| 14 | Norstephalagine | 36.64 |
| 15 | Morusyunnansin F | 49.75 |
| 16 | Typharin | 48.26 |
| 17 | Diospongin B | 31.74 |
| 18 | (3S,7S)-5,6-dehydro-4"-de-O- methylcentrolobine | 41.01 |
| 19 | Mahanimbine | 16.34 |
| 20 | Isolobelanine | 41 |
| 21 | 3-[2-(1,3-benzodioxol-5-yl)-7-methoxy-1- benzofuran-5-yl]propylbenzoate | 36.03 |
| 22 | Umbelliprenin | 11.77 |
| 23 | 5-(γ,γ-dimethylallyl)-oxyresveratrol | 7.58 |

**Table S8.** Summary of MD simulation conditions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| System | No. | Mutation | Ligand | Ion | *NWater* | Time(ns) |
| WT\_4G1Q | 1 | Wild type | RPV | 210 Na+/213 Cl- | 71691 | 30 |
| 2 | C1 | 210 Na+/213 Cl- | 71640 |
| 3 | C2 | 210 Na+/213 Cl- | 71694 |
| 4 | C3 | 210 Na+/213 Cl- | 71697 |
| MT\_2ZE2 | 5 | L100I  K103N | RPV | 213 Na+/214 Cl- | 74879 | 30 |
| 6 | C1 | 213 Na+/214 Cl- | 74880 |
| 7 | C2 | 213 Na+/214 Cl- | 74885 |
| MT\_2HND | 8 | K101E | NVP | 218 Na+/220 Cl- | 76432 | 30 |
| 9 | C1 | 218Na+/220 Cl- | 76410 |
| 10 | C2 | 218Na+/220 Cl- | 76411 |

**Table S9.** The structure of top 3 hits and control existing drugs.

|  |  |  |
| --- | --- | --- |
| Control | Top 3 hits | |
| RPV: Rilpivirine | C1: Pallidisetin A | C2: Pallidisetin B |
| 180px-Rilpivirine_svg |  |  |
| NVP : Nevirapine | C3: ((E)-2-(4-methylpenta-1,3-dienyl)anthraquinone | |
| hy-10570 |  | |

**Table S10.** Hydrogen bond frequency for Pallidisetin A (PA) and Pallidisetin B (PB)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| System | Ligand | H-bonda | Donor | Accepter | H-bond  occupancy |
| WT | PA | Distance1 | PA:H1 | LYS101:O | 83.43% |
| Distance2 | PA:H1 | PRO236:O | 12.00% |
| PB | Distance1 | PB:H1 | LYS101:O | 85.53% |
| Distance2 | LYS103:HN | PB:O2 | 86.80% |
| K101E | PA | Distance1 | PA:H1 | GLU101:O | 38.83% |
| Distance2 | PA:H1 | PRO236:O | 58.70% |
| PB | Distance1 | PA:H1 | GLU101:O | 46.90% |
| Distance2 | PA:H1 | PRO236:O | 52.13% |
| L100I/K103N | PA | Distance1 | PA:H1 | LYS101:O | 59.23% |
| Distance2 | PA:H1 | PRO236:O | 40.20% |
| Distance3 | ASN103:HD2 | PA:O2 | 23.73% |
| PB | Distance1 | PB:H1 | LYS101:O | 14.57% |
| Distance2 | PB:H1 | HIS235: | 28.33% |

aH-bond cutoff distance: 3.5 Å