HIV-1 TAR RNA QSAR test dataset (taken from Cai et al., 2022)

|  |  |  |
| --- | --- | --- |
| **HIV\_QSAR\_SMILES** | **Mol\_ID** | **pKd** |
| NC[C@@H]1O[C@H](O[C@@H]2[C@@H](CO)O[C@@H](O[C@@H]3[C@@H](O)  [C@H](N)C[C@H](N)[C@H]3O[C@H]3O[C@H](CO)[C@@H](O)[C@H](O) [C@H]3N)[C@@H]2O)[C@H](N)[C@@H](O)[C@@H]1O | HIV\_QSAR\_1 | 5.43651891460559 |
| CN[C@H]1[C@H](O)[C@@H](O)[C@H](CO)O[C@H]1O[C@H]1[C@H]  (O[C@H]2[C@H](O)[C@@H](O)[C@H](NC(N)=N)[C@@H](O) [C@@H]2NC(N)=N)O[C@@H](C)[C@]1(O)C=O | HIV\_QSAR\_2 | 4.95860731484177 |
| NC[C@H]1O[C@H](O[C@@H]2[C@@H](N)C[C@@H](N)[C@H](O[C@H]3O[C@H]  (CO)[C@@H](O)[C@H](N)[C@H]3O)[C@H]2O)[C@H](N)C[C@@H]1O | HIV\_QSAR\_3 | 5.44977164694491 |
| CNC(C)C1CCC(N)C(OC2C(N)CC(N)C(OC3OCC(C)(O)C(NC)C3O)C2O)O1 | HIV\_QSAR\_4 | 5.42021640338319 |
| NC[C@H]1O[C@H](O[C@@H]2[C@@H](N)C[C@@H](N)[C@H](O)[C@H]2O) [C@H](N)[C@@H](O)[C@@H]1O | HIV\_QSAR\_5 | 5.58737147945562 |
| NC[C@H]1O[C@H](O[C@@H]2[C@@H](N)C[C@@H](N)[C@H](O[C@H]3O[C@H]  (CO)[C@@H](O)[C@H](N)[C@H]3O)[C@H]2O)[C@H](O)[C@@H](O)[C@@H]1O | HIV\_QSAR\_6 | 4.91904899562052 |
| CN(C)C1=C(Cl)N=C(C(=O)NC(N)=[NH2+])C(N)=N1 | HIV\_QSAR\_7 | 3.48469985672782 |
| NCCNC1=C(Cl)N=C(C(=O)NC(N)=[NH2+])C(N)=N1 | HIV\_QSAR\_8 | 5.24511677174783 |
| CN(C)C1=C(N=C(C(=O)NC(N)=[NH2+])C(N)=N1)C1=CC=C(C=C1)C#N | HIV\_QSAR\_9 | 3.04667530361081 |
| CN1CCN(CCNC2=C(Cl)N=C(C(=O)NC(N)=[NH2+])C(N)=N2)CC1 | HIV\_QSAR\_10 | 4.43203309317685 |
| NC(=[NH2+])NC(=O)C1=NC(C2=CC=CC=C2)=C(NCCC2=NC3=C(N2)C=CC=C3)N=C 1N | HIV\_QSAR\_11 | 3.62598525970809 |
| NC(=[NH2+])NC(=O)C1=NC(C2=CN=CN=C2)=C(N=C1N)N1CCCC1 | HIV\_QSAR\_12 | 2.70524000796455 |
| CN1CCN(CCNC2=C(N=C(C(=O)NC(N)=[NH2+])C(N)=N2)C2=CC=CC=C2)CC1 | HIV\_QSAR\_13 | 4.63469925136201 |
| NC(=[NH2+])NC(=O)C1=NC(=C(NCCC2=CNC3=C2C=CC=C3)N=C1N)C1=CC=NC=C 1 | HIV\_QSAR\_14 | 4.48818345869691 |
| NC(=[NH2+])NC(=O)C1=NC(=C(NCCC2=CNC3=C2C=CC=C3)N=C1N)C1=CN=CN=C 1 | HIV\_QSAR\_15 | 4.48858459279123 |
| NC(=[NH2+])NC(=O)C1=NC(=C(NCCC2=CNC3=C2C=CC=C3)N=C1N)C1=CN=C(NC CCN2C=CN=C2)N=C1 | HIV\_QSAR\_16 | 4.57438427542307 |
| NC(=[NH2+])NC(=O)C1=NC(=C(NCCC2=CNC3=C2C=CC=C3)N=C1N)C1=CN=C(N= C1)N1CCCC1 | HIV\_QSAR\_17 | 4.90187566170576 |
| [I-].[I-].C[N+]1=C(SC2=C1C=CC=C2)\C=C1/C=CN(CCC[N+](C)(C)C)C2=C1C=CC=C2 | HIV\_QSAR\_18 | 6.17848647159523 |
| CN(C)CCCNC(=N)C1=CC(=CC=C1)C1=CC=C(O1)C1=CC(=CC=C1)C(=N)NCCCN(C) C | HIV\_QSAR\_19 | 6.94692155651658 |
| CN(C)CCCNC(=N)C1=CC=C(C=C1)C1=CC=C(O1)C1=CC=C(C=C1)C(=N)NCCCN(C) C | HIV\_QSAR\_20 | 7.10127481841051 |
| NC(=N)C1=CC=C(C=C1)C1=CC=C(O1)C1=CC=C(C=C1)C(N)=N | HIV\_QSAR\_21 | 5.59911678445164 |
| [Br-].CC[N+]1=C(C2C=C(N)C=CC2C2=CC=C(N)C=C12)C1=CC=CC=C1 | HIV\_QSAR\_22 | 5.95860731484177 |
| [H][N+]1=C(NC2=C1C=C(C=C2)N1CC[N+]([H]) (C)CC1)C1=CC2=C(NC(=[N+]2[H])C2=CC=C(O)C=C2)C=C1 | HIV\_QSAR\_23 | 5.141102042768 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC=C3N=CC=NC3=C2)C=C1)C(=O)NC( N)=[NH2+] | HIV\_QSAR\_24 | 4.34345143906787 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC=C(C=C2)N2C=CN=C2)C=C1)C(=O)N C(N)=[NH2+] | HIV\_QSAR\_25 | 4.05827991762937 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC3=C(NC=N3)C=C2)C=C1)C(=O)NC(N)  =[NH2+] | HIV\_QSAR\_26 | 3.77408364750598 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC=C(C=C2)N2CCOCC2)C=C1)C(=O)NC (N)=[NH2+] | HIV\_QSAR\_27 | 4.48731560378284 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC=NN2)C=C1)C(=O)NC(N)=[NH2+] | HIV\_QSAR\_28 | 3.125309492778 |
| CN(C)C1=NC(N)=C(N=C1C1=CC=C(CNC2=CC=CC=C2)C=C1)C(=O)NC(N)=[NH2+] | HIV\_QSAR\_29 | 3.2134792693533 |
| CC1=CC=C(C=C1)S([O-])(=O)=O.CN1\C(SC2=CC=CC=C12)=C\C1=CC=[N+] (C)C2=CC=CC=C12 | HIV\_QSAR\_30 | 5.06348625752111 |
| N=C(NCCC1=CC=CN=C1)C2=CC(C3=CC=C(C4=CC(C(NCCC5=CC=CN=C5)=N)=CC  =C4)O3)=CC=C2 | HIV\_QSAR\_31 | 5.41987367458842 |
| N=C(NCCN1CCNCC1)C2=CC(C3=CC=C(C4=CC(C(NCCN5CCNCC5)=N)=CC=C4)O3) | HIV\_QSAR\_32 | 5.93930215964639 |

|  |  |  |
| --- | --- | --- |
| **HIV\_QSAR\_SMILES** | **Mol\_ID** | **pKd** |
| =CC=C2 |  |  |
| N=C(NCCN1CCN(C)CC1)C2=CC(C3=CC=C(C4=CC(C(NCCN5CCN(C)CC5)=N)=CC= C4)O3)=CC=C2 | HIV\_QSAR\_33 | 6.24488773360493 |
| N=C(NCC1=CC=CC=C1Cl)C(C=C2)=CC=C2C3=CC=C(C4=CC=C(C(NCC5=CC=CC=C 5Cl)=N)C=C4)O3 | HIV\_QSAR\_34 | 4.24123945609002 |
| [NH2+]=C(N[C@H]1CCC[NH+] (CC2=CC=CC=C2)C1)C1=CC=C(C=C1)C1=CC=C(O1)C1=CC=C(C=C1)C(=[NH2+])N[ C@H]1CCC[NH+](CC2=CC=CC=C2)C1 | HIV\_QSAR\_35 | 5.95467702121334 |
| CN(C)C1=CC2=NC3=C(C=CC(=C3)N(C)C)C=C2C=C1 | HIV\_QSAR\_36 | 5.24641694110709 |
| N=C(NCCC1=CC=CC=C1)C2=CC(C3=CC=C(C4=CC(C(NCCC5=CC=CC=C5)=N)=CC  =C4)O3)=CC=C2 | HIV\_QSAR\_37 | 5.42021640338319 |
| N=C(NCCC1=CC=CC=C1)C(C=C2)=CC=C2C3=CC=C(C4=CC=C(C(NCCC5=CC=CC= C5)=N)C=C4)O3 | HIV\_QSAR\_38 | 5.41680122603138 |
| N=C(NCC1=CC=C(Cl)C=C1)C(C=C2)=CC=C2C3=CC=C(C4=CC=C(C(NCC5=CC=C(Cl  )C=C5)=N)C=C4)O3 | HIV\_QSAR\_39 | 5.36451625318509 |
| N=C(NC1CCN(CC2=CC=CC=C2)CC1)C(C=C3)=CC=C3C4=CC=C(C5=CC=C(C(NC6C CN(CC7=CC=CC=C7)CC6)=N)C=C5)O4 | HIV\_QSAR\_40 | 6.20830935097988 |
| [NH2+]=C(NCCC1CC[NH+] (CC2=CC=CC=C2)CC1)C1=CC=C(C=C1)C1=CC=C(O1)C1=CC=C(C=C1)C(=[NH2+])N CCC1CC[NH+](CC2=CC=CC=C2)CC1 | HIV\_QSAR\_41 | 6.1681302257195 |
| N=C(NCCC1=CC=CN=C1)C1=CC(=CC=C1)N=NNC1=CC(=CC=C1)C(=N)NCCC1=CC  =CN=C1 | HIV\_QSAR\_42 | 5.65757731917779 |
| N=C(NC1CCN(CC2=CC=CC=C2)CC1)C1=CC=C(N\N=N\ C2=CC=C(C=C2)C(=N)NC2CCN(CC3=CC=CC=C3)CC2)C=C1 | HIV\_QSAR\_43 | 6.08354605145007 |
| N=C(NCCC1CCN(CC2=CC=CC=C2)CC1)C1=CC=C(N\N=N\ C2=CC=C(C=C2)C(=N)NCCC2CCN(CC3=CC=CC=C3)CC2)C=C1 | HIV\_QSAR\_44 | 6.09528445472132 |