SUPPLEMENTARY MATERIAL

Classification of Active and Weakly Active ST Inhibitors of HIV-1 Integrase Using a Support Vector Machine

Aixia Yan*, Shouyi Xuan and Xiaoying Hu

State Key Laboratory of Chemical Resource Engineering, Department of Pharmaceutical Engineering, P.O. Box 53, Beijing University of Chemical Technology, 15 BeiSanHuan East Road, Beijing 100029, P.R. China

Table S1. List of All 40 Molecule Descriptors Used in this Study, Obtained from MOE 2008 Version

No.	Descriptor	Description
1	petitjean	Value of (diameter - radius) / diameter. Largest value in the distance matrix [Petitjean 1992]. If r_i is the largest matrix entry in row i of the distance matrix D, then the radius is defined as the smallest of the r_i [Petitjean 1992].
2	VDistEq	If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of $\log_2 m - p_i \log_2 p_i / m$ where p_i is the number of distance matrix entries equal to i.
3	wienerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in [Balaban 1979].
4	BCUT_PEOE_1	The BCUT descriptors [Pearlman 1998] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(b_{ij}) where b_{ij} is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are the 1/3-ile eigenvalues.
5	BCUT_PEOE_2	The BCUT descriptors [Pearlman 1998] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(b_{ij}) where b_{ij} is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are the 2/3-ile eigenvalues.
6	BCUT_SLOGP_2	The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge. The resulting eigenvalues are the 2/3-ile eigenvalues.
7	BCUT_SMR_2	The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge. The resulting eigenvalues are the 2/3-ile eigenvalues.
8	GCUT_PEOE_2	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are the 2/3-ile eigenvalues.
9	GCUT_PEOE_3	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are the largest eigenvalues.
10	GCUT_SMR_1	The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge. The resulting eigenvalues are the 1/3-ile eigenvalues.
11	chiral_u	The number of unconstrained chiral centers.
12	a_nC	Number of carbon atoms.
13	a_nF	Number of fluorine atoms.
14	a_nN	Number of nitrogen atoms.
15	a_nS	Number of sulfur atoms.
16	VAdjEq	Vertex adjacency information (equality): $-(1-f)\log_2(1-f) - f \log_2 f$ where $f = (n^2 - m) / n^2$, n is the number of heavy atoms and m is the number of heavy-heavy bonds. If f is not in the open interval (0,1), then 0 is returned.
17	balabanJ	Balaban's connectivity topological index [Balaban 1982].
18	PEOE_PC+	Total positive partial charge: the sum of the positive q_i.
19	PEOE_RPC+	Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i.
20	PEOE_RPC-	Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i.
21	PEOE_VSA+2	Sum of v_i where q_i is in the range [0.10,0.15).
22	PEOE_VSA+4	Sum of v_i where q_i is in the range [0.20,0.25).
23	PEOE_VSA+6	Sum of v_i where q_i is greater than 0.3.
24	PEOE_VSA-4	Sum of v_i where q_i is in the range [-0.25,-0.20).

(Table S1) contd…..

No.	Descriptor	Description
25	PEOE_VSA_FPOL	Fractional polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
26	PEOE_VSA_FPOS	Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.
27	PEOE_VSA_FPPOS	Fractional positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
28	PEOE_VSA_PNEG	Total negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2. The v_i are calculated using a connection table approximation.
29	PEOE_VSA_POS	Total positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative. The v_i are calculated using a connection table approximation.
30	PEOE_VSA_PPOS	Total positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2. The v_i are calculated using a connection table approximation.
31	opr_brigid	The number of rigid bonds from [Oprea 2000].
32	a_acc	Number of hydrogen bond acceptor atoms (not counting acidic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH).
33	a_don	Number of hydrogen bond donor atoms (not counting basic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH).
34	vsa_don	Approximation to the sum of VDW surface areas of pure hydrogen bond donors (not counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH) ($Å^2$).
35	SlogP_VSA3	Sum of v_i such that L_i is in (0,0.1].
36	SMR_VSA1	Sum of v_i such that R_i is in (0.11,0.26].
37	SMR_VSA2	Sum of v_i such that R_i is in (0.26,0.35].
38	SMR_VSA3	Sum of v_i such that R_i is in (0.35,0.39].
39	SMR_VSA7	Sum of v_i such that $R_i > 0.56$.
40	TPSA	Polar surface area ($Å^2$) calculated using group contributions to approximate the polar surface area from connection table information only. The parameterization is that of Ertl *et al.* [Ertl 2000].

Table S2. List of Inter-Correlations Between 40 MOE 2D Descriptors and Activity

	A^0	D^1	D^2	D^3	D^4	D^5	D^6	D^7	D^8	D^9	D^{10}	D^{11}	D^{12}	D^{13}	D^{14}	D^{15}	D^{16}	D^{17}	D^{18}	D^{19}	D^{20}	D^{21}	D^{22}	D^{23}	D^{24}	D^{25}	D^{26}	D^{27}	D^{28}	D^{29}	D^{30}	D^{31}	D^{32}	D^{33}	D^{34}	D^{35}	D^{36}	D^{37}	D^{38}	D^{39}
D^1	-0.203	1																																					
D^2	0.195	-0.357	1																																				
D^3	0.188	-0.314	0.404	1																																			
D^4	0.181	-0.554	0.340	0.521	1																																		
D^5	-0.180	0.826	-0.230	-0.128	-0.422	1																																	
D^6	0.180	-0.283	0.132	-0.024	-0.049	-0.232	1																																
D^7	0.177	-0.329	-0.072	0.213	0.216	-0.312	0.356	1																															
D^8	0.171	-0.254	0.104	0.637	0.669	-0.147	-0.112	0.223	1																														
D^9	0.169	-0.305	-0.242	0.219	0.263	-0.320	0.231	0.549	0.461	1																													
D^{10}	0.161	-0.216	0.104	-0.231	-0.114	-0.266	0.486	0.217	-0.203	0.278	1																												
D^{11}	-0.160	0.768	-0.247	-0.331	-0.541	0.751	-0.271	-0.418	-0.389	-0.434	-0.025	1																											
D^{12}	0.159	-0.574	0.175	0.005	0.245	-0.722	0.300	0.244	0.099	0.222	0.225	-0.595	1																										
D^{13}	0.154	-0.820	0.344	0.365	0.716	-0.816	0.229	0.473	0.386	0.404	0.142	-0.815	0.602	1																									
D^{14}	0.150	-0.680	0.179	0.070	0.374	-0.856	0.221	0.284	0.160	0.380	0.234	-0.656	0.832	0.744	1																								
D^{15}	0.148	-0.594	0.122	0.507	0.573	-0.554	0.184	0.520	0.501	0.591	-0.065	-0.656	0.374	0.760	0.510	1																							
D^{16}	0.145	-0.599	0.367	0.563	0.658	-0.552	0.081	0.444	0.430	0.393	-0.182	-0.626	0.296	0.775	0.483	0.785	1																						
D^{17}	-0.143	-0.144	-0.057	0.181	0.111	-0.110	0.007	0.280	0.166	0.255	-0.072	-0.213	0.339	0.286	0.221	0.359	0.289	1																					
D^{18}	0.141	0.024	0.022	0.440	0.151	0.051	-0.045	0.053	0.406	0.302	-0.045	-0.009	-0.199	-0.017	-0.030	0.072	0.062	-0.081	1																				
D^{19}	0.141	-0.370	-0.028	-0.082	0.048	-0.591	0.239	0.189	0.069	0.402	0.220	-0.431	0.625	0.392	0.794	0.259	0.146	0.084	0.209	1																			
D^{20}	0.141	-0.190	0.282	0.514	0.814	0.090	-0.138	0.053	0.657	0.157	-0.206	-0.186	-0.139	0.289	-0.077	0.318	0.386	0.032	0.262	-0.247	1																		
D^{21}	0.139	0.092	-0.020	0.150	-0.092	0.028	0.050	-0.035	0.148	0.140	0.104	0.090	-0.167	-0.156	-0.001	-0.150	-0.169	-0.197	0.806	0.335	-0.020	1																	
D^{22}	-0.138	-0.026	0.038	-0.411	-0.121	-0.007	0.040	-0.100	-0.431	-0.346	0.020	0.051	0.147	-0.012	-0.005	-0.128	-0.078	0.051	-0.822	-0.196	-0.185	-0.689	1																
D^{23}	-0.138	0.113	-0.200	-0.238	-0.213	-0.026	0.045	0.079	-0.221	-0.134	-0.095	0.100	-0.006	-0.008	0.005	0.021	0.134	0.097	-0.199	0.023	-0.304	-0.086	0.148	1															
D^{24}	-0.137	-0.121	0.084	0.175	-0.135	-0.072	0.137	0.110	-0.243	-0.115	-0.035	-0.088	-0.106	0.083	-0.047	-0.012	0.268	0.063	-0.072	-0.104	-0.197	-0.072	0.057	0.253	1														
D^{25}	0.132	-0.228	0.121	-0.028	-0.135	-0.280	0.495	0.360	-0.227	0.313	0.218	-0.299	0.212	0.221	0.200	0.305	0.325	0.050	-0.144	0.264	-0.256	-0.091	0.114	0.199	0.454	1													
D^{26}	0.126	0.077	-0.065	0.323	0.031	0.061	-0.105	0.042	0.290	0.288	-0.044	0.061	-0.231	-0.093	-0.025	0.022	0.020	-0.087	0.889	0.219	0.123	0.838	-0.810	-0.052	-0.067	-0.121	1												
D^{27}	0.126	-0.705	0.163	0.148	0.344	-0.868	0.199	0.292	0.115	0.265	0.204	-0.661	0.414	0.712	0.590	0.507	0.486	-0.013	0.022	0.414	-0.130	0.032	-0.075	0.042	0.177	0.317	-0.002	1											
D^{28}	0.124	-0.651	0.564	0.361	0.723	-0.496	0.229	0.216	0.376	0.097	0.007	-0.582	0.336	0.746	0.434	0.528	0.689	0.131	-0.030	0.026	0.475	-0.161	0.036	-0.073	-0.032	-0.028	-0.104	0.400	1										
D^{29}	0.122	-0.250	0.440	0.473	0.725	0.090	0.012	0.044	0.551	0.058	-0.163	-0.206	-0.129	0.308	-0.084	0.302	0.432	0.036	0.180	-0.317	0.899	-0.064	-0.108	-0.263	-0.158	-0.245	0.055	-0.141	0.698	1									
D^{30}	-0.121	0.284	0.122	0.064	0.108	0.479	-0.223	-0.152	0.030	-0.350	-0.193	0.332	-0.365	-0.231	-0.514	-0.152	-0.032	0.025	-0.322	-0.810	0.362	-0.450	0.296	-0.020	-0.048	-0.348	-0.343	-0.518	0.134	0.441	1								
D^{31}	0.117	-0.177	0.057	0.014	0.058	-0.262	0.085	0.048	0.051	0.106	0.099	-0.197	0.115	0.196	0.153	0.155	0.121	-0.057	0.052	0.125	-0.068	0.059	-0.114	-0.011	0.001	0.097	0.057	0.319	0.107	-0.051	-0.186	1							
D^{32}	0.117	-0.719	0.162	0.098	0.445	-0.773	0.302	0.456	0.154	0.308	0.148	-0.759	0.670	0.831	0.694	0.635	0.555	0.217	-0.302	0.358	-0.004	-0.347	0.219	0.021	0.029	0.321	-0.349	0.693	0.531	0.013	-0.240	0.185	1						
D^{33}	0.115	-0.346	0.061	0.081	0.261	-0.171	0.313	0.426	0.168	0.278	0.047	-0.482	0.307	0.407	0.163	0.435	0.285	0.152	-0.334	-0.015	0.193	-0.440	0.284	-0.124	-0.080	0.276	-0.420	0.134	0.300	0.228	0.075	0.043	0.675	1					
D^{34}	0.112	-0.301	0.685	0.235	0.301	-0.252	0.084	-0.068	0.205	-0.204	-0.027	-0.271	0.341	0.334	0.234	0.175	0.278	-0.003	-0.081	0.075	0.189	-0.090	0.084	-0.102	-0.173	0.047	-0.152	0.190	0.490	0.311	0.050	0.096	0.334	0.251	1				
D^{35}	0.111	-0.001	0.127	0.608	0.319	0.122	-0.138	0.055	0.549	0.273	-0.187	-0.062	-0.274	0.032	-0.135	0.136	0.213	-0.038	0.862	-0.002	0.486	0.597	-0.717	-0.214	-0.005	-0.165	0.739	-0.040	0.097	0.404	-0.085	0.007	-0.279	-0.209	0.013	1			
D^{36}	-0.109	0.134	-0.186	-0.572	-0.374	-0.041	0.064	-0.015	-0.517	-0.248	0.096	0.186	0.120	-0.132	0.033	-0.241	-0.142	-0.063	-0.639	-0.022	-0.501	-0.384	0.661	0.549	0.111	0.205	-0.484	-0.027	-0.205	-0.458	0.078	-0.042	0.097	-0.002	-0.078	-0.708	1		
D^{37}	0.109	-0.626	0.112	0.042	0.324	-0.854	0.098	0.146	0.125	0.229	0.184	-0.604	0.684	0.665	0.871	0.363	0.301	0.049	0.056	0.750	-0.157	0.102	-0.065	-0.112	-0.091	0.021	0.013	0.735	0.336	-0.201	-0.583	0.188	0.637	0.016	0.170	-0.087	-0.051	1	
D^{38}	-0.107	-0.077	-0.176	-0.317	-0.137	-0.220	-0.076	0.029	-0.343	-0.129	-0.009	-0.011	0.204	0.079	0.184	0.010	0.047	0.134	-0.572	-0.142	-0.354	-0.462	0.439	0.411	0.146	0.103	-0.383	0.111	-0.066	-0.371	0.180	0.003	0.308	0.058	-0.186	-0.565	0.611	0.069	1
D^{39}	0.103	-0.139	0.321	0.135	0.268	-0.096	0.029	0.063	-0.010	-0.146	0.066	0.037	-0.039	0.227	-0.030	0.126	0.202	-0.094	-0.060	-0.208	0.276	-0.169	0.126	-0.052	-0.012	0.001	-0.140	0.106	0.260	0.291	0.233	0.056	0.087	0.064	0.158	-0.007	-0.016	-0.025	-0.012	1
D^{40}	0.101	-0.460	0.027	0.365	0.515	-0.456	-0.135	0.275	0.520	0.495	-0.027	-0.506	0.467	0.575	0.596	0.484	0.397	0.402	0.193	0.399	0.280	0.033	-0.215	-0.277	-0.239	-0.304	0.170	0.264	0.345	0.194	-0.202	0.054	0.420	0.171	0.047	0.204	-0.294	0.540	0.035	-0.081

A^0: Activity; D^1: PEOE_RPC-; D^2: SMR_VSA2; D^3: a_don; D^4: PEOE_VSA_PPOS; D^5: VAdjEq; D^6: PEOE_VSA-4; D^7: PEOE_VSA+2; D^8: PEOE_VSA+4; D^9: SMR_VSA1; D^{10}: a_nF; D^{11}: PEOE_RPC+; D^{12}: GCUT_PEOE_3; D^{13}: PEOE_PC+; D^{14}: weinerPol; D^{15}: a_acc; D^{16}: TPSA; D^{17}: chiral_u; D^{18}: BCUT_PEOE_2; D^{19}: opr_brigid; D^{20}: PEOE_VSA_FPPOS; D^{21}: BCUT_SLOGP_2; D^{22}: BCUT_PEOE_1; D^{23}: a_nS; D^{24}: vsa_don; D^{25}: a_nN; D^{26}: BCUT_SMR_2; D^{27}: VDistEq; D^{28}: PEOE_VSA_PNEG; D^{29}: PEOE_VSA_FPOL; D^{30}: balabanJ; D^{31}: petitjean; D^{32}: PEOE_VSA_POS; D^{33}: PEOE_VSA_FPOS; D^{34}: SlogP_VSA3; D^{35}: GCUT_PEOE_2; D^{36}: GCUT_SMR_1; D^{37}: a_nC; D^{38}: SMR_VSA7; D^{39}: PEOE_VSA+6; D^{40}: SMR_VSA3.

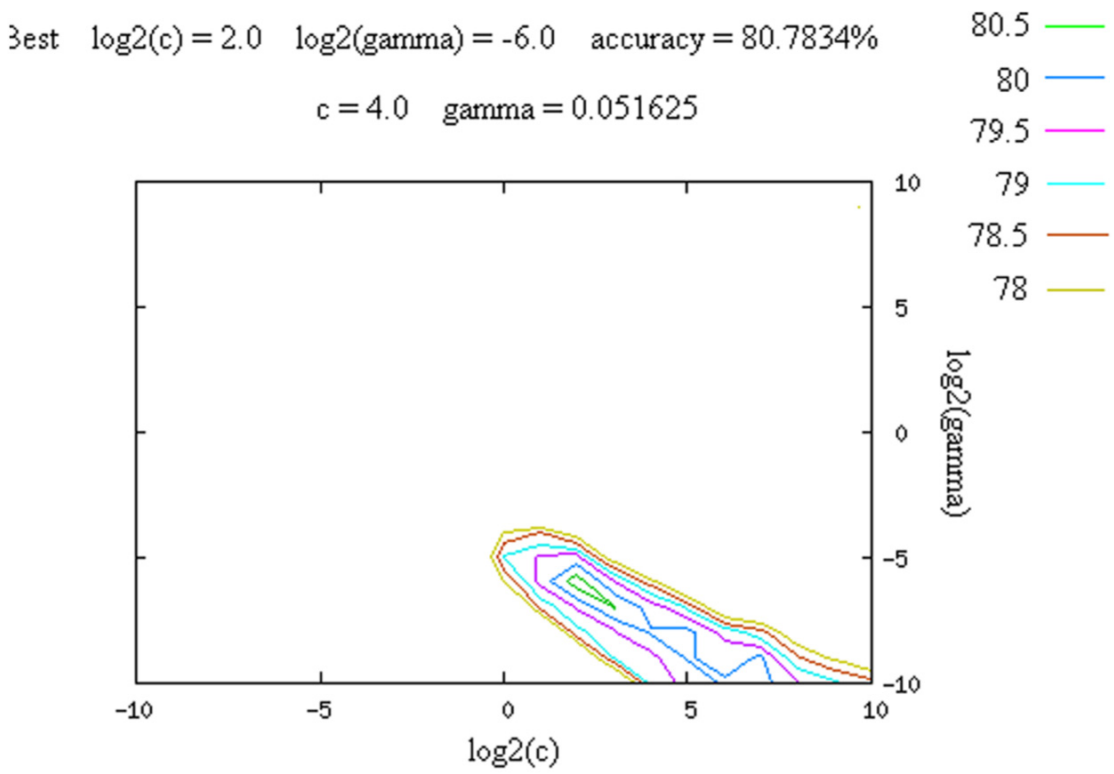

Fig. (S1). The SVM drill process figure of Model 1 built with MACCS fingerprints.

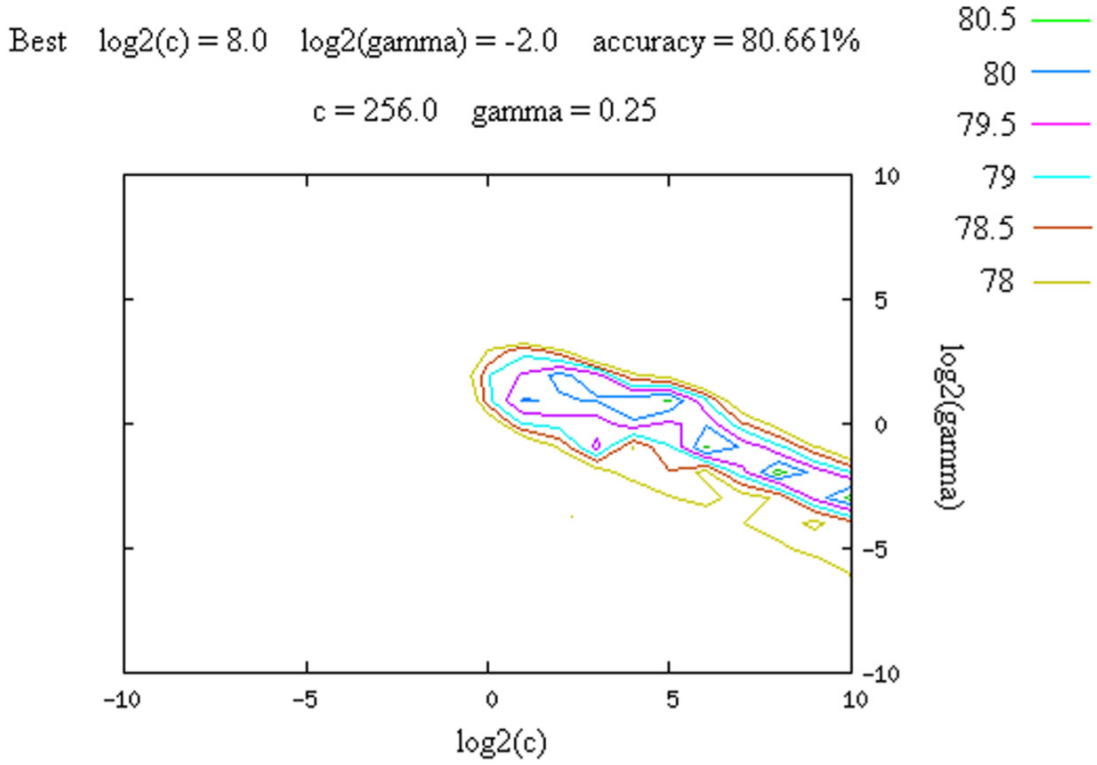

Fig. (S2). The SVM drill process figure of Model 2 built with 40 MOE descriptors.