Supporting Information:

MS # JM050043W

M-Score: A Knowledge-Based Potential Scoring Function Accounting For Protein Atom Mobility

Chao-Yie Yang, Renxiao Wang and Shaomeng Wang*

Departments of Internal Medicine, Pharmacology and Medicinal Chemistry,

University of Michigan,

1500 E. Medical Center Dr.,

Ann Arbor, Michigan 48109-0934

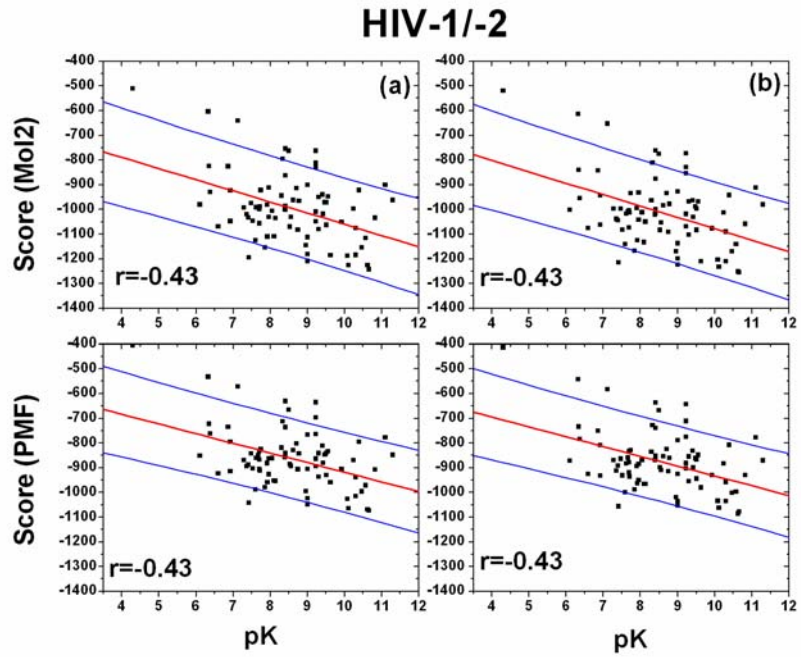
**Correspondence should be addressed to Shaomeng Wang;*

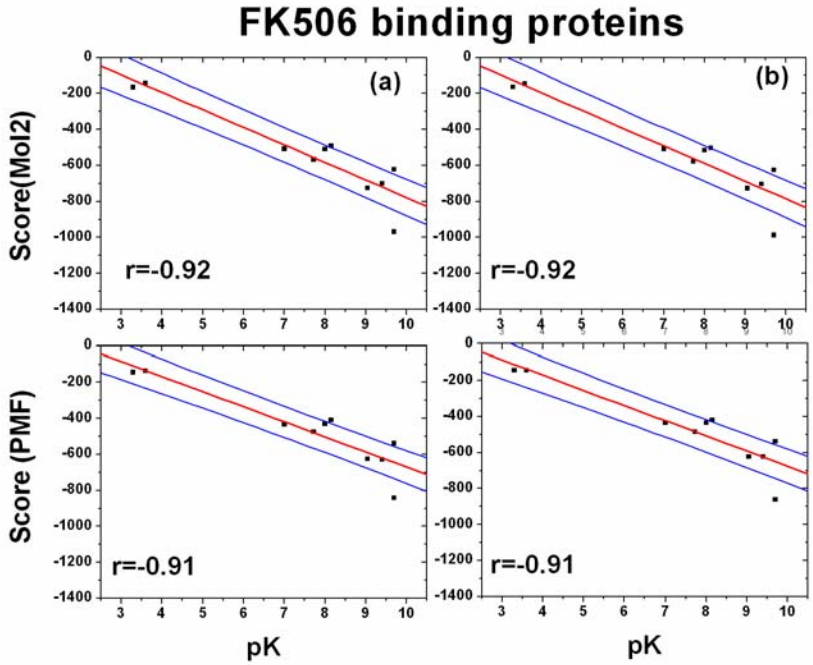
Phone: 734-615-0362.

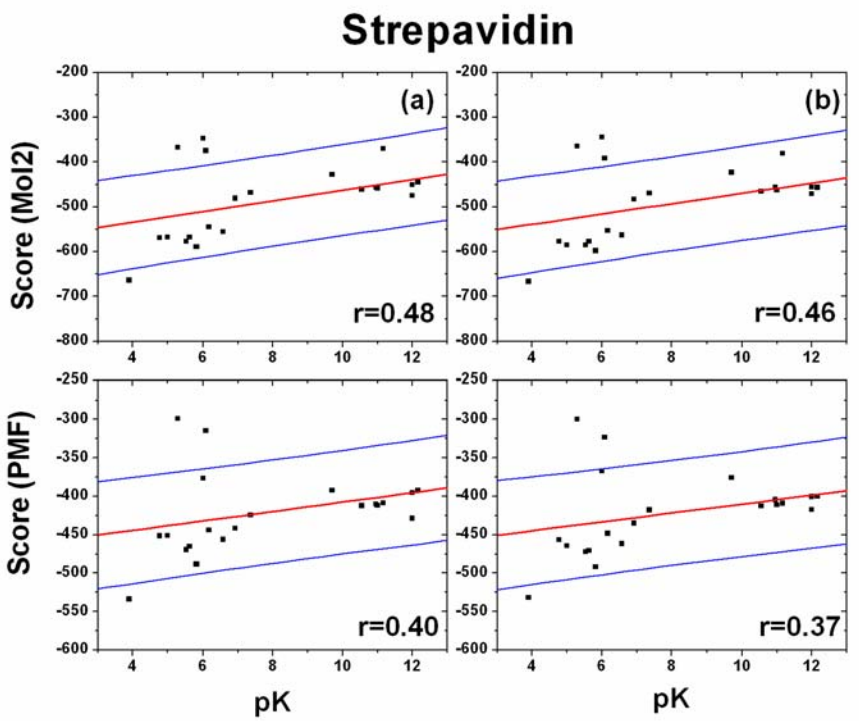
Fax:734-647-9647.

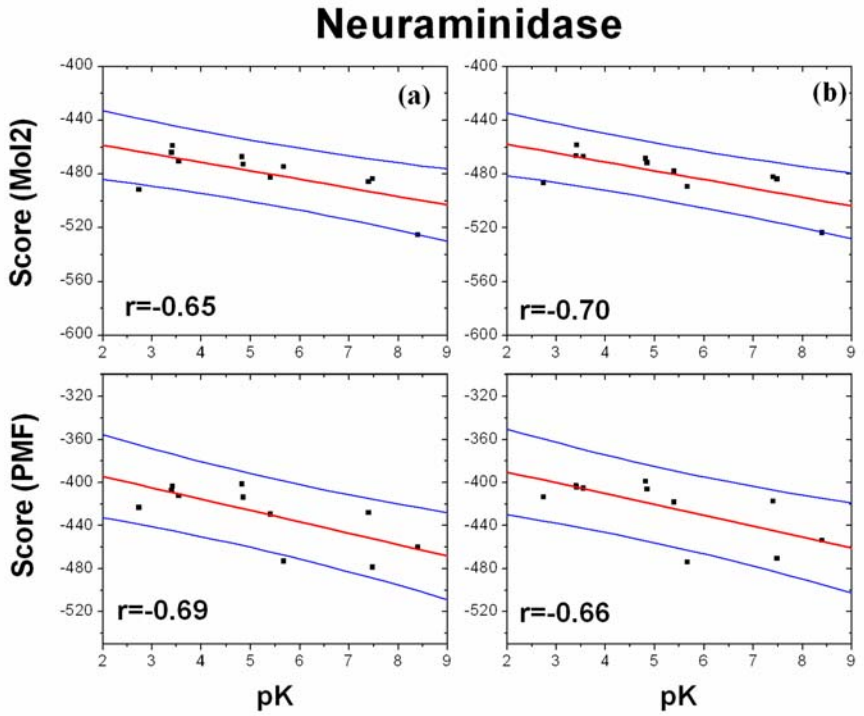
email: shaomeng@umich.edu

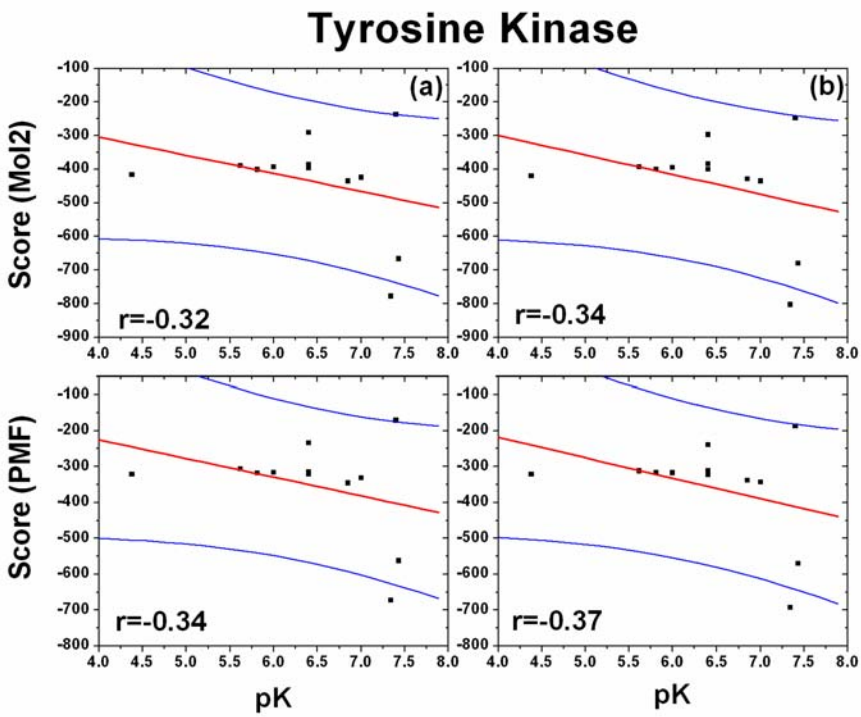
Figure S1. Examples: HIV-1/-2 inhibitor (the largest single protein family set), FK506-Binding protein (the best case), Strepavidin (the worst case), Neuraminidase (slightly improvement with B-factor augmentation), Tyrosine-Kinase and Tyrosine-Phosphatase (protein families with opposite reactions against Tyrosine). Score (Mol2) and Score (PMF) denote the atom types of Mol2 and PMF for the ligands are used.






Strepavidin
(a)
(b)
Score (Mol2)
Score (PMF)
r=0.48
r=0.46
r=0.40
r=0.37
pK
pK


Neuraminidase
(a)
(b)
Score (Mol2)
Score (PMF)
r=-0.65
r=-0.70
r=-0.69
r=-0.66
pK
pK


Tyrosine Kinase
(a)
(b)
Score (Mol2)
Score (PMF)
r=-0.32
r=-0.34
r=-0.34
r=-0.37
pK
pK

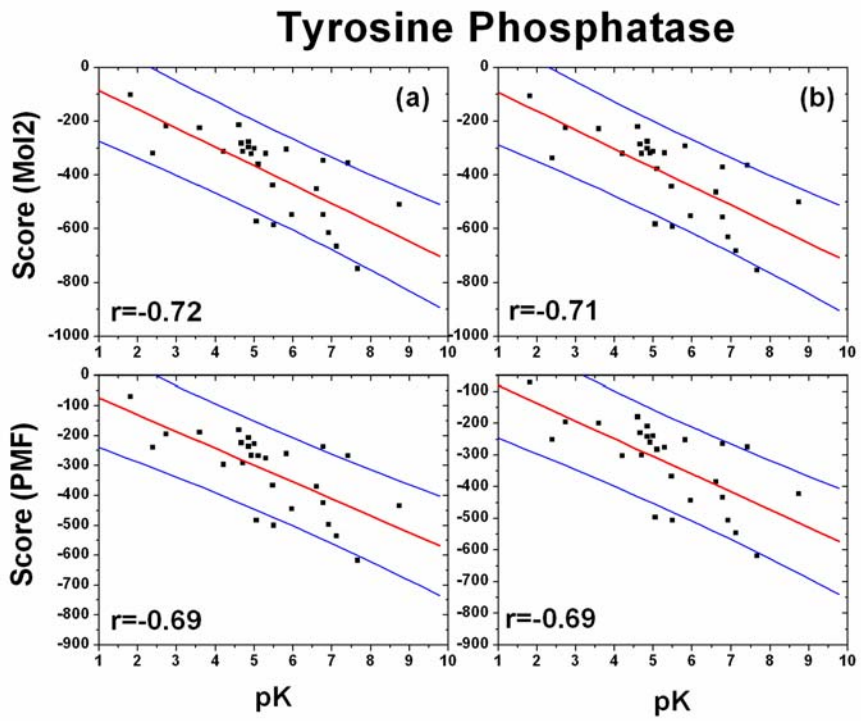

Tyrosine Phosphatase
(a)
(b)
Score (Mol2)
Score (PMF)
r=-0.72
r=-0.71
r=-0.69
r=-0.69
pK
pK

Table S1. Numbers of atom pairs derived from 2331 potein-ligand complexes using the atom types defined in Table 1 for proteins and PMF atom types for ligands with a cutoff distance of 10 Å.

	Cα	C(bk)	Cβ	C.3 (sch)	C.2 (N,Q)	C.2 (E,D)	C.cat (R)	C.ar	N.am	N.ar (W,H)	Npolar (N,Q)	N.4 (R,K)	O(bk)	O.co2	O (polar)	O.2 (N,Q)	S(M)	S (C)	Metal
CF	135570	134251	122479	146721	8727	4407	5776	150893	137198	13041	14343	14660	131447	26591	26856	17338	4454	3137	1261
CP	162246	158011	149494	157371	15579	9580	10825	200722	163614	24695	26850	28591	151324	50696	34540	30683	3756	2808	2467
cF	114759	114031	99692	116189	7165	3456	4088	111324	116708	12540	11272	10558	110547	21896	20252	14193	3197	3795	966
cP	50417	50034	43984	49183	3264	1997	1715	42791	51446	5057	4910	4969	48092	9242	8997	5703	1600	1506	382
C3	22006	22006	20088	24859	1421	681	1124	23485	22239	1915	2521	2810	21268	3884	4081	2672	1008	602	210
CW	68383	66636	62410	71660	5448	2988	3584	73533	68871	7886	9221	9714	64107	16494	14276	10620	1672	1745	731
CO	21654	21300	20294	23130	1987	1149	1820	21017	21597	2503	3861	4526	20216	5124	5120	3427	525	549	271
CN	16407	16000	14907	15136	1292	858	743	16905	16620	1695	2092	1891	15592	4235	3830	2436	398	608	160
C0	422	430	368	392	0	0	0	540	435	0	0	0	409	69	54	0	0	0	0
NC	393	385	337	295	0	0	0	525	390	0	0	0	390	78	96	0	0	0	0
NP	21089	20736	18509	19962	1601	895	886	18208	21699	2166	2481	2563	19830	4595	3966	2913	623	665	260
NA	1505	1436	1322	1519	132	69	66	1917	1471	209	203	174	1417	332	316	235	55	0	0
ND	72133	70806	65595	72834	5692	3200	3715	76692	72810	7943	9423	9491	67920	18022	15569	11256	1620	2176	622
NR	14059	13789	12175	14057	856	591	447	10613	14440	1097	1292	1328	13133	2549	2268	1497	433	254	56
N0	205	207	189	250	0	0	0	344	225	0	0	0	205	0	0	0	0	0	0
NS	784	733	727	834	0	0	0	921	762	203	72	56	708	175	138	85	0	0	0
OC	87312	85115	80154	90009	7785	4540	7256	75767	87618	9811	15049	19712	81156	22397	20292	14164	1835	1766	1655
OA	41607	40485	38591	45613	3452	1862	2277	49001	41935	4962	5768	6028	39118	10253	9045	6712	959	909	393
OE	25904	25163	24052	24517	2740	1606	1829	34333	25799	4050	4601	4764	24110	8305	5138	5111	604	399	339
OR	702	700	628	860	0	0	0	552	698	68	0	0	667	100	103	0	0	0	0
OS	22412	21685	20344	22619	1879	1129	1728	21119	22775	2939	3673	4836	20383	5991	5080	3619	465	377	569
OD	70019	68387	63709	64506	6792	4594	4281	83264	71107	11514	11229	11317	65150	22481	13909	13184	1830	1263	1214
P	16392	15823	14766	16307	1335	858	1429	12747	16214	1783	2794	4034	14707	4783	3758	2806	279	264	466
SA	2271	2208	2073	2383	230	145	139	2513	2317	468	384	381	2140	521	475	333	0	0	62
SD	6266	6227	5615	5886	486	236	275	7983	6400	965	770	685	5873	1338	1442	888	165	151	75
F	2881	2857	2639	3228	163	82	91	2652	2924	254	264	256	2811	515	512	320	74	118	0
Cl	1254	1260	1113	1354	60	0	0	1088	1286	87	86	87	1217	164	192	120	0	0	0
Br	615	601	525	654	0	0	0	502	641	0	0	52	576	100	120	54	0	0	0
I	686	662	607	819	0	0	0	508	692	0	0	56	639	72	142	0	0	0	0

Table S2. Numbers of atom pairs for different atom pairs included in the knowledge-based potential function constructed based on the Tripos Mol2 atom types for both protein and ligand.

	C.2	C.3	C.ar	C.cat	N.am	N.pl3	N.2	N.4	O.2	O.co2	O.3	S.3	Metal
C.3	430228	909492	303028	17233	339661	62778	27738	9976	320128	80689	64392	14854	3845
C.2	150631	334421	98667	6530	121931	22628	9239	3861	114601	25527	23475	6099	1212
C.ar	217546	474474	126546	5808	178556	21640	13523	3849	168775	31215	29273	10102	1348
C.1	567	1182	456	0	451	0	0	0	425	69	54	0	0
C.cat	5231	10652	3036	104	4345	448	209	0	4077	840	812	305	0
N.3	5638	11975	3884	214	4340	825	432	135	4078	1075	887	188	0
N.am	53995	118060	38292	2395	42645	8462	3462	1355	40012	10066	8859	1647	408
N.2	19270	40931	10696	571	15918	2131	1231	500	14617	3163	2849	894	159
N.ar	17719	40291	8832	447	15300	1663	851	343	13928	2549	2268	687	56
N.1	285	644	298	0	237	51	0	0	217	0	0	0	0
N.4	14409	30805	9297	557	11512	1917	850	338	11153	2979	2548	695	101
N.pl3	36069	76067	19632	958	30352	3599	1545	666	27817	5934	4942	1771	236
O.3	167790	327105	105580	7675	127582	27558	12823	4832	116998	35990	23321	4816	2026
O.2	63447	138978	45631	2450	50580	8978	4291	1505	47425	11173	9954	2041	497
O.co2	118957	257475	60578	7256	95706	23443	8115	4926	88976	22413	20292	3601	1655
S.3,S.2,P.3	31410	66209	16801	1741	24970	5661	2226	1361	22715	6241	5286	862	555
S.o, S.o2	2959	5750	1617	102	2161	378	380	58	2022	415	389	62	0
F	3661	8748	2258	91	3097	315	199	87	2966	515	512	192	0
Cl	1560	3721	930	0	1342	102	68	0	1281	164	192	74	0
Br	752	1794	436	0	667	0	0	0	605	100	120	0	0
I	791	2112	438	0	710	56	0	0	663	72	142	0	0

Table S3. Numbers of atom pairs in different atom pairs included in the knowledge-based potential function constructed based on the atom types of PMF.

	CF	CP	cF	cP	C0	CN	NC	ND	NR	OC	OA	OD	SA	SD	Metal
CF	228289	311211	120158	30735	12862	14024	20302	149422	9518	26600	140321	26856	4454	3137	1261
CP	254995	372399	149199	51523	24439	26132	38847	187686	17217	50721	167123	34540	3756	2808	2467
cF	190026	271115	87532	27262	11077	10831	15264	131283	9918	22786	121689	20951	3312	3897	1020
cP	104057	155764	43125	14232	6497	6420	9187	75309	5433	13061	68753	12180	2023	2105	599
C3	31506	42085	16391	3597	1491	2261	3374	19707	1035	3033	18617	3418	915	499	148
CW	86003	116726	45368	13626	6204	7352	10707	57008	4412	12705	51927	11057	1227	1147	522
CO	35220	50692	15573	5444	2457	4273	6271	24257	1959	5128	22331	5120	525	549	271
CN	24498	38103	12893	4012	2064	1884	2637	18444	1217	4237	16820	3830	398	608	160
C0	652	967	431	109	0	0	0	462	0	69	425	54	0	0	0
NC	520	888	411	114	0	0	0	430	0	78	408	96	0	0	0
NP	8862	13036	5022	1527	627	698	1018	6501	523	1267	5822	1127	158	129	58
NA	2371	3448	1468	449	164	161	245	1681	131	333	1556	316	55	0	0
ND	109939	156363	56370	16918	8344	8977	12676	76416	5200	17140	69507	14757	1546	1909	564
NR	50770	76875	19488	6161	3362	3238	4593	37445	2300	6765	33503	5907	970	1056	316
N0	427	596	294	77	0	0	56	297	0	0	269	57	0	0	0
NS	1330	1767	567	354	83	0	74	830	189	175	757	138	0	0	0
OC	136819	202367	55169	20598	10868	18445	26673	97402	8115	22413	88976	20292	1835	1766	1655
OA	70951	95418	37835	11166	4986	5656	8225	46962	3506	10253	42706	9045	959	909	393
OE	40699	59226	25797	8536	3992	4280	6419	30049	2746	8308	26832	5138	604	399	339
OR	1020	1344	374	114	0	0	54	609	56	91	575	93	0	0	0
OS	34467	51077	15025	5928	2878	4551	6480	25116	2463	5996	22097	5023	455	377	569
OD	106012	161379	60388	22876	10954	10303	15257	81576	8334	22490	71965	13909	1830	1263	1214
P	24648	37552	9178	3569	2351	3852	5395	17858	1572	4786	16017	3758	279	264	466
SA	5328	7724	2343	1168	364	443	630	3853	580	750	3455	683	0	55	87
SD	7925	12095	5296	1689	520	544	823	6030	454	1120	5265	1234	148	130	0
F	5024	6564	2096	556	247	271	347	3152	199	515	2966	512	74	118	0
Cl	2135	2846	873	215	82	84	117	1361	68	164	1281	192	0	0	0
Br	999	1387	384	118	51	53	61	675	0	100	605	120	0	0	0
I	1158	1590	405	103	0	64	70	725	0	72	663	142	0	0	0

Table S4. PDB entries of 17 protein families.

Protein Names	Number of PDB entries	PDB IDs
FK506 Binding Protein	**10**	1FKB, 1FKF, 1FKH, 1D7I, 1D7J, 1FKG, 1FKI, 1J4R, 1PBK, 1A7X
Thermolysin	**12**	1QF0, 1QF1, 1QF2, 1TLP, 1TMN, 2TMN, 4TLN, 4TMN, 5TLN, 5TMN, 1OS0, 1THL
Trypsin	**63**	1F0T, 1F0U, 1G3B, 1G3D, 1G3E, 1K1I, 1K1J, 1K1L, 1K1M, 1K1N, 1O2P, 1O2R, 1O2X, 1OSS, 1PPC, 1PPH, 1QB1, 1QB6, 1QB9, 1QBN, 1QBO, 1TNG, 1TNH, 1TNI, 1TNJ, 1TNK, 1TNL, 1XUG, 2BZA, 1CE5, 1LQE, 1OYQ, 1BRA, 1GI1, 1O2H, 1O2K, 1O2N, 1O2O, 1O2Q, 1O2T, 1O2Z, 1O30, 1O35, 1O36, 1O3C, 1O3G, 1O3H, 1O3K, 1O3L, 1C1R, 1C2D, 1C5P, 1C5Q, 1C5S, 1C5T, 1GHZ, 1GI4, 1GI6, 1GJ6, 1H4W, 1J14, 1J16, 1J17
Urokinase-Type Plasminogen Activator	**14**	1F5K, 1F5L, 1GI9, 1GJ9, 1GJB, 1O3P, 1EJN, 1GI7, 1GI8, 1GJ7, 1GJ8, 1GJA, 1GJC, 1GJD
Tyrosine Phosphatase	**27**	1ONY, 1PH0, 1PXH, 1PYN, 1NO6, 1NWL, 1NZ7, 1ONZ, 1NL9, 1NNY, 1QXK, 1PA9, 1G7F, 1G7G, 1BZC, 1BZH, 1BZJ, 1D1P, 1D5R, 1DG9, 1C83, 1C84, 1C86, 1C87, 1C88, 1ECV, 1KAV
Neuraminidase	**11**	1F8B, 2QWB, 2QWC, 2QWD, 2QWE, 2QWF, 2QWG, 2SIM, 1F8C, 1F8D, 1F8E
Ribonuclease T1, A	**13**	1RGK, 6RNT, 1JN4, 1O0F, 1O0M, 1O0N, 1RSK, 1AFK, 1AFL, 1HI3, 1HI4, 1O0O, 1QHC
Thrombin	**37**	1O2G, 1G30, 1G32, 1D3D, 1D3P, 1ETS, 1ETT, 1GHV, 1GHW, 1GHY, 1GJ4, 1KTS, 1NM6, 1TMT, 1TOM, 1UVT, 7KME, 1D4P, 1D6W, 1D9I, 1GJ5, 1NT1, 1OYT, 1QBV, 1BCU, 1FPC, 1A4W, 1BHX, 1C1U, 1C1V, 1C4U, 1C4V, 1C5N, 1C5O, 1JWT, 1K21, 1K22
Carbonic Anhydrase II	**39**	1BCD, 1CIL, 1CNX, 1CNY, 1G1D, 1G45, 1G46, 1G48, 1G4J, 1G4O, 1G52, 1G53, 1G54, 1H4N, 1OKL, 1OKN, 1ZSB, 2H4N, 1A42, 1AVN, 1BN1, 1BN3, 1BN4, 1BNN, 1BNQ, 1BNT, 1BNU, 1BNV, 1BNW, 1CIM, 1CIN, 1CNW, 1IF7, 1IF8, 1YDA, 1YDB, 1YDD, 1I9N, 1I9P
Antibody Fab Fragment	**31**	1IGJ, 1A4K, 1JGL, 1MFD, 2PCP, 1FL3, 1C5C, 1KEL, 1A0Q, 1FLR, 43CA, 1GAF, 1AJ7, 1HYX, 1CT8, 1IND, 1L7S, 1D6V, 1YEI, 1YEJ, 1I7Z, 2CGR, 1DL7, 1MJ7, 1MJJ, 1MH5, 1F4X, 1PZ5, 1M7D, 1M7I, 1TET
Protocatechuate 3,4-Dioxygenase	**10**	3PCB, 3PCC, 3PCE, 3PCF, 3PCG, 3PCH, 3PCJ, 3PCK, 3PCN, 1EOC
HIV-1, HIV-2 Protease	**79**	1A94, 1AAQ, 1G2K, 1G35, 1HBV, 1HOS, 1HPO, 1HPS, 1HPV, 1HPX, 1HSG, 1HSH, 1HVH, 1HVI, 1HVJ, 1HVK, 1HVL, 1HVR, 1HWR, 1HXB, 1HXW, 1K6C, 1K6P, 1K6T, 1K6V, 1ODY, 1OHR, 1PRO, 1QBR, 1QBS, 1QBT, 1QBU, 1SBG, 2BPV, 2BPY, 3AID, 5HVP, 7HVP, 7UPJ, 8HVP, 1A30, 1AJV, 1AJX, 1B6J, 1B6K, 1B6L, 1B6M, 1B6N, 1B6O, 1B6P, 1BDQ, 1BV7, 1BV9, 1BWA, 1BWB, 1C6Y, 1C70, 1D4K, 1D4L, 1D4S, 1D4Y, 1DMP, 1HIH, 1HIV, 1IZH, 1IZI, 1MTR, 1DIF, 1LZQ, 1A9M, 1MES, 1HVS, 1MET, 1MEU, 1HII, 5UPJ, 6UPJ, 1IVP, 1KZK
Oligo-Peptide Binding Protein	**28**	1B05, 1B32, 1B3F, 1B3G, 1B3L, 1B40, 1B46, 1B4Z, 1B51, 1B52, 1B58, 1B5I, 1B5J, 1B9J, 1JET, 1QKA, 1QKB, 2OLB, 2RKM, 1B1H, 1B2H, 1B3H, 1B4H, 1B5H, 1B6H, 1B7H, 1JEU, 1JEV
Tyrosine Kinase	**12**	1A07, 1A08, 1A1B, 1A1C, 1A1E, 1IS0, 1FPU, 1B55, 2FGI, 1LKL, 1LKK, 1BHF
Endothiapepsin	**15**	1EED, 1ENT, 1EPO, 1EPP, 1EPQ, 2ER6, 2ER9, 3ER3, 4ER1, 4ER2, 5ER2, 1GVW, 1GVX, 1GVU, 5ER1
Coagulation Factor X, Xa	**15**	1LPK, 1G2L, 1LPG, 1MQ5, 1MQ6, 1NFU, 1NFW, 1NFY, 1EZQ, 1F0R, 1F0S, 1FJS, 1KSN, 1XKA, 1NFX
Strepavidin	**21**	1SLD, 1SLE, 1SRG, 1STR, 1STS, 1SWK, 1SWN, 1SWP, 1SWR, 1VWF, 1SLG, 1SRI, 1VWL, 1VWN, 2IZL, 1DF8, 1SWG, 1N43, 1N9M, 1NC9, 1NDJ,

Table S5. The scoring potential functions are provided by two files, “mscore.index” and “mscore.pot”. The first file lists the indexes of all atom pairs between protein and ligand atoms. The numbers of the first and second columns in this file correspond to the protein and ligand atom types respectively. The second file lists the scoring potential functions for each atom pair in accordance with the order listed in the first file. The second file contains two columns where the first column is the distance between each atom pair in unit of 0.1 Å and the second column is the score. The range of distance provided is between 1 and 10 Å. The correspondence between the indexes in the first file and the atom types for protein and ligand atoms are listed in the following.

Index	Protein Atom Type	Ligand Atom Type
0	Cα or Ca	C.3
1	C(bk)	C.2
2	Cβ or Cb	C.ar
3	C.3(sch)	C.1
4	C.2(N,Q)	C.cat
5	C.2(E,D)	N.3
6	C.cat(R)	N.am
7	C.ar	N.2
8	N.am	N.ar
9	N.ar(W,H)	N.1
10	Npolar(N,Q)	N.4
11	N.4(R,K)	N.pl3
12	O(bk)	O.3
13	O.co2	O.2
14	O(polar)	O.co2
15	O.2(N,Q)	S.3, S.2, P.3
16	S(M)	S.o, S.o2
17	S(C)	F
18		Cl
19	Metal	Br
20		I