

## Supporting Information

# SFCscore<sup>RF</sup>: A Random-Forest-Based Scoring Function for Improved Affinity Prediction of Protein-Ligand Complexes

David Zilian and Christoph A. Sotriffer

---

Table S1

PDB IDs of the 1005 complexes used for training SFCscore<sup>RF</sup>

1lgs 1a0q 1a1c 1a1e 1a42 1a4k 1a4r 1a4w 1a7x 1a94 1a9m 1aaq 1abf 1add 1adl 1ado  
1af2 1af6 1afk 1af1 1agm 1ai4 1ai7 1aid 1aj6 1aj7 1ajn 1ajv 1ajx 1alw 1amw 1apb  
1apv 1atl 1awi 1axz 1b05 1b0h 1b11 1b1h 1b2h 1b32 1b38 1b3f 1b3g 1b3h 1b3l 1b40  
1b46 1b4h 1b4z 1b51 1b52 1b55 1b58 1b5h 1b5i 1b5j 1b6h 1b6k 1b74 1b8n 1b8y 1bai  
1bap 1bcd 1bdq 1bgq 1bhf 1bhx 1bn1 1bn3 1bn4 1bnn 1bnq 1bnt 1bnu 1bnv 1bnw 1bp0  
1bq4 1br6 1bv7 1bv9 1bwa 1bwb 1bxq 1bxr 1byk 1bzc 1bzh 1bzj 1bzy 1c1r 1c1u 1c1v  
1c3x 1c4u 1c4v 1c5c 1c5n 1c5o 1c5p 1c5q 1c5s 1c5t 1c5x 1c5y 1c5z 1c70 1c83 1c84  
1c86 1c87 1c88 1caq 1cbx 1ce5 1cea 1ceb 1cet 1cgl 1ch8 1cil 1cim 1cin 1ciz 1cnw  
1cnx 1cny 1cps 1ct8 1ctt 1ctu 1d2e 1d3d 1d3p 1d4k 1d4l 1d4p 1d4y 1d5r 1d6v 1d6w  
1d7i 1d9i 1db1 1dfo 1dhj 1dif 1d17 1dmp 1dqn 1dqx 1drj 1drk 1dud 1dzk 1e1x 1e2l  
1e3v 1e4h 1e6q 1e6s 1ec9 1ecq 1ecv 1efy 1egh 1eix 1ejn 1elc 1eld 1ele 1elr 1ent  
1epo 1erb 1ew8 1ew9 1ex8 1ez9 1ezq 1f0r 1f0s 1f0t 1f0u 1f2o 1f2p 1f3e 1f4x 1f57  
1f5l 1f8a 1f9g 1fao 1fch 1fcy 1fdq 1fiv 1fj4 1fjs 1fkf 1fkg 1fkh 1fkx 1fl3 1fm9  
1fo0 1fpc 1fq5 1fv0 1fwu 1fwv 1fzm 1fzo 1fzq 1g1d 1g2l 1g2o 1g30 1g32 1g35 1g36  
1g3b 1g3c 1g3d 1g3e 1g45 1g46 1g48 1g4j 1g4o 1g52 1g53 1g54 1g7g 1g7v 1g85 1g98  
1g9r 1ga8 1gaf 1gah 1gai 1gar 1gcz 1ghv 1ghw 1ghy 1ghz 1gi1 1gi4 1gi6 1gi7 1gi8  
1gj4 1gj6 1gj7 1gjb 1gjc 1gnm 1gnn 1gno 1gny 1gpn 1grp 1gt1 1gul 1gvw 1gvx 1gww  
1gyx 1gz3 1gz4 1h0a 1h1d 1h1h 1h1p 1h1s 1h2t 1h4w 1h6h 1h9z 1hbv 1heg 1hi3 1hi5  
1h1h 1h1i 1h1k 1hmr 1hms 1hmt 1hn2 1hos 1hp0 1hpo 1hps 1hpx 1hqg 1hsh 1hvh  
1hvi 1hvj 1hvk 1hvl 1hvr 1hvs 1hwr 1hxb 1hxw 1hyo 1i00 1i5r 1i80 1i9l 1i9m 1i9n  
1i9o 1i9p 1i9q 1icj 1if8 1igb 1igj 1ii5 1iih 1iiq 1ik4 1ikt 1imx 1izh 1izi 1j01  
1j14 1j4r 1jak 1jao 1jcx 1jd5 1jet 1jeu 1jev 1jfh 1jfh 1jlr 1jmf 1jmg 1jmi 1jn2  
1jn4 1joc 1jq8 1jqe 1jqy 1jt1 1jwt 1jyq 1jzs 1k1i 1k1j 1k1l 1k1m 1k1n 1k1y 1k21  
1k22 1k4h 1kav 1kc7 1kdk 1kel 1kjr 1kl1 1kmy 1koj 1kpm 1ksn 1kyv 1kzn 1l6m 1l83  
1laf 1lag 1lah 1lbf 1lee 1lf2 1lf9 1lgt 1lgw 1li2 1li3 1li6 1lke 1lkk 1lkl 1lor  
1lox 1lpg 1lpk 1lpz 1lrt 1lyb 1lyx 1lzq 1m0o 1m13 1m1b 1m2p 1m2r 1m2x 1m48 1m4h  
1m5w 1m6p 1m7d 1m7i 1m7y 1m83 1m9n 1mai 1me7 1me8 1mes 1met 1meu 1mfa 1mfd 1mfi

1mh5 1mhw 1mj7 1mjj 1mmp 1mmq 1mmr 1moq 1mq5 1mrn 1mrs 1mrw 1mrx 1msm 1msn 1mtr  
1mu6 1mu8 1mue 1nlm 1nlt 1n3i 1n46 1n4h 1n4k 1n51 1n5r 1n8v 1ndv 1nf8 1nfu 1nfw  
1nfx 1nh0 1njc 1njd 1njj 1njs 1nl9 1nm6 1no6 1nq7 1nt1 1nu3 1nvr 1nvs 1nw4 1nw7  
1nwl 1nz7 1o0f 1o0m 1o0n 1o0o 1o2g 1o2h 1o2j 1o2k 1o2n 1o2o 1o2q 1o2s 1o2w 1o2x  
1o2z 1o30 1o33 1o36 1o38 1o3d 1o3h 1o3i 1o3j 1o3k 1o7o 1o86 1o8b 1o9d 1oai 1oar  
1oba 1obx 1ocq 1od8 1odj 1ody 1oe8 1ofz 1ogg 1ogx 1ohr 1oif 1okl 1oko 1olx 1ony  
1onz 1ork 1os0 1os5 1oss 1owd 1owe 1owh 1oyq 1oyt 1oz0 1p19 1p28 1pa9 1pb8 1pbk  
1pdz 1pfu 1pgp 1pkx 1pme 1pot 1ppc 1pph 1ppi 1ppk 1ppl 1pr1 1pro 1pu7 1pu8 1pvn  
1px4 1pxn 1pyn 1pzi 1pzo 1pzp 1q1g 1q54 1q63 1q65 1q66 1q84 1q8u 1q8w 1q91 1q95  
1qan 1qaw 1qbl 1qb6 1qb9 1qbn 1qbo 1qbq 1qbr 1qbs 1qbt 1qbu 1qbv 1qca 1qf0 1qf1  
1qf2 1qft 1qhc 1qi0 1qin 1qiw 1qji 1qka 1qkb 1qpb 1qsc 1qsk 1qy1 1qy2 1qy5 1r0p  
1r0x 1rlh 1rlj 1r4w 1r6n 1r9l 1rbo 1rbp 1rd4 1rdj 1rdl 1rdt 1rej 1rek 1rgk 1rgl  
1rle 1ro6 1ro7 1rp7 1rpj 1rql 1rzx 1s63 1sb1 1sbg 1sc8 1sdt 1sdu 1sdv 1sgu 1sgx  
1sh9 1siv 1sld 1sle 1sln 1sqn 1sqo 1sqt 1srg 1sri 1ssq 1stc 1str 1sts 1sw2 1swg  
1swr 1szd 1t31 1t32 1t4v 1t5f 1t7d 1t7j 1ta2 1ta6 1tcw 1tcx 1thz 1tkb 1tlp 1tng  
1tnh 1tni 1tnj 1tnk 1tnl 1tom 1tq4 1ts1 1tuf 1tvo 1tx7 1txr 1tys 1ucn 1udt 1ugp  
1uho 1uj5 1uj6 1ulg 1umw 1uou 1upf 1ur9 1usn 1utc 1utj 1utl 1utn 1uv6 1uvt 1uw6  
1uwf 1ux7 1uxa 1uxb 1uz8 1v0k 1v11 1v1m 1v2j 1v2k 1v2l 1v2n 1v2q 1v2r 1v2s 1v2t  
1v2u 1v2w 1vj9 1vja 1vjc 1vjd 1vjj 1vkj 1vwl 1vwn 1vyf 1vyg 1vyq 1w0y 1w1d 1w1g  
1w3j 1w3l 1w4o 1w4p 1w4q 1w5v 1w5w 1w5x 1w5y 1w6y 1w7g 1w7x 1w8l 1w8m 1w96 1wc1  
1wcq 1wdn 1we2 1wei 1wht 1wkm 1wml 1ws1 1ws5 1wvc 1wvj 1x38 1x39 1x8j 1x8r 1x8t  
1xbb 1xbo 1xd0 1xff 1xgi 1xh9 1xhy 1xjd 1xk5 1xk9 1xka 1xkk 1xpz 1xq0 1xt8 1xzx  
1y2f 1y2g 1y3g 1y3n 1y3p 1y6r 1yc4 1yda 1ydb 1ydd 1ydr 1yds 1yei 1yej 1yfb 1yid  
1ykp 1yqj 1yqy 1yvm 1z1h 1z1r 1z3v 1z4n 1z4o 1z6e 1z6s 1z71 1z9g 1zdp 1zgi 1zky  
1zog 1zoh 1zp8 1zpa 1zsf 1zsr 2201 2a31 2a4m 2a5b 2a5c 2a8g 2ada 2adm 2adu 2afw  
2aj8 2am4 2amt 2ans 2aoc 2aod 2aoe 2aog 2aqu 2arm 2avm 2avo 2avq 2avs 2avv 2b4l  
2bal 2bfq 2bfr 2bjm 2bmk 2bmz 2bo4 2bpv 2bpy 2bq7 2bqv 2br1 2br6 2bsu 2bt9 2bza  
2c1p 2c3j 2c3l 2c4v 2c80 2c97 2ca8 2cbu 2cbv 2cbz 2cc7 2ccb 2ccc 2ces 2cex 2cf8  
2cf9 2cht 2csn 2dl1 2dri 2dua 2dw7 2er6 2erz 2euk 2eum 2evl 2ewa 2ewb 2ez7 2flg  
2f2h 2f34 2f35 2f6t 2f7o 2f7p 2f81 2f8g 2f8i 2fgi 2fgu 2fgv 2fkf 2fqo 2fwf 2fxv  
2fzg 2fzk 2fzz 2g00 2gh7 2gl0 2gst 2gsu 2gv6 2gv7 2gvj 2gvv 2gyi 2gzl 2h4g 2h4k  
2h4n 2h5a 2h6q 2hb3 2hf8 2hh5 2hs1 2hs2 2hu6 2hxm 2i0g 2i3h 2i3i 2i80 2ihq 2iko  
2iuz 2ivu 2izl 2j47 2j75 2j79 2j7b 2j7d 2j7e 2j7f 2j7g 2j7h 2jxr 2mas 2olb 2pcp  
2qwc 2qwf 2sim 2tmn 2toh 2tpi 2ypi 3aid 3gst 3jdw 3kiv 3pcb 3pcc 3pcf 3pcg 3pck  
3pcn 3tlh 456c 4er1 4fiv 4rsk 4std 4tmk 4ts1 5abp 5er2 5fiv 5std 5tln 5tmn 5tmp  
5yas 6abp 6cpa 6fiv 6tim 7abp 7hvp 7std 7upj 830c 8abp 966c 9abp

**Table S2**

List of files included in the SFCscore\_descriptors.zip archive. All files are comma-separated-value (csv) files with a header line. Each data row corresponds to an individual protein-ligand complex.

File name	data rows	columns	description of column no.						
			1	2	3-68	69	70	71	72
Training_set.csv	1005	71	protein file identifier	ligand file identifier	SFCscore descriptors	PDB ID	exp. pK <sub>i</sub>	cluster identifier	--
PDBbind_test_set.csv	195	71	protein file identifier	ligand file identifier	SFCscore descriptors	PDB ID	exp. pK <sub>i</sub>	cluster identifier	--
CSAR2010_test_set.csv	332	72	ligand number	protein filename	SFCscore descriptors	PDB ID	exp. pK <sub>i</sub>	ligand identifier	CSAR set ID
CSAR_chk_set.csv	36	69	ligand identifier	protein filename	SFCscore descriptors	exp. pIC <sub>50</sub>	--	--	--
CSAR_erk_set.csv	39	69	ligand identifier	protein filename	SFCscore descriptors	exp. pK <sub>i</sub>	--	--	--
CSAR_lpxc_set.csv	8	69	ligand identifier	protein filename	SFCscore descriptors	exp. pK <sub>d</sub>	--	--	--