

Supporting Information

All in One: Cavity detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening.

Viet Khoa Tran-Nguyen^{†#}, Franck Da Silva^{†#}, Guillaume Bret[†] and Didier Rognan^{†*}

[†] Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, 67400 Illkirch, France

[#] Both authors equally contributed to this work.

^{*} To whom correspondence should be addressed (phone: +33 3 68 85 42 35, fax: +3 33 68 85 43 10, email: rognan@unistra.fr)

Table S1. sc-PDB Diverse Set of 213 protein-ligand complexes.

Cluster	PDB ID	Ligand ID	Protein name	sc-PDB entries in cluster	DPI value, Å
0	10gs	VWW	Glutathione S-transferase P	18	0.350
1	1k1x	IMP	Adenylosuccinate synthetase	9	0.386
2	2r3a	SAM	Histone-lysine N-methyltransferase SUV39H2	25	0.160
3	2r3f	SC8	Cyclin-dependent kinase 2	6	0.094
4	3e5h	GNP	Ras-related protein Rab-28	171	0.062
5	13pk	ADP	Phosphoglycerate kinase, glycosomal	6	0.451
6	2fde	385	Protease	81	0.442
7	1v3s	ATP	Signaling protein	31	0.142
8	2fdp	FRP	Beta-secretase 1	66	0.398
10	1v45	3DG	Purine nucleoside phosphorylase	11	0.408
12	3orf	NAD	Dihydropteridine reductase	58	0.208
14	3orn	3OR	Dual specificity mitogen-activated protein kinase kinase 1	11	0.473
15	3oro	AGS	Serine/threonine protein kinase	65	0.353
16	2r4b	GW7	Receptor tyrosine-protein kinase erbB-4	9	0.558
18	3e65	XXZ	Nitric oxide synthase, inducible	8	0.171
20	3orz	BI4	3-phosphoinositide-dependent protein kinase 1	10	0.255
21	1klk	PMD	Dihydrofolate reductase	9	n.a.
22	1a28	STR	Progesterone receptor	19	0.155
24	2r4f	RIE	3-hydroxy-3-methylglutaryl-coenzyme A reductase	17	0.118
28	1a2n	TET	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	9	0.266
32	1v79	FR7	Adenosine deaminase	6	0.756
34	2r4t	ADP	Glycogen synthase	13	0.117
36	2feq	34P	Prothrombin	76	0.519
37	2r59	PHO	Leukotriene A-4 hydrolase	7	0.136
38	1a42	BZU	Carbonic anhydrase 2	9	n.a.
39	2r5c	C6P	Kynurenine aminotransferase	23	0.224
42	3otf	CMP	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4	18	0.509
45	1knr	FAD	L-aspartate oxidase	125	0.392
46	1knu	YPA	Peroxisome proliferator-activated receptor gamma	7	0.470
52	3ou2	SAH	SAM-dependent methyltransferase	174	0.077
53	1kol	NAD	Glutathione-independent formaldehyde dehydrogenase	75	0.096
54	2r6h	FAD	NADH:ubiquinone oxidoreductase, Na translocating, F subunit	32	0.435
55	2r6j	NDP	Eugenol synthase 1	6	0.098
58	1kor	ANP	Argininosuccinate synthase	21	0.125
59	2r6w	LLB	Estrogen receptor	17	0.219
68	2r7m	AMP	5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate synthetase	28	0.213
70	1a4z	NAD	Aldehyde dehydrogenase, mitochondrial	31	0.529
73	1v9n	NDP	Malate dehydrogenase	10	0.206
79	3ow3	SMY	cAMP-dependent protein kinase catalytic subunit alpha	8	0.135
80	3e7x	AMP	D-alanine--poly(phosphoribitol) ligase subunit 1	14	0.315
83	3e87	G95	RAC-beta serine/threonine-protein kinase	8	0.427
85	1vbm	YSA	Tyrosine--tRNA ligase	9	0.418
87	1vc2	NAD	Glyceraldehyde 3-phosphate dehydrogenase	52	0.877
88	3owa	FAD	Acyl-CoA dehydrogenase	22	0.132
89	3owb	BSM	Heat shock protein HSP 90-alpha	21	0.140
90	1vcf	FMN	Isopenentenyl-diphosphate delta-isomerase	8	0.507
92	2r8o	T5X	Transketolase 1	32	0.059
96	1kp8	ATP	60 kDa chaperonin	21	0.225
100	3e8x	NAP	BH1520 protein	17	0.115
102	1kpg	SAH	Cyclopropane mycolic acid synthase 1	33	0.239
104	1a80	NDP	2,5-diketo-D-gluconic acid reductase A	30	0.328
105	3e92	G6A	Mitogen-activated protein kinase 14	21	0.152
107	1vdc	FAD	Thioredoxin reductase 1	28	0.213
111	3e9h	KAA	Lysine--tRNA ligase	10	0.278
116	4c4f	7CE	Dual specificity protein kinase TTK	8	0.179
118	3ox4	NAD	Alcohol dehydrogenase 2	7	0.252
119	1kqb	FMN	Oxygen-insensitive NAD(P)H nitroreductase	18	n.a.
120	2r97	FMN	NAD(P)H dehydrogenase (quinone)	12	0.243
122	4c58	824	Cyclin-G-associated kinase	7	0.201
125	4c5o	FAD	Putative monooxygenase	30	0.818
126	4c61	LMM	Tyrosine-protein kinase JAK2	7	0.228
128	3oy1	589	Mitogen-activated protein kinase 10	7	0.198
130	3oy3	XY3	Tyrosine-protein kinase ABL1	28	0.196
133	1kqm	ANP	Myosin heavy chain, striated muscle	10	0.581
135	2r9r	NAP	Voltage-gated potassium channel subunit beta-2	10	0.239
136	1kqn	NAD	Nicotinamide mononucleotide adenylyltransferase 1	6	0.199
151	4c8g	C5P	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	9	0.190
156	2fky	N2T	Kinesin-like protein KIF11	17	0.319
162	4ca6	3EF	Angiotensin-converting enzyme	21	0.529
173	3p0n	BPU	Tankyrase-2	12	0.121
180	1vhn	FMN	tRNA-dihydrouridine synthase	28	0.081
181	2rd2	QSI	Glutamine--tRNA ligase	6	0.376
183	3ebh	BES	M1 family aminopeptidase	12	0.238
184	1vhw	ADN	Purine nucleoside phosphorylase DeoD-type 1	18	0.077
185	3p19	NAP	Putative blue fluorescent protein	91	0.177
189	4ccb	OFG	ALK tyrosine kinase receptor	8	0.203
192	3p23	ADP	Serine/threonine-protein kinase	39	0.400
195	2foi	JPA	Enoyl-acyl carrier reductase	22	0.475
199	1adc	PAD	Alcohol dehydrogenase E chain	17	n.a.

200	4cdg	ADP	Bloom syndrome protein	16	0.359
202	4cdq	7VR	Polyprotein	6	0.063
203	3p3c	3P3	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	8	0.035
212	2ftf	ILB	Dihydroorotate dehydrogenase (quinone), mitochondrial	14	0.226
223	3eei	MTM	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	8	0.125
225	3eej	53R	Strain CBS138 chromosome J complete sequence	13	0.254
228	3p5s	AVU	CD38 molecule	7	0.186
232	3efq	714	Farnesyl pyrophosphate synthase	6	0.215
234	2rh1	CAU	Beta-2 adrenergic receptor	6	0.273
238	3p7n	FMN	Sensor histidine kinase	8	0.316
241	1aj2	2PH	Dihydropteroate synthase	6	n.a.
242	2fsn	ADP	Archaeal actin homolog	15	0.550
246	2fsv	NAP	NAD(P) transhydrogenase subunit beta	9	0.221
247	1kyi	ATP	ATP-dependent protease ATPase subunit HslU	6	0.472
249	3p88	P88	Bile acid receptor	8	0.447
252	2fto	TMP	Thymidylate synthase	10	0.170
255	3p8x	ZYD	Vitamin D3 receptor	66	0.099
256	1akw	FMN	Flavodoxin	9	0.043
258	3ehg	ATP	Sensor histidine kinase DesK	26	0.120
259	3p8z	36A	RNA-directed RNA polymerase NS5	12	0.114
261	1kyx	CRM	6,7-dimethyl-8-ribityllumazine synthase	20	0.358
263	1am1	ADP	ATP-dependent molecular chaperone HSP82	14	0.148
264	3ehx	BDL	Macrophage metalloelastase	13	0.165
267	2fv9	002	Disintegrin and metalloproteinase domain-containing protein 17	6	0.248
269	2fvc	888	Genome polyprotein	12	0.207
274	4d86	ADP	Poly [ADP-ribose] polymerase 14	6	0.159
276	2rkg	AB1	Pol protein	11	0.144
278	3p9j	P9J	Aurora kinase A	11	0.961
279	2rku	R78	Serine/threonine-protein kinase PLK1	13	0.142
281	2rl5	2RL	Vascular endothelial growth factor receptor 2	13	0.342
295	4d9t	OJG	Ribosomal protein S6 kinase alpha-3	7	0.273
296	4d9w	X32	Thermolysin	13	0.043
303	1l2t	ATP	Uncharacterized ABC transporter ATP-binding protein MJ0796	12	0.156
307	1l4e	RBZ	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	8	0.187
314	2g1n	1IG	Renin	8	n.a.
319	2udp	UPP	UDP-glucose 4-epimerase	11	0.129
322	1aqb	RTL	Retinol-binding protein 4	6	0.106
323	3elj	GS7	Mitogen-activated protein kinase 8	10	0.121
325	3elm	24F	Collagenase 3	6	0.145
333	4dc3	2FA	Putative adenosine kinase	14	0.242
336	3en4	KS1	Proto-oncogene tyrosine-protein kinase Src	22	0.407
351	2uu0	LK3	UDP-N-acetylmuramoylalanine--D-glutamate ligase	8	0.428
354	1aux	AGS	Synapsin-1	7	0.374
355	3eos	PK2	Queuine tRNA-ribosyltransferase	12	0.109
356	3epp	SFG	mRNA cap guanine-N7 methyltransferase	14	0.445
360	3ept	FDA	Putative FAD-monooxygenase	10	0.519
364	4dfp	OL7	DNA polymerase I, thermostable	9	0.158
369	4dgm	AGI	Casein kinase II subunit alpha	8	0.143
378	1b0h	LYS_LYS_ALN	Periplasmic oligopeptide-binding protein	10	0.140
379	3pd3	A3T	Threonine--tRNA ligase	6	0.183
383	3eqp	T95	Activated CDC42 kinase 1	7	0.482
388	1b0p	TPP	Pyruvate-flavodoxin oxidoreductase	7	0.372
389	3erk	SB4	Mitogen-activated protein kinase 1	7	0.239
400	1vrw	NAD	Enoyl-ACP reductase	15	0.297
401	4dk5	OKO	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform	7	0.476
407	1vso	AT1	Glutamate receptor ionotropic, kainate 1	9	0.125
411	4dko	OLM	Envelope glycoprotein gp160	6	0.163
412	1b3d	S27	Stromelysin-1	16	0.643
419	3peh	IBD	Endoplasmic homolog, putative	7	0.738
427	2ga2	A19	Methionine aminopeptidase 2	6	0.209
431	2uyy	NA7	Putative oxidoreductase GLYR1	13	0.277
432	4dlk	ATP	Phosphoribosylaminoimidazole carboxylase, ATPase subunit	15	0.199
434	1vtk	TMP	Thymidine kinase	9	0.553
444	1lhn	AON	Sex hormone-binding globulin	8	n.a.
453	1lik	ADN	Adenosine kinase	6	0.241
454	2v0i	UD1	Bifunctional protein GImU	10	0.105
459	1b9i	PXG	Putative UDP-kanosamine synthase aminotransferase subunit	7	0.155
475	3pjg	UGA	UDP-glucose 6-dehydrogenase	6	0.265
478	2v1u	ADP	ORC1-type DNA replication protein 1	11	0.521
488	3ewr	APR	Non-structural protein 3	6	0.219
501	4dqw	ATP	Inosine-5'-monophosphate dehydrogenase	13	0.318
502	3plq	RP2	cAMP-dependent protein kinase type I-alpha regulatory subunit	6	0.390
503	4dr9	BB2	Peptide deformylase	16	0.148
511	1w05	W05	Isopenicillin N synthase	20	0.199
512	4drx	GTP	Tubulin alpha chain	10	0.170
521	3exh	TPP	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	19	0.319
532	1bif	AGS	6-phosphofructo-2-kinase	13	0.201
535	1bjy	CTC	Tetracycline repressor protein class D	7	0.368
538	1lvg	SGP	Guanylate kinase	7	0.184
540	2glx	NDP	1,5-anhydro-D-fructose reductase	7	0.223
551	3eyg	MI1	Tyrosine-protein kinase JAK1	16	0.142
567	1boo	SAH	Modification methylase Pvull	17	0.334

586	2gqt	FAD	UDP-N-acetylenolpyruvoylglucosamine reductase	14	0.078
598	2v6g	NAP	3-oxo-Delta(4,5)-steroid 5-beta-reductase	6	0.179
609	3ptq	NFG	OSIGBa0135C13.7 protein	8	0.438
616	3f3y	4OA	Bile salt sulfotransferase	6	0.355
621	2gtb	AZP	Orf1ab polyprotein	6	0.229
627	2v95	HCY	Corticosteroid-binding globulin	6	0.166
632	4dya	OMF	Nucleocapsid protein	6	1.304
640	1w7k	ADP	Dihydrofolate synthase	10	0.185
649	4e0i	FAD	Mitochondrial FAD-linked sulfhydryl oxidase ERV1	10	0.503
650	3f82	353	Hepatocyte growth factor receptor	8	0.475
665	1c1c	612	Reverse transcriptase/ribonuclease H	6	0.910
677	3pzb	NAP	Aspartate-semialdehyde dehydrogenase	9	0.157
685	1c30	ADP	Carbamoyl-phosphate synthase large chain	6	0.141
689	3fbu	COA	Acetyltransferase, GNAT family	7	0.144
704	3q0u	LL3	HTH-type transcriptional regulator EthR	6	0.100
759	1cbf	SAH	Cobalt-precorrin-4 C(11)-methyltransferase	6	0.224
765	2ha8	SAH	Probable methyltransferase TARBP1	11	0.088
768	2vfz	UPF	N-acetyllactosaminide alpha-1,3-galactosyltransferase	9	0.583
773	4e7z	ADP	Unconventional myosin-VI	10	0.258
785	1wkg	POI	Acetylorlithine/acetyl-lysine aminotransferase	11	0.312
801	4eaw	ONQ	RNA-directed RNA polymerase	9	0.277
802	4eb3	OO3	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	7	0.141
826	3flk	NAD	D-malate dehydrogenase [decarboxylating]	8	0.168
885	2vna	NAP	Prostaglandin reductase 2	6	0.179
889	3qcf	NXY	Receptor-type tyrosine-protein phosphatase gamma	7	0.461
903	1mp3	TTP	Glucose-1-phosphate thymidyltransferase	6	n.a.
969	3fy4	FAD	(6-4)DNA photolyase	11	0.858
978	2hsd	NAD	3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase	6	n.a.
984	3qgz	ADN	Histidine triad nucleotide-binding protein 1	7	0.025
1031	2vww	7X2	Ephrin type-B receptor 4	7	0.178
1061	3g5e	Q74	Aldose reductase	7	0.103
1073	3qov	ADP	Phenylacetate-coenzyme A ligase	8	0.180
1099	2w0j	ZAT	Serine/threonine-protein kinase Chk2	7	0.165
1163	1xoi	288	Glycogen phosphorylase, liver form	6	0.214
1202	3r04	UNQ	Serine/threonine-protein kinase pim-1	7	0.099
1260	3gjq	TRP_GLU_HIS_	Caspase-3	7	n.a.
1265	2we3	DUT	Deoxyuridine 5'-triphosphate nucleotidohydrolase	6	0.211
1271	1xwk	GDN	Glutathione S-transferase Mu 1	7	0.682
1310	4fhh	OU3	Vitamin D3 receptor A	9	0.282
1418	2wqo	VGK	Serine/threonine-protein kinase Nek2	6	0.204
1440	3rll	RLL	Androgen receptor	10	0.133
1453	4fsm	HK1	Serine/threonine-protein kinase Chk1	12	0.178
1505	1o6h	W37	Squalene--hopene cyclase	7	0.374
1717	4gfd	OYB	Thymidylate kinase	6	0.120
1719	4gfn	SUY	DNA gyrase subunit B	10	0.136
1801	4gpj	OQ1	Bromodomain-containing protein 4	6	0.116
1845	4gv2	5ME	Poly [ADP-ribose] polymerase 3	19	1.312
2170	3iub	FG2	Pantothenate synthetase	6	0.071
2615	4jd4	JDM	Dihydroorotate dehydrogenase (fumarate)	10	0.059
2716	1sqb	AZO	Cytochrome b	6	0.549
2898	4kfn	1QR	Nicotinamide phosphoribosyltransferase	8	0.109
3197	3zcm	PX3	Integrase	13	0.088

n.a.: not available. The Diffraction precision index (DPI) cannot be calculated due to insufficient parameters

The Diffraction Precision Index (DPI) is calculated according to Kumar et al. Journal of Applied Crystallography 48.3 (2015): 939-942.

Table S2. AstexDiverseSet of 85 protein-ligand complexes

PDB ID	Ligand ID	Protein Name	DPI, Å
1g9v	RQ3	Hemoglobin alpha chain	0.146
1gkc	NFH	92 kda type iv collagenase	0.316
1gm8	SOX	Penicillin g acylase beta subunit	0.181
1gpk	HUP	Acetylcholinesterase	0.140
1hnn	SKF	Phenylethanolamine n-methyltransferase	0.292
1hp0	AD3	Inosine-adenosine-guanosine-preferring nucleoside hydrolase	0.273
1hq2	PH2	6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase	0.047
1hvy	D16	Thymidylate synthase	0.212
1hwi	115	Hmg-coa reductase	0.291
1hww	SWA	Alpha-mannosidase ii	0.150
1ia1	TQ3	Dihydrofolate reductase	0.117
1ig3	VIB	Thiamin pyrophosphokinase	0.144
1j3j	CP6	Bifunctional dihydrofolate reductase-thymidylate synthase	0.318
1jd0	AZM	Carbonic anhydrase xii	0.077
1jje	BYS	Imp-1 metallo beta-lactamase	0.158
1jla	TNK	Hiv-1 rt a-chain	0.502
1k3u	IAD	Tryptophan synthase alpha chain	0.088
1ke5	LS1	Cell division protein kinase 2	0.338
1kzk	JE2	Protease	0.029
1l2s	STC	Beta-lactamase	0.153
1l7f	BCZ	Neuraminidase	0.099
1lpz	CMB	Blood coagulation factor xa	0.497
1lrh	NLA	Auxin-binding protein 1	0.173
1m2z	DEX	Glucocorticoid receptor	0.936
1meh	MOA	Inosine-5'-monophosphate dehydrogenase	0.142
1mmv	3AR	Nitric-oxide synthase, brain	0.212
1mzc	BNE	Protein farnesyltransferase beta subunit	0.133
1n1m	A3M	Dipeptidyl peptidase iv soluble form	0.812
1n2j	PAF	Pantothenate synthetase	0.137
1n2v	BDI	Queuine trna-ribosyltransferase	0.215
1n46	PFA	Thyroid hormone receptor beta-1	0.329
1nav	IH5	Hormone receptor alpha 1, thra1	0.297
1of1	SCT	Thymidine kinase	0.144
1of6	DTY	Phospho-2-dehydro-3-deoxyheptonate aldolase, tyrosine-inhibited	0.238
1opk	P16	Proto-oncogene tyrosine-protein kinase abl1	0.122
1oq5	CEL	Carbonic anhydrase ii	0.085
1owe	675	Urokinase-type plasminogen activator	0.133
1oyt	FSN	Thrombin heavy chain	0.094
1p2y	NCT	Cytochrome p450-cam	0.369
1p62	GEO	Deoxycytidine kinase	0.129
1pmn	984	Mitogen-activated protein kinase 10	0.291
1q1g	MTI	Uridine phosphorylase putative	0.191
1q41	IXM	Glycogen synthase kinase-3 beta	0.156
1q4g	BFL	Prostaglandin g/h synthase 1	0.139
1r1h	BIR	Neprilysin	0.195
1r55	097	Adam 33	0.112
1r58	AO5	Methionine aminopeptidase 2	0.198

1r9o	FLP	Cytochrome p450 2c9	0.172
1s19	MC9	Vitamin d3 receptor	0.181
1s3v	TQD	Dihydrofolate reductase	0.156
1sg0	STL	Nrh dehydrogenase [quinone] 2	0.086
1sj0	E4D	Estrogen receptor	0.206
1sq5	PAU	Pantothenate kinase	0.283
1sqn	NDR	Progesterone receptor	0.082
1t40	ID5	Aldose reductase	0.118
1t46	STI	v-kit hardy-zuckerman 4 feline sarcoma viral oncogene homolog	0.083
1t9b	1CS	Acetolactate synthase, mitochondrial	0.157
1tow	CRZ	Fatty acid-binding protein, adipocyte	0.249
1tt1	KAI	Glutamate receptor, ionotropic kainate 2	0.161
1tz8	DES	Transthyretin	0.125
1u1c	BAU	Uridine phosphorylase	0.334
1u4d	DBQ	Activated cdc42 kinase 1	0.239
1uml	FR4	Adenosine deaminase	0.896
1unl	RRC	Cyclin-dependent kinase 5	0.250
1uou	CMU	Thymidine phosphorylase	0.604
1v0p	PVB	Cell division control protein 2 homolog	0.193
1v48	HA1	Purine nucleoside phosphorylase	0.282
1v4s	MRK	Glucokinase isoform 2	0.298
1vcj	IBA	Neuraminidase	0.406
1w1p	GIO	Chitinase b	0.275
1w2g	THM	Thymidylate kinase tmk	0.199
1x8x	SO4	Tyrosyl-trna synthetase	0.170
1xm6	5RM	Camp-specific 3',5'-cyclic phosphodiesterase 4b	0.139
1xoq	ROF	Camp-specific 3',5'-cyclic phosphodiesterase 4d	0.122
1xoz	CIA	Cgmp-specific 3',5'-cyclic phosphodiesterase	0.063
1y6b	AAX	Vascular endothelial growth factor receptor 2	0.188
1ygc	905	Coagulation factor vii	0.126
1yqy	915	Lethal factor	0.323
1yv3	BIT	Myosin ii heavy chain	0.138
1yvf	PH7	Hcv ns5b polymerase	0.390
1ywr	LI9	Mitogen-activated protein kinase 14	0.185
1z95	198	Androgen receptor	0.147
2bm2	PM2	Human beta2 tryptase	0.297
2br1	PFP	Serine/threonine-protein kinase chk1	0.143
2bsm	BSM	Heat shock protein hsp90-alpha	0.173

The Diffraction Precision Index (DPI) is calculated according to Kumar et al.
Journal of Applied Crystallography 48.3 (2015): 939-942

Table S3. 10 DUD-E entries

Protein Name	PDB ID	Ligand ID	Pharmacophoric features	Actives	Decoys
<i>G Protein-Coupled receptors</i>					
Adenosine A2A receptor (AA2AR)	3pwh	ZMA	35	482	31,500
Beta2 adrenergic receptor (ADRB2)	3ny8	JRZ	42	231	15,000
<i>Nuclear hormone receptors</i>					
Androgen receptor (ANDR)	2am9	TES	33	269	14,350
Glucocorticoid receptor (GCR)	1p93	DEX	36	258	15,000
<i>Other enzymes</i>					
Adenosine deaminase (ADA)	1a41	DCF	19	282	16,900
Prostaglandin G/H synthase 2 (PGH2)	3ln1	CEL	44	104	6,958
<i>Proteases</i>					
Angiotensin-converting enzyme (ACE)	3zqz	SLC	28	139	8,700
Renin (RENI)	3sfc	S53	43	293	16,450
<i>Protein kinases</i>					
Fibroblast growth factor receptor 1 (FGFR1)	3tt0	07J	23	93	5,450
RAC-alpha protein kinase (AKT1)	4ekl	ORF	41	435	123,150

Table S4. Shaper2 forcefield for aligning ligand atoms to cavity features

```
#####  
#      DEFINE      #  
#####  
##### define degree (independent of explicit/implicit)  
DEFINE hd1 [X1H0,X2H1,X3H2,X4H3,X5H4,X6H5]  
DEFINE hd2 [X2H0,X3H1,X4H2,X5H3,X6H4]  
DEFINE hd3 [X3H0,X4H1,X5H2,X6H3]  
DEFINE hd4 [X4H0,X5H1,X6H2]  
##### hydrophobic  
DEFINE php [#6,#16&$hd2&!$(S=*)#35,#53;RO;!$(~[!#1;!#6;!$(#16,$hd2)])]  
DEFINE thp [$php;$hd1]  
DEFINE hp [$php;!$hd1]  
DEFINE ehp [$hp;!$(#hp)]  
##### acceptors  
DEFINE ACamine [N;!$(N*=[!#6]);!$(N~[!#6;!#1]);!$(Na);!$(N#*);!$(N=*)]  
DEFINE ACphosphate [O;$hd1;$O~P(~O)~O]  
DEFINE ACcarboxylate [O;$hd1;$O[C;!$(N)=O),$O=[C;!$(N)]O;$hd1]]  
DEFINE ACwater [OH2]  
DEFINE AChet6N [nH0;X2;$n1aaaa1]  
DEFINE ACphosphinyl [O;$O=P);!$(O=P~O)]  
DEFINE ACSulphoxide [O;$O=[S;!$(S(~O)~O)];$(#6)[#6]]]  
#DEFINE ACprimaryAmine [$ACamine;$hd1;!X4] #leave off for implicit charge  
DEFINE AChet5N [nH0;X2;$n1aaaa1]  
DEFINE ACthiocarbonyl [S;X1;$S=[#6]]  
DEFINE AChydroxyl [O;$hd1;$O-[C;!$(C=*)]]  
DEFINE ACSulphate [O;$hd1;$O~S(~O)~O]  
#DEFINE ACtertiaryAmine [$ACamine;$hd3] #leave off for implicit charge  
DEFINE ACamide [O;$O=[#6][#7];!$(O=[#6][#7][#7,#8,#16]]]  
DEFINE ACcarbamate [O;$O=[#6][#7][#8]]  
DEFINE ACurea [O;$O=[#6][#7][#7]]  
DEFINE ACester [O;$O=[#6][#8]*;!$(O=[#6][#7,#8,#16][#8]*)]  
DEFINE ACnitrile [N;$hd1;$N#C]  
DEFINE ACimine [N;!$hd3;$N=(C)C,$N=[#6];!$(N=[#6][#7,#8;!$(S=O)])]  
DEFINE ACKetone [O;$hd1;$O=[#6;$([H2]),$([H1]-[#6]),$(#6)[#6]]]  
#DEFINE ACsecondaryAmine [$ACamine;$hd2;!X4] #leave off for implicit charge  
DEFINE ACphenol [O;$hd1;$Oa]  
DEFINE ACether [O;$([#6;!$(=[O,S,N])][#6;!$(=[O,S,N])]]]  
DEFINE ACprimaryAniline [N;$Na);$hd1]  
DEFINE ACnitro [O;$hd1;$O~N~O;$hd1]]  
DEFINE AChet5O [o;X2;$o1cccc1,$o1cccc1);!$(#[6]=O)]  
DEFINE ACSulphone [O;$O=[S;$S(~O)(~O)(#6,#7)[#6]]]  
### strong acceptors  
DEFINE strongAcceptor [$ACphosphate,$ACcarboxylate,$ACwater,$AChet6N,$ACphosphinyl]  
### moderate acceptors  
#DEFINE moderateAcceptor  
[$ACSulphoxide,$ACprimaryAmine,$AChet5N,$ACthiocarbonyl,$AChydroxyl,$ACSulphate,$ACtertiary  
Amine,$ACamide,$ACcarbamate,$ACurea]  
DEFINE moderateAcceptor  
[$ACSulphoxide,$AChet5N,$ACthiocarbonyl,$AChydroxyl,$ACSulphate,$ACamide,$ACcarbamate,$AC  
urea]  
### weak acceptors
```



```

#DEFINE weakAcceptor
[$ACnitrile,$ACimine,$ACketone,$ACsecondaryAmine,$ACester,$ACphenol,$ACether,$ACprimaryAnil
ine,$ACnitro,$ACHet5O,$ACsulphone]
DEFINE weakAcceptor
[$ACnitrile,$ACimine,$ACketone,$ACester,$ACphenol,$ACether,$ACprimaryAniline,$ACnitro,$ACHet
5O,$ACsulphone]
##### donors
DEFINE Damine [N;!$(N*=[!#6]);!$(N~[!#6;!#1]);!$(Na);!$(N#*);!$(N=*)]
DEFINE Dhet5NH [nH;$n1aaaa1),$n1aaaaa1]
DEFINE DNpH [NH,H2,H3;+]
#DEFINE DacidOH [OH1;$O-[C,S,P]=[O,S]] #leave off for implicit protonation
DEFINE Dhydroxyl [OH1;$hd1;$O-C;!$(OC=[O,N,S])]
DEFINE Dwater [OH2]
DEFINE DprimaryAmide [N;$hd1;$NC=O),$NS=O]
DEFINE DanilineNH [NH1,NH2;$hd2;$Nc;!$(NS(=O)=O)]
DEFINE DamidineNH [NH1,NH2;$N~C~N),$N~C(~N)~N]
DEFINE DsecondaryAmide [#7;$hd2;$* [#6,#16]=O;!$(N(a)S=O)]
DEFINE DanilineNH2 [N;$hd1;$Nc]
DEFINE DhydraN [NH1,NH2,NH3;$hd1&$NN[#6]),$hd2&$N(N)[#6]]
DEFINE DimineNH [NH1;$N=C]
DEFINE DphenylOH [OH1;$Oc]
DEFINE DprimaryAmine [$Damine;$hd1]
DEFINE DsecondaryAmine [$Damine;$hd2]
### strong donors
#DEFINE strongDonor [$Dhet5NH,$DNpH,$DacidOH,$Dhydroxyl]
DEFINE strongDonor [$Dhet5NH,$DNpH,$Dhydroxyl]
### moderate donors
DEFINE moderateDonor
[$Dwater,$DprimaryAmide,$DanilineNH,$DamidineNH,$DsecondaryAmide,$DanilineNH2]
### weak donors
DEFINE weakDonor [$DhydraN,$DimineNH,$DphenylOH,$DprimaryAmine,$DsecondaryAmine]
##### anion intermediate
DEFINE negHet [#8,#16;$hd1]
DEFINE terminalHet [#7,#8,#16;$hd1]
DEFINE ANarylsulfonamide [N;$N(a)S(=O)(=O)*]
DEFINE ANmalonic [C;!$hd4;$C(C=[O,S])C=[O,S]]
DEFINE ANarylthiol [S;$hd1;$Sa]
DEFINE ANhalideion [I,Br,Cl,F;!H0,-]
DEFINE ANhydroxylamine [O;$hd1;$ON~C),$O[n+]),$O=n);!$(ONC=[S,O,N])]
##### cation intermediate
DEFINE CATnonewN [#7;!$(NC=O);!$(NS(=O)=O)]
DEFINE CATguanidine [$CATnonewN]!:[#6](!:[$CATnonewN])!:[$CATnonewN]
DEFINE CATguanidineC [#6]~[$CATguanidine]
DEFINE CATamine [N;!$(N*=[!#6]);!$(N~[!#6;!#1]);!$(Na);!$(N=*);!$(N#*);!$([#7;X0])]
##### Zn intermediates
DEFINE hydroxamate O=[CX3]N[O-]
DEFINE reverseHydrox O=[CH][NX3][O-]
#
#
#####
#          TYPES          #
#####

```

```

TYPE donor
TYPE acceptor
TYPE cation
TYPE anion
TYPE rings
TYPE hydrophobe
TYPE metal
TYPE donac
#
#
#####
#      PATTERNS      #
#####
##### rings
PATTERN rings [R]~1~[R]~[R]~[R]1
PATTERN rings [R]~1~[R]~[R]~[R]~[R]1
PATTERN rings [R]~1~[R]~[R]~[R]~[R]~[R]1
PATTERN rings [R]~1~[R]~[R]~[R]~[R]~[R]~[R]1
### hydrophobic
# terminal hp
PATTERN hydrophobe [$thp]~*(~[$thp])~[$thp]          #triple
PATTERN hydrophobe [$thp][!$(*(~[$thp])(~[$thp])~[$thp]);!$(*=[N,S,O])[$thp] #double
PATTERN hydrophobe [$thp;!$(*~*~[$thp]);$(*~[$php])]    #single
PATTERN hydrophobe [$thp;#35,#53]          #large
# non-terminal hp
PATTERN hydrophobe [$ehp][$hp][$hp][$ehp]
PATTERN hydrophobe [$ehp]([ehp])[$hp][$ehp]
PATTERN hydrophobe [$hp]([ehp])([ehp])[$ehp]
PATTERN hydrophobe [$ehp][$hp][$hp][$hp][$ehp]
PATTERN hydrophobe [$ehp][$hp][$hp;$(*[$hp][$hp][$ehp])]
### donor/acceptor patterns
PATTERN acceptor [$strongAcceptor,$moderateAcceptor,$weakAcceptor]
PATTERN donor [$strongDonor,$moderateDonor,$weakDonor]
### anion/cation patterns
# cations
PATTERN cation [$CATnonewN]!:[#6;!$(C(N)(N)N)](!:[$CATnonewN])!:[$CATnonewN] #guanidine
PATTERN cation [$CATnonewN]!:[#6;!$([$CATguanidineC]);!$(C(N)N)](!:[$CATnonewN] #amidine
PATTERN cation n:1cncc1          #azole
PATTERN cation [$CATamine]
# anions
PATTERN anion [$negHet][#6X3]~[$terminalHet]          #carboxylate
PATTERN anion [$negHet][#16X4](~[$terminalHet])~[$terminalHet] #sulfonate
PATTERN anion [$negHet][#15X4](=O)[$negHet,$terminalHet] #phosphonate
PATTERN anion [n;$hd2]1[n;$hd2][n;$hd2][n;$hd2]c1    #tetrazole
PATTERN anion [$ANarylsulfonamide,$ANmalonic,$ANarylthiol,$ANhalideion,$ANhydroxylamine]
#####
#      Type Patterns      #
#####
##### metal binders #####
#### Pattern Ca_Mg & ZN
PATTERN metal [#8;-]
PATTERN metal [#16;-]

```

```
#### Pattern Ca_Mg
# PATTERN metal [nh0]1aaaa1
##### Pattern Zn
PATTERN metal [nh0;-]
PATTERN metal [#7;-]
PATTERN metal [#8;-;!$([$hydroxamate,$reverseHydrox])]
PATTERN metal O=[CX3]N[O-]
PATTERN metal O=[CH][NX3][O-]
PATTERN metal [S]([N,-1])(=[O])(=[O])
PATTERN metal O=C[ND2][O-]
#####
#   CAVITY Type Patterns                               #
#####
PATTERN donor [14#7]
PATTERN acceptor [14#8]
PATTERN cation [15#7]
PATTERN anion [17#8]
PATTERN rings [15#6]
PATTERN hydrophobe [13#6]
#PATTERN donac [15#8]
PATTERN metal [54#30]
#
#
#####
#           INTERACTIONS                               #
#####
INTERACTION rings rings attractive gaussian weight=1.0 radius=1.0
INTERACTION hydrophobe hydrophobe attractive gaussian weight=1.0 radius=1.0
INTERACTION donor donor attractive gaussian weight=1.0 radius=1.0
INTERACTION donac donac attractive gaussian weight=1.0 radius=1.0
INTERACTION acceptor acceptor attractive gaussian weight=1.0 radius=1.0
INTERACTION cation cation attractive gaussian weight=1.0 radius=1.0
INTERACTION anion anion attractive gaussian weight=1.0 radius=1.0
INTERACTION metal metal attractive gaussian weight=10.0 radius=1.0
#
INTERACTION rings hydrophobe attractive gaussian weight=1.0 radius=1.0
INTERACTION donor cation attractive gaussian weight=1.0 radius=1.0
INTERACTION acceptor anion attractive gaussian weight=1.0 radius=1.0
INTERACTION donac donor attractive gaussian weight=1.0 radius=1.0
INTERACTION donac acceptor attractive gaussian weight=1.0 radius=1.0
INTERACTION metal anion attractive gaussian weight=10.0 radius=1.0
INTERACTION metal acceptor attractive gaussian weight=10.0 radius=1.0
INTERACTION metal donac attractive gaussian weight=10.0 radius=1.0
#
INTERACTION anion cation repulsive gaussian weight=1.0 radius=1.0
INTERACTION metal hydrophobe repulsive gaussian weight=10.0 radius=1.0
INTERACTION metal rings repulsive gaussian weight=10.0 radius=1.0
INTERACTION metal cation repulsive gaussian weight=10.0 radius=1.0
INTERACTION metal donor repulsive gaussian weight=10.0 radius=1.0
#
INTERACTION hydrophobe donor repulsive gaussian weight=1.0 radius=1.0
INTERACTION hydrophobe acceptor repulsive gaussian weight=1.0 radius=1.0
```

INTERACTION hydrophobe donac repulsive gaussian weight=1.0 radius=1.0
INTERACTION hydrophobe cation repulsive gaussian weight=1.0 radius=1.0
INTERACTION hydrophobe anion repulsive gaussian weight=1.0 radius=1.0