Supplementary Information:

Binding Site Characterization – Similarity, Promiscuity, and Druggability

Christiane Ehrt, Tobias Brinkjost, Oliver Koch

MedChemComm, 2019, DOI: 10.1039/C9MD00102F

Contents:

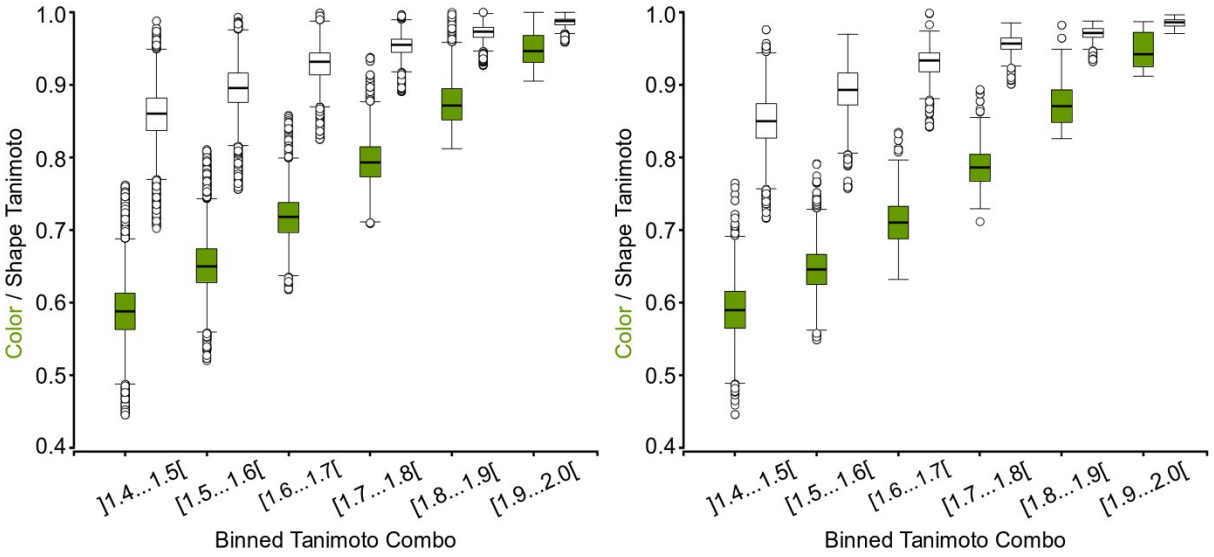


Fig. S1 Boxplots of the Shape Tanimoto (white) and Color Tanimoto (green) for the binned Tanimoto Combo similarities of the ROCS[1]-based sc-PDB ligand comparisons. The left panel shows the similarities for all similarities (similar ligands), while the right panel presents the similarities for the subset of similar ligands binding to unrelated proteins.

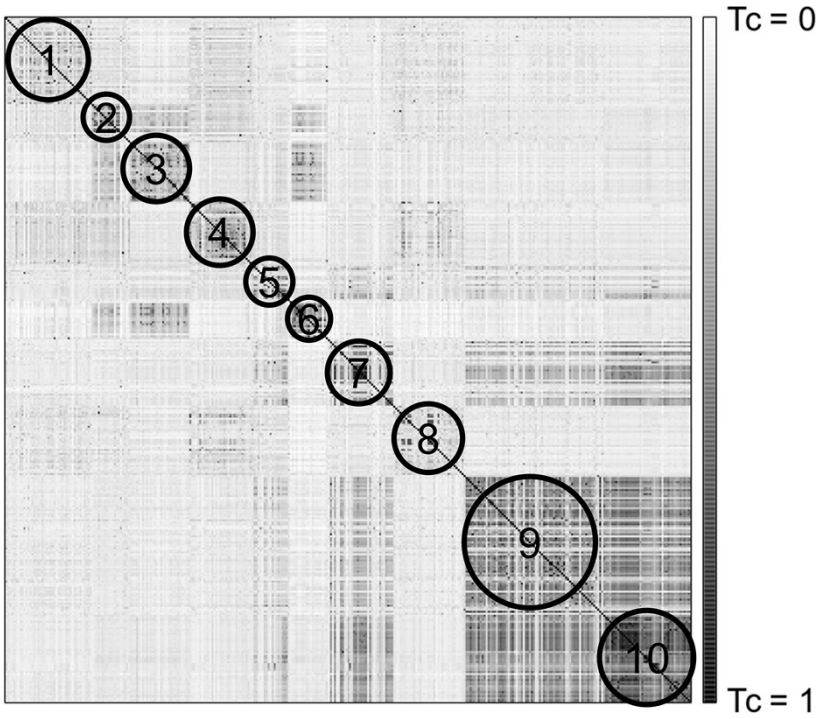


Fig. S2 Similarity matrix of clusters of sc-PDB ligands with high shape and physicochemical similarities. The coloring corresponds to the molecular 2D similarity based on the Tanimoto coefficients (Tc) between the ECFP4 fingerprints (black: Tc = 1, white: Tc = 0).

Overrepresented Target Classes		Mean # Targets (PDB)		Mean # Targets (ChEMBL)	
cluster 1	Enzymes, Proteases	cluster 10	189.93	cluster 4	18.83
cluster 2	Nuclear Receptors, Transcription Factors	cluster 7	151.42	cluster 3	11.77
cluster 3	Nuclear Receptors, Transcription Factors	cluster 9	21.54	cluster 1	8.70
cluster 4	Nuclear Receptors, Transcription Factors, Protein Kinases	cluster 6	20.05	cluster 2	7.45
cluster 5	Enzymes, Transmembrane Transporters	cluster 5	13.11	cluster 9	7.23
cluster 6	Nuclear Receptors, Transcription Factors	cluster 4	8.58	cluster 7	4.95
cluster 7	Proteases, Enzymes	cluster 3	6.03	cluster 8	4.13
cluster 8	Proteases, Enzymes	cluster 1	4.86	cluster 5	4.11
cluster 9	GPCRs, Transcription Factors, Enzymes	cluster 2	3.05	cluster 10	2.43
cluster 10	Ion Channels, Transmembrane Transporters, Enzymes	cluster 8	1.60	cluster 6	2.18

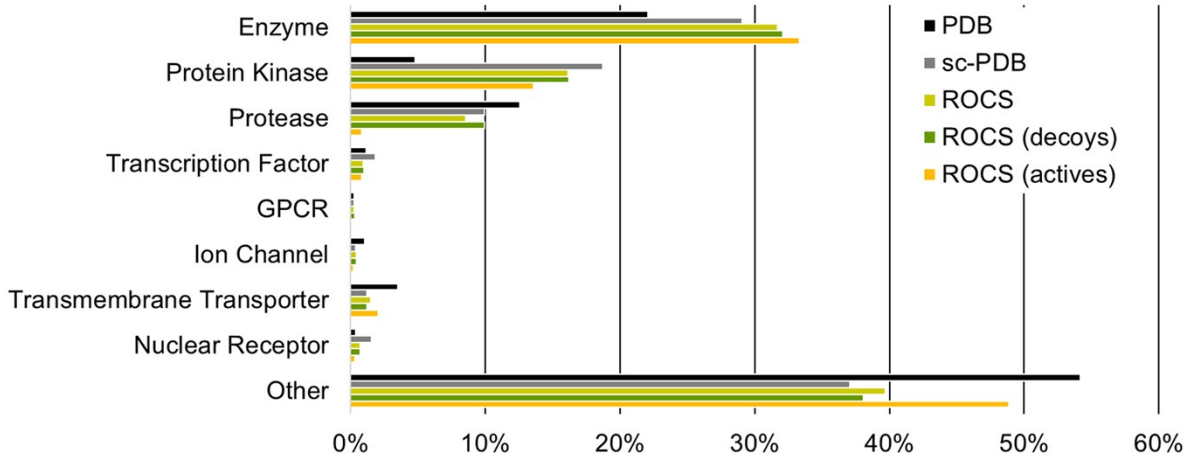


Fig. S3 Distribution of targets in the ROCS dataset. Top: The overrepresented target families in the ligand clusters are shown on the left, while the mean number of targets per ligand cluster is given on the right. Bottom: The distribution of target families in the PDB, the sc-PDB, and the ROCS dataset.

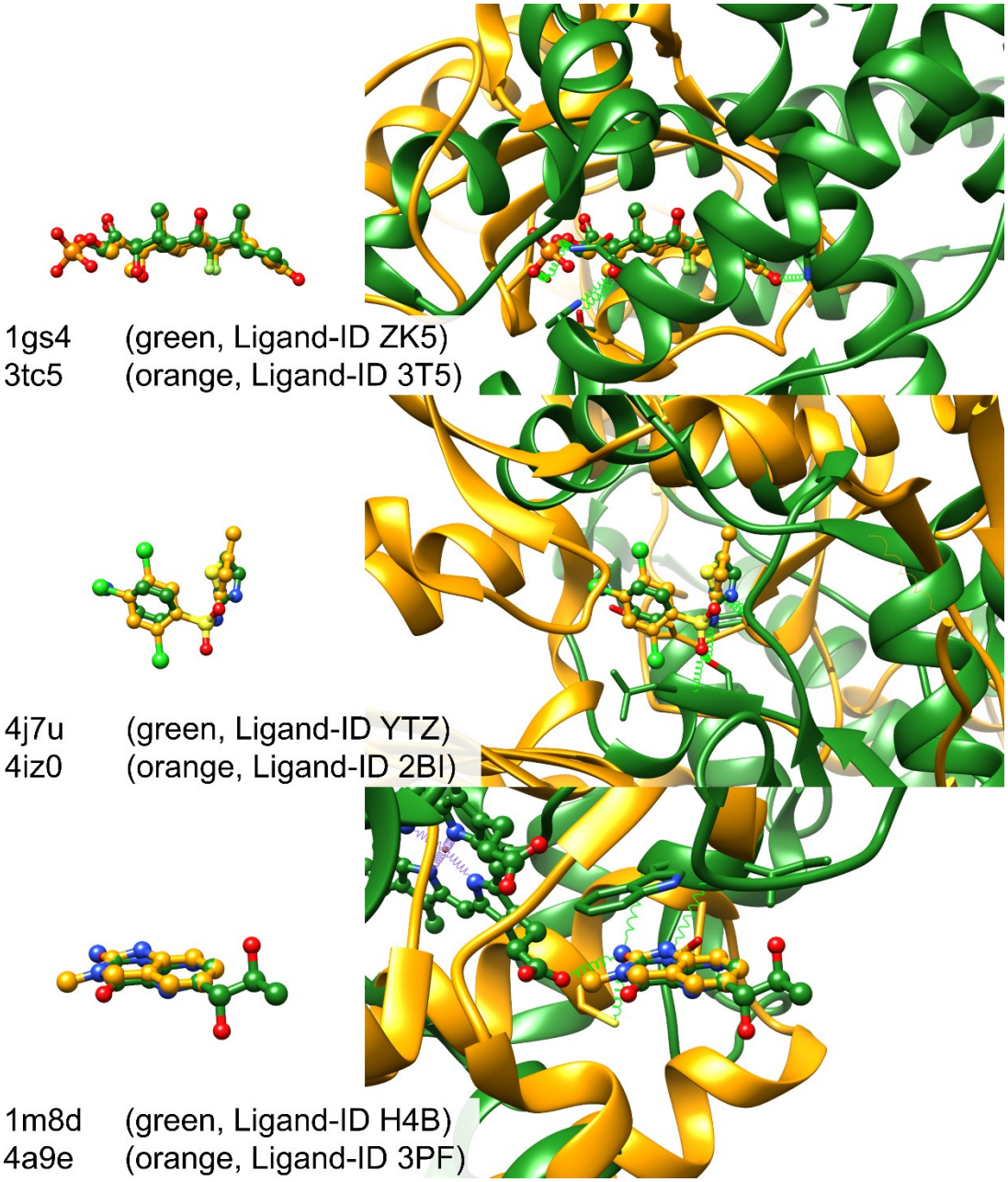


Fig. S4 Ligand-based alignments of protein binding sites binding to similar ligands in similar conformations. The ligand alignments are presented on the left while the corresponding protein alignments are presented on the right. Green springs indicate hydrogen bonds between the protein and the ligand for ligand atoms which can be found in both ligands.

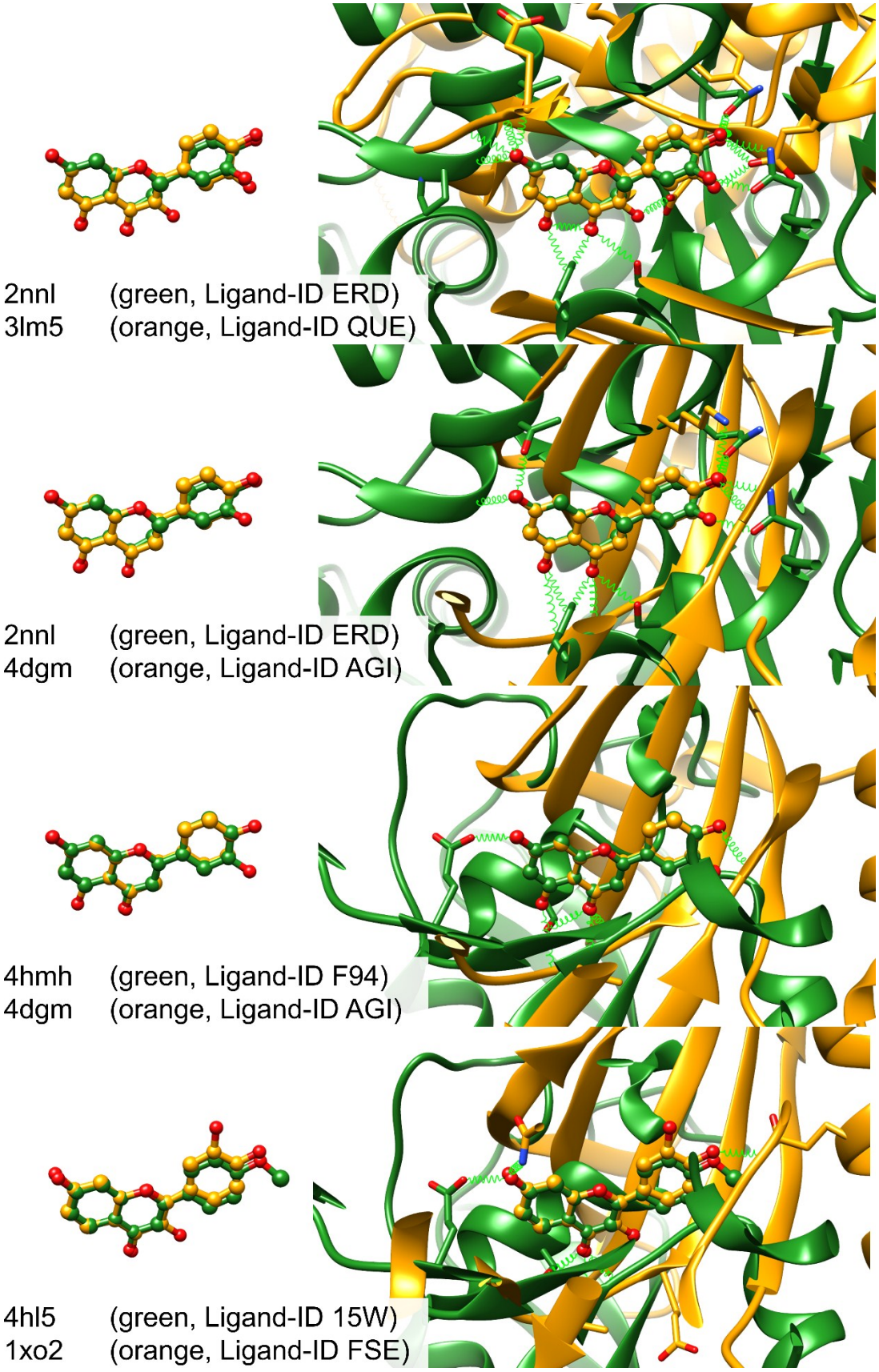


Fig. S5 Ligand-based alignments of protein binding sites binding to similar flavone-related ligands in similar conformations. The ligand alignments are presented on the left while the corresponding protein alignments are presented on the right. Green springs indicate hydrogen bonds between the protein and the ligand for ligand atoms which can be found in both ligands.

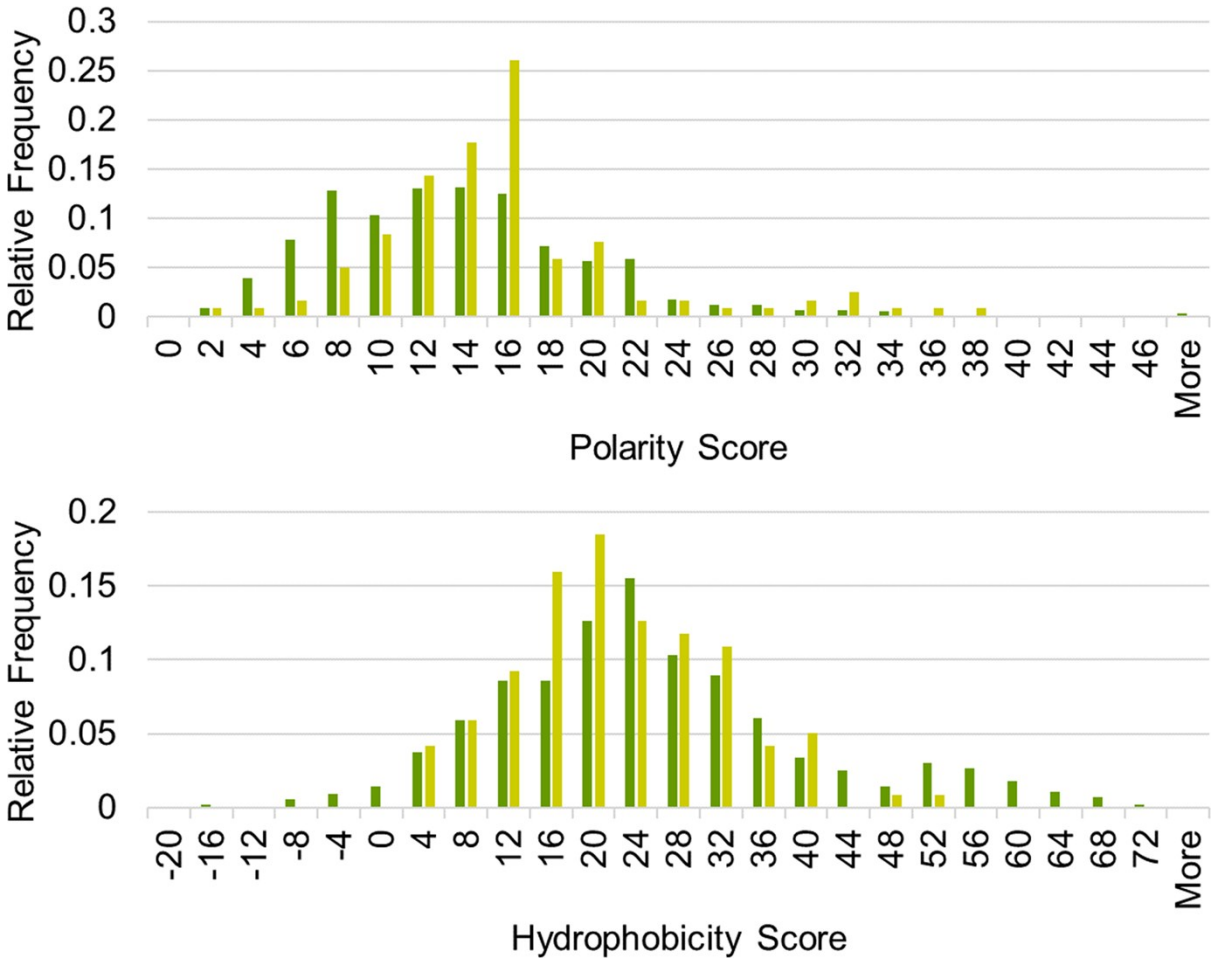
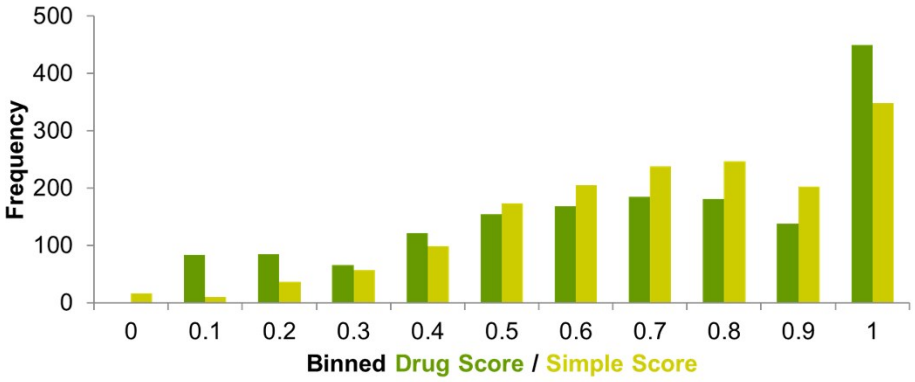


Fig. S6 Distribution of the dpocket[2]-derived polarity and hydrophobicity scores for structures which are exclusively found in pairs with high shape and low physicochemical similarity (dark green bars) or in pairs with high shape and color similarity (light green bars).



	Drug Score		Simple Score	
	arithmetic mean	standard deviation	arithmetic mean	standard deviation
cluster 1	0.64	0.30	0.71	0.23
cluster 2	0.76	0.23	0.80	0.19
cluster 3	0.87	0.15	0.85	0.14
cluster 4	0.73	0.24	0.75	0.20
cluster 5	0.63	0.24	0.65	0.21
cluster 6	0.83	0.17	0.82	0.14
cluster 7	0.60	0.28	0.62	0.24
cluster 8	0.71	0.24	0.73	0.19
cluster 9	0.66	0.26	0.69	0.21
cluster 10	0.62	0.31	0.65	0.25

Fig. S7 DoGSite[3]-derived druggability for the sites in unrelated proteins binding to similar ligands in the ROCS dataset. Top: Distribution of the druggability scores for all sites. Bottom: Mean and standard deviation of the druggabilities for the ten clusters of ligands (see also Fig. 1 in the original publication).

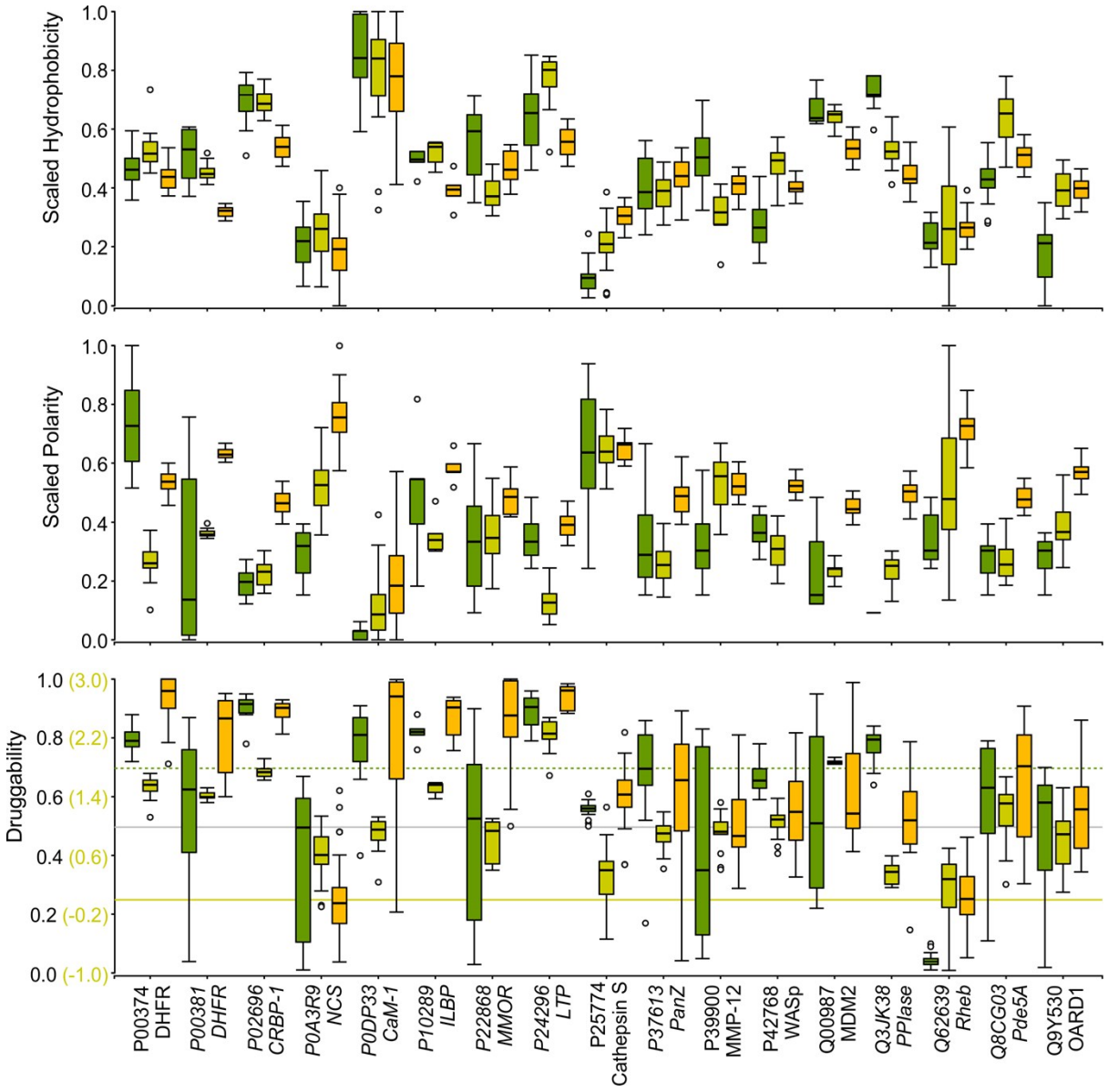


Fig. S8 Box plots of the scaled hydrophobicity (top) and polarity (center) descriptors derived from analyses with fpocket[2] (dark green boxes), VolSite[4] (light green boxes), and DoGSite[3] (yellow boxes) together with the obtained druggability scores (bottom) for the NMR dataset. The cut-offs to distinguish between druggable and non-druggable sites are given as grey lines. The dashed line represents a stricter threshold for druggability predictions with fpocket. The threshold for VolSite predictions is 0 (light green line).

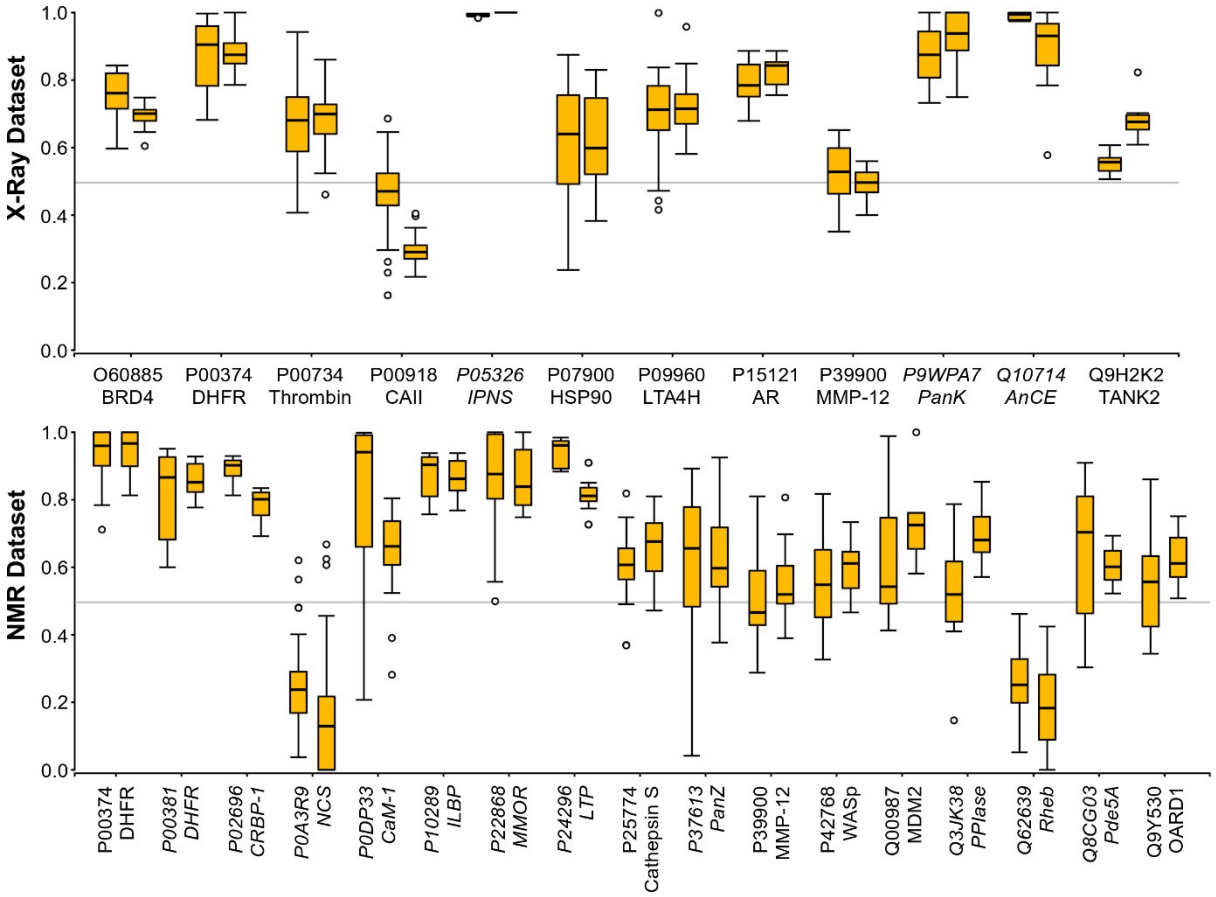


Fig. S9 Boxplots of the DrugScore (left) and the SimpleScore (right) calculated with DoGSiteScorer[3] for the X-ray (top) and NMR (bottom) dataset.

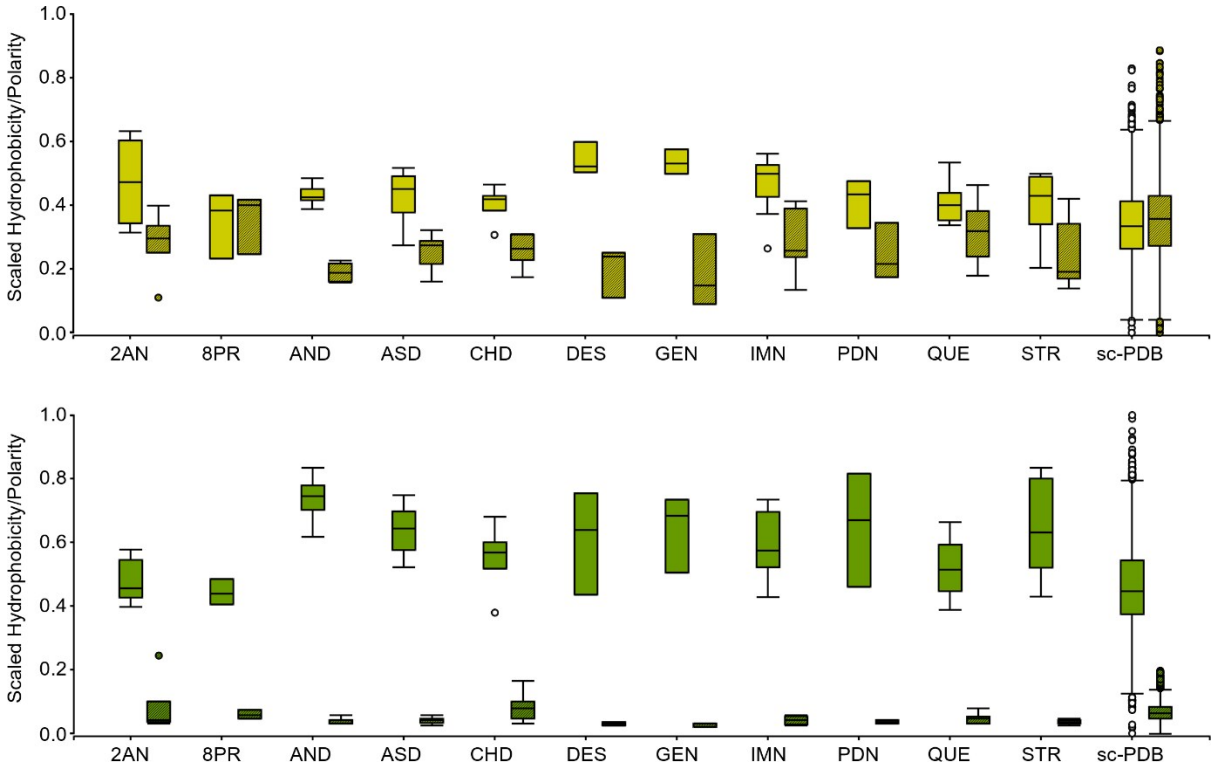


Fig. S10 Boxplots of the scaled hydrophobicity and polarity values for the dissimilar, mainly hydrophobic binding sites in complex with different ligand types. The boxes with a solid fill represent the hydrophobicity and the hatch pattern-filled boxes the polarity of the sites. The upper chart presents the results of the VolSite[4] analysis (light green) and the lower chart presents the results of the fpocket[2] analysis (dark green). The corresponding ligand-IDs are given on the X-axis. The boxplot of the hydrophobicity and polarity in the complete sc-PDB is also given for comparison purposes.

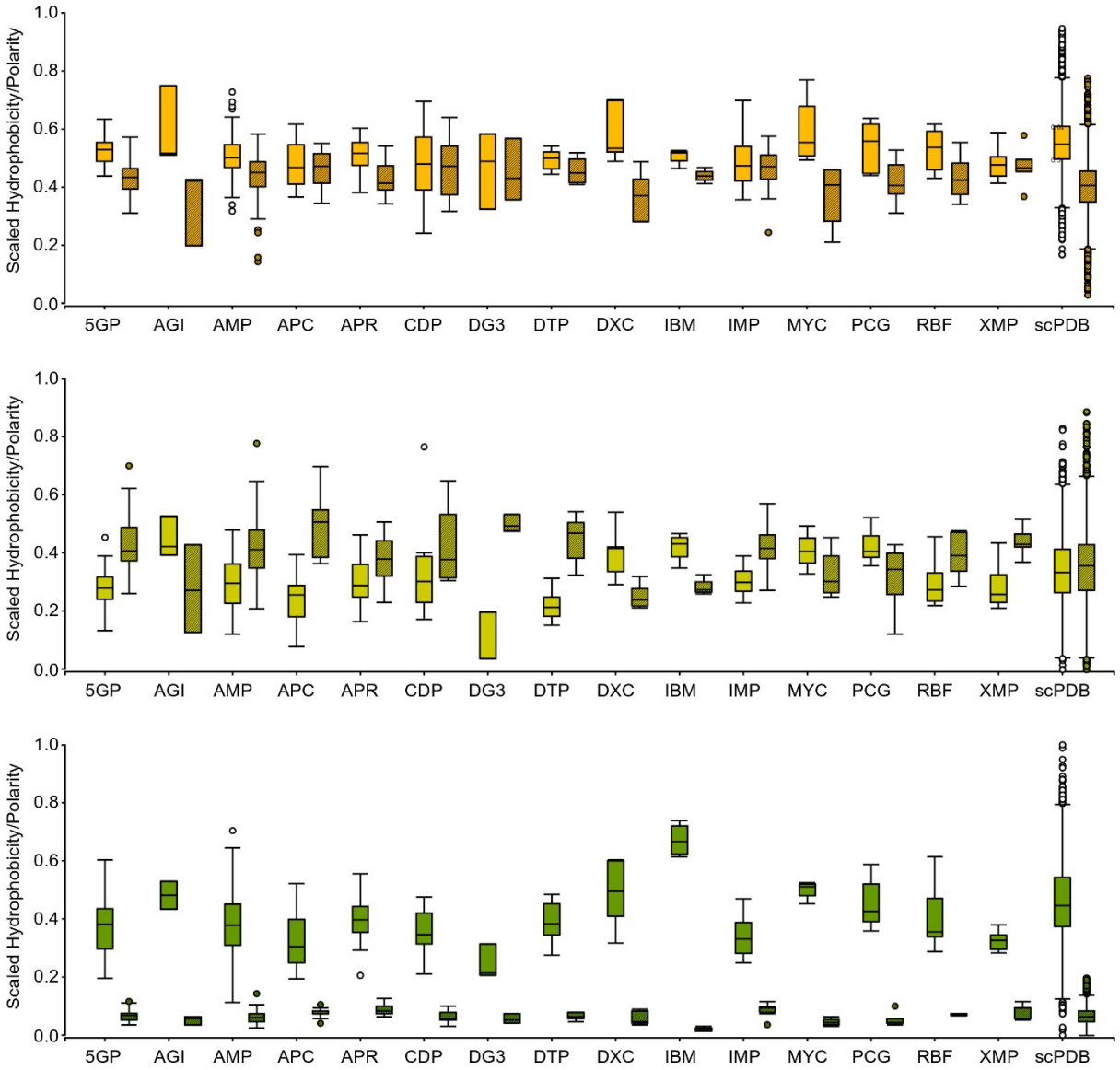


Fig. S11 Boxplots of the scaled hydrophobicity and polarity values for the dissimilar binding sites of average hydrophobicity in complex with different ligand types. The boxes with a solid fill represent the hydrophobicity and the hatch pattern-filled boxes the polarity of the sites. The results obtained with DoGSite[3] (top, yellow), VolSite[4] (center, light green), and fpocket[2] (bottom, dark green) are presented. The corresponding ligand-IDs are given on the X-axis. The boxplot of the hydrophobicity and polarity in the complete sc-PDB is also given for comparison purposes.

Table S1 The ten largest enzyme groups (according to the first two digits of the EC nomenclature) in the PDB, the sc-PDB, and the ROCS dataset. The ten largest classes per dataset are highlighted by an asterisk.

EC Number	ROCS (actives)	ROCS (decoys)	ROCS	sc-PDB	PDB	Description
1.1	8.88*	9.54*	9.16*	7.06*	3.12*	dehydrogenases that act on primary alcohols, secondary alcohols, and hemi-acetals, e.g., L-lactate dehydrogenase
1.14	2.50	4.10*	3.79*	3.03*	3.53*	enzymes that act on two hydrogen donors with incorporation or reduction of molecular oxygen, e.g., nitric-oxide synthase
1.2	2.84*	2.30	2.05	1.70	0.92	enzymes that oxidize aldehydes, e.g., glyceraldehyde-3-phosphate dehydrogenase
1.3	2.59*	3.35*	3.13*	3.14*	1.27	enzymes that introduce a double bond by dehydrogenation of a C-C single bond, e.g., enoyl-[acyl-carrier-protein] reductase
1.5	1.21	2.56	2.38	2.86*	1.04	enzymes that dehydrogenate secondary amines introducing a C-N double bond, e.g., dihydrofolate reductase
1.8	2.59*	1.19	1.06	0.71	0.57	enzymes that act either on inorganic substrates or organic thiols, e.g., dihydrolipoyl dehydrogenase
2.1	9.31*	3.44*	4.56*	4.58*	2.88	enzymes that transfer one-carbon groups, e.g., hydroxymethyl-, formyl-, and related transferases
2.4	3.88*	5.03*	4.70*	4.54*	3.47*	glycosyl transferases, e.g., ADP-ribosyltransferase (polymerizing)
2.5	1.12	2.99*	2.80*	2.98*	2.19	enzymes that transfer alkyl or related groups, e.g., 6,7-dimethyl-8-ribityllumazine synthase
2.7	28.79*	26.48*	27.89*	30.75*	15.14*	enzymes that transfer phosphor-containing groups, e.g., serine/threonine protein kinase
3.1	4.31*	4.10*	4.04*	3.90*	8.03*	enzymes that hydrolyze ester bonds, e.g., 3',5'-cyclic-nucleotide phosphodiesterase
3.2	0.43	2.09	1.82	1.68	7.06*	glycosylases, e.g., lysozyme
3.4	0.78	10.75*	9.47*	10.86*	17.41*	enzymes that hydrolyze peptide bonds, e.g., thrombin
3.5	1.29	1.85	1.66	1.95	3.58*	enzymes that hydrolyze amides, amidines, and other C-N bonds, but not peptide bonds, e.g., β-lactamase
3.6	6.38*	1.64	2.53	2.42	3.37*	enzymes that act on anhydrides, e.g., phosphatases
4.2	0.60	2.89*	2.57*	2.16	3.19*	C-O lyases, e.g., carbonic anhydrase
6.3	6.29*	1.85	2.42	2.56	1.96	enzymes that form C-N bonds, e.g., glutamine synthetase

Table S2 Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
2ydo.A| ADORA2A| *Homo sapiens*| P29274|ADN	2aa0 Adenosine kinase Q9TVW2	0.88	0.40	0.16	CHEMBL477
	1lik Adenosine kinase Q9TVW2	0.87	0.39	0.16	CHEMBL477
	2zoq MAP kinase 3 P27361	0.86	0.39	0.15	CHEMBL496
	3kw2 Ribosomal RNA small subunit methyltransferase E B2RH75	0.85	0.41	0.15	
	1efp Electron transfer flavoprotein B P38975	0.85	0.47	0.12	
	1yi4 Ser/Thr-protein kinase pim-1 P11309	0.81	0.38	0.14	CHEMBL455264
	1muo Aurora kinase A O14965	0.77	0.34	0.14	CHEMBL2010872
	2eva MAP kinase kinase kinase 7 O43318	0.77	0.37	0.14	CHEMBL477
	2ivt Tyr-protein kinase receptor Ret P07949	0.76	0.38	0.13	CHEMBL2010872
	1bx4 Adenosine kinase P55263	0.73	0.37	0.12	CHEMBL302376
	2ivs Tyr-protein kinase receptor Ret P07949	0.69	0.32	0.12	CHEMBL2010872
1t7r.A| AR| *Pan troglodytes*| O97775|DHT	1lho Sex hormone-binding globulin P04278	1.05	0.63	0.14	
	3klp Estradiol 17-beta-dehydrogenase 1 P14061	0.84	0.37	0.16	
	1jtv Estradiol 17-beta-dehydrogenase 1 P14061	0.83	0.39	0.15	
	1lhv Sex hormone-binding globulin P04278	0.82	0.45	0.12	

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
2qpy.A| AR| *Mus musculus*| P19091|DHT	1lhu Sex hormone-binding globulin P04278	0.99	0.54	0.15	CHEMBL103, CHEMBL27769
	1lho Sex hormone-binding globulin P04278	0.93	0.49	0.15	CHEMBL103, CHEMBL27769
	1lhv Sex hormone-binding globulin P04278	0.84	0.42	0.14	CHEMBL103, CHEMBL27769
	1d2s Sex hormone-binding globulin P04278	0.83	0.42	0.14	CHEMBL103, CHEMBL27769
1e3g.A| AR| *Homo sapiens*| P10275|R18	1lho Sex hormone-binding globulin P04278	1.13	0.63	0.17	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhu Sex hormone-binding globulin P04278	1.10	0.65	0.15	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhv Sex hormone-binding globulin P04278	0.99	0.42	0.19	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1fdu Estradiol 17-beta-dehydrogenase 1 P14061	0.77	0.32	0.15	CHEMBL150, CHEMBL28
	1dht Estradiol 17-beta-dehydrogenase 1 P14061	0.77	0.34	0.14	CHEMBL150, CHEMBL28
	1equ Estradiol 17-beta-dehydrogenase 1 P14061	0.76	0.29	0.16	CHEMBL150, CHEMBL28
	3km0 Estradiol 17-beta-dehydrogenase 1 P14061	0.75	0.35	0.13	CHEMBL150, CHEMBL28
	3klp Estradiol 17-beta-dehydrogenase 1 P14061	0.74	0.33	0.14	CHEMBL150, CHEMBL28

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
1e3g.A| AR| *Homo sapiens*| P10275|R18	1xf0 3-alpha-HSD type II, brain P42330	0.74	0.32	0.14	CHEMBL2023820, CHEMBL717
	1qyx Estradiol 17-beta-dehydrogenase 1 P14061	0.74	0.31	0.14	CHEMBL150, CHEMBL28
	1jtv Estradiol 17-beta-dehydrogenase 1 P14061	0.71	0.34	0.13	CHEMBL150, CHEMBL28
	2ipj HAKRD P52895	0.58	0.28	0.10	CHEMBL2023820, CHEMBL717
2amb.A| AR| *Homo sapiens*| P10275|17H	1lho Sex hormone-binding globulin P04278	0.98	0.45	0.18	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1d2s Sex hormone-binding globulin P04278	0.93	0.38	0.18	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhv Sex hormone-binding globulin P04278	0.83	0.42	0.14	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1qyx Estradiol 17-beta-dehydrogenase 1 P14061	0.80	0.36	0.14	CHEMBL150, CHEMBL28
	1qyw Estradiol 17-beta-dehydrogenase 1 P14061	0.79	0.36	0.14	CHEMBL150, CHEMBL28
	1fdu Estradiol 17-beta-dehydrogenase 1 P14061	0.79	0.35	0.14	CHEMBL150, CHEMBL28
	3klp Estradiol 17-beta-dehydrogenase 1 P14061	0.76	0.37	0.13	CHEMBL150, CHEMBL28
	1jtv Estradiol 17-beta-dehydrogenase 1 P14061	0.75	0.35	0.13	CHEMBL150, CHEMBL28

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
2oz7.A| AR| *Homo sapiens*| P10275|CA4	1d2s Sex hormone-binding globulin P04278	0.89	0.47	0.14	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1qyw Estradiol 17-beta-dehydrogenase 1 P14061	0.85	0.34	0.17	CHEMBL150, CHEMBL28
	1qyx Estradiol 17-beta-dehydrogenase 1 P14061	0.74	0.36	0.13	CHEMBL150, CHEMBL28
	1jtv Estradiol 17-beta-dehydrogenase 1 P14061	0.74	0.37	0.12	CHEMBL150, CHEMBL28
3l3z.A| AR| *Homo sapiens*| P10275|DHT	1lhu Sex hormone-binding globulin P04278	1.18	0.68	0.17	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lho Sex hormone-binding globulin P04278	1.11	0.63	0.16	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhv Sex hormone-binding globulin P04278	0.99	0.42	0.19	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1fdu Estradiol 17-beta-dehydrogenase 1 P14061	0.90	0.27	0.21	CHEMBL150, CHEMBL28
	1dht Estradiol 17-beta-dehydrogenase 1 P14061	0.81	0.32	0.16	CHEMBL150, CHEMBL28
	1qyx Estradiol 17-beta-dehydrogenase 1 P14061	0.77	0.35	0.14	CHEMBL150, CHEMBL28
	1jtv Estradiol 17-beta-dehydrogenase 1 P14061	0.77	0.38	0.13	CHEMBL150, CHEMBL28

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
3l3z.A| AR| *Homo sapiens*| P10275|DHT	3km0 Estradiol 17-beta-dehydrogenase 1 P14061	0.77	0.35	0.14	CHEMBL150, CHEMBL28
	3klp Estradiol 17-beta-dehydrogenase 1 P14061	0.76	0.36	0.13	CHEMBL150, CHEMBL28
	1xf0 3-alpha-HSD type II, brain P42330	0.75	0.33	0.14	CHEMBL2023820, CHEMBL717
	1qyw Estradiol 17-beta-dehydrogenase 1 P14061	0.71	0.36	0.12	CHEMBL150, CHEMBL28
3zqt.A| AR| *Homo sapiens*| P10275|TES	1lho Sex hormone-binding globulin P04278	0.99	0.54	0.15	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhu Sex hormone-binding globulin P04278	0.98	0.53	0.15	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1d2s Sex hormone-binding globulin P04278	0.88	0.45	0.14	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	1lhv Sex hormone-binding globulin P04278	0.88	0.41	0.16	CHEMBL103, CHEMBL135, CHEMBL27769, CHEMBL386630, CHEMBL440283, CHEMBL717
	3klp Estradiol 17-beta-dehydrogenase 1 P14061	0.80	0.34	0.15	CHEMBL150, CHEMBL28
	1fdu Estradiol 17-beta-dehydrogenase 1 P14061	0.75	0.35	0.13	CHEMBL150, CHEMBL28
	1xf0 3-alpha-HSD type II, brain P42330	0.73	0.34	0.13	CHEMBL2023820, CHEMBL717

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
3zqt.A| AR| *Homo sapiens*| P10275|TES	3km0 Estradiol 17-beta-dehydrogenase 1 P14061	0.72	0.37	0.12	CHEMBL150, CHEMBL28
	1qyx Estradiol 17-beta-dehydrogenase 1 P14061	0.70	0.37	0.11	CHEMBL150, CHEMBL28
	1qyw Estradiol 17-beta-dehydrogenase 1 P14061	0.67	0.34	0.11	CHEMBL150, CHEMBL28
1j99.A| SULT2A1| *Homo sapiens*| Q06520|AND	1eup 6-deoxyerythronolide B hydroxylase Q00441	0.89	0.39	0.17	
2pw3.B| PDE4D| *Homo sapiens*| Q08499|CMP	1ykd Adenylate cyclase P94182	0.82	0.39	0.14	
1ptw.D| PDE4D| *Homo sapiens*| Q08499|AMP	3t9f diphosphoinositol-pentakisphosphate kinase 2 O43314	0.83	0.40	0.14	
	2q8m Fructose-1,6-bisphosphatase class 1 P0A993	0.82	0.35	0.16	
4jvl.B| SULT1E1| *Homo sapiens*| P49888|EST	1lhu Sex hormone-binding globulin P04278	0.82	0.28	0.18	
1xv9.B| NR1I3| *Homo sapiens*| Q14994|CI2	3s79 Cytochrome P450 19A1 P11511	0.92	0.32	0.20	CHEMBL104
3n8y.A| PTGS1| *Ovis aries*| P05979|DIF	3r6i 3-alpha-HSD type II, brain P42330	0.73	0.34	0.13	CHEMBL1034, CHEMBL154, CHEMBL2323476, CHEMBL2323522, CHEMBL509, CHEMBL563, CHEMBL6
4oty.A| PTGS2| *Mus musculus*| Q05769|LUR	3r6i 3-alpha-HSD type II, brain P42330	0.79	0.38	0.14	CHEMBL2323476, CHEMBL2323508, CHEMBL503179, CHEMBL509, CHEMBL563, CHEMBL6

Table S2 ***(continued)*** Typical antitargets in the similar cavity pairs of the ROCS dataset. Their characteristics, the proteins of pockets in complex with similar ligands, the SiteHopper scores, and known common inhibitors from the ChEMBL database are provided. Binding site matches are shown if the SiteHopper PatchScore was above 0.82 or we found at least one common inhibitor in the CHEMBL database.

PDB ID.Chain| Gene Name| Organism| UniProt Acc| Ligand ID	**PDB ID Name UniProt Acc**	**PatchScore**	**Shape-Tanimoto**	**Color-Tanimoto**	**ChEMBL Ids**
2d06.B| SULT1A1| *Homo sapiens*| P50225|EST	1lhw Sex hormone-binding globulin P04278	0.71	0.35	0.12	CHEMBL135, CHEMBL691
	1lhn Sex hormone-binding globulin P04278	0.70	0.39	0.10	CHEMBL135, CHEMBL691

Table S3 AUC values and AUC confidence intervals for the ROC curves of the analyzed binding site comparison methods for the ROCS dataset. The pairwise AUC differences and the corresponding p-values were calculated according to DeLong and co-workers[6].

	Cavbase	FuzCav (PDB)	FuzCav	Grim (PDB)	Grim	IsoMIF	KRIPO	PocketMatch	ProBiS	RAPMAD
AUC	0.58	0.57	0.57	0.77	0.75	0.97	0.93	0.79	0.52	0.60
CI	0.58-0.59	0.56-0.57	0.57-0.58	0.77-0.78	0.74-0.75	0.97-0.97	0.93-0.93	0.79-0.80	0.52-0.53	0.60-0.61
	0.00	-0.02	-0.01	0.19	0.16	0.39	0.34	0.21	-0.06	0.02
Cavbase	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.02	0.00	0.00	0.20	0.18	0.40	0.36	0.22	-0.05	0.03
FuzCav (PDB)	0.00	1.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.01	0.00	0.00	0.20	0.17	0.40	0.35	0.22	-0.05	0.03
FuzCav	0.00	0.38	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.19	-0.20	-0.20	0.00	-0.03	0.20	0.15	0.02	-0.25	-0.17
Grim (PDB)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.16	-0.18	-0.17	0.03	0.00	0.23	0.18	0.04	-0.22	-0.14
Grim	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	-0.39	-0.40	-0.40	-0.20	-0.23	0.00	-0.05	-0.18	-0.45	-0.37
IsoMIF	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	-0.34	-0.36	-0.35	-0.15	-0.18	0.05	0.00	-0.14	-0.40	-0.32
KRIPO	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	-0.21	-0.22	-0.22	-0.02	-0.04	0.18	0.14	0.00	-0.27	-0.19
PocketMatch	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	0.06	0.05	0.05	0.25	0.22	0.45	0.40	0.27	0.00	0.08
ProBiS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	-0.02	-0.03	-0.03	0.17	0.14	0.37	0.32	0.19	-0.08	0.00
RAPMAD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
VolSite/	-0.11	-0.13	-0.13	0.08	0.05	0.27	0.23	0.09	-0.18	-0.10
Shaper (PDB)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VolSite/	-0.10	-0.11	-0.11	0.09	0.07	0.29	0.25	0.11	-0.16	-0.08
Shaper	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.09	-0.11	-0.10	0.10	0.07	0.30	0.25	0.11	-0.15	-0.07
Shaper (PDB)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.08	-0.10	-0.10	0.11	0.08	0.30	0.26	0.12	-0.14	-0.06
Shaper	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.13	0.11	0.11	0.32	0.29	0.51	0.47	0.33	0.07	0.14
SiteAlign	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.03	0.01	0.02	0.22	0.19	0.42	0.37	0.24	-0.03	0.05
SiteEngine	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.38	-0.40	-0.40	-0.19	-0.22	0.00	-0.04	-0.18	-0.45	-0.37
SiteHopper	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
	0.03	0.01	0.01	0.22	0.19	0.41	0.37	0.23	-0.03	0.04
SMAP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.03	-0.04	-0.04	0.16	0.13	0.36	0.32	0.18	-0.09	-0.01
TIFP (PDB)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	0.01	-0.01	0.00	0.20	0.17	0.40	0.35	0.21	-0.05	0.03
TIFP	0.03	0.08	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	-0.20	-0.22	-0.21	-0.01	-0.04	0.19	0.14	0.00	-0.26	-0.18
TM-align	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00

Table S3 ***(continued)*** AUC values and AUC confidence intervals for the ROC curves of the analyzed binding site comparison methods for the ROCS dataset. The pairwise AUC differences and the corresponding p-values were calculated according to DeLong and co-workers[6].

	VolSite/ Shaper (PDB)	VolSite/ Shaper	Shaper (PDB)	Shaper	SiteAlign	SiteEngine	SiteHopper	SMAP	TIFP (PDB)	TIFP	TM-align
AUC	0.70	0.68	0.68	0.67	0.46	0.56	0.97	0.56	0.61	0.58	0.79
CI	0.69-0.70	0.67-0.69	0.67-0.68	0.66-0.67	0.45-0.46	0.55-0.56	0.93-0.97	0.55-0.56	0.61-0.62	0.57-0.58	0.78-0.79
Cavbase	0.11	0.10	0.09	0.08	-0.13	-0.03	0.38	-0.03	0.03	-0.01	0.20
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
FuzCav (PDB)	0.13	0.11	0.11	0.10	-0.11	-0.01	0.40	-0.01	0.04	0.01	0.22
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00
FuzCav	0.13	0.11	0.10	0.10	-0.11	-0.02	0.40	-0.01	0.04	0.00	0.21
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.00
Grim (PDB)	-0.08	-0.09	-0.10	-0.11	-0.32	-0.22	0.19	-0.22	-0.16	-0.20	0.01
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grim	-0.05	-0.07	-0.07	-0.08	-0.29	-0.19	0.22	-0.19	-0.13	-0.17	0.04
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IsoMIF	-0.27	-0.29	-0.30	-0.30	-0.51	-0.42	0.00	-0.41	-0.36	-0.40	-0.19
	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
KRIPO	-0.23	-0.25	-0.25	-0.26	-0.47	-0.37	0.04	-0.37	-0.32	-0.35	-0.14
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PocketMatch	-0.09	-0.11	-0.11	-0.12	-0.33	-0.24	0.18	-0.23	-0.18	-0.21	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29
ProBiS	0.18	0.16	0.15	0.14	-0.07	0.03	0.45	0.03	0.09	0.05	0.26
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RAPMAD	0.10	0.08	0.07	0.06	-0.14	-0.05	0.37	-0.04	0.01	-0.03	0.18
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
VolSite/ Shaper (PDB)	0.00	-0.02	-0.02	-0.03	-0.24	-0.14	0.27	-0.14	-0.09	-0.12	0.09
	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VolSite/ Shaper	0.02	0.00	0.00	-0.01	-0.22	-0.12	0.29	-0.12	-0.07	-0.10	0.11
	0.00	1.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Shaper (PDB)	0.02	0.00	0.00	-0.01	-0.22	-0.12	0.29	-0.12	-0.07	-0.10	0.11
	0.00	0.40	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Shaper	0.03	0.01	0.01	0.00	-0.21	-0.11	0.30	-0.11	-0.06	-0.09	0.12
	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SiteAlign	0.24	0.22	0.22	0.21	0.00	0.10	0.51	0.10	0.15	0.12	0.33
	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
SiteEngine	0.14	0.12	0.12	0.11	-0.10	0.00	0.41	0.00	0.06	0.02	0.23
	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.41	0.00	0.00	0.00
SiteHopper	-0.27	-0.29	-0.29	-0.30	-0.51	-0.41	0.00	-0.41	-0.36	-0.39	-0.18
	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
SMAP	0.14	0.12	0.12	0.11	-0.10	0.00	0.41	0.00	0.05	0.02	0.23
	0.00	0.00	0.00	0.00	0.00	0.41	0.00	1.00	0.00	0.00	0.00
TIFP (PDB)	0.09	0.07	0.07	0.06	-0.15	-0.06	0.36	-0.05	0.00	-0.04	0.18
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
TIFP	0.12	0.10	0.10	0.09	-0.12	-0.02	0.39	-0.02	0.04	0.00	0.21
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
TM-align	-0.09	-0.11	-0.11	-0.12	-0.33	-0.23	0.18	-0.23	-0.18	-0.21	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table S4 AUC values and AUC confidence intervals for the ROC curves of the analyzed binding site comparison methods for the reduced ROCS dataset (only site pairs with a SiteHopper[5] PatchScore of more than 0.82 are classified as active site pairs). The pairwise AUC differences and the corresponding p-values were calculated according to DeLong and co-workers[6].

	Cavbase	FuzCav (PDB)	FuzCav	Grim (PDB)	Grim	IsoMIF	KRIPO	PocketMatch	ProBiS	RAPMAD
AUC	0.69	0.58	0.58	0.78	0.76	0.98	0.97	0.88	0.59	0.54
CI	0.68-0.70	0.57-0.59	0.57-0.59	0.77-0.79	0.75-0.77	0.98-0.99	0.96-0.97	0.87-0.88	0.58-0.60	0.53-0.55
Cavbase	0.00 1.00	-0.11 0.00	-0.11 0.00	0.09 0.00	0.07 0.00	0.29 0.00	0.28 0.00	0.19 0.00	-0.10 0.00	-0.15 0.00
FuzCav (PDB)	0.11 0.00	0.00 1.00	0.00 0.81	0.20 0.00	0.18 0.00	0.40 0.00	0.38 0.00	0.30 0.00	0.01 0.22	-0.04 0.00
FuzCav	0.11 0.00	0.00 0.81	0.00 1.00	0.20 0.00	0.17 0.00	0.40 0.00	0.38 0.00	0.29 0.00	0.01 0.33	-0.04 0.00
Grim (PDB)	-0.09 0.00	-0.20 0.00	-0.20 0.00	0.00 1.00	-0.03 0.00	0.20 0.00	0.18 0.00	0.09 0.00	-0.19 0.00	-0.24 0.00
Grim	-0.07 0.00	-0.18 0.00	-0.17 0.00	0.03 0.00	0.00 1.00	0.23 0.00	0.21 0.00	0.12 0.00	-0.17 0.00	-0.22 0.00
IsoMIF	-0.29 0.00	-0.40 0.00	-0.40 0.00	-0.20 0.00	-0.23 0.00	0.00 1.00	-0.02 0.00	-0.11 0.00	-0.40 0.00	-0.45 0.00
KRIPO	-0.28 0.00	-0.38 0.00	-0.38 0.00	-0.18 0.00	-0.21 0.00	0.02 0.00	0.00 1.00	-0.09 0.00	-0.38 0.00	-0.43 0.00
PocketMatch	-0.19 0.00	-0.30 0.00	-0.29 0.00	-0.09 0.00	-0.12 0.00	0.11 0.00	0.09 0.00	0.00 1.00	-0.29 0.00	-0.34 0.00
ProBiS	0.10 0.00	-0.01 0.22	-0.01 0.33	0.19 0.00	0.17 0.00	0.40 0.00	0.38 0.00	0.29 0.00	0.00 1.00	-0.05 0.00
RAPMAD	0.15 0.00	0.04 0.00	0.04 0.00	0.24 0.00	0.22 0.00	0.45 0.00	0.43 0.00	0.34 0.00	0.05 0.00	0.00 1.00
VolSite/ Shaper (PDB)	-0.01 0.28	-0.12 0.00	-0.12 0.00	0.09 0.00	0.06 0.00	0.29 0.00	0.27 0.00	0.18 0.00	-0.11 0.00	-0.16 0.00
VolSite/ Shaper	0.02 0.00	-0.09 0.00	-0.09 0.00	0.11 0.00	0.09 0.00	0.32 0.00	0.30 0.00	0.21 0.00	-0.08 0.00	-0.13 0.00
Shaper (PDB)	-0.01 0.31	-0.12 0.00	-0.12 0.00	0.09 0.00	0.06 0.00	0.29 0.00	0.27 0.00	0.18 0.00	-0.11 0.00	-0.16 0.00
Shaper	0.01 0.05	-0.10 0.00	-0.09 0.00	0.11 0.00	0.08 0.00	0.31 0.00	0.29 0.00	0.20 0.00	-0.09 0.00	-0.14 0.00
SiteAlign	0.17 0.00	0.06 0.00	0.06 0.00	0.26 0.00	0.23 0.00	0.46 0.00	0.44 0.00	0.35 0.00	0.06 0.00	0.01 0.02
SiteEngine	0.03 0.00	-0.08 0.00	-0.08 0.00	0.12 0.00	0.10 0.00	0.32 0.00	0.31 0.00	0.22 0.00	-0.07 0.00	-0.12 0.00
SiteHopper	-0.31 0.00	-0.42 0.00	-0.42 0.00	-0.22 0.00	-0.24 0.00	-0.02 0.00	-0.04 0.00	-0.12 0.00	-0.41 0.00	-0.46 0.00
SMAP	-0.04 0.00	-0.15 0.00	-0.15 0.00	0.05 0.00	0.03 0.00	0.25 0.00	0.23 0.00	0.15 0.00	-0.14 0.00	-0.19 0.00
TIFP (PDB)	0.06 0.00	-0.05 0.00	-0.05 0.00	0.15 0.00	0.13 0.00	0.36 0.00	0.34 0.00	0.25 0.00	-0.04 0.00	-0.09 0.00
TIFP	0.10 0.00	-0.01 0.31	-0.01 0.44	0.20 0.00	0.17 0.00	0.40 0.00	0.38 0.00	0.29 0.00	0.00 0.86	-0.05 0.00
TM-align	-0.24 0.00	-0.35 0.00	-0.35 0.00	-0.15 0.00	-0.17 0.00	0.06 0.00	0.04 0.00	-0.05 0.00	-0.34 0.00	-0.39 0.00

Table S4 ***(continued)*** AUC values and AUC confidence intervals for the ROC curves of the analyzed binding site comparison methods for the reduced ROCS dataset (only site pairs with a SiteHopper[5] PatchScore of more than 0.82 are classified as active site pairs). The pairwise AUC differences and the corresponding p-values were calculated according to DeLong and co-workers[6].

	VolSite/ Shaper (PDB)	VolSite/ Shaper	Shaper (PDB)	Shaper	SiteAlign	SiteEngine	SiteHopper	SMAP	TIFP (PDB)	TIFP	TM-align
AUC	0.70	0.67	0.70	0.68	0.52	0.66	1.00	0.73	0.63	0.59	0.93
CI	0.69-0.71	0.66-0.68	0.69-0.71	0.67-0.68	0.52-0.53	0.65-0.67	1.00-1.00	0.72-0.74	0.62-0.64	0.58-0.60	0.92-0.93
Cavbase	0.01 0.28	-0.02 0.00	0.01 0.31	-0.01 0.05	-0.17 0.00	-0.03 0.00	0.31 0.00	0.04 0.00	-0.06 0.00	-0.10 0.00	0.24 0.00
FuzCav (PDB)	0.12 0.00	0.09 0.00	0.12 0.00	0.10 0.00	-0.06 0.00	0.08 0.00	0.42 0.00	0.15 0.00	0.05 0.00	0.01 0.31	0.35 0.00
FuzCav	0.12 0.00	0.09 0.00	0.12 0.00	0.09 0.00	-0.06 0.00	0.08 0.00	0.42 0.00	0.15 0.00	0.05 0.00	0.01 0.44	0.35 0.00
Grim (PDB)	-0.09 0.00	-0.11 0.00	-0.09 0.00	-0.11 0.00	-0.26 0.00	-0.12 0.00	0.22 0.00	-0.05 0.00	-0.15 0.00	-0.20 0.00	0.15 0.00
Grim	-0.06 0.00	-0.09 0.00	-0.06 0.00	-0.08 0.00	-0.23 0.00	-0.10 0.00	0.24 0.00	-0.03 0.00	-0.13 0.00	-0.17 0.00	0.17 0.00
IsoMIF	-0.29 0.00	-0.32 0.00	-0.29 0.00	-0.31 0.00	-0.46 0.00	-0.32 0.00	0.02 0.00	-0.25 0.00	-0.36 0.00	-0.40 0.00	-0.06 0.00
KRIPO	-0.27 0.00	-0.30 0.00	-0.27 0.00	-0.29 0.00	-0.44 0.00	-0.31 0.00	0.04 0.00	-0.23 0.00	-0.34 0.00	-0.38 0.00	-0.04 0.00
PocketMatch	-0.18 0.00	-0.21 0.00	-0.18 0.00	-0.20 0.00	-0.35 0.00	-0.22 0.00	0.12 0.00	-0.15 0.00	-0.25 0.00	-0.29 0.00	0.05 0.00
ProBiS	0.11 0.00	0.08 0.00	0.11 0.00	0.09 0.00	-0.06 0.00	0.07 0.00	0.41 0.00	0.14 0.00	0.04 0.00	0.00 0.86	0.34 0.00
RAPMAD	0.16 0.00	0.13 0.00	0.16 0.00	0.14 0.00	-0.01 0.02	0.12 0.00	0.46 0.00	0.19 0.00	0.09 0.00	0.05 0.00	0.39 0.00
VolSite/ Shaper (PDB)	0.00 1.00	-0.03 0.00	0.00 0.93	-0.02 0.00	-0.17 0.00	-0.04 0.00	0.30 0.00	0.03 0.00	-0.07 0.00	-0.11 0.00	0.23 0.00
VolSite/ Shaper	0.03 0.00	0.00 1.00	0.03 0.00	0.01 0.26	-0.14 0.00	-0.01 0.19	0.33 0.00	0.06 0.00	-0.04 0.00	-0.08 0.00	0.26 0.00
Shaper (PDB)	0.00 0.93	-0.03 0.00	0.00 1.00	-0.02 0.00	-0.17 0.00	-0.04 0.00	0.30 0.00	0.03 0.00	-0.07 0.00	-0.11 0.00	0.23 0.00
Shaper	0.02 0.00	-0.01 0.26	0.02 0.00	0.00 1.00	-0.15 0.00	-0.02 0.01	0.32 0.00	0.05 0.00	-0.05 0.00	-0.09 0.00	0.25 0.00
SiteAlign	0.17 0.00	0.14 0.00	0.17 0.00	0.15 0.00	0.00 1.00	0.14 0.00	0.48 0.00	0.21 0.00	0.10 0.00	0.06 0.00	0.40 0.00
SiteEngine	0.04 0.00	0.01 0.19	0.04 0.00	0.02 0.01	-0.14 0.00	0.00 1.00	0.34 0.00	0.07 0.00	-0.03 0.00	-0.07 0.00	0.27 0.00
SiteHopper	-0.30 0.00	-0.33 0.00	-0.30 0.00	-0.32 0.00	-0.48 0.00	-0.34 0.00	0.00 1.00	-0.27 0.00	-0.37 0.00	-0.41 0.00	-0.07 0.00
SMAP	-0.03 0.00	-0.06 0.00	-0.03 0.00	-0.05 0.00	-0.21 0.00	-0.07 0.00	0.27 0.00	0.00 1.00	-0.10 0.00	-0.14 0.00	0.20 0.00
TIFP (PDB)	0.07 0.00	0.04 0.00	0.07 0.00	0.05 0.00	-0.10 0.00	0.03 0.00	0.37 0.00	0.10 0.00	0.00 1.00	-0.04 0.00	0.30 0.00
TIFP	0.11 0.00	0.08 0.00	0.11 0.00	0.09 0.00	-0.06 0.00	0.07 0.00	0.41 0.00	0.14 0.00	0.04 0.00	0.00 1.00	0.34 0.00
TM-align	-0.23 0.00	-0.26 0.00	-0.23 0.00	-0.25 0.00	-0.40 0.00	-0.27 0.00	0.07 0.00	-0.20 0.00	-0.30 0.00	-0.34 0.00	0.00 1.00

Text S1 Discussion of the TM-align performance.

A striking observation for the ROCS dataset of similar and dissimilar cavity pairs is the obviously high early enrichment for TM-align[7] which also uses Cα atom coordinates to superpose structures. Partially, it can be explained by the rigorous binding site definition based on the distance to the ligands whose shape is highly similar to their "active" counterpart. On the other hand, the algorithm also relies on Cα atom coordinates and an initial secondary structures alignment. It can also be excluded that the underlying proteins show a high overall similarity as TM-align was used to filter out those protein pairs. The analysis of the active pair with the highest TM-score in the dataset revealed the explanation for this observation. Although both proteins differ in their overall structure when taking into account more than one chain, they share a similar protein chain belonging to the CATH[8] homology superfamily (P-loop containing nucleotide triphosphate hydrolase 3.40.50.300 for the structures with the PDB-IDs 4hdq (chain B) and 3tnf (chain A)). Two protein chains of the structure with the PDB-ID 3tnf are involved in ligand binding and were therefore retained for binding site comparison. Nevertheless, this general observation within our dataset does not hamper the results as both binding sites differ with respect to the underlying residues. The question arises of whether similar issues could be observed for the decoy structure pairs. Therefore, we analyzed the decoy pairs with high TM-scores. No such case was found within these pairs.

References

1. ROCS. Santa Fe, NM: OpenEye Scientific Software. Available from: URL: http://www.eyesopen.com.

2. Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. J Med Chem 2010; 53(15):5858–67.

3. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. J Chem Inf Model 2012; 52(2):360–72.

4. Desaphy J, Azdimousa K, Kellenberger E, Rognan D. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. J Chem Inf Model 2012; 52(8):2287–99.

5. Batista J, Hawkins PCD, Tolbert R, Geballe MT. SiteHopper - Aunique tool for binding site comparison. J Cheminform 2014; 6(Suppl 1):P57.

6. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988; 44(3):837–45.

7. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 2005; 33(7):2302–9.

8. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL et al. CATH: Comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res 2015; 43(Database issue):81.