Supporting Information

Understanding and predicting druggability. A high-throughput method for detection of drug binding sites.

Peter Schmidtke, Xavier Barril

Contents:

List of protein name abbreviations used in Figure 3.

Table S1. Summary of prediction results per protein.

List of protein name abbreviations used in Figure 3.

AR : Androgen Receptor

PR : Progesterone Receptor

MCR : Mineral Corticoid Receptor

GCR : Glucocorticoid Receptor

SHB : Sex Hormone Binding globulin

ER : Estrogen Receptor

THRA : Thyroid Hormone Receptor

DHFR : Dihydrofolate Reductase

MDM2 : Mouse Double Minute 2-Tumor Protein

EGF : Epidermial Growth Factor Receptor

P38 : P38 Map Kinase

COX2 : Cyclooxygenase 2

DHFR-TS : Bifunctional Dihydrofolate Reductase-Thymidylate Synthase

PPARg : Peroxisome Proliferator Activated Receptor gamma

B2AR : Human Beta2 Adrenergic Receptor

PLA2 : Phospholipase A2

CYP450_51 : Cytochrome P450 51

3AHD : 3Alpha-Hydroxysteroid Dehydrogenase type 3

PDE5A : Phosphodiesterase 5

EAR : Enoyl Reductase

CDK2 : Cell Division Protein Kinase 2

TK : Tyrosine Kinase

PTP-1B : Protein Tyrosine Phosphatase 1B

PDE4D : Phosphodiesterase 4D

HDAC8 : Histone Deacetylase 8

FX : Factor XA

HMG_CoA : 3-Hydroxy-3-Methylglutaryl-coenzyme A Reductase

PHH : P-Hydroxybenzoate Hydrolase

GST : Glutathione S-Transferase

P4M : Phenylalanine-4-Monooxygenase

CYP450_SU1 : Cytochrome P450 105A1

HemK : N5-glutamine methyltransferase

DNH : Deoxyuridine Nucleotide Hydrolase

FDS : Farnesyl Diphosphate Synthase

CCAT : cytosolic Branched Chain Aminotransferase

NPRT : Nicotinate phosphoribosyltransferase

PISPLP : Phosphatidylinositol spec. phospholipase

EBDGA : Exo-beta-D-glucosaminidase

TPK : Tyrosine protein kinase BTK

ACE : Angiotensin Converting Enzyme

PBP : Penicillin Binding Protein

IMPDH : Inosine Monophosphate Dedhydrogenase

Table S1. Summary of prediction results per protein.

	Holo cavities					Apo cavities				
Protein Name	No. of cavities	Mean score	Std. dev	Mean score of top10%	Confi-dence[1]	No. of cavities	Mean score	Std. dev	Mean score of top10%	Confi-dence[1]
druggable										
Carbonic Anhydrase	76	0.56	0.31	0.86	0.93	61	0.42	0.30	0.82	0.83
DHFR-TS	3	0.73	0.06	0.77	0.90	-				
DHFR	39	0.69	0.14	0.85	0.91	1	0.73	0.00	0.73	0.00
Thymidine Kinase	29	0.46	0.19	0.76	0.87	2	0.43	0.06	0.47	0.87
SHB	6	0.84	0.03	0.88	0.96	-				
PDE4D	54	0.40	0.28	0.78	0.84	-				
Mineralcorticoid Receptor	25	0.87	0.09	0.95	0.94	-				
Glucocorticoid Receptor	6	0.86	0.07	0.96	0.91	-				
Androgen Receptor	38	0.95	0.01	0.97	0.99	-				
Progesterone Receptor	9	0.94	0.03	0.97	0.97	-				
Estrogen Receptor	77	0.83	0.10	0.96	0.95	2	0.86	0.01	0.87	0.97
CYP P450 - 121	13	0.62	0.21	0.87	0.82	-				
HIV RT - NNRTI	38	0.75	0.09	0.90	0.94	-				
HIV1 Protease	24	0.65	0.14	0.76	0.90	1	0.44	0.00	0.44	0.00
PPARγ	43	0.70	0.12	0.86	0.93	2	0.50	0.04	0.52	0.93
ADAM33	1	0.08	0.00	0.08	0.00	1	0.15	0.00	0.15	0.00
Acetylcholinesterase	42	0.60	0.26	0.80	0.84	4	0.68	0.11	0.74	0.85
HMG CoA Reductase	27	0.12	0.00	0.42	0.93	-				
DNA gyrase B	3	0.43	0.28	0.75	0.53	-				
Plasminogen Kringle 4	8	0.31	0.15	0.44	0.84	1	0.29	0.00	0.29	0.00
p-hydroxybenzoate Hydroxylase	12	0.05	0.02	0.08	0.98	-				
COX2	9	0.72	0.11	0.83	0.89	-				
Phospholipase A2	19	0.62	0.19	0.79	0.85	4	0.77	0.04	0.80	0.94

Beta-2-adrenergic Receptor/T4 lysozyme chimera	1	0.69	0.00	0.69	0.00	2	0.25	0.21	0.39	0.56
Enoyl-[acyl-carrier-protein] reductase	22	0.49	0.26	0.85	0.81	2	0.47	0.07	0.47	0.85
Coagulation factor X,	20	0.12	0.18	0.62	0.87	-				
Cardiac Troponin C	2	0.76	0.01	0.76	0.99	-				
3 Alpha Hydroxysteroid Dehydrogenase	12	0.69	0.27	0.89	0.76	-				
Beta Lactamase	59	0.11	0.12	0.45	0.93	20	0.11	0.09	0.34	0.93
PDE-5A	17	0.61	0.22	0.89	0.82	2	0.74	0.13	0.83	0.73
THRA protein	2	0.82	0.01	0.83	0.97	-				
Cyclophilin C	4	0.33	0.24	0.66	0.66	-				
Aldose Reductase	67	0.81	0.08	0.87	0.96	5	0.90	0.06	0.86	0.93
cABL Kinase	17	0.63	0.19	0.79	0.85	-				
CDK2	50	0.59	0.27	0.83	0.84	2	0.32	0.20	0.46	0.59
Epidemial Growth Factor	3	0.73	0.10	0.83	0.83					
MDM2	5	0.73	0.18	0.88	0.77					
P38 Map Kinase	43	0.73	0.21	0.92	0.87	2	0.49	0.27	0.91	0.44
Renin	37	0.68	0.20	0.88	0.95	1	0.88	0.00	0.88	0.00
HDAC	10	0.36	0.28	0.72	0.73					
Topoisomerase 2	1	0.17	0.00	0.17	0.00					
Xanthine Oxydase	4	0.10	0.02	0.12	0.97					
Alpha Beta Tubulin	1	0.30	0.00	0.30	0.00					
PTP-1B (druggable site)	2	0.44	0.07	0.49	0.85					
non druggable										
cytosolic Branched Chain Aminotransferase 1	1	0.24	0.00	0.24	0.00	-				

Farnesyl Diphosphate Synthase	3	0.26	0.29	0.59	0.51	-				
Glutathione S-Transferase	1	0.90	0.00	0.90	0.00	-				
Nicotinate phosphoribosyltransferase	1	0.10	0.00	0.10	0.00	-	0.00	0.00	0.00	0.00
Exo-beta-D-glucosaminidase	2	0.05	0.04	0.07	0.93	-	0.00	0.00	0.00	0.00
Phenylalanine-4-Monooxygenase	1	0.80	0.00	0.80	0.00	-	0.00	0.00	0.00	0.00
CYP P450 – SU1	1	0.76	0.00	0.76	0.00	-	0.00	0.00	0.00	0.00
HIV RT Nucl. site	-	0.00	0.00		0.00	-	0.00	0.00	0.00	0.00
Cathepsin K	14	0.30	0.25	0.70	0.79	-	0.00	0.00	0.00	0.00
PTP-1B (non druggable site)	6	0.37	0.26	0.60	0.69	-	0.00	0.00	0.00	0.00
Caspase 1	8	0.03	0.01	0.04	0.99	-	0.00	0.00	0.00	0.00
HIV Integrase	1	0.01	0.00		0.00	-	0.00	0.00	0.00	0.00
HemK	7	0.45	0.22	0.67	0.76	-	0.00	0.00	0.00	0.00
IP Phosphatase	1	0.02	0.00	0.02	0.00	-	0.00	0.00	0.00	0.00
Tyrosine protein kinase BTK	2	0.03	0.00	0.03	0.00	-	0.00	0.00	0.00	0.00
Deoxyuridine nucleotide hydrolase	5	0.27	0.19	0.60	0.76	-	0.00	0.00	0.00	0.00
Glucoamylase	1	0.73	0.00	0.73	0.00	-	0.00	0.00	0.00	0.00
Phosphatidylinositol spec. phospholipase	1	0.08	0.00	0.08	0.00	3	0.05	0.03	0.07	0.96
Endoglucanase C	1	0.09	0.00	0.09	0.00	-	0.00	0.00	0.00	0.00
Malate dehydrogenase	2	0.06	0.04	0.09	0.92	-	0.00	0.00	0.00	0.00
Prodrug binding										
IMPDH	6	0.04	0.01	0.35	0.99	-	0.00	0.00	0.00	0.00
Neuraminidase	22	0.08	0.10	0.34	0.92	3	0.05	0.03	0.07	0.95
Thrombin	30	0.51	0.18	0.64	0.88	-	0.00	0.00	0.00	0.00

Penicillin Binding Protein	1	0.25	0.00	0.25	0.00	-	0.00	0.00	0.00	0.00
Angiotensin converting enzyme	2	0.39	0.47	0.72	0.01	-	0.00	0.00	0.00	0.00

[1]The confidence level is a measure of possibility of accurate prediction. The confidence level is calculated as difference between 1 and the ratio of the standard deviation of the drug score and the log of the number structures incremented by 1. Thus a confidence level close to 1 signifies that drug score yielded very stable results on multiple structures. The less structures one uses or the more the prediction gets variable in one protein structural family, the lower the confidence.