Supporting Information

CholMine: Determinants and Prediction of Cholesterol and Cholate Binding Across Non-homologous Protein Structures

Nan Liu[†,‡], Jeffrey R. Van Voorst[‡,§,¶] John B. Johnston[§], and Leslie A. Kuhn[‡,§,*]

[†]Department of Chemistry, [‡]Department of Computer Science and Engineering, and [§]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824-1319 USA. [¶]Current address: Symantec Corporation, 2815 Cleveland Ave. North, Minneapolis, MN 55113

*Corresponding author. E-mail: KuhnL@msu.edu

Table S1. 140 non-homologous protein sites binding diverse ligands, containing one cholesterol binding site (in PDB entry 1LRI) and no cholate sites

PDB code	Ligand	Source	Res. (Å)	R-factor	Protein name
1R8S	GDP	B. taurus	1.46	0.159	ADP-ribosylation factor 1
1QXY	M2C	S. aureus	1.04	0.144	Methionyl aminopeptidase
1ECM	TSA	E. coli	2.2	0.192	Endo-oxabicyclic transition state analogue
1KYF	AAchain	M. musculus	1.22	0.154	Alpha-adaptin c
1I24	UPG	A. thaliana	1.2	0.192	Sulfolipid biosynthesis protein sqd1
1AWQ	His-Ala-Gly-Pro-Ile-Ala	H. sapiens	1.58	0.343	Cyclophilin A
1PUJ	GNP	B. subtilis	2.0	0.216	Conserved hypothetical protein ylqf
4UBP	HAE	S. pasteurii	1.55	0.151	Urease, chain A
1CHM	CMS	P. putida	1.9	0.177	Creatine amidinohydrolase
1KEK	HTL	D. africanus	1.9	0.178	Pyruvate-ferredoxin oxidoreductase
1EFY	BZC	G. gallus	2.2	0.194	Poly (ADP-ribose) polymerase
1MSK	SAM	E. coli k12	1.8	0.198	Cobalamin-dependent methionine synthase
1EVL	TSB	E. coli	1.55	0.215	Threonyl-trna synthetase
1JC9	NAG	T. tridentatus	2.01	0.183	Techylectin-5a
1DL5	SAH	T. maritima	1.8	0.182	Protein-l-isoaspartate o-methyltransferase
1GX5	GTP	H. c virus (isolate bk)	1.7	0.193	RNA-directed RNA polymerase
1GK8	CAP	C. reinhardtii	1.4	0.149	Ribulose-1,5 bisphosphate carboxylase larg
1FK5	OLA	Z. mays	1.3	0.135	Nonspecific lipid-transfer protein
1O7N	IND	P. putida	1.4	0.19	Naphthalene 1,2-dioxygenase alpha subunit
1M15	ARG	L. polyphemus	1.2	0.125	Arginine kinase
1KMV	LII	H. sapiens	1.05	0.13	Dihydrofolate reductase
1F20	NAP	R. norvegicus	1.9	0.186	Nitric-oxide synthase
1MXT	FAE	S. sp.	0.95	0.11	Cholesterol oxidase
1GS5	NLG	E. coli	1.5	0.2088	Acetylglutamate kinase
1QD1	FON	S. scrofa	1.7	0.191	Formiminotransferase-cyclodeaminase
1C96	FLC	B. taurus	1.81	0.225	Mitochondrial aconitase

1K3Y	GTX	H. sapiens	1.3	0.148	Glutathione s-transferase a1
1T2D	NAD	P. falciparum	1.1	0.143	L-lactate dehydrogenase
1JET	Lys-Ala-Lys	S. typhimurium	1.2	0.229	Oligopeptide binding protein
1P7T	ACO	E. coli str. k12 substr.	1.95	0.197	Malate synthase G
1KGQ	NPI	M. bovis	2.0	0.179	Tetrahydrodipicolinate N-Succinyltransferase
1DMH	LIO	A. sp.	1.7	0.185	Catechol 1,2-dioxygenase
1XVA	SAM	E. coli	2.2	0.196	Glycine N-methyltransferase
1B37	FAD	Z. mays	1.9	0.199	Polyamine oxidase
1B5E	DCM	E. phage t4	1.6	0.189	Deoxycytidylate hydroxymethylase
1LTZ	HBL	C. violaceum	1.4	0.159	Phenylalanine-4-hydroxylase
1K5N	AAchain	H. sapiens	1.09	0.123	Major histocompatibility complex HLA-B*2709
1H16	DTL	E. coli	1.53	0.145	Formate acetyltransferase 1
1NKI	PPF	P. aeruginosa	0.95	0.148	Probable fosfomycin resistance protein
1G6S	S3P	E. coli	1.5	0.149	EPSP synthase
1LRI	CLR	P. cryptogea	1.45	0.161	Beta-elicitin cryptogein
1R1H	BIR	H. sapiens	1.95	0.211	Neprilysin
1AMU	PHE	B. brevis	1.9	0.213	Gramicidin synthetase 1
1L8B	MGP	M. musculus	1.8	0.224	Eukaryotic translation initiation factor 4E
1PFV	2FM	E. coli	1.7	0.186	Methionyl-tRNA synthetase
1M0K	RET	H. salinarum	1.43	0.134	Bacteriorhodopsin
1UZE	EAL	H. sapiens	1.82	0.188	Angiotensin converting enzyme
1AF7	SAH	S. typhimurium	2.0	0.2	Chemotaxis receptor methyltransferase CheR
1G72	PQQ	M. methylotrophus	1.9	0.161	Methanol dehydrogenase heavy subunit
1QZ5	KAB	O. cuniculus	1.45	0.17	Actin, alpha skeletal muscle
1DTD	Glu	H. sapiens	1.65	0.187	Carboxypeptidase A2
1JHG	TRP	E. coli	1.3	0.127	Trp operon repressor
1CCW	TAR	C. cochlearium	1.6	0.137	Glutamate mutase
1MQO	CIT	B. cereus	1.35	0.222	Beta-lactamase II
1QMG	DMV	S. oleracea	1.6	0.196	Acetohydroxy-acid isomeroreductase
1UFY	MLI	T. thermophilus	0.96	0.11	Chorismate mutase
1KJQ	ADP	E. coli	1.05	0.19	Phosphoribosylglycinamide formyltransferase 2
1CIP	GNP	R. norvegicus	1.5	0.213	GI-alpha-1 subunit

1AYL	OXL	E. coli	1.8	0.195	Phosphoenolpyruvate carboxykinase
1GTE	IUR	S. scrofa	1.65	0.181	Dihydropyrimidine dehydrogenase
1MRJ	ADN	T. kirilowii	1.6	0.173	Alpha-trichosanthin
1PZ4	PLM	A. aegypti	1.35	0.187	Sterol carrier protein 2
1R4U	OXC	A. flavus	1.65	0.157	Uricase
1RQW	TAR	T. daniellii	1.05	0.127	Thaumatin I
2TCT	CTC	E. coli	2.1	0.18	Tetracycline repressor
1VJJ	GDP	H. sapiens	1.9	0.205	Glutamine glutamyltransferase
1PQ7	ARG	F. oxysporum	0.8	0.109	Trypsin
1CZA	G6P	H. sapiens	1.9	0.213	Hexokinase type I
1O2D	NAP	T. maritima	1.3	0.139	Alcohol dehydrogenase, iron-containing
1F0L	APU	C. diphtheriae	1.55	0.188	Diphtheria toxin
1TW6	AAchain	H. sapiens	1.71	0.156	Baculoviral IAP repeat-containing protein 7
2DPM	SAM	S. pneumoniae	1.8	0.238	Adenine-specific methyltransferase
1KA1	A3P	S. cerevisiae	1.3	0.134	Halotolerance protein Hal2
1F5N	GNP	H. sapiens	1.7	0.226	Interferon-induced guanylate-binding protein 1
1HQS	CIT	B. subtilis	1.55	0.202	Isocitrate dehydrogenase
1NVV	GNP	H. sapiens	2.18	0.208	Transforming protein p21/h-ras-1
1UNQ	ITS	H. sapiens	0.98	0.154	Rac-alpha serine/threonine kinase
1KRH	FAD	A. sp.	1.5	0.242	Benzoate 1,2-dioxygenase reductase
1M0W	3GC	S. cerevisiae	1.8	0.172	Glutathione synthetase
1UCD	URA	M. charantia	1.3	0.2	Ribonuclease MC
1HYO	HBU	M. musculus	1.3	0.181	Fumarylacetoacetate hydrolase
1DKX	AAchain	E. coli	2.0	0.206	Substrate binding domain of DNAK
1SOX	MTE	G. gallus	1.9	0.175	Sulfite oxidase
1LB6	AAchain	H. sapiens	1.8	0.203	TNF receptor-associated factor
1I1Q	TRP	S. typhimurium	1.9	0.219	Anthranilate synthase comp. I
1ND4	KAN	K. pneumoniae	2.1	0.206	Aminoglycoside 3'-phosphotransferase
1EU1	MGD	R. sphaeroides	1.3	0.121	Dimethyl sulfoxide reductase
1BX4	ADN	H. sapiens	1.5	0.192	Protein (adenosine kinase)
1NOX	FMN	T. thermophilus	1.59	0.19	NADH oxidase
1HP1	ATP	E. coli	1.7	0.176	5'-nucleotidase

1LKK	AAchain	H. sapiens	1.0	0.133	Human p56 tyrosine kinase
1B4U	DHB	S. paucimobilis	2.2	0.161	Protocatechuate 4,5-dioxygenase
1GZ8	MBP	H. sapiens	1.3	0.153	Cell division protein kinase 2
1EYQ	NAR	M. sativa	1.85	0.237	Chalcone-flavonone isomerase
1TX4	GDP	H. sapiens	1.65	0.169	P50-rhogap
1US0	LDT	H. sapiens	0.66	0.0938	Aldose reductase
1UXY	EPU	E. coli	1.8	0.202	MURB
1J09	ATP	T. thermophilus	1.8	0.199	Glutamyl-tRNA synthetase
1D3V	ABH	R. norvegicus	1.7	0.157	Arginase
1KPF	AMP	H. sapiens	1.5	0.209	Protein kinase C interacting protein
1UUY	PPI	A. thaliana	1.45	0.163	Molybdopterin biosynthesis CNX1
1OUW	MLT	C. sepium	1.37	0.153	Lectin
1HFE	FCY	D. vulgaris	1.6	0.158	Fe-only hydrogenase
1JAK	IFG	S. plicatus	1.75	0.176	Beta-N-acetylhexosaminidase
1UIO	HPR	M. musculus	2.4	0.203	Adenosine deaminase
1P6O	HPY	S. cerevisiae	1.14	0.112	Cytosine deaminase
1KOL	NAD	P. putida	1.65	0.171	Formaldehyde dehydrogenase
1OAI	AAchain	H. sapiens	1.0	0.149	Nuclear RNA export factor
1FCY	564	H. sapiens	1.3	0.134	Retinoic acid receptor
1F3L	SAH	R. norvegicus	2.03	0.209	Protein arginine methyltransferase PRMT3
1N62	MCN	O. carboxidovorans	1.09	0.144	Carbon monoxide dehydrogenase small chain
1QJA	AAchain	H. sapiens	2.0	0.214	14-3-3 Protein zeta
1G2L	T87	H. sapiens	1.9	0.237	Coagulation factor X
2SLI	SKD	M. decora	1.8	0.185	Intramolecular trans-sialidase
1A9X	ORN	E. coli	1.8	0.191	Carbamoyl phosphate synthetase (large chain)
1TBB	ROL	H. sapiens	1.6	0.187	CAMP-specific 3',5'-cyclic phosphodiesterase 4D
1O7Q	UDP	B. taurus	1.3	0.1155	N-acetyllactosaminide
1RLZ	NAD	H. sapiens	2.15	0.199	Deoxyhypusine synthase
1U4G	HPI	P. aeruginosa	1.4	0.18	Elastase
1TL2	NAG	T. tridentatus	2.0	0.162	Tachylectin-2
1RKD	RIB	E. coli	1.84	0.221	Ribokinase
1Q79	3AT	B. taurus	2.15	0.205	Poly(a) polymerase alpha
1PP9	SMA	B. taurus	2.1	0.25	Ubiquinol-cytochrome c reductase complex core protein
1E8G	FCR	P. simplicissim.	2.1	0.218	Vanillyl-alcohol oxidase
1L5O	2MP	S. enterica	1.6	0.174	CobT

1OEW	Ser-Thr	C. parasitica	0.9	0.121	Endothiapepsin
1H8E	ALF	B. taurus	2.0	0.201	Bovine mitochondrial F1-ATPase
1BGV	GLU	C. symbiosum	1.9	0.173	Glutamate dehydrogenase
1USC	FMN	T. thermophilus	1.24	0.203	Putative styrene monooxygenase small comp.
1MGP	PLM	T. maritima	2.0	0.202	Hypothetical protein tm841
1QNF	HDF	S. elongatus	1.8	0.197	Photolyase
1C1D	NAD	R. sp.	1.25	0.195	L-phenylalanine dehydrogenase
1UW6	NCT	L. stagnalis	2.2	0.22386	Acetylcholine-binding protein
1G55	SAH	H. sapiens	1.8	0.21	DNA cytosine methyltransferase DNMT2
1LUG	SUA	H. sapiens	0.95	0.119	Carbonic anhydrase II
1UF5	CDT	A. sp.	1.6	0.178	N-carbamyl-d-amino acid amidohydrolase
1V7R	CIT	P. horikoshii	1.4	0.202	Hypothetical protein ph1917
1D0C	INE	B. taurus	1.65	0.213	Bovine endothelial nitric oxide synthase heme domain
5CSM	TRP	S. cerevisiae	2.0	0.186	Chorismate mutase
1P5D	G1P	P. aeruginosa	1.6	0.157	Phosphomannomutase

Table S2. Putative cholesterol binding sites in class A GPCRs[1]

PDB code (motif matched)	Source	Res. (Å)	Protein name	Alignment RMSD[β] with respect to PDB 2RH1
2RH1 (strict-CCM[α])	H. sapiens	2.4 Å	Beta adrenoceptor type 2 (ADRB2)	0.00 Å
3EML (strict-CCM[α])	H. sapiens	2.6 Å	Adenosine type 2A receptor (ADORA2A)	0.43 Å
3PBL (strict-CCM[α])	H. sapiens	2.9Å	Dopamine vertebrate type 3 receptor (DRD3)	0.48 Å
2KS9 (strict-CCM[α])	H. sapiens	NMR	Vertebrate tachykinin receptor (TACR1)	0.32 Å
2Y00 (CCM[α])	M. gallopavo	2.5 Å	Beta adrenoceptor type 1 (ADRB1)	0.29 Å
3RZE (CCM[α])	H. sapiens	3.1 Å	Histamine type 1 receptor	0.59 Å
1U19	B. taurus	2.2 Å	Rhodopsin	0.63 Å
2Z73	T.pacificus	2.5 Å	Rhodopsin	0.71 Å
3ODU	H. sapiens	2.5 Å	C-X-C chemokine receptor type 4 (CXCR4)	2.48 Å
3V2W	H. sapiens	3.35 Å	Sphingosine-1-phosphate receptor (EDG)	0.63 Å
3UON	H. sapiens	3.0 Å	M2 Human muscarinic acetylcholine receptor	0.44 Å
4DJH	H. sapiens	2.9 Å	κ-opioid receptor	0.67 Å

[α] strict-CCM: R/K -$(X)_{2-6}$-I/V/L-$(X)_3$-W/Y on one helix and F/Y on the neighboring helix[13]; CCM: R/K -$(X)_{2-6}$-I/V/L-$(X)_3$-W/Y[13]. Entries without motif notations belong to class A GPCRs but were not included in reference 12 or Table 2 predictions in the present manuscript.

[β]The alignment RMSD is based on relative positions of backbone atoms (N, C_α, C and O) of residues within 9 Å of cholesterol.

[1]Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V. P.; Chien, E. Y.; Velasquez, J.; Kuhn, P.; Stevens, R. C. A Specific Cholesterol Binding Site is Established by the 2.8 Å Structure of the Human B2-Adrenergic Receptor. *Structure* **2008**, *16*, 897–905.

Table S3. Diverse non-cholesterol, non-cholate lipid binding sites

PDB code	Ligand	Source	Res. (Å)	R-factor	Protein Name
1A28	Progesterone	H. sapiens	1.8 Å	0.191	Progesterone receptor
2AA6	Progesterone	H. sapiens	2.0 Å	0.197	Mineralocorticoid receptor
2ABA	Progesterone	E. cloacae	1.0 Å	0.129	Pentaerythritol tetranitrate reductase
2HZQ	Progesterone	H. sapiens	1.8 Å	0.189	Apolipoprotein D
1AQU	Estradiol	M. musculus	1.6 Å	0.218	Estrogen sulfotransferase
1E6W	Estradiol	R. norvegicus	1.7 Å	0.184	Short chain 3-hydroxyacyl-CoA dehydrogenase
1JGL	Estradiol	M. musculus	2.2 Å	0.199	Ig kappa-chain
1LHU	Estradiol	H. sapiens	1.8 Å	0.204	Sex hormone-binding globulin
3OLL	Estradiol	H. sapiens	1.5 Å	0.177	Estrogen receptor beta
2AM9	Testosterone	H. sapiens	1.6 Å	0.191	Androgen receptor
1J96	Testosterone	H. sapiens	1.2 Å	0.181	3-Alpha-hydroxysteroid dehydrogenase type 3
3KDM	Testosterone	H. sapiens	1.5 Å	0.181	Immunoglobulin light chain
4EIY	Oleic acid	H. sapiens	1.8 Å	0.176	Adenosine receptor A2a
3DDL	PX4	S. ruber	1.90 Å	0.247	Xanthorhodopsin
3DDL	PCW	S. ruber	1.90 Å	0.247	Xanthorhodopsin
2Z73	PC1	T. pacificus	2.50 Å	0.188	Rhodopsin
3UTW	MC3	H. sp.	2.40Å	0.206	Bacteriorhodopsin
3UTV	MC3	H. sp.	2.06Å	0.197	Bacteriorhodopsin

Table S4. Sites in 109 low-homology bacterial membrane protein sites analyzed as potential false positive cases for cholesterol (CLR) or cholate (CHD) binding. Sites predicted to match the CholMine cholesterol or cholate site conserved interactions are noted in the third column. The last column indicates whether the crystallographic ligand at the prediction site (second column) was of lipid or lipid-like (L), drug-like (D), polar (P), or intermediate character (e.g., P/L for a polar lipid group). 73% of the sites contained lipids or partly lipidic molecules.

PDB entry	Ligand site analyzed	Prediction (CLR, CHD, or neither)	Crystal structure ligand type
1LGH	LYC A97	CLR,CHD	L
1M56	PEH A2009	CLR,CHD	L
1QD5	BOG A500	CLR	L
1U7G	BOG A400	CLR	L
1YC9	BOG A1001	CLR	L
2ERV	CXE A300	CLR	L
2YEV	5PL A900	CLR	L
3GP6	SDS A163	CLR	L
3RKO	LFA L614	CLR,CHD	L
4H44	7PH C303	CLR	L
4IL6	DGD C515	CLR,CHD	L
1B12	1PN B1001	---	D
1CWV	CIT A994	---	P
1EHK	BNG A901	---	L
1J79	NCD A950	---	P
1JB0	BCR A4001	---	L
1K4C	F09 A2001	---	L
1KMO	HTO A759	---	L
1KQF	MGD A1018	---	D/P
1LDF	GOL A476	---	P

1NKZ	RG1 A404	---	L
1Q16	MD1 A1300	---	D/P
1QFG	DDQ A1100	---	L
1QJP	C8E A1172	---	L
1UJW	GP1 A801	---	P
1UYN	CXE X2085	CHD	L
1XEZ	BOG A999	---	L
1XIO	RET A301	---	L
1XKW	LDA A2001	---	L
1Y4Z	MD1 A1800	---	D/P
2A65	LEU A601	---	D
2BL2	UMQ A1162	---	L
2BS2	FAD A1656	---	D/P
2GSK	LDA A800	---	L
2GSM	DMU A5001	---	L
2GUF	MPG A701	---	L
2HDI	LDA A664	---	L
2IWV	TAM B1289	---	D
2J58	OCT A600	---	L
2NS1	BOG A601	---	L
2O4V	C8E A1295	---	L
2OQO	EPE A244	---	D
2POR	C8E A545	CHD	L
2QCU	TAM A805	---	D
2QI9	1PE C800	---	L
2SQC	C8E A632	---	L

2VDF	OCT A1254	---	L
2VPZ	MGD A1765	---	D/P
2VQG	MRD B1097	---	D
2WDQ	CBE C1130	CHD	D
2WIE	CVM A102	---	L/D
2WJN	MQ7 M1328	---	L
2WJR	EPE A1217	---	D
2WSW	CM5 A1505	---	L/D
2X27	C8E X1216	---	L
2X2V	DPV A200	---	L
2X55	C8E A1293	---	L
2XCI	PG4 A1353	---	D/P
2XOV	BNG A503	---	L
2YHC	URE A1234	---	P
2YNK	OCT A1001	---	L
2ZFG	C8E A342	---	L
3B9W	BOG A408	---	L
3BS0	C8E A501	---	L
3CSL	GOL A867	---	P
3DDL	UNL A1402	---	L
3DWN	LDA A502	CHD	L
3DWO	C8E X453	CHD	L
3DZM	C8E A209	---	L
3FID	CXE A304	---	L
3HB3	LMT A568	---	L
3HYW	DCQ A500	CHD	L

3JQO	MPD D1	---	D
3KDS	NHX E998	CHD	D
3KLY	BOG A281	---	L
3L1L	BNG A447	---	L
3L7I	EDO B731	---	L/P
3M71	BOG A315	---	L
3OUF	MPD A501	---	L/P
3QE7	URA A430	---	D/P
3QRA	C8E A1	---	L
3RLB	VIB A191	---	D
3RLF	UMQ E5004	---	L
3RQW	ACH A323	---	P/D
3RVY	PX4 A4001	CHD	L
3SZV	C8E A385	---	L
3TIJ	URI A419	---	D/P
3USE	GOL L605	---	G
3V8X	C8E A1001	---	L
3WO6	OLC A302	---	L
4AFK	78M A1510	---	L
4DVE	BTN A201	---	D/L
4E1S	OLB A502	CHD	L
4EHW	MPD A402	---	D/L
4GBY	BNG A505	---	L
4GEY	DMU A510	---	L
4IKV	PG4 A613	---	L/P
4JR9	GYP A501	---	P

4MT4	3PK A1008	---	L
4N7W	MPG A402	---	L
4NHR	PEG A301	---	P/L
4NM9	FAD A2001	---	P
4NV5	U10 A501	CHD	L
4P1X	MPD A401	---	P/L
4PR7	OCT A301	---	L
4Q35	LDA A2004	---	L
4QNC	MYS A104	---	L
2J7A	LMT C1005	---	L
3WU2	SQD A412	---	L