Supporting Information

Oberai et al. 10.1073/pnas.0906390106

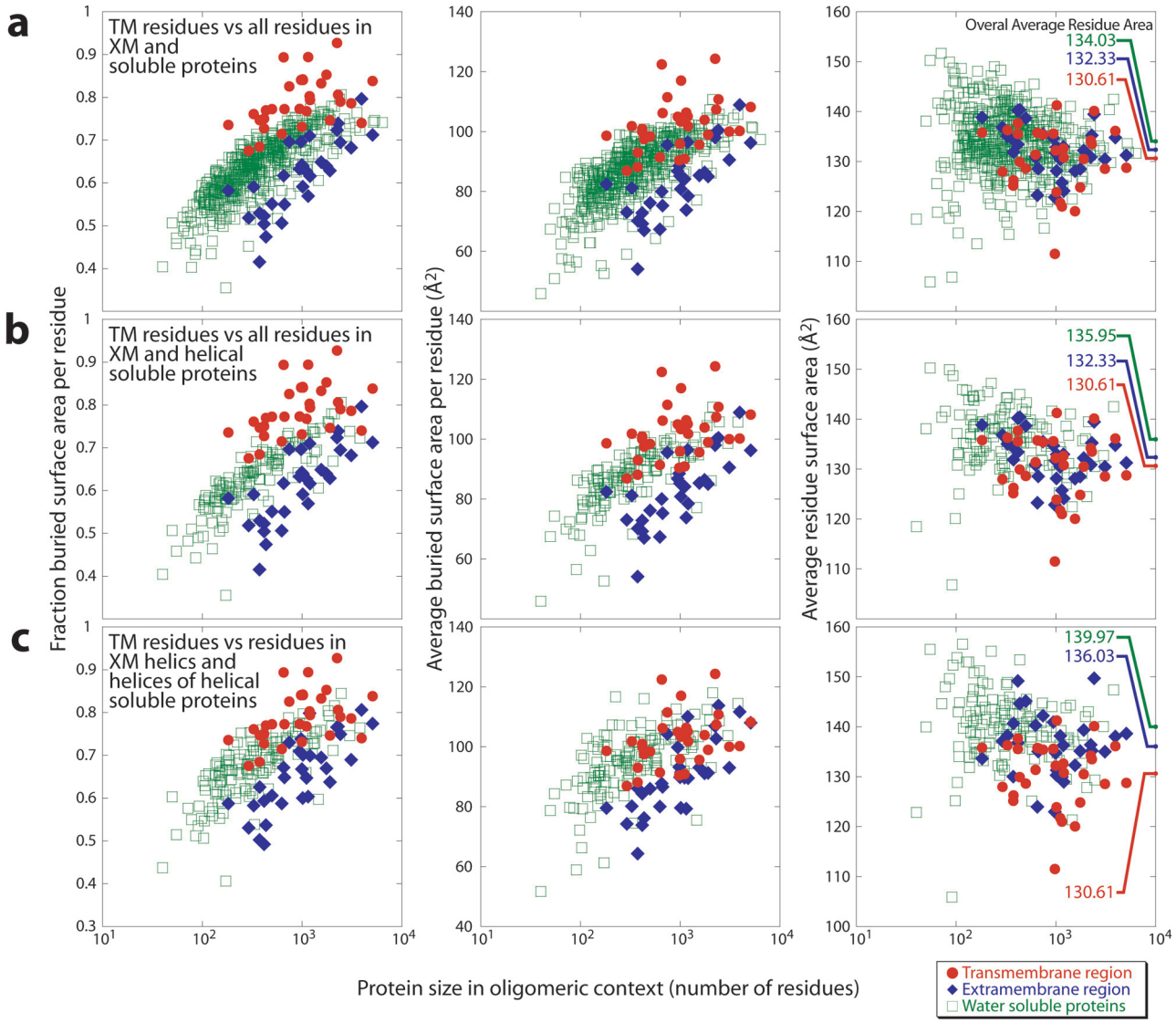

Fig. S1. More extensive burial of smaller residues in the transmembrane (TM) region compared with the extramembrane (XM) region of membrane proteins and soluble proteins. (*A*) Comparison of burial and size trend for residues in the TM and XM regions of membrane proteins and water-soluble proteins. (*B*) Comparison of burial and size trend for residues in the TM regions and XM regions of membrane proteins and helical water-soluble proteins. (*C*) Comparison of burial and size trend for residues in the TM region of helical membrane proteins, helices in the XM region of membrane proteins, and helices of helical water-soluble proteins. In each comparison, the fraction buried surface area (*Left*), average buried surface area (*Center*), and the average residue surface area for each protein (*Right*) are shown as a function of the size of the protein quaternary structure. The overall average sizes, in terms of surface area $(A°)^2$), for residues in the different regions of membrane and soluble proteins are listed on the *y*-axis of each panel on the right, color coded according to the region of the protein, red for the transmembrane region, blue for the extramembrane region, and green for water soluble proteins.

Table S1. Conservation scores for the transmembrane (TM) and extramembrane (XM) regions of 21 chains from 19 helical membrane proteins. Shown are scores for the overall protein and for different levels of burial, divided into quartiles: 1 (0–52% buried), 2 (52.1–81.3% buried), 3 (81.4–97.3% buried), and 4 (97.4–100% buried)

PDB	TM-1	XM-1	TM-2	XM-2	TM-3	XM-3	TM-4	XM-4	TM-all	XM-all
1dxrL,M	0.296305	0.222822	0.34599	0.313049	0.399967	0.396724	0.525527	0.459542	0.425028	0.3645
1ehkA	0.199539	0.248401	0.326694	0.312138	0.383285	0.501342	0.510493	0.53082	0.418585	0.401242
1 gzmA	0.464341	0.536739	0.64071	0.549307	0.69675	0.608558	0.792465	0.72313	0.663549	0.584056
1l7vA	0.098776	0.121704	0.282213	0.279717	0.423054	0.361636	0.461246	0.597844	0.337573	0.265995
1ldfA	0.234858	0.167305	0.32347	0.181379	0.430394	0.437722	0.495069	0.35701	0.414289	0.254564
1m0lA	0.344908	0.235004	0.299357	0.258367	0.424457	0.372982	0.608916	0.361141	0.43086	0.281379
1okcA	0.244696	0.180098	0.439857	0.322401	0.496763	0.412846	0.495532	0.470482	0.416844	0.282618
1otsA	0.173443	0.157438	0.207157	0.213695	0.330494	0.264434	0.508889	0.455589	0.375449	0.247942
1pw4A	0.301328	0.280996	0.3406	0.451844	0.438463	0.448966	0.464577	0.444444	0.395841	0.354883
1q16C	0.317565	0.351298	0.404779	0.414255	0.492464	0.608554	0.68116	0.721429	0.472163	0.512668
1q90B,D	0.410703	0.405503	0.441088	0.511445	0.579315	0.642908	0.620099	0.702981	0.546047	0.560058
1rwtB	0.283394	0.435005	0.439983	0.479176	0.615746	0.53954	0.59478	0.578442	0.567921	0.488342
1u7 gA	0.054883	0.155036	0.262759	0.197441	0.299348	0.366371	0.43874	0.444563	0.351924	0.242558
1wpgA	0.141784	0.158355	0.202454	0.27818	0.335087	0.408048	0.290434	0.565309	0.246082	0.321421
1zcdA	0.265961	0.240172	0.411253	0.351812	0.577676	0.431479	0.644919	0.691898	0.485044	0.339335
2a65A	0.194267	0.191029	0.191773	0.221084	0.301876	0.336162	0.394519	0.493177	0.309973	0.286671
2cfqA	0.503081	0.418334	0.605101	0.479837	0.624325	0.641374	0.673874	0.782161	0.604639	0.489257
2 gifA	0.284758	0.259915	0.39479	0.368243	0.592823	0.445265	0.635944	0.524079	0.503077	0.386634
2nwlA	0.202072	0.24317	0.296256	0.288103	0.436737	0.454238	0.466691	0.471693	0.374573	0.336733
OVER-ALL	0.264035	0.263596	0.360857	0.340604	0.467317	0.456797	0.542309	0.546091	0.438919	0.368466

Table S2. Disease-associated nonsynonymous SNPs that were mapped onto membrane protein structures, based on regions that could be well aligned

Protein structure (PDB ID)	Disease variant location	Residues associated with disease variants	Fraction residue area buried	Disease gene (OMIM ID, % ID if applicable), disease mutation, disease*
Rhodopsin (1GZM)	Trans-membrane domain	F45	0.42	Rhodopsin (180380, 100%) F45L, retinitis
		G51	1	Rhodopsin (180380, 100%) G51R, retinitis
		P53	0.53	Rhodopsin (180380, 100%) P53R, retinitis
		T58	1	Rhodopsin (180380, 100%) T58R, retinitis
		V87	1	Rhodopsin (180380, 100%) V87D, retinitis
		G89	0.63	Rhodopsin (180380, 100%) G89D, retinitis
		G90	0.98	Rhodopsin (180380, 100%) G90D, retinitis
		T94	1	Rhodopsin (180380, 100%) T94I, night blindness
		C110	0.98	Rhodopsin (180380, 100%) C110Y, retinitis
		G114	1	Rhodopsin (180380, 100%) G114D, retinitis
		A164	1	Rhodopsin (180380, 100%) A164E, retinitis
		P171	0.95	Rhodopsin (180380, 100%) P171S, retinitis
		M207	1	Rhodopsin (180380, 100%) M207R, retinitis
		H211	1	Rhodopsin (180380, 100%) H211P, retinitis
		P267	0.56	Rhodopsin (180380, 100%) P267L, retinitis
		A292	0.99	Rhodopsin (180380, 100%) A292E, retinitis
		K296	0.98	Rhodopsin (180380, 100%) K296E, retinitis
	Soluble domain	N15	0.43	Rhodopsin (180380, 100%) N15S, retinitis
		T17	0.12	Rhodopsin (180380, 100%) T17M, retinitis
		P23	1	Rhodopsin (180380, 100%) P23H, retinitis
				Rhodopsin (180380, 100%) P23A, retinitis
		G106	0.71	Rhodopsin (180380, 100%) G106W, retinitis
				Rhodopsin (180380, 100%) G106R, retinitis
		R135	0.94	Rhodopsin (180380, 100%) R135L, retinitis
				Rhodopsin (180380, 100%) R135W, retinitis
		E150	0.43	Rhodopsin (180380, 100%) E150K, retinitis
		Y178	0.94	Rhodopsin (180380, 100%) Y178C, retinitis
		G182	0.93	Rhodopsin (180380, 100%) G182S, retinitis
		D190	1	Rhodopsin (180380, 100%) D190G, retinitis
				Rhodopsin (180380, 100%) D190N, retinitis
				Rhodopsin (180380, 100%) D190Y, retinitis
Ca^{2+} ATPase#	Trans-membrane domain	C268	0.55 ± 0.4	AT2A2 (108740, 84%) C268F, Darrier disease
		A305	1 ± 0.004	AT2C1 (604384, 32%) A304T, Hailey-Hailey disease
		N768	0.95 ± 0.009	AT2A2 (108740, 84%) N767S, Darrier disease
		G770	1 ± 0.02	AT2A2 (108740, 84%) G769R, Darrier disease
		P789	0.93 ± 0.04	AT2A1 (108730, 100%) P789L, Brody myopathy
	Soluble domain	G23	0.91 ± 0.01	AT2A2 (108740, 84%) G23E, Darrier disease
		R131	0.88 ± 0.01	AT2A2 (108740, 84%) R131Q, Darrier disease
		V459	0.86 ± 0.07	AT2B2 (108733, 31%) V587M, deafness
		C525	0.9 ± 0.027	AT2C1 (604384, 32%) C490F, Hailey-Hailey disease
		C561	0.73 ± 0.05	AT2A2 (108740, 84%) C560R, Darrier disease
		P603	0.78 ± 0.03	AT2A2 (108740, 84%) P602L, acrokeratosis
Voltage-gated K^+ channel Kv1.2 (2R9R)	Trans-membrane domain	V170	0.75	Kv1.1 (176260, 80%) V174F, episodic ataxia
		I173	0.99	Kv1.1 (176260, 80%) I177R, episodic ataxia
		F180	1	Kv1.1 (176260, 80%) F184C, episodic ataxia
		T227	0.84	Kv1.1 (176260, 80%) T226R, myokymia1
				Kv1.1 (176260, 80%) T226C, myokymia1
				Kv1.1 (176260, 80%) T226A, myokymia1
		R240	0.96	Kv1.1 (176260, 80%) R239S, episodic ataxia
		R296	0.95	Kv3.3 (176264, 45%) R420H, spinocerebral ataxia
		E323	0.9	Kv1.1 (176260, 80%) E325D, episodic ataxia
		V402	0.89	Kv1.1 (176260, 80%) V404I, episodic ataxia
	Soluble domain	A243	0.94	Kv1.1(176260, 80%) A242P, myokymia1
		P245	0.56	Kv1.1(176260, 80%) P244H, myokymia1
		F250	0.98	Kv1.1(176260, 80%) F249I, episodic ataxia
		V406	0.85	Kv1.1(176260, 80%) V408A, episodic ataxia
Voltage-gated K^+ channel Kv1.2 (2R9R)	Trans-membrane domain	A162	0.984	KCNQ1, C122Y, long QT
		I173	0.987	KCNQ1, V133I, long QT
		S176	0.988	KCNQ1, C136F, long QT
		I177	0.971	KCNQ1, L137F, long QT

Protein structure (PDB ID)	Disease variant location	Residues associated with disease variants	Fraction residue area buried	Disease gene (OMIM ID, % ID if applicable), disease mutation, disease*
		F180	1	KCNQ1, S140G, atrial fibrillation
		T184	1	KCNQ1, T144A, long QT
		F223	0.926	KCNQ1, F157C, long QT
		E226	0.967	KCNQ1, E160K, long QT
		S234	0.819	KCNQ1, G168R, long QT
		L238	0.465	KCNQ1, V172M, long QT
		R240	0.955	KCNQ1, R174C, long QT
				KCNQ1, R174H, long QT
				KCNQ1, R174P, long QT
		D259	0.954	KCNQ1, D202N, long QT
				KCNQ1, D202H, long QT
		V261	0.726	KCNQ1, I204M, long QT
				KCNQ1, I204F, long QT
		F301	0.655	KCNQ1, L239P, long QT
		S304	0.852	KCNQ1, D242N, long QT
				KCNQ1, D242E, long QT
		R305	0.764	KCNQ1, R243C, long QT
				KCNQ1, R243P, long QT
				KCNQ1, R243H, long QT
		S320	1	KCNQ1, H258N, long QT
				KCNQ1, H258R, long QT
		M321	0.207	KCNQ1, R259H, long QT
				KCNQ1, R259C, long QT
				KCNQ1, R259L, long QT
		E323	0.899	KCNQ1, E261D, long QT
				KCNQ1, E261K, long QT
		L324	0.765	KCNQ1, L262V, long QT
		I328	0.451	KCNQ1, L266P, long QT
		L331	0.951	KCNQ1, G269S, long QT
				KCNQ1, G269D, long QT
		G334	0.802	KCNQ1, G272D, long QT
Voltage-gated K^+ channel Kv1.2 (2R9R)	Trans-membrane domain	V335	0.888	KCNQ1, L273F, long QT
				KCNQ1, L273R, long QT
		I336	0.962	KCNQ1, I274V, sudden infant death syndrome
		L337	0.73	KCNQ1, F275S, long QT
		S339	1	KCNQ1, S277L, long QT
				KCNQ1, S277W, long QT
		S340	0.994	KCNQ1, Y278H, long QT
		V342	1	KCNQ1, V280A, long QT
				KCNQ1, V280E, long QT
		Y343	0.916	KCNQ1, Y281C, long QT
		P358	0.709	KCNQ1, A300T, long QT
		A360	1	KCNQ1, A302V, long QT
				KCNQ1, A302T, long QT
		W362	0.951	KCNQ1, W304R, long QT
		W363	0.991	KCNQ1, W305S, long QT
		A364	1	KCNQ1, G306V, long QT
				KCNQ1, G306R, long QT
		V365	0.872	KCNQ1, V307L, long QT
		V366	0.946	KCNQ1, V308D, long QT
		S367	0.994	KCNQ1, T309I, long QT
				KCNQ1, T309R, long QT
		M368	0.988	KCNQ1, V310I, long QT
		T369	0.975	KCNQ1, T311I, long QT
		G383	1	KCNQ1, G325R, long QT
		A399	0.689	KCNQ1, A341E, long QT
				KCNQ1, A341V, long QT
		L400	0.953	KCNQ1, L342F, long QT
		P401	0.992	KCNQ1, P343S, long QT
				KCNQ1, P343L, long QT
				KCNQ1, P343R, long QT
		V402	0.892	KCNQ1, A344E, long QT

Protein structure (PDB ID)	Disease variant location	Residues associated with disease variants	Fraction residue area buried	Disease gene (OMIM ID, % ID if applicable), disease mutation, disease*
				KCNQ1, A344V, long QT
		P403	0.586	KCNQ1, G345E, long QT
				KCNQ1, G345R, long QT
Voltage-gated K^+ channel Kv1.2 (2R9R)	Soluble domain	P186	0.639	KCNQ1, E146K, long QT
		C244	0.917	KCNQ1, A178T, long QT
				KCNQ1, A178P, long QT
		P245	0.562	KCNQ1, G179S, long QT
		F250	0.976	KCNQ1, F193L, long QT
		F251	0.351	KCNQ1, A194P, long QT
		L310	0.895	KCNQ1, W248R, long QT
		I312	0.99	KCNQ1, L250H, long QT
		L313	0.788	KCNQ1, L251P, long QT
		T316	1	KCNQ1, V254M, long QT
				KCNQ1, V254L, long QT
		E346	1	KCNQ1, E284K, long QT
		E349	0.508	KCNQ1, A287E, long QT
		T370	0.848	KCNQ1, T312I, long QT
		V371	0.992	KCNQ1, I313M, long QT
		G272	1	KCNQ1, G314R, long QT
				KCNQ1, G314A, long QT
				KCNQ1, G314D, long QT
				KCNQ1, G314C, long QT
				KCNQ1, G314S, long QT
		Y373	0.959	KCNQ1, Y315S, long QT
				KCNQ1, Y315C, long QT
		G374	0.25	KCNQ1, G316R, long QT
				KCNQ1, G316E, long QT
		D375	0.481	KCNQ1, D317G, long QT
				KCNQ1, D317N, long QT
		M376	0.883	KCNQ1, K318N, long QT
		P378	0.987	KCNQ1, P320A, long QT
		T380	0.796	KCNQ1, T322A, long QT
				KCNQ1, T322M, long QT
		S407	0.217	KCNQ1, S349W, long QT
				KCNQ1, S349P, long QT
		N408	0.736	KCNQ1, G350R, long QT
		F409	0.995	KCNQ1, F351S, long QT
		Y411	0.412	KCNQ1, L353P, long QT
		R415	0.417	KCNQ1, Q357R, long QT
Voltage-gated K^+ channel Kv1.2 (2R9R)	Trans-membrane domain	R299	0.987	hERG, K525N, long QT
		K302	0.943	hERG, R528P, long QT
		R305	0.764	hERG, R531Q, long QT
		L326	0.869	hERG, L552S, long QT
		F332	0.527	hERG, A558P, long QT
		I333	0.93	hERG, L559H, long QT
		V335	0.888	hERG, A561V, long QT
				hERG, A561T, long QT
				hERG, A561P, long QT
		I336	0.962	hERG, H562P, long QT
		F338	0.856	hERG, L564P, long QT
		S340	0.994	hERG, C566S, long QT
		V342	1	hERG, W568C, long QT
				hERG, W568R, long QT
		Y343	0.916	hERG, Y569H, long QT
		A345	0.526	hERG, I571L, long QT
				hERG, I571V, long QT
		I357	1	hERG, Y611H, long QT
		P358	0.709	hERG, V612L, long QT
		D359	0.468	hERG, T613M, long QT
		A360	1	hERG, A614V, long QT
		F361	0.662	hERG, L615V, long QT
				hERG, L615F, long QT

Protein structure (PDB ID)	Disease variant location	Residues associated with disease variants	Fraction residue area buried	Disease gene (OMIM ID, % ID if applicable), disease mutation, disease*
		S367	0.994	hERG, S621R, long QT
				hERG, S621N, long QT
		M368	0.988	hERG, L622F, long QT
		T369	0.975	hERG, T623I, long QT
		I381	0.124	hERG, N635I, long QT
		G383	1	hERG, E637K, long QT
				hERG, E637D, long QT
		K384	0.705	hERG, K638E, long QT
		V386	0.834	hERG, F640L, long QT
		G287	0.927	hERG, S641F, long QT
		C390	1	hERG, V644F, long QT
		A391	0.929	hERG, M645V, long QT
				hERG, M645L, long QT
		V402	0.892	hERG, F656, long QT
Voltage gated K^+ channel Kv1.2 (2R9R)	Soluble domain	K308	0.359	hERG, R534C, long QT
				hERG, R534L, long QT
		E346	1	hERG, G572C, long QT
				hERG, G572D, long QT
				hERG, G572S, long QT
				hERG, G572R, long QT
		G349	0.508	hERG, E575G, long QT
		S356	1	hERG, R582C, long QT
				hERG, R582L, long QT
		V371	0.992	hERG, V625E, long QT
		G372	1	hERG, G626V, long QT
				hERG, G626S, long QT
				hERG, G626A, long QT
		Y373	0.959	hERG, F627L, long QT
		G374	0.25	hERG, G628V, long QT
				hERG, G628S, long QT
		D375	0.481	hERG, N629D, long QT
				hERG, N629S, long QT
				hERG, N629K, long QT
		M376	0.883	hERG, V630L, long QT
				hERG, V630A, long QT
		V377	0.903	hERG, S631A, long QT
		P378	0.987	hERG, P632S long QT
		T379	0.411	hERG, N633D, long QT
				hERG, N633S, long QT
		V406	0.849	hERG, S660L, long QT
KCNQ1 (Kv7.1) C-terminus (3BJ4)	Soluble domain	G584	0	KCNQ1, G584S, long QT
		N586	0.127	KCNQ1, N586D, long QT
		T587	0.327	KCNQ1, T587M, Long QT
		G589	0.473	KCNQ1, G589D, long QT
		A590	0.727	KCNQ1, A590T, long QT
		R591	0.801	KCNQ1, R591H, long QT
		R594	0.652	KCNQ1, R594Q, long QT

*KCNQ1 and hERG mutation data were obtained from Jackson HA, Accili EA (2008) Evolutionary analyses of KCNQ1 and HERG voltage-gated potassium channel sequences reveal location-specific susceptibility and augmented chemical severities of arrhythmogenic mutations. *BMC Evolutionary Biology* 8:188. Their alignment data with Kv1.2 were obtained from Smith JA, Vanoye CG, George AL, Jr., Meiler J, Sanders CR (2007) Structural models for the KCNQ1 voltage-gated potassium channel. *Biochemistry* 46(49):14141–14152 and from Wynia-Smith SL, Gillian-Daniel AL, Satyshur KA, Robertson GA (2008) hERG gating microdomains defined by S6 mutagenesis and molecular modeling. *J Gen Physiol* 132(5):507–520.

#Average for 5 different conformations of Ca^{2+} ATPase: 1wpg, 1wpe, 1su4, 1t5s, and 2agv.

Table S3. Helical membrane proteins whose oligomeric structures are used in the study

PDB	Resolution (Å)	No. of residues	Source	Oligomer State	No. transmembrane helices	Transmembrane fraction	Description
1dxr	2.0	1187	*R. viridis*	Hetero-tetramer	11	0.2	Photosynthetic reaction center
1ehk	2.4	743	*T. thermophilus*	Hetero-trimer	15	0.45	Cytochrome c oxidase
1fft	3.5	943	*E. coli*	Hetero-trimer	19	0.42	Ubiquinol oxidase
1 gzm	2.7	329	*B. taurus*	Monomer	7	0.5	Rhodopsin: rhodopsin-like GPCR
1jb0	2.5	2243	*S. elongatus*	Hetero-dodecamer	32	0.31	Photosynthetic reaction centers: photosystem I
1l7v	3.2	1110	*E. coli*	Hetero-tetramer	22	0.32	ABC transporter BtuCD
1ldf	2.1	1016	*E. coli*	Homo-tetramer	32	0.55	Glycerol uptake facilitator (GLPF): aquaporin-like
1m0l	1.47	665	*H. salinarium*	Homo-trimer	21	0.66	Bacteriorhodopsin: rhodopsin-like
1okc	2.2	292	*B. taurus*	Monomer	6	0.5	Mitochondrial ADP/ATP carrier
1ots	2.5	1752	*M. musculus*	Hetero-hexamer	28	0.52	ClC chloride channel
1pw4	3.3	434	*E. coli*	Monomer	12	0.61	Glycerol-3-phosphate transporter (MFS)
1q16	1.9	3952	*E. coli*	Hetero-hexamer	10	0.06	Respiratory nitrate reductase: cytochrome b-like
1q90	3.1	1904	*C. reinhardtii*	Hetero-18mer	26	0.32	Cytochrome b6f: cytochrome b-like
1rh5	3.2	498	*M. jannaschii*	Hetero-trimer	12	0.41	Translocase SecY protein-conducting channel
1rwt	2.7	654	*S. oleracea*	Homo-trimer	9	0.29	Light-harvesting complex II photosynthesis
1u7 g	1.4	1149	*E. coli*	Homo-trimer	33	0.6	Ammonia channel AmtB
1wpg	2.3	994	*O. cuniculus*	Monomer	10	0.18	Ca-ATPase (P-type ATPase)
1yce	2.4	979	*I. tartaricus*	Homo-undecamer	22	0.57	F-type Na-ATPase
1yew	2.8	2424	*M. capsulatus*	Hetero-nonamer	42	0.28	Methane monooxygenase: oxidoreductase
1zcd	3.4	376	*E. coli*	Monomer	14	0.6	Sodium/proton antiporter (NhaA)
2a65	1.65	1018	*A. aeolicus*	Homo-dimer	28	0.5	Neurotransmitter Na^+/Cl- dependent transporter LeuT
2ahy	2.4	416	*B. cereus*	Homo-tetramer	12	0.57	Potassium voltage-gated ion channel
2axt	3.0	5078	*T. elongatus*	Hetero-34mer	72	0.28	Photosystem II: Photosynthetic reaction centers
2bl2	2.1	1560	*E. hirae*	Homo-decamer	40	0.65	V-type Na ATPase: hydrolase
2bs2	1.78	2302	*W. succinogenes*	Hetero-hexamer	10	0.09	Oxidoreductase: quinol-fumarate reductase
2cfq	2.95	417	*E. coli*	Monomer	12	0.66	Lactose permease (MFS)
2 gfp	3.5	375	*E. coli*	Monomer	12	0.70	Multidrug transporter EmrD (MFS)
2 gif	2.9	3108	*E. coli*	Homo-trimer	36	0.23	AcrB Acriflavine multidrug resistance transporter
2ic8	2.1	182	*E. coli*	Monomer	6	0.57	Protease glpG: Rhomboid protein
2nwl	3.0	1203	*P. horikoshii*	Homo-trimer	39	0.5	Proton glutamate symport protein
2oar	3.5	625	*M. tuberculosis*	Homo-pentamer	10	0.45	Large conductance mechanosensitive channel MscL