Title: X-ray diffraction reveals the intrinsic difference in the physical properties of membrane and soluble proteins

Authors: Xavier Robert[1a], Josiane Kassis-Sahyoun[1a], Nicoletta Ceres[1a], Juliette Martin[a], Michael R. Sawaya[b], Randy J. Read[c], Patrice Gouet[a], Pierre Falson[a], and Vincent Chaptal[a*]

Legends supplemental-information

Fig. S1: comparison of anisotropy between soluble and membrane proteins. **a/** Calculation on amplitudes. Plot of all the entries on a scatter plot. Black bars show the average anisotropy of the sample and brackets display the 95% confidence interval of the population mean. Student's T-test with Welch correction (two-tailed p-value < 0.0001), and Mann-Whitney difference of medians (two-tailed p-value < 0.0001). **b/** Calculation on intensities. Plot of all the entries on a scatter plot. Black bars show the average anisotropy of the sample and brackets display the 95% confidence interval of the population mean. Student's T-test with Welch correction (two-tailed p-value < 0.0001), and Mann-Whitney difference of medians (two-tailed p-value < 0.0001).
The probability distribution function of these two types of proteins can be fitted by a Weibull distribution : y = a*b*x^(b-1)*exp(-a*x^b) ;

	On amplitudes (F)	On Intensities (I)
Soluble proteins	a = 0.15 ; b= 0.83	a = 0.05 ; b= 1.07
Membrane proteins	a = 0.05 ; b= 0.93	a = 0.02 ; b= 1.09

Fig. S2: comparison of crystal contacts between soluble and membrane proteins. **a/** Graph of anisotropy (calculated on amplitudes) as a function of crystal contact ratio for soluble (blue) and membrane proteins (red). Each dot represent an entry. The black dots represent the 95th percentile for each contact ratio bin, and has been fitted with a double exponential: y = a*exp(b*x)+c*exp(d*x)
Soluble proteins:
Coefficients (with 95% confidence bounds):
a = 20.39 (17.13, 23.65)
b = -0.2679 (-0.4945, -0.04127)
c = 44.4 (40.61, 48.19)
d = -10.21 (-12.1, -8.327)
Membrane proteins:
Coefficients (with 95% confidence bounds):
a = 79.22 (59.35, 99.1)
b = -3.639 (-4.653, -2.626)
c = 19.61 (-7.676, 46.89)
d = -55.22 (-249.9, 139.4)
b/ Scatter plot of soluble proteins in blue and membrane proteins in red, with black bars representing the mean of the sample and brackets displaying the 95% confidence interval of the population mean. Student's T-test with Welch correction (two-tailed p-value < 0.0001), and Mann-Whitney difference of medians (two-tailed p-value < 0.0001).

Fig. S3: comparison of solvent content between soluble and membrane proteins. **a/** Graph of anisotropy (calculated on amplitudes) as a function of solvent content for soluble (blue) and membrane proteins (red). Each dot represent an entry. The black dots represent the 95th percentile for each solvent content bin, and has been fitted with a Gaussian fit: y=a*exp(-((x-b)/c)^2);
Soluble proteins:
Coefficients (with 95% confidence bounds):
a = 64.06 (58.86, 69.27)
b = 85.56 (76.75, 94.37)
c = 44.24 (35.19, 53.28)
Membrane proteins:

Coefficients (with 95% confidence bounds):
a = 79.35 (71.89, 86.81)
b = 68.39 (66.56, 70.21)
c = 21.19 (18.44, 23.94)

b/ Scatter plot of soluble proteins in blue and membrane proteins in red, with black bars representing the mean of the sample and brackets displaying the 95% confidence interval of the population mean. Student's T-test with Welch correction (two-tailed p-value < 0.0001), and Mann-Whitney difference of medians (two-tailed p-value < 0.0001).

Fig. S4: comparison of resolution between soluble and membrane proteins. **a/** Graph of anisotropy (calculated on amplitudes) as a function of resolution for soluble (blue) and membrane proteins (red). Each dot represent an entry. **b/** Scatter plot of soluble proteins in blue and membrane proteins in red, with black bars representing the mean of the sample and brackets displaying the 95% confidence interval of the population mean. Student's T-test with Welch correction (two-tailed p-value < 0.0001), and Mann-Whitney difference of medians (two-tailed p-value < 0.0001).

Fig. S5: Power-law fits of anisotropy as a function of resolution. **a/** The four panels display the fits on amplitudes shown in Fig. 2A for comparison. The red curves corresponding to membrane proteins are always above the soluble proteins (blue). **b/** fits calculated on intensities as in Fig. 2B. Bellow are equations defining the percentile-fits of anisotropy (calculated on amplitudes or intensities) as a function of resolution.

	Soluble proteins		**Membrane proteins**	
	AnisoB(F) =	AnisoB(I) =	AnisoB(F) =	AnisoB(I) =
99th percentile	$6.68 \times Reso^{2.22} + 4.28$	$15.3 \times Reso^{1.71}$	-	-
95th percentile	$3.64 \times Reso^{2.46} + 3.84$	$7.84 \times Reso^{1.93}$	$5.07 \times Reso^{2.32} + 1.93$	$8.63 \times Reso^{2.11}$
75th percentile	$2.16 \times Reso^{2.40} + 1.39$	$4.25 \times Reso^{2.07}$	$2.95 \times Reso^{2.34}$	$4.96 \times Reso^{2.16}$
50th percentile	$0.94 \times Reso^{2.55} + 1.01$	$2.4 \times Reso^{2.13} + 0.08$	$1.57 \times Reso^{2.44} + 0.11$	$2.89 \times Reso^{2.28}$
25th percentile	$0.36 \times Reso^{2.45} + 0.42$	$1.35 \times Reso^{2.03}$	$1.32 \times Reso^{1.78}$	$1.18 \times Reso^{2.62}$

c/ Two types of membrane protein crystallization. Type I crystals: layers of membrane proteins stacked side by side, with hydrophobic surfaces providing crystals contacts, are stacked one on another; Type II crystals: detergent molecules shield hydrophobic regions of membrane proteins and crystal contacts involve only polar heads; The hydrophobic region of the protein is depicted here in green, and extra-membranous parts in orange. Fig. adapted from (24). **d/** Examples of proteins belonging to different classes of membrane proteins according to their insertion in the membrane. The proteins are depicted in blue cartoon representation, the membrane is shown in orange. All proteins are on the same scale. Alpha: Alpha-helical transmembrane proteins. GpA: Glycophorin A, *PDB code (2kpe);* LacY: Lactose permease, *PDB code (2y5y);* Atm1: *PDB code (4mrs);* A2A receptor: Adenosine receptor, *PDB code (3pwh);* Aqp0: Aquaporin 0, *PDB code (2z5x). Beta:* Beta-barrel transmembrane proteins: OprM: Outer, *PDB code (3d5k);* mVDAC-1: channel, *PDB code (3emn). Mono:* Monotopic membrane proteins: MoaB: Monoamine oxidase B, *PDB code (1s3e);* PGH2S1: Prostaglandine H2 synthase 1, *PDB code (1q4g).*

Fig. S6: **a/** comparison of anisotropy as a function of insertion into the membrane. ANOVA of the three types of membrane proteins as defined by Stephen White's database. The three samples have been analyzed for samples of the same resolution range (2.5-3 Å) to avoid a bias due to the effect of resolution. No significant difference could be found using ANOVA or Kruskal-Wallis analysis. Average ALPHA = 22.75, BETA

= 24.01 and MONO = 17.27. **b/** Comparison of diffraction anisotropy between ATPases, Electron-transfer proteins, transporters, channels and receptors. The comparison has been conducted using the ANOVA with Tukey's post-Hoc analysis. The three samples have been analyzed with the same number of entries in each category, after randomization of the transporters, channels and receptors. No statistical differences have been observed. Means: ATPases = 22.06 Å^2, E-transfer = 22.84 Å^2, transporters = 25.51 Å^2, channels = 24.62 Å^2, receptors = 20.54 Å^2. **c/** Influence of ligands on diffraction anisotropy of membrane protein crystals. The horizontal bar represents the mean and the brackets the 95% confidence interval of the mean. Student's T-test with Welch correction (two-tailed p-value = 0.0058), and Mann-Whitney difference of medians (two-tailed p-value = 0.037). **d/** Anisotropy of macromolecule crystals over the years. For each year, the average anisotropy is shown as a red horizontal bars; note that the scale (right) has been magnified to show more clearly the values.

Figure S1

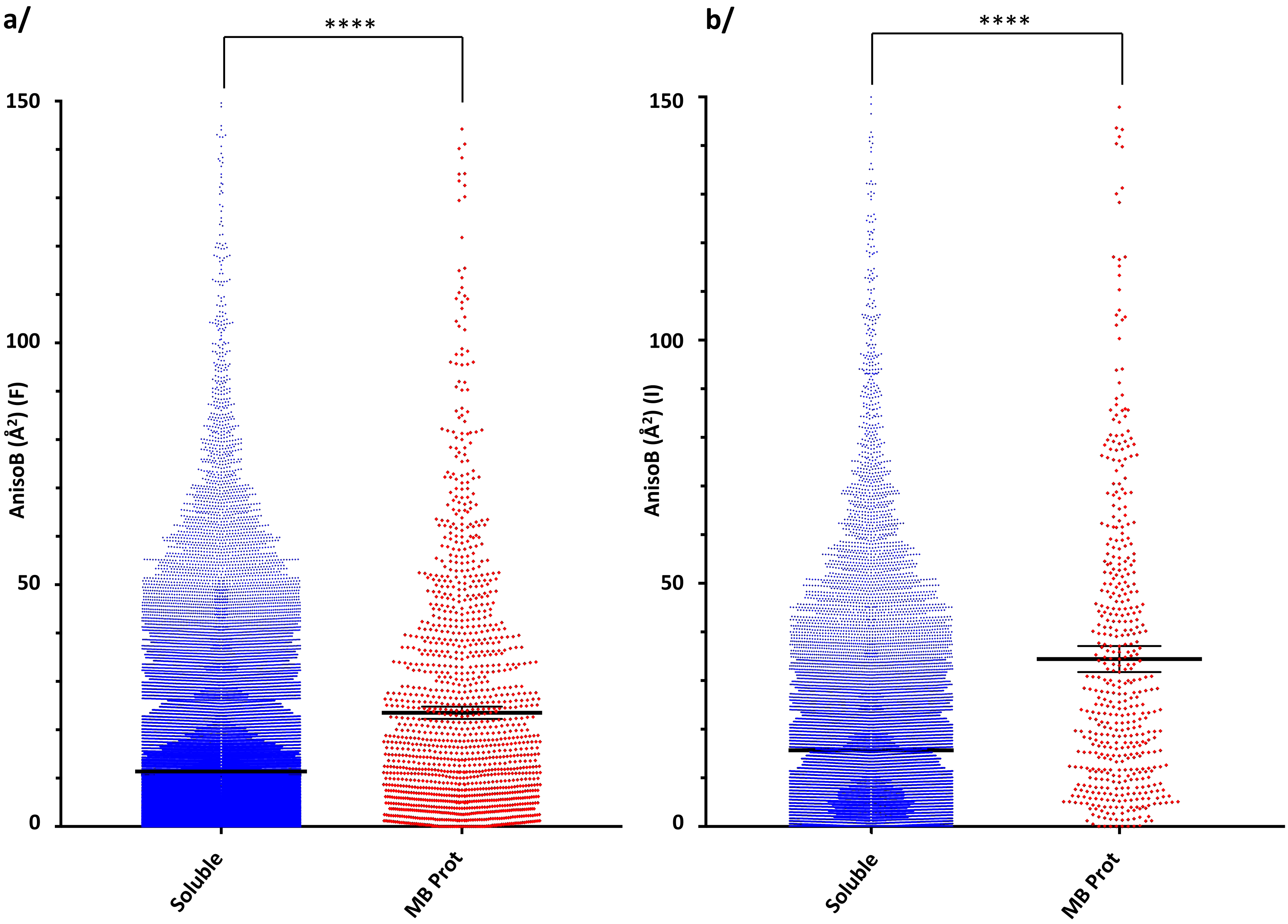

a/
****
150
100
50
0
AnisoB (Å²) (F)
Soluble
MB Prot
b/
****
150
100
50
0
AnisoB (Å²) (I)
Soluble
MB Prot

Figure S2

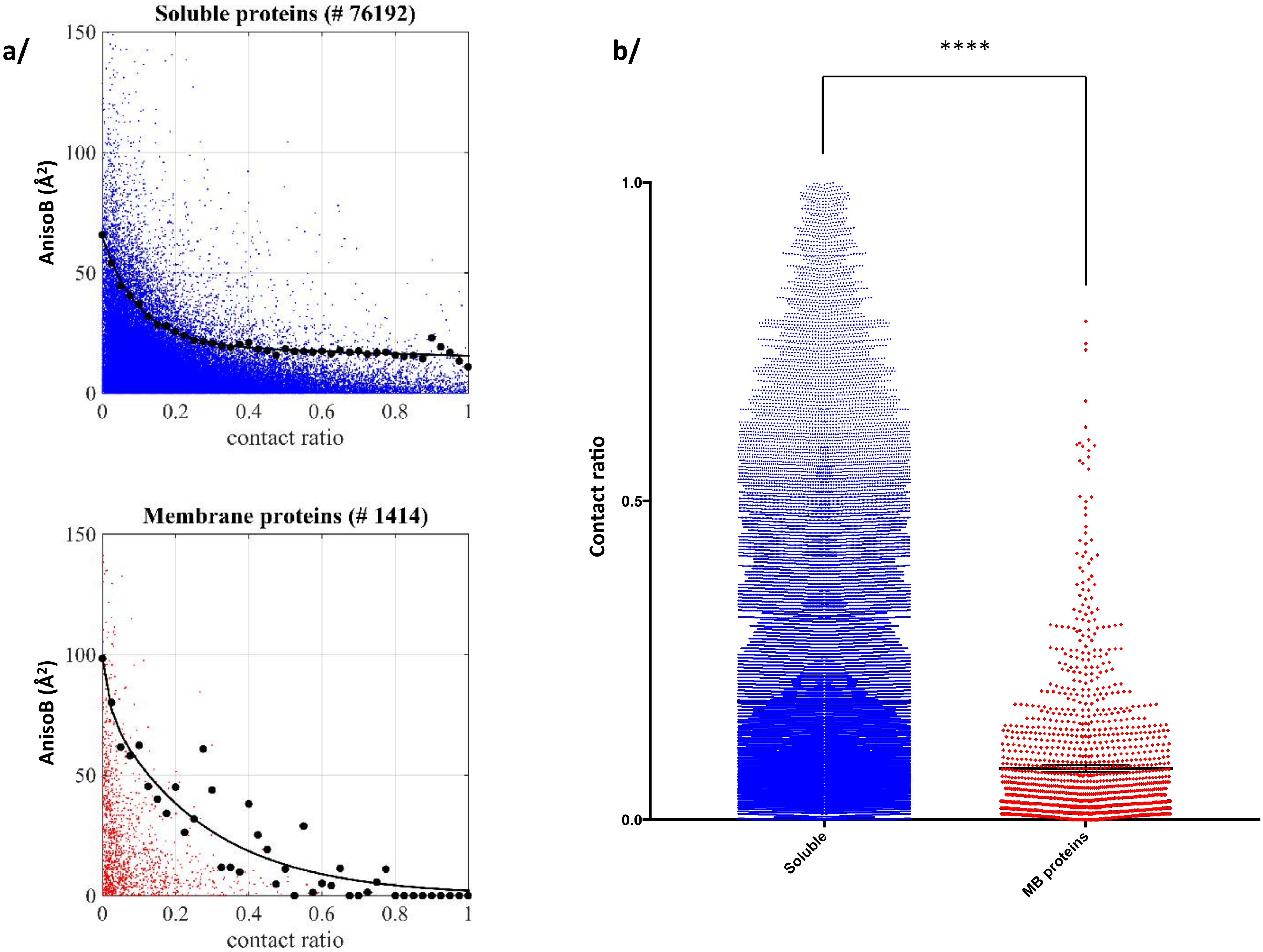
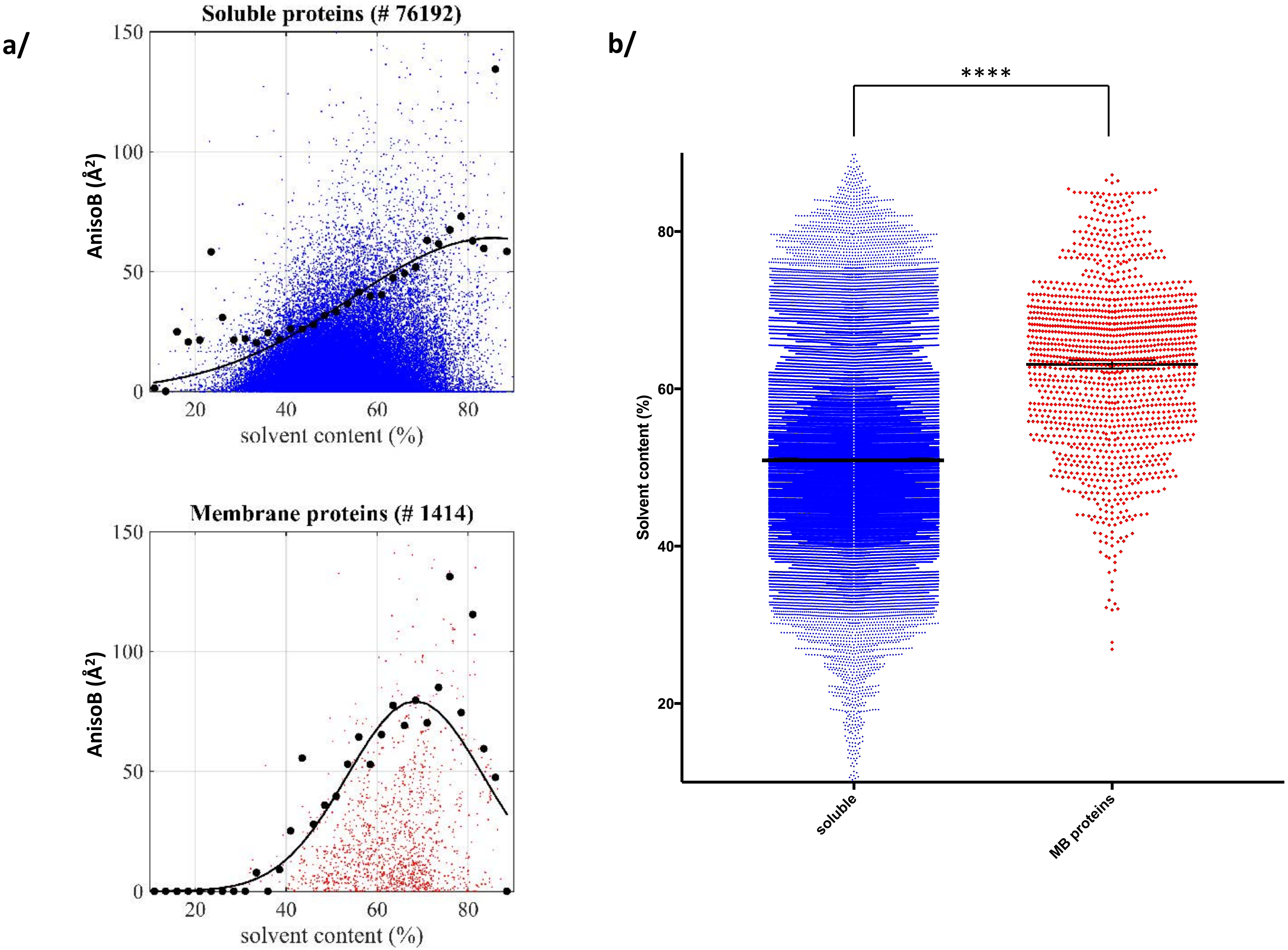

a/
Soluble proteins (# 76192)
150
100
50
0
AnisoB (Å²)
0
0.2
0.4
0.6
0.8
1
contact ratio
Membrane proteins (# 1414)
150
100
50
0
AnisoB (Å²)
0
0.2
0.4
0.6
0.8
1
contact ratio
b/
****
1.0
0.5
0.0
Contact ratio
Soluble
MB proteins

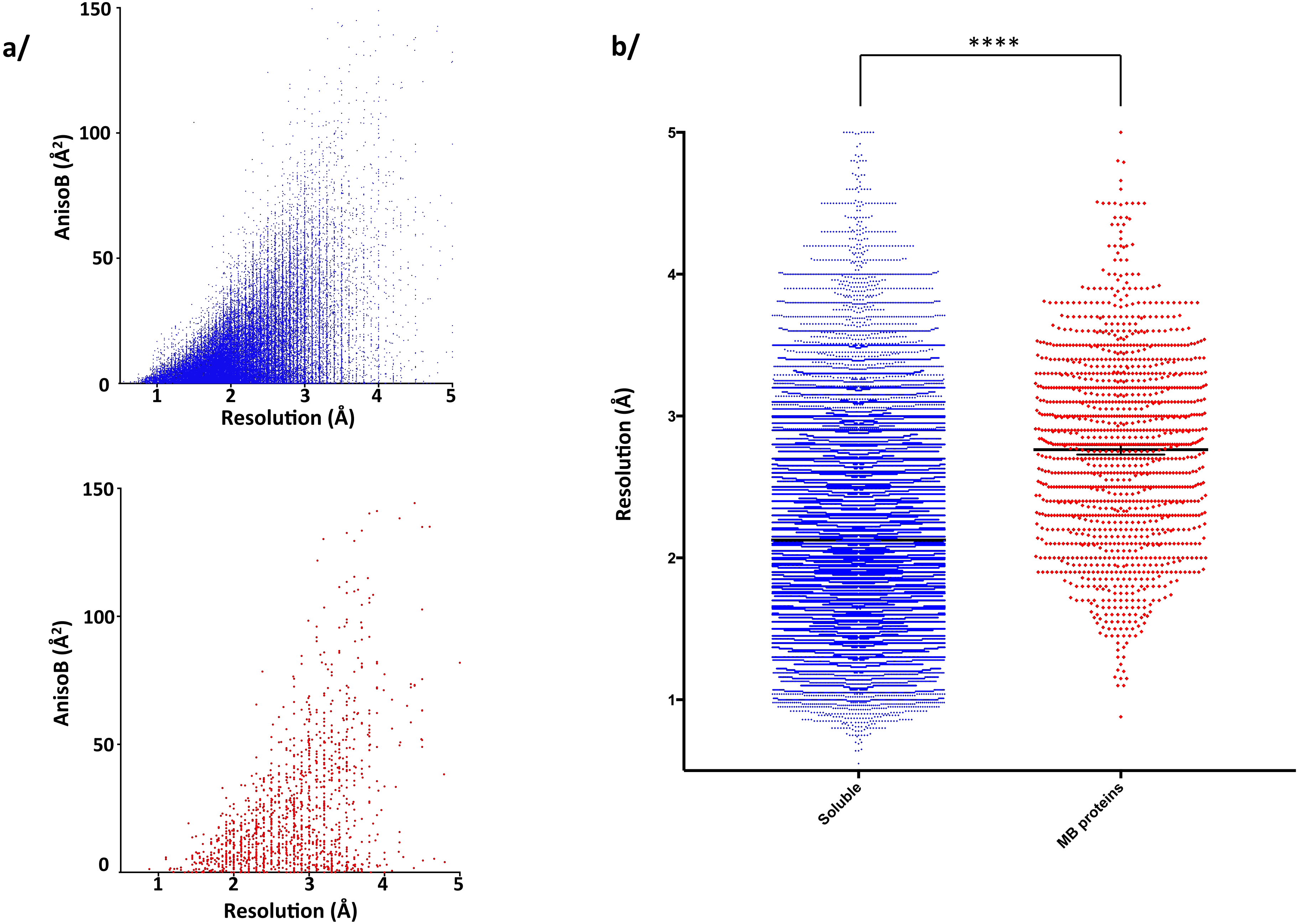
Figure S3



Soluble proteins (# 76192)

Membrane proteins (# 1414)

Figure S4


a/
150
100
50
0
AnisoB (Å^2)
1
2
3
4
5
Resolution (Å)
150
100
50
0
AnisoB (Å^2)
1
2
3
4
5
Resolution (Å)
b/
****
5
4
3
2
1
Resolution (Å)
Soluble
MB proteins

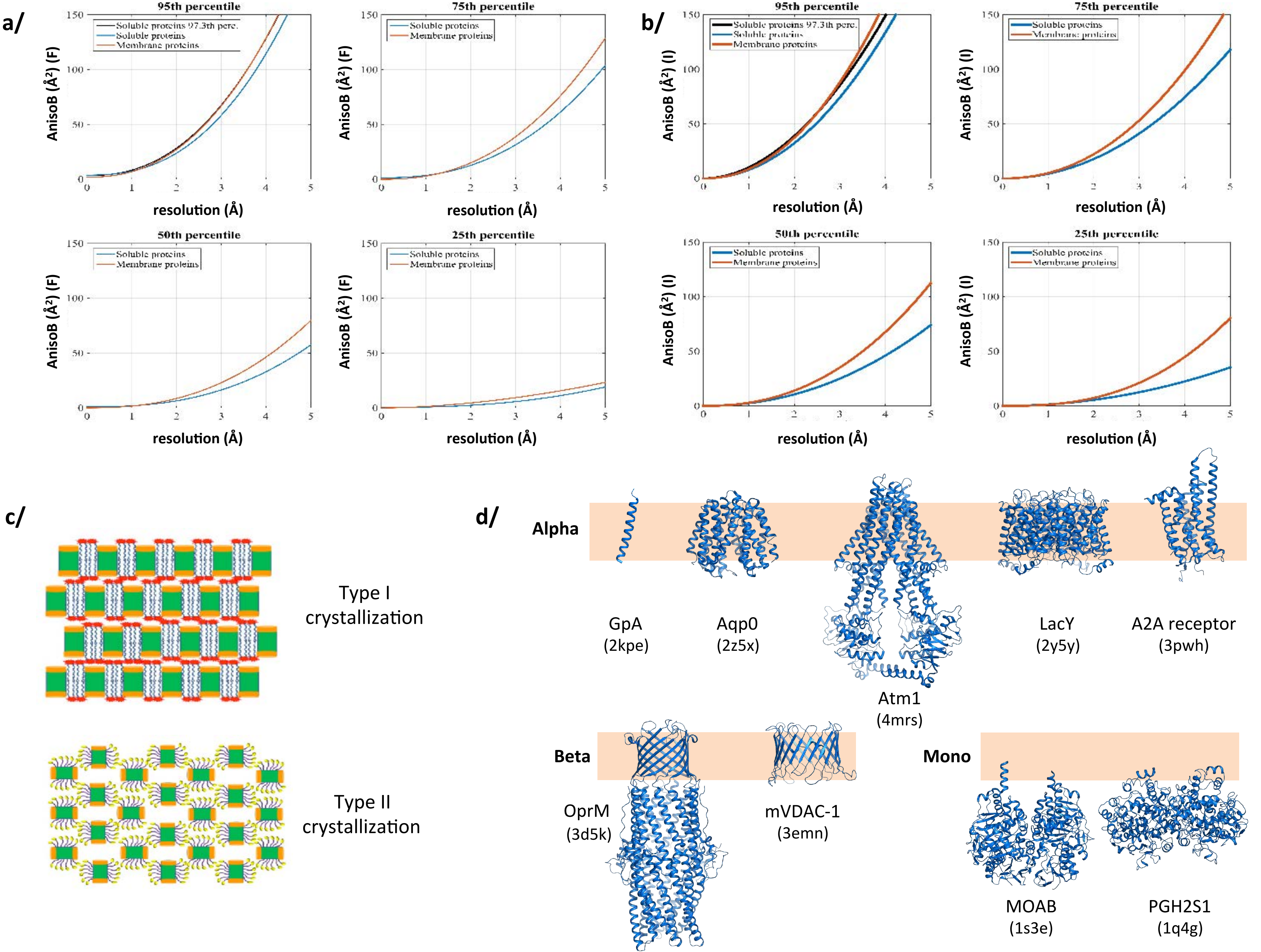

a/
95th percentile
Soluble proteins 97.3th perc.
Soluble proteins
Membrane proteins
75th percentile
50th percentile
25th percentile
AnisoB (Å²) (F)
resolution (Å)
0
50
100
150
1
2
3
4
5
b/
AnisoB (Å²) (I)
c/
Type I crystallization
Type II crystallization
d/
Alpha
GpA (2kpe)
Aqp0 (2z5x)
Atm1 (4mrs)
LacY (2y5y)
A2A receptor (3pwh)
Beta
OprM (3d5k)
mVDAC-1 (3emn)
Mono
MOAB (1s3e)
PGH2S1 (1q4g)

Fig. S6

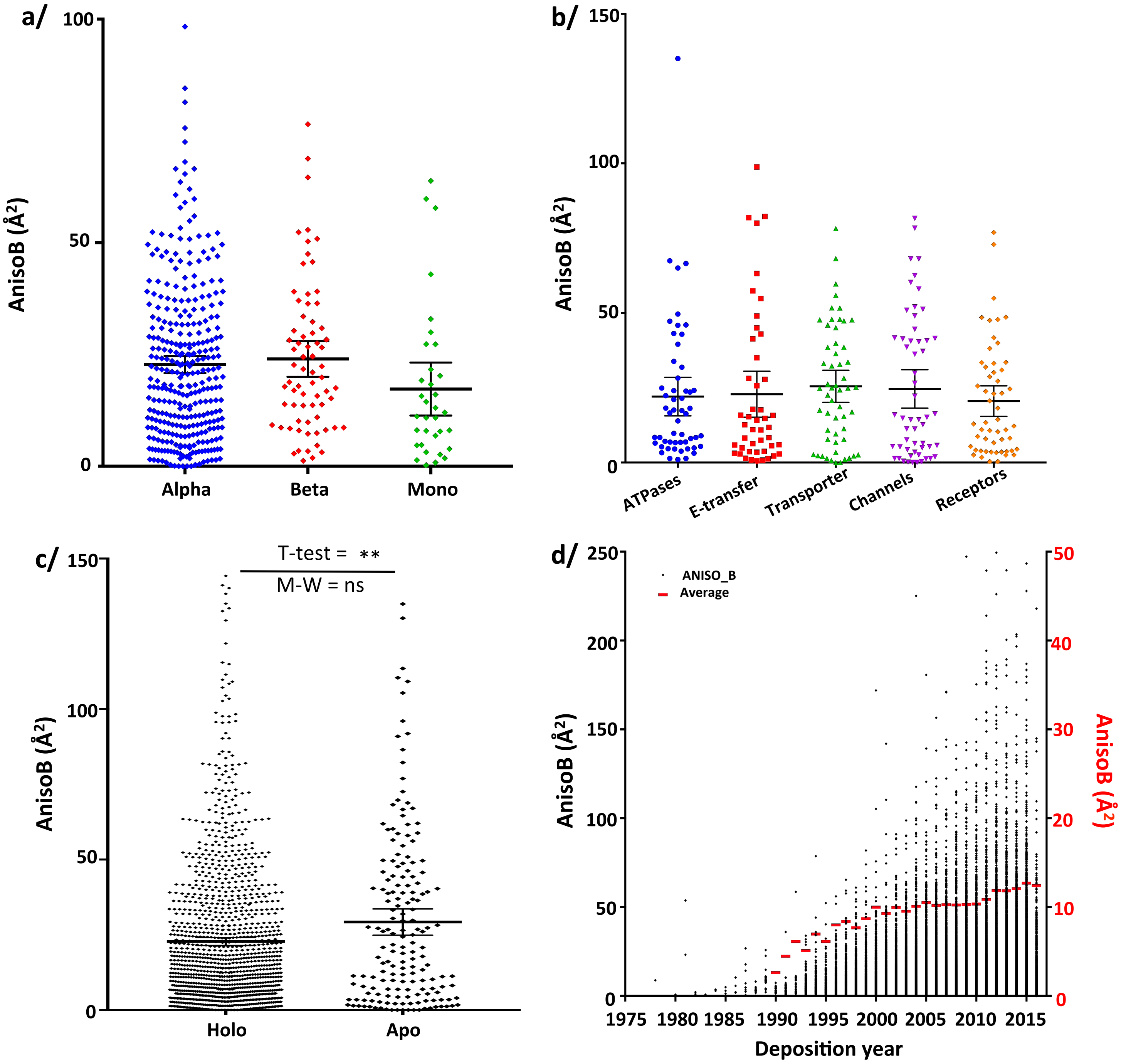

a/
100
50
0
AnisoB (Å²)
Alpha
Beta
Mono
b/
150
100
50
0
AnisoB (Å²)
ATPases
E-transfer
Transporter
Channels
Receptors
c/
150
100
50
0
AnisoB (Å²)
T-test = **
M-W = ns
Holo
Apo
d/
250
200
150
100
50
0
AnisoB (Å²)
ANISO_B
Average
50
40
30
20
10
0
AnisoB (Å²)
1975
1980
1985
1990
1995
2000
2005
2010
2015
Deposition year

Table S1: General statistics

		Resolution (Å)		Ratio1		Solvent content (%)		Anisotropy (Å^2)	
	Number of entries	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Soluble proteins	76192 structure factors	2.126 (2.122 – 2.13)	2.04	0.1855 (0.1844 – 0.1867)	0.136	51.11 (51.04 – 51.18)	50.16	10.72 (10.63 – 10.81)	6.39
	22985 intensites							*15.68 (15.57 – 15.79)*	*10.62*
Membrane proteins	1415 structure factors	2.763 (2.729 – 2.797)	2.8	0.08 (0.075 – 0.085)	0.042	63.11 (62.58 – 63.65)	64.22	23.53 (22.26 – 24.8)	15.15
	483 intensities							*34.43 (33.07 – 35.79)*	*25.83*
Membrane proteins in detergent	1241	2.79 (2.75 – 2.82)	2.8	0.078 (0.072 – 0.083)	0.037	64.19 (63.63 – 64.75)	65.35	24.1 (22.69 – 25.52)	15.1
Membrane proteins in LCP	173	2.61 (2.53 – 2.69)	2.6	0.096 (0.081 – 0.111)	0.069	55.64 (54.45 – 56.82)	55.83	19.53 (17.25 – 21.82)	15.46
ALPHA	1138	2.83 (2.79 – 2.87)	2.9	0.075 (0.069 – 0.08)	0.036	63.88 (63.27 – 64.48)	65.23	24.71 (25.68 – 26.02)	15.87
BETA	200	2.79 (2.41 – 2.57)	2.5	0.12 (0.099 – 0.13)	0.074	60.42 (59.19 – 61.67)	60.98	19.24 (16.82 – 21.66)	13.85
MONO	90	2.54 (2.63 – 2.54)	2.4	0.062 (0.052 – 0.072)	0.048	58.56 (56.92 – 60.21)	59.62	16.38 (12.89 – 19.87)	12.35
Transporters	272	2.99 (2.92 – 3.07)	3	0.049 (0.041 – 0.058)	0.025	65.04 (63.8 – 66.27)	67.3	31.46 (27.94 – 34.97)	24.36
Channels	225	2.78 (2.68 – 2.87)	2.94	0.11 (0.09 – 0.13)	0.038	66.25 (65.25 – 68.05)	67.87	24.65 (21.22 – 28.09)	14.6
Receptors	99	2.84 (2.74 – 2.93)	2.8	0.068 (0.055 – 0.081)	0.058	60.96 (59.01 – 62.9)	58.92	21.13 (17.4 – 24.86)	16.48
ATPases	54	2.99 (2.82 – 3.17)	2.95	0.042 (0.026 – 0.057)	0.024	63.43 (60.75 – 66.11)	64.86	22.06 (15.6 – 28.51)	16.45
Electron-transfer	46	2.87 (2.64 – 3.1)	2.93	0.032 (0.017 – 0.048)	0.019	67.47 (65.3 – 69.65)	68.16	22.84 (15.15 – 30.52)	12.19

The number in parenthesis represent the 95% confidence interval of the mean.

Table S2: List of proteins crystallized both in detergent and LCP.

	Crystallization in detergent		Crystallization in LCP	
Protein name	**PDB entry**	AnisoB (Å^2)	**PDB entry**	AnisoB (Å^2)
A2A adenosine receptor	**2YDV**	8.677	**3EML**	0.346
	2YDO	12.22	**4EIY**	2.476
	4UHR	11.333	**3QAK**	6.627
	3VG9	48.519		
	4UG2	10.018		
	3VGA	35.389		
AlgE alginate export protein	**4AZL**	17.505	**4AFK**	5.878
	3RBH	27.378	**4XNL**	27.483
	4TQU	29.299	**4XNK**	26.597
	4TQV	134.91	**4B61**	19.264
apo BtuB cobalamin transporter	**3M8B**	0.11	**2GUF**	25.8
	3M8D	0.13		
	2GSK	18.343		
	2YSU	40.021		
Bacterial homologue of ASBT (ASBTNM) with bound taurocholate	**3ZUX**	5.056	**4N7W**	14.94
	3ZUY	0	**4N7X**	20.815
Bacteriorhodopsin (BR)	**3T45**	6.384	**3NSB**	12.097
			4FPD	13.463
			4X32	3.924
			3NS0	12.746
			4X31	0.614
β2 adrenergic receptor	**3KJ6**	26.647	**2RH1**	20.474
	2R4R	39.25	**3D4S**	28.681
	2R4S	31.907	**3PDS**	41.713
			4GBR	71.171
			3NY8	30.945
			3NY9	28.964
			3SN6	1.636
			3NYA	33.284
			3P0G	72.903
Cytochrome C Oxidase aa3	**3HB3**	10.872	**2YEV**	14.112
	3EHB	11.017		
	3X2Q	14.736		
	2GSM	5.872		
	3MK7	3.132		
FraC eukaryotic pore-forming toxin from sea anemone	**4TSL**	4.517	**4TSY**	54.846
	3VWI	6.88		
	3W9P	9.453		
	4TSN	3.863		
	4TSO	17.297		
	4TSQ	0.03		

	4TSP **3LIM**	5.41 8.863		
M2 proton channel (AM2)	**3C9J** **3LBW** **3BKD**	15.278 3.422 16.197	**4QK7** **4QKC**	5.13 5.86
MgtE Mg^{2+} Transporter	**2YVY** **2YVZ** **2YVX** **2ZY9**	15.728 62.254 27.04 57.848	**4U9N** **4U9L** **4WIB**	11.824 11.204 25.899
NarK nitrate/nitrite exchanger	**4JR9** **4JRE**	7.759 9.593	**4U4W** **4U4V** **4U4T**	35.519 28.219 15.38
NTS1 neurotensin receptor in complex with neurotensin	**4BUO** **3ZEV** **4BV0** **4BWB**	28.104 28.75 25.37 56.492	**4XES** **4XEE** **4GRV**	31.809 17.772 16.476
PepTSo Oligopeptide-proton symporter (POT family)	**4TPJ** **4TPH** **4TPG** **2XUT**	1.667 2.599 2.262 8.682	**4UVM**	29.504
PepTSt Oligopeptide-proton symporter (POT family)	**4APS**	4.774	**4XNJ** **4XNI**	17.906 14.703
Photosynthetic Reaction Center	**4N7L** **4N7K**	3.077 5.987	**2J8C** **2J8D** **2WJM** **2WJN** **4CAS**	1.424 7.643 5.821 0.588 8.182
RPE65 visual cycle retinoid isomerase	**3FSN** **4F30** **4F3D**	13.538 21.878 7.892	**4F3A** **4F2Z**	1.774 59.784
Stearoyl-coenzyme A desaturase (SCD1) in complex with substrate	**4ZYO**	25.788	**4YMK**	16.714
Translocator protein (TSPO)	**4RYJ**	56.148	**4UC1** **4UC2** **4UC3** **4RYN** **4RYQ** **4RYM** **4RYI**	8.056 19.401 12.382 10.555 6.76 52.38 23.376
Zn^{2+}-transporting four-helix bundle	**4P6J** **4P6L**	0.461 0	**4P6K**	8.604

Table S3: Membrane proteins for which structures are available, Apo or in complex with a ligand, with more than two entries in each category.

	Apo		Holo	
Protein name	**PDB entry**	AnisoB (Å^2)	**PDB entry**	AnisoB (Å^2)
AcrB, bacterial multi-drug efflux transporter	**2HQC**	1.112	**2RDD**	33.109
	2HQG	3.455	**4DX5**	8.284
	2HQF	12.778	**4DX7**	20.154
	2HQD	0.685	**2W1B**	19.04
	4CDI	0.462	**3AOC**	10.979
	3AOA	6.957	**2HRT**	5.384
	2DHH	45.89	**2GIF**	14.027
			4U96	7.434
			4DX6	9.1
			2J8S	24.021
			4U95	10.926
			4U8Y	11.016
			4U8V	14.985
			4C48	0
			2DRD	29.286
			3AOD	6.593
			2F1M	21.537
			3W9H	6.745
			3AOB	12.665
BamB, component of the Bam β-barrel assembly machine	**3Q54**	14.964	**4XGA**	7.826
	3PRW	3.439	**3Q7N**	14.372
	2HY3	3.377	**3Q7M**	3.302
	3Q7O	0	**3P1L**	7.98
β2 adrenergic receptor	**2R4R**	39.25	**2RH1**	20.474
	2R4S	31.907	**4GBR**	71.171
			3PDS	41.713
			3NY8	30.945
			3NY9	28.964
			3NYA	33.284
			3D4S	28.681
			3SN6	1.636
			3P0G	72.903
			3KJ6	26.647
CusA, metal-ion efflux pump	**3NE5**	17.835	**3OPO**	108.412
	3OOC	86.491	**3OW7**	114.948
	3K07	16.984		
CusC, heavy metal discharge outer membrane protein	**4K7K**	13.891	**4K7R**	10.675
	4K34	9.828	**3PIK**	9.999
DgkA, diacylglycerol	**3ZE5**	11.112	**3ZE3**	5.192
	3ZE4	40.456	**4UXZ**	4.733

kinase (DAGK)			**4UYO**	16.36
			4UXX	21.676
			4UXW	10.549
Fadl, long-chain fatty acid transporter	**2R88**	21.716	**3DWO**	5.995
	2R89	36.168	**3PGS**	6.187
			3PGU	6.78
			3PGR	32.298
			3PF1	16.135
			3DWN	22.64
			2R4N	25.179
			2R4O	38.693
			2R4P	16.967
			2R4L	30.545
GlpG, rhomboid-family intramembrane protease	**2NRF**	38.781	**4QO2**	11.812
	3ODJ	15.986	**4QNZ**	7.445
	5F5D	26.075	**4QO0**	11.409
			3UBB	18.036
			5F5K	35.322
			5F5B	19.882
			5F5J	24.204
			3B44	5.611
			3B45	5.662
			2IC8	13.493
			2XTU	4.349
			2O7L	17.439
			2XOV	4.385
			2XOW	18.331
			3TXT	44.921
			2IRV	2.07
			2XTV	3.989
			2NR9	5.519
H^+/Cl^- Exchange Transporter	**4FG6**	61.648	**2EZ0**	39.96
	2EXW	47.669	**4FTP**	50.685
	2EXY	46.241	**3NMO**	11.16
			4MQX	72.237
			4ENE	32.318
Heterodimeric ABC exporter TM287-TM288	**4Q7M**	5.56	**3QF4**	81.445
	4Q7K	13.359	**4Q4A**	60.68
	4Q4H	2.125	**4Q7L**	2.492
KcsA, Potassium channel H^+ gated	**3EFD**	1.716	**2ITD**	5.428
	3EFF	14.359	**3STL**	14.732
	3PJS	27.158	**2W0F**	7.076
			4UUJ	2.231

			2JK5	15.103
			2BOC	18.26
			2BOB	11.207
			2NLJ	23.302
			3GB7	0.095
			3IGA	11.934
			3OR6	16.416
			3OR7	8.807
			3HPL	18.782
			3FB6	19.991
			3F7Y	11.341
			3F5W	10.761
			3F7V	10.964
			3FB5	17.709
			3STZ	19.731
			2ITC	14.718
			3FB7	0.813
NTS1, neurotensin receptor	**4BV0**	25.37	**4XEE**	17.772
	4BWB	56.492	**4XES**	31.809
			4GRV	16.476
			4BUO	28.104
			3ZEV	28.75
OmpF, outer membrane protein F	**2ZLD**	68.801	**3FYX**	9.673
	3K1B	1.694	**3O0E**	62.567
	4D5U	69.748	**2ZFG**	11.059
	3K19	2.144	**3HW9**	7.401
			3NSG	76.49
			3HWB	50.273
ELIC, Prokaryotic pentameric ligand-gated ion channel			**4TWD**	57.217
	2VL0	36.962	**3RQW**	37.942
	4TWH	67.004	**4TWF**	71.84
			3RQU	43.626
			3ZKR	62.23
GLIC, Prokaryotic pentameric ligand-gated ion channel	**3EHZ**	91.848	**4IL9**	33.678
	3EI0	109.179	**4ILA**	61.213
	4NPQ	58.614	**4HFI**	36.234
			4ILB	35.295
			4ILC	53.295
			4IL4	81.197
			3IGQ	6.434
			4QH1	85.773
			4QH5	58.982
			3P4W	44.481
			4F8H	68.071
			3EAM	10.055
			3P50	50.889

			4NPP	50.097
			2XQ9	103.447
			2XQ4	129.462
			2XQ3	109.097
			2XQ7	111.443
			2XQA	104.485
			2XQ8	109.741
V1-ATPase	**3GQB**	1.378	**3W3A**	22.354
	3A5D	3.992	**3A5C**	4.63
			3V6I	5.432
			3K5B	47.236