

***Pseudo-Symmetric Assembly of Protodomains as a Common Denominator in the Evolution of Polytopic Helical Membrane Proteins***

Philippe Youkharibache<sup>1,\*</sup>, Alexander Tran<sup>2</sup>, and Ravinder Abrol<sup>2,\*</sup>

<sup>1</sup> Cancer Data Science Lab, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

<sup>2</sup> Department of Chemistry and Biochemistry, California State University, Northridge, CA, USA

**SUPPLEMENTARY INFORMATION**

MFS	5EQI	<a href="https://d55qc.app.goo.gl/wcE7nXES2yhE4CQr6">https://d55qc.app.goo.gl/wcE7nXES2yhE4CQr6</a>
TRIC	5WUF 3KCU	<a href="https://d55qc.app.goo.gl/XUkPY7iPmR8LVJ8c9">https://d55qc.app.goo.gl/XUkPY7iPmR8LVJ8c9</a> <a href="https://d55qc.app.goo.gl/srgkH59q3ADhRSTs8">https://d55qc.app.goo.gl/srgkH59q3ADhRSTs8</a>
AQP1	3NE2	<a href="https://d55qc.app.goo.gl/pN9hZKMbiS4VML5H6">https://d55qc.app.goo.gl/pN9hZKMbiS4VML5H6</a>
SemiSWEET Apo vs. Ligand	4QNC 4QND	<a href="https://d55qc.app.goo.gl/i9vWJfzXEcl8q2A69">https://d55qc.app.goo.gl/i9vWJfzXEcl8q2A69</a>
SWEET vs. SemiSWEET	5CTH 4QND	<a href="https://d55qc.app.goo.gl/9NfEK6KmidELHYt1A">https://d55qc.app.goo.gl/9NfEK6KmidELHYt1A</a> Use keyboard "a" letter or "Alternate" command to visualize alternatively the 2 aligned structures of SWEET (PDB:5CTH) vs. 3TMH monomer A of SemiSWEET (PDB:4QND) (aligned only on 3TMH protodomain 1 ) Use "a" alternate command to alternate between structures
PnuC	4QTN	<a href="https://d55qc.app.goo.gl/S1atQgdGptdt7bCs6">https://d55qc.app.goo.gl/S1atQgdGptdt7bCs6</a>
GPCR C	4OR2	<a href="https://d55qc.app.goo.gl/ZFChMCsBRER3uSbD7">https://d55qc.app.goo.gl/ZFChMCsBRER3uSbD7</a>
GPCR Aa	5G53	<a href="https://d55qc.app.goo.gl/igoFR7sh9hZdjK1L9">https://d55qc.app.goo.gl/igoFR7sh9hZdjK1L9</a>
Rhodopsin	1GZM	<a href="https://d55qc.app.goo.gl/KtAe6nkkSJ7fm3aHA">https://d55qc.app.goo.gl/KtAe6nkkSJ7fm3aHA</a>
Rhodopsin active-inactive	6CMO 1GZM	<a href="https://d55qc.app.goo.gl/KLJg7G6Zx3g9hju4A">https://d55qc.app.goo.gl/KLJg7G6Zx3g9hju4A</a>
pLGIC Acetylcholine Receptor	6CNK	<a href="https://icn3d.page.link/Knzy">https://icn3d.page.link/Knzy</a>

**Table S1 - 3D visualization links** - using iCn3D

	TM1	TM2	2b	TM3	3TMH	RMS (Å)
sequence length aligned (5CTH)	20	18		25	63	
<b>SWEET 123 567 (# res. identical)</b>	3	5		5	13	
%Id	15%	28%		20%	21%	1.36
<b>vs. SemiSWEET (4QNC/4QND)</b>	0/6(*)	3		4	7/13	
%Id	0/30%	17%		16%	11/21%	1.98
length aligned (4QTN)	13	42		23	78	
<b>PnuC 123 vs 567</b>	1	6		4	11	
%Id	8%	14%		17%	14%	1.26
length aligned (5WUF)	23	23		16	62	
<b>TriC 123 vs 456</b>	8	10		4	22	
%Id	35%	43%		25%	35%	1.53
length aligned (4FC4)	21	18	13	23	75	
<b>FocA 123 vs 456</b>	4	3	1	8	16	
%Id	19%	17%	8%	35%	21%	1.37
length aligned (5DYE)	24	25	10	19	78	
<b>Aquaporin 123 vs 456</b>	4	5	4	4	17	
%Id	17%	20%	40%	21%	22%	1.92
length aligned (3C02)	24	25	14	19	82	
<b>Aquaglyceroporin 123 vs 456</b>	5	5	3	4	17	
%Id	21%	20%	21%	21%	21%	1.92
length aligned (4FC4-1 / 5DYE-2)	21	17	12	23	73	
<b>FocA vs. Aquaporin</b>	3	4	0	8	15	
%Id	14%	24%	0%	35%	21%	2.94
length aligned (5EQI)	18	18		18	54	
<b>MFS 123 vs 456 (protos1-2)</b>	4	3		2	9	
%Id	22%	17%		11%	17%	1.70
<b>vs 789 (protos1-3)</b>	2/4(**)	1/6(**)		4		
%Id	11/22%(**)	6/33%(**)		22%	13/26%	1.88/2.97(**)
<b>vs 10-11-12 (protos1-4)</b>	3	1		2		
%Id	17%	6%		11%	11%	2.32
<b>Average</b>	19%	21%	17%	22%	20%	1.81
Min	8%	6%	0%	11%	11%	1.26
Max	35%	43%	40%	35%	35%	2.94

**Table S2 - Protodomain alignment statistics for SWEET, PnuC, TriC, FoCA, Aquaporin, and MFS**

(\*) a shift of 3 in sequence will match SPLA sequence in SWEET proto2, part of the FxSP motif in TM1/5 conserved in SWEET protos

(\*\*) A 4 protodomains optimized multiple structure alignment. A shift in sequence vs optimized structure alignment would for example increase the sequence match from 1 to 6 in TM2 for an overall RMS change from 1.88 Å to 2.97 Å

GPCRs	TM1/5	TM2/6	TM3/7	3TMH	RMS (A)	Symmetry	TM1/4	TM2/5	TM3/6	3TMH	RMS (A)	Symmetry
length aligned	21	29	27	77			21	29	27			
<b>1F88 123 vs 567 (# res identical)</b>	4	4	7	15		Rhodopsin	2	4	2			
%Id	<b>19%</b>	<b>14%</b>	<b>26%</b>	<b>19%</b>	<b>3.24</b>	SYM	<b>10%</b>	<b>14%</b>	<b>7%</b>	<b>10%</b>	<b>7.73</b>	NO SYM
length aligned	28	21	21	70			28	21	21			
<b>4OR2 123 vs 567</b>	5	0	4	9		Class C	3	1	0			
%Id	<b>18%</b>	<b>0%</b>	<b>19%</b>	<b>13%</b>	<b>3.36</b>	SYM	<b>11%</b>	<b>5%</b>	<b>0%</b>	<b>5%</b>	<b>4.89</b>	NO SYM
length aligned	22	21	24	67			22	21	24			
<b>4MBS 123 vs 567</b>	6	0	3	9		Class A	3	4	1			
%Id	<b>27%</b>	<b>0%</b>	<b>13%</b>	<b>13%</b>	<b>2.62</b>	SYM	<b>14%</b>	<b>19%</b>	<b>4%</b>	<b>12%</b>	<b>6.78</b>	NO SYM
length aligned	28	29	28	85								
<b>4GRV 123 vs 567</b>	6	2	7	15		Class A						
%Id	<b>21%</b>	<b>7%</b>	<b>25%</b>	<b>18%</b>	<b>3.31</b>	SYM						
length aligned	22	25	20	67								
<b>3RZE 123 vs 567</b>	4	0	4	8		Class A						
%Id	<b>18%</b>	<b>0%</b>	<b>20%</b>	<b>12%</b>	<b>3.38</b>	SYM						
length aligned	24	24	18	66								
<b>3OE9 123 vs 567</b>	5	3	1	9		Class A						
%Id	<b>21%</b>	<b>13%</b>	<b>6%</b>	<b>14%</b>	<b>3.38</b>	SYM						
length aligned	24	26	20	70			16	20	20	56		
<b>4JKV 123 vs 567</b>	6	2	2	10		Class F	2	3	3	8		
%Id	<b>25%</b>	<b>8%</b>	<b>10%</b>	<b>14%</b>	<b>2.39</b>	SYM	<b>13%</b>	<b>15%</b>	<b>15%</b>	<b>14%</b>	<b>5.23</b>	NO SYM
length aligned	25	22	22	69			16	28	21	65		
<b>5EE7 123 vs 567</b>	2	2	3	7		Class B	1	3	1	5		
%Id	<b>8%</b>	<b>9%</b>	<b>14%</b>	<b>10%</b>	<b>2.56</b>	SYM	<b>6%</b>	<b>11%</b>	<b>5%</b>	<b>7%</b>	<b>6.49</b>	NO SYM
length aligned	23	16	25	64								
<b>4E1Y 123 vs 567</b>	1	2	4	7		Na+ binding						
%Id	<b>4%</b>	<b>13%</b>	<b>16%</b>	<b>11%</b>	<b>2.68</b>	SYM						
length aligned	17	24	17	58								
<b>1UAZ ABC vs EFG</b>	3	6	0	9		Bacteriorhodopsin						
%Id	<b>18%</b>	<b>25%</b>	<b>0%</b>	<b>16%</b>	<b>2.56</b>	SYM						
Average (GPCRs)	<b>18%</b>	<b>7%</b>	<b>16%</b>	<b>14%</b>	<b>2.99</b>		<b>11%</b>	<b>13%</b>	<b>4%</b>	<b>9%</b>	<b>6.47</b>	
Min	4%	0%	6%	10%	2.39	SYM	10%	5%	0%	0%	4.89	NO SYM
Max	27%	14%	26%	19%	3.38		14%	19%	0%	7%	7.73	

**Table S3 - Protodomain alignment statistics for GPCRs (and Bacteriorhodopsin).** [Left] Sequence Identity and RMSD of the two 3TMH TM123/TM567 protodomains matching symmetrically and individual matching transmembrane helices. [Right] Protodomain 1 TM123 vs 3TMH TM456 for comparison (not related symmetrically).

Odds Ratio	Motif1: S3.39		Motif1: D2.50	
	Motif2: NS7.46		Motif2: FxxxW6.48	
GPCR-Class	A $\alpha$	A $\gamma$	A $\alpha$	A $\gamma$
Humans	76.8	12.3	24.6	-
Mammals	29.0	21.3	9.6	2.6
Vertebrates	52.4	16.9	3.1	2.6

**Table S4:** Odds-Ratio of three pairs of co-evolutionary related residues/motifs, calculated using the GPCR-SAS server [<http://lmc.uab.cat/gpcrsas/>, Tamayo et al. (2018) *PLoS One*, 13(7):e0199843] for available GPCRs from humans, mammals, and vertebrates. Numbers following the motifs are the BW numbers of the only residue or last residue in the motif.



## Protodomain alignments (pairwise, familywise and across two families FocA and Aquaporin)

```

FocA 1/2ab/3 4FC4_A | IgFWVSSAMAGAYVGLGIIILIFTLgnlId~psvrpLVMGATFGIALTLVIAAGseLFTGHTMFLTLCvkagtis~hgq wailPQTWLGNLVGVSFVALLYSWGGgsI
FocA 4/5ab/6 4FC4_A | tvLFFKGAALCNWLVCLAIWMAIRte~gTAKFLAIWCLLAFIASGyeHSVANMTLFALESwfgghsday~tIagighnLLWVTLGNTLSGVVFMGLGYWYATpks
FocA 1/2ab/3 3KLZ_A | ykSFLLAISAGIQIGIAFVYTVVttgahd pygvtkLLGGLAFSLGLILVVITGgeLFTSSVLI LVAKAsgkissw~keIvrnWTVVYFNGLCGSI~ILVFI MLATRqf
FocA 4/5ab/6 3KLZ_A | lqAFALGLMCNlLVCLAVWMTFSars~ITDKVMVLI LPVAMFVSSGfeHCIANMFQVPMAlgikyfapesfwa tganiaqyadlnfnvfnvnnLIPVTLGNIVGGG~VFVGMWYWLlYlk
FocA 1/2ab/3 3KCV_A | lktFYLAITAGVFISIAFVFIYITattgtgt pfg akLVGGICFSLGLILCVVCGadLFTSTVLI VVAKAsgritw~gqlaknWLVVYFNGLVGAL~LFVLLMWLSGey
FocA 4/5ab/6 3KCV_A | IeAVCLGILANLMVCLAVWMSYSGrs~IMDKAFIMVLPVAMFVASGfeHSIANMFMI PMGivirdfaspefwtavgsapenfshltv nftdnLIPVTIGNIIGGG~LLVGLTYWVlylr

Aqp1 1/2ab/3 5I32_A | IaSLRAYLAEFISTLLFVAGVGSAlayakltsda~al dtpglvAI AVCHGFALFVAVAI GANISGGHVNPVTFGLAVGgqi~tviTGVFYWI AQLLGSTAAACFLKYYVTggIav
Aqp1 4/5ab/6 5I32_A | IgSIEGVVMEIITFALVYTYATAAdpkkg~slgtIAPLAIGLIVGANILAAGPFSGGSMNPARSFGPAVAag~dfsgHWVYVWVGLIGGGLAGLIYGNVFGsse
Aqp1 1/2ab/3 3NE2_A | tLAKRFTAEEVVGTFILVFFGPGAAVitl iangadkpnfnigigalggIgdwfaIGMAFALAAAVIYSLGRISGAHINPAVTIALWSigrf~pgrEVVPPYIYVIGPIVGAVAALYNYLAKe
Aqp1 4/5ab/6 3NE2_A | igYGAFLTEAIGTFLMLVIMGVAVderapppFAGLVIGLTVGGIITIGNITGSSLNPARTFGPYLGDsl ginlwqfPPIYVIGPIVGAVAALYNYLAKe
Aqp1 1/2ab/3 5DYE_A | vaFLKAVFAEFLATLIFVFFGLGSALKwpsa~lptIIQIALAFGLAIGTLAQALGPVSSGGINPAITLALLVGNqis~sIRAFFYVAAQLVGAIAAGAGIYGVAPInar
Aqp1 4/5ab/6 5DYE_A | ttQQGAMVVELILTFQLALCIFASTDsrrt~epvgSPALSIGLSVTLGHLVGIYFTGCSMNPARSFGPAVVnr~fspAHWVFWVWVGLVGAVALAALYFYLLfpnsI

FocA 1/2ab/3 4FC4_A | IgFWVSSAMAGAYVGLGIIILIFTLgnlId~psvrpLVMGATFGIALTLVIAAGseLFTGHTMFLTLCvkagtis~hgq wailPQTWLGNLVGVSFVALLYSWGGgsI
FocA 4/5ab/6 4FC4_A | tvLFFKGAALCNWLVCLAIWMAIRte~gTAKFLAIWCLLAFIASGyeHSVANMTLFALESwfgghsday~tIagighnLLWVTLGNTLSGVVFMGLGYWYATpks
FocA 1/2ab/3 3KLZ_A | ykSFLLAISAGIQIGIAFVYTVVttgahd pygvtkLLGGLAFSLGLILVVITGgeLFTSSVLI LVAKAsgkissw~keIvrnWTVVYFNGLCGSI~ILVFI MLATRqf
FocA 4/5ab/6 3KLZ_A | lqAFALGLMCNlLVCLAVWMTFSars~ITDKVMVLI LPVAMFVSSGfeHCIANMFQVPMAlgikyfapesfwa tganiaqyadlnfnvfnvnnLIPVTLGNIVGGG~VFVGMWYWLlYlk
FocA 1/2ab/3 3KCV_A | lktFYLAITAGVFISIAFVFIYITattgtgt pfg akLVGGICFSLGLILCVVCGadLFTSTVLI VVAKAsgritw~gqlaknWLVVYFNGLVGAL~LFVLLMWLSGey
FocA 4/5ab/6 3KCV_A | IeAVCLGILANLMVCLAVWMSYSGrs~IMDKAFIMVLPVAMFVASGfeHSIANMFMI PMGivirdfaspefwtavgsapenfshltv nftdnLIPVTIGNIIGGG~LLVGLTYWVlylr
Aqp1 1/2ab/3 5I32_A | IaSLRAYLAEFISTLLFVAGVGSAlayakltsda~al dtpglvAI AVCHGFALFVAVAI GANISGGHVNPVTFGLAVGgqi~tviTGVFYWI AQLLGSTAAACFLKYYVTggIav
Aqp1 4/5ab/6 5I32_A | IgSIEGVVMEIITFALVYTYATAAdpkkg~slgtIAPLAIGLIVGANILAAGPFSGGSMNPARSFGPAVAag~dfsgHWVYVWVGLIGGGLAGLIYGNVFGsse
Aqp1 1/2ab/3 3NE2_A | tLAKRFTAEEVVGTFILVFFGPGAAVitl iangadkpnfnigigalggIgdwfaIGMAFALAAAVIYSLGRISGAHINPAVTIALWSigrf~pgrEVVPPYIYVIGPIVGAVAALYNYLAKe
Aqp1 4/5ab/6 3NE2_A | igYGAFLTEAIGTFLMLVIMGVAVderapppFAGLVIGLTVGGIITIGNITGSSLNPARTFGPYLGDsl ginlwqfPPIYVIGPIVGAVAALYNYLAKe
Aqp1 1/2ab/3 5DYE_A | vaFLKAVFAEFLATLIFVFFGLGSALKwpsa~lptIIQIALAFGLAIGTLAQALGPVSSGGINPAITLALLVGNqis~sIRAFFYVAAQLVGAIAAGAGIYGVAPInar
Aqp1 4/5ab/6 5DYE_A | ttQQGAMVVELILTFQLALCIFASTDsrrt~epvgSPALSIGLSVTLGHLVGIYFTGCSMNPARSFGPAVVnr~fspAHWVFWVWVGLVGAVALAALYFYLLfpnsI
Aqp1 1/2ab/3 3C02_A | vrrEFIGEFLGTFLVLMFLGEGATAnfhtg~lsgdwyKLLCLGWGLAVFFGLVLSaklsgahINLAVSIGLSINkfd~lkkIPVYFFAQLLGAFA~VGTSTVYGLYhg
Aqp1 4/5ab/6 3C02_A | tgAFFNELILTLGILLVLI LVVVDenl cg~kfhIIKLSVVVGLIILCIGITFGngtfgaINPSRDLGSRFLSIaygkdt~f tkdnfyFWVPLVAPCVGVSFVCFQFYDKVICp

```

**Figure S2 - Multiple alignment of FocA and Aquaporin protodomains. FocA** (PDB: 4FC4, 3KLZ, 3KCV), **AQP** (PDB: 3NE2, 5I32, 5DYE, 3C02)

**Protodomains RMSD optimized within the FocA family** 1.88 Å (4FC4), 1.17 Å/2.11 Å (3KLZ), 1.55 Å/2.06 Å (3KCV), highlighting for each pair of protodomains in a structure the conserved residues in RED. Some residues are conserved at the domain level, and at the family level. FocA show a partial “internal” conservation between TM3 and TM6 at the family level, with especially a [G]NxxG[G] motif, while within each individual representative, internal” homology is significant. A larger sequence alignment of the Pfam01226 family confirms that motif.

**Protodomains RMSD optimized within the Aquaporin family** 1.37 Å (5I32), 0.80 Å/1.20 Å (3NE2), , 0.78 Å/1.58 Å (5DYE), highlighting for each pair of protodomains in a structure the conserved residues in RED. Some residues are conserved at the domain level, and at the family level. Aquaporins has retained a higher internal homology between protodomains than FoCA, especially with its NPA motif in TM2b and TM5b at the family level.

**Protodomains RMSD across FoCA and Aquaporin families.** The first protodomain of **FocA vs. other FocA protodomains** 1.88 Å (4FC4), 1.18 Å/2.05 Å (3KLZ), 1.54 Å/2.00 Å (3KCV) **vs. AQP protodomains** 2.17 Å/2.82 Å (5I32), 2.25 Å/2.43 Å (3NE2), 2.33 Å/2.94 Å (5DYE), 2.21 Å/2.71 Å (3C02). While the structure match is very good between any representative of these two families FocA and Aquaporin, the sequence match is rather poor overall, yet the FoCA TM3 motif region shows some overlapping homology with AQP [G/A]xxx[G/S][G/A/S] motif. See Figure S10 for sequence vs. structure similarities.







```

2RH1_A | g GIVMSLIVLAI VFGNVLVITAI AKFerl ~~~qtVTNYFITS LACADLVMLG LavvPFGaahil k ~~~~~ wtfgnfWCEFWTSIDVLCVTSAS IETLCVIAVDRYFAITspfkyyq~~
3UON_A | f iVLVAGSLSLVTIIGNI LVMYSIKVNrhl ~~~qtVNNYFLFSLACADLIIGVfs NLYtlytvi g ~~~~~ ywplgpvVCDLWLALDYVVSNASVMNLLIISFDRYFCVTkplityp~~
4MQS_A | f iVLVAGSLSLVTIIGNI LVMYSIKVNrhl ~~~qtVNNYFLFSLACADLIIGVfs NLYtlytvi g ~~~~~ ywplgpvVCDLWLALDYVVSNASVMNLLIISFDRYFCVTkplityp~~
35N6_R | g GIVMSLIVLAI VFGNVLVITAI AKFerl ~~~qtVTNYFITS LACADLVMLG LavvPFGaahil k ~~~~~ wtfgnfWCEFWTSIDVLCVTSAS IETLCVIAVDRYFAITspfkyyq~~
5G53_A | v yITVE LAIAVLA I LGNVLVCWAVWLNsnl ~~~qnVTNYFVVS LAAAD I LVGVla i PFAitistg ~~~~~ fcaahGCLFIACFVLLVLAQSSIFSLAIAIDRYIAIRiplyrn~~
4N6H_A | a iTALYSAVCAVGLGNVLMVFGIVRYtk ~~~ktATNIYIFNLALADALATSt~IPFQsakyle ~~~~~ twpfgelCKAVLSIDYNNMFTSIFTLTMSVDRYIAVChpvkall~
5IU4_A | v yITVE LAIAVLA I LGNVLVCWAVWLNsnl ~~~qnVTNYFVVS LAAAD I LVGVla i PFAitistg ~~~~~ fcaahGCLFIACFVLLVLAQSSIFSLAIAIDRYIAIRiplyrn~~
4PHU_A | l sFGLYAAAFALGFP LNVLAI R GATAHAr ~~~rltPSAVYALNLGCSDLLLTVs~IPLKave lasg ~~~~~ awplpasLCPVFAVAH FAPLYAGGGFLAALSAAARYLGAAfpi gyq~~
4ZJ8_A | v lIAAYVAVFVVALVGNLTVLAVWRNhh ~~~rtVTNYFIVNLSLADVLTVAicIPASllydite ~~~~~ swl fghaLCKVlPYLQAVSVSVAVLTLSFIALDRWYAICHp l f~~
45VY_A | v lIAGYIIVFVALIGNVLCVAVWKNhh ~~~rtVTNYFIVNLSLADVLTVAicIPASllydite ~~~~~ twf fggqLCKVlPYLQAVSVSVAVLTLSFIALDRWYAICHp l f~~
5DSG_A | f iATVTGSLSLVTVVGNILVMLSIKVNrql ~~~qtVNNYFLFSLACADLIIGVfs NLYtlytvi g ~~~~~ ywplgavVCDLWLALDYVVSNASVMNLLIISFDRYFCVTkplityp~~
5CXV_A | f iGITGLLSLATVTGNLLVLSFKVNTel ~~~ktVNNYFLLS LACADLIIGTfs NLYtlytvi g ~~~~~ hwalgtlACDLWLALDYVASQASVMNLLIISFDRYFSVTrp l syr~~
3VZV_A | l tSVVFLICCFIILENI FVLLTIWKTKkf ~~~hrPMYFIGNLALSDLLAGVa~yTANlilsgat ~~~~~ awplpasLCPVFAVAH FAPLYAGGGFLAALSAAARYLGAAfpi gyq~~
5UO9_A | a iAVLSTLGTFTVLENLLVLCVILHSrsl ~~~rcrPSYHFIGSLAVADLLGsvi~FVYSfidfhvf ~~~~~ hrkdsrnVFLKLGVTASFTASVGSFLAAIDRYISIHrp l ayk~~
5GLH_A | i nTVVSCLVFVLGIIGNSTLLYIYKknc ~~~rnGNPILIASLALGDLHLHIVaiPINvykllae ~~~~~ dwpfgaeMCKLVPFIQASVGITVLSLALSIDRYRAVAsw s rik~~
2K5B_A | l wAAAYTVIVT SVVGNVVMWII LAHkr ~~~ktVNNYFLNLAFAEASMAAfnTVNftya vhn ~~~~~ ewvfgnaMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVhafa l~~
3VWZ_A | f vPSVYTGvFVSLP LNIMAIVVFLK kv ~~~kkPAVVYMLHLATADVLFVsv~IPFKisyfsgs ~~~~~ dwqfgseLCRFVTAAYFCNMYSIILMTVSIIDRF LAVVYqsl~~
4IAR_A | l lVMLLALITLATTLSNAFVIATVYRTrkl ~~~htPANYLIASLAVTDLVLSI l v PIST ytvgtg ~~~~~ rwtlqqvVCDLWLALDYVVSNASVMNLLIISFDRYFCVTkplityp~~
5UEN_A | a yIGIEVLIALVSPGNVLVWAVKVNqal ~~~rdATFCFIVSLAVADVAVGAlviPLAilinig ~~~~~ pqtyfhTCLMVACPVLLTQSSILALLAIAVDRYLRVKip l ryk~~
4ZUD_A | i PTLYSIIFVVGIFGNSLVVIVYFY kl ~~~ktVASVFLNLAALADLCFLlt~IPLWavytaey ~~~~~ rwpfgnyLCKIASASVSNLYASVFLTCLSIDRYLAIVhpk s r~~
4K5Y_A | v aAIINYLGHCI SLVALLVA FVLFLRArsi ~~~rcLRNIHANLIAAFILRNAt~wFVvqlt spe ~~~~~ wvfgnaMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVhafa l~~
3RZE_A | p lVVVLTICLVTVGNLLLVLYAVRSErkl ~~~htVGNLYIVLSVADLIVGAvv PMNilylls ~~~~~ kws lgrPLCLFWSMDYVASTASIFSVFILCIDRYRSVQqplryl~~
3PBL_A | y yALSICALILAIVFGNGLVCMVAVLKEral ~~~qtTNYLVSLAVADLLVATlv PWWvylevtgg ~~~~~ vwnfsr iCCDVFTLDVMMCTASIWNLCAISIDRYTAVV pvh yqh g~~
4DJH_A | i iTAVYSVVFVGLVGNLVMFVIRYtk ~~~ktATNIYIFNLALADALVTTt~PFQstvylln ~~~~~ swpfgdvLCKI VLSIDYNNMFTSIFTLTMSVDRYIAVChpvkall~
5TVN_A | w aALLIMVI IPTIGNTLVI LAVSLEkkl ~~~qyATNYFLMSLAVADLLVGLfv PIALliti fea ~~~~~ wplplvLCPAWLFLDLVFTASIWHLCAISVDRYIAIKkpiqan~~
5T1A_A | l iPPLYSLVFI FGFVGNMLVLLI LINCkkl ~~~kcLTDIYLLNLAISD LFLlt~IPLWahsaane ~~~~~ wvfgnaMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVhafa l~~
5VEW_A | f iYIYTVGYALSFSALVIA SAILLGFrl ~~~hctRNYIHLNLFASFILRALc~vFFKdaalkw gsg ~~~~~ dgllsyqdsIACRVLVLLXQCVAAANYWLLVEGVYLYTLAfnife l r~~
5EEZ_A | s iQVMYTVGYSLSLAALLLALAILGGLskl ~~~hctANAIHANLFLS FVLKASa~VLFldgl l r trysqkiedd lsvs twl s d g a v a ACRVAAVFMQYGVAN YCWLLVEGLYLNLLGInife l r~~
4XNV_A | y lPAVYLVFIIIGFLGN SVAIWMFVFH kp ~~~wsGI SVYMFNLALADFLYVLTlpAlifyyfnkt ~~~~~ dwifgdMCKLQRFIFHVNLVYGSILFTCI SAHRYSGVvyp l ksl~~
5UNF_A | a iPILYYIIFVIGFLVNI VVVTLCQQkqp ~~~kkVSSYIIFNLAVADLLLAt~IPLWatyysyr ~~~~~ dwl fgpvMCKVFGSFLTNMFASIFFITCMSVDRYQSVIyp l sq~~
4MBS_A | l iPPLYSLVFI FGFVGNMLVLLI LINYkrl ~~~ksMTDIYLLNLAISD LFLlt~vFWahyaaaq ~~~~~ wd f g n t M C Q L L T G L Y F I G F F S G I F F I I L L T I D R Y L A V V h a v f a l ~~~
5LWE_A | f iPPLYWLVFIVGALGN S L V I L V Y W Y C a r a ~~~ktADMFLNLAISD LFLlt~vFWahyaaaq ~~~~~ wk f q t F M C K V V N S M Y K M N F Y S C V L L I M C I C V D R Y I A I A q a r a h t ~~~
5NDD_A | f iPIVYTI V F V V A L P S N G M A L W V F L F R t k k ~~~kaPAVIYMANLALADLLSVIw~fPKia yhihgn ~~~~~ nwi ygeaLCNVLIGFFYANMYSIILFTCLSVQRAWEI Vnp ghs~~
3ODU_A | f iPTIYSIIFLTVGVGNLVI LVMGYQkkl ~~~rsMTDKYRLHLSVADLLFVIt~LPFwadvava ~~~~~ nwyfgn fLCKAVHVIYTNVLYSSVWILAFISLDRYLAIVhatsnsq~~
4Z35_A | l vMGLGITVCFIMLANL LVMVAIYVNr f ~~~hfPIYLLMANLAAADFFAGLa~yFYL fntgpn ~~~~~ trrltvsTWLLRQGLIDTSLTASVANLLAIAIERHITVFr q l h t ~~~
4N4W_A | d HSYIAAFGAVTGLCTLF L A T F V A D w r n ~~~snrYPAVILFYVNA CFVGSIGwIAQF dgareivcradg~t r l g e p t s n e t l S C V I I F V I V Y Y A L M A G V V V F V V L T Y A W H T S F K A I g t t y ~~~
4PXZ_A | l f P L L Y T V L F F V G L I T N G L A M R I F F Q I r s ~~~ksNFIIFLKNTVISD LLMl L t f p F K I l s d a k l g ~~~~~ t g p l r t f V C Q V T S V I F Y F T M Y I S I S F L G L I T I D R Y Q K T T r p f k t s ~~~
1GZM_A | l AAYMFLIMLGFPI N F L T L Y V T Q H k k l ~~~rtPLNYI L L N L A V A D L F M V F g g f T T l y t s l h g ~~~~~ y f v f g p t G C N L E G F F A T L G G E I A L W S L V V L A I E R Y V V V C k p s n f ~~~
4OR2_A | i eSIIAIAFSCGLGLVTLFVTLIFVLYrdtpv vksSSRELCYIILAGIFLGYVc~pFTLiakp ~~~~~ ttSCYLQRLLVGLSSAMCYSALVTKTNRARI Lagskkkic t~~
4009_A | p aPIAAVVFACGLLATL F V T V V F I Y r d t p v v k s S R E L C Y I I L A G I C L G Y L c ~ t F X L i a k p ~~~~~ k q i Y C Y L Q R I G I G L S P A M S Y A L V T K T Y R A A R I L A s k k n i f e ~~~
4Z9G_A | v aAIINYLGHCI SLVALLVA FVLFLRArsi ~~~rcLRNIHANLIAAFILRNAt~wFVvqlt spe ~~~~~ wvfgnaMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVhafa l~~

```

TM1

TM2

TM3

```

2RH1_A ~slltknKARVILMWVIVSGLTSFLPIQ hwyathqea inc~~~~~yaeetccd f ftnqaYAIASSIVSFYVPLVIMVFVYSRVFqeakrqlnife
3UON_A ~vkr t tkMAGMMIAAAWVLSFILWAPAILfwqfivgvr tved~~~~~gecyiqf f snaaVTFGTAIAAFYLPVIIMTVLYWHISrasksrinife
4MQS_A ~vkr t tkMAGMMIAAAWVLSFILWAPAILfwqfivgvr tved~~~~~gecyiqf f snaaVTFGTAIAAFYLPVIIMTVLYWHISrasksr ikkdk
35N6_R ~slltknKARVILMWVIVSGLTSFLPIQ hwyathqea inc~~~~~yaeetccd f ftnqaYAIASSIVSFYVPLVIMVFVYSRVFqearqlqkid
5G53_A ~glvtgtRAKGI I AICWVLSFAIGLTPMLgwnncgqpkgekahsqgc~~~gegqvac l f e d v v p n y M V Y F N F F A C V L V P L L L M L G V Y L R I F l a a r r q l k q e
4NGH_A ~dfr t p a K A K L I N I C I W V L A S G V G V P I M V a v t r p r d g a v v c ~~~~~l q f p s p s w y w d t v T K I C V F L F A F V V P I L I I T V C Y G L M L I r l r s v r l l s ~
5IU4_A ~glvtgtRAAGI I AICWVLSFAIGLTPMLgwnncgqpkgekahsqgc~~~gegqvac l f e d v v p n y M V Y F N F F A C V L V P L L L M L G V Y L R I F a a a r r q l a d l e
4PHU_A ~a f r p c Y S W G V C A A I W A L V L C H L G L V F G l e a p g g l d h s n t s l g i n t p v n g s p v c l e a w d p a s a g g A R F S L S L L L F F L P L A I T A F C F V G C L r a l a r g s n i f e
4ZJ8_A ~~k s t a r R A R G S I L G I W A V S L A I M V P Q A A v e c s s v l p e l a n r t r l ~~~~~f s v c d e r w a d d l y p k i Y H S C F F I V T Y L A P L G L M A M A Y F Q I F r k l w g r q g i d c
450V_A ~~k s t a k R A R N S I V I I W I V S C I I M I P Q A I v e c s t v f p g l a n k t t l ~~~~~f t v c d e r w g g e i y p k Y H I C F F L V T Y M A P L C L M V L A Y L Q I F r k l w c r q g i d c
5DSG_A ~a r r t t k M A G L M I A A A W V L S F V L W A P A I L f w q f v v g k r t v p d ~~~~~n q c f i q f l s n p a V T F G T A I A A F Y L P V V I M T V L Y I H I S l a s r s r v n i f e
5CXV_A ~a k r t p r A A A L M I G L A W L V S F V L W A P A I L f w q y l v g e r t v l a ~~~~~g q c y i q f l s q p i I T F G T A M A A F Y L P V T V M C T L Y W R I Y r e t e n r n i f e
3V2Y_A ~~g s n n f R L F L L I S A C W V I S L I L G G L P I M g w n c i s a l s s ~~~~~c s t v l p l y h k H Y I L F C T T V F T L L L L S I V I L Y C R I Y s l v r t r n i f e
5U09_A ~r i v t r p K A V V A F C L M W T I A I V I A V L P L L g w n c e k l q s v ~~~~~c s d i f p h i d e T Y L M F W I G V T S V L L L F I V Y A Y M Y I L w k a g i d c s f w n
5GLH_A ~g i g v p k W T A V E I V L I W V Y S V L A V P E A I g f d i t d y k g s y l r i c ~~~~~l l h p v q k t a f q y a t a K D W W L F S F Y F C L P L A I T A F F Y T L M T c e l r k n i f e
2K5B_A ~~r l s a t A T K V V I C V I W V L A L L L A F P Q G Y s t t e t p s r v v c ~~~~~i e w e p h n k i y e k v Y H I C V T V L I Y F L P L L V I G Y A Y T V V G i t l w a s e i p g d
3VW7_A ~s w r t l g R A S F T C L A I W A L A I A G V V P L L L k e q t i q v p g l i t t c ~~~~~h d v l s e t l l e g y y a y Y F S A F S A V F F F V P L I I S T V C Y V S I I r c l s s a n i f e
4IAR_A ~a k r t p k R A A V M I A L V W V F S I S I S L P P F F w r q a k a e e e v s ~~~~~c v v n t d h i l Y T V Y S T V G A F Y P T L L L I A L Y G R I Y v e a r s r i a d l e
5UEN_A ~v v t p r A A A V A I A G C W I L S F V V G L T P M F g w n n l s a v e r a w a a a g s ~~~~~g e p v i k c e f e k v i s e y M V Y F N F F V W V L P P L L M V L I Y L E V F y l i r k q l a d l e
4ZUD_A ~l r r t I V A K V T C I I I W L L A G L A S L P A I I h r n v f f i e n t n i t v c ~~~~~a f h y e s q n s t l p i g L G L T K N I L G F L P F L I I L T S Y T L I W k a l k k a y e i ~
4K5Y_A w d a y d r I R A W M F I C I G W G V P F P I I V A W A I g k l y d n e k ~~~~~c w a g k r p g v Y T D Y I Y Q G P M A L V L L I N F I F L F N I V r i l t k l r a s ~
3RZE_A ~k y r t k t R A S A T I L G A W F L S F L W V I P I L G w n h f q q t s v r r e ~~~~~d k c e t d f y d v t w F K V M T A I I N F Y L P T L L M L W F Y A K I Y k a v r q h c n i f e
3PBL_A t g q s c r R V A L M I T A V W L A F A V S C P L L F g f n t t g d p t ~~~~~v c s i s n p d F V I Y S S V S F Y L P F G V T V L Y A R I Y v v l k q r r r k n i
4DJH_A ~d f r t p l K A K I I N I C I W L L S S S V G I S A I V l g g t k v r e d v d v i e c ~~~~~s l q f p d d d y s w w d l f M K I C V F I F A F V I P V L I I I V C Y T L M I I r l k s v r l l s g
5TVN_A ~q y n s r a T A F I K I T V V W L I S I G I A I P V P I k g i e t d v d n p n n ~~~~~i t c v l t k e r f g d F M L F G S L A A F F T P L A I M I V T Y F L T h a l q k k a a d l e
5T1A_A ~k a r t v t F G V V T S V I T W L V A V F A S V P G I I f t k x q k e d s v ~~~~~y v c g p y f p r g W N N F H T I M R N I L G L V L P L L I M V I C Y s g i s r a s k s r i
5VEW_A d a y s e q w I F R L Y V A I G W G V P L L F V V P W G I v k y l y e d e g ~~~~~c w t r n s n N Y W L I R L P I L F A C I V N F L I F V R V I c i v v s k l k a n ~
5EE7_A d a y p e r s F F S L Y L G I G W G A P A L F V V P W A V v k c l f e n v q ~~~~~c w t s n d n G F W W I L R F P V F L A I L I N F F I F V R I V q l l v a k l r a r ~
4XNV_A ~g r l k k k N A I C I S V L V W L I V V V A I S P I L F y s g t g v r k n k t i t c ~~~~~y d t s d e y l r s y f i Y S M C T T V A M F C V P L V L I L G C Y G L I V r a l i y k k y t
5UNF_A ~~r r n p w Q A S Y I V P L V W C M A C L S S L P T F Y f r d v r t i e y l g v n a c ~~~~~i a f p e k y a q w s a g I A L M K N I L G F I I P L I F I A T C Y F G I R k h l k t n s y ~
4MBS_A ~k a r t v t F G V V T S V I T W V V A V F A S L P N I I f t r s q k e g l h y t c s ~~~~~s h f p y s q q w k n F Q T L K I V I L G L V L P L L M V M I C Y S G I I k t l l r k y t
5LWE_A w r e k r l l Y S K M V C F T I W V L A A A L C I P E I L y s q i k e e s g i a i c t v ~~~~~y p s d e s t k l k s a v l a L K V I L G F F L P F V V M A C C Y T I I I H T L i q a k k ~~~~~
5NDD_A ~~r k k a n I A I G I S L A I W L L I L L V T I P L Y V v k t i f i p a l q i t t c ~~~~~h d v l p e q l l v g d f n Y F L S L A I G V F L P A F L T A S A Y V L M I r a l a d l e d n w e
3ODL_A ~r p r k l l A E K V V Y V G V I P A L L L T I P D F I f a n v s e a d d r y i ~~~~~c d r f y p n d l w v v V F Q F Q H I M V G L I L P G I V I L S C Y C I I s k l s h s g s n i
4Z35_A ~~r s n r R V V V V I V V I W T M A I V M G A I P S V g w n c i c d i e n ~~~~~c s n a p l y s d S Y L V F W A I F N L V T F V V M V V L Y A H I F g y v a d l e d n w e
4N4W_A ~~q p l s g K T S Y F H L L T W S L P F V L T V A I L A v a q v d g d s v s g ~~~~~i c f v g y k n y r Y R A G F V L A P I G L V L I V G G Y F L I R G V t l f s i k s n h p
4PXZ_A ~n p k n l I G A K I L S V V I W A F M F L L S L P N M I l t n r q p r d k n v k c ~~~~~s f l k s e g l v w h e I V N Y I C Q V I F W I N F L I V I V C Y T L I T k e l y r s y v r t a
1GZM_A ~~r f g e n H A I M G V A F T W M A L A C A A P P L V g w s r y i e p e g q c s c g ~~~~~i d y y t p h e e t n e s F V I Y M F V V H F I P L I V I F F C Y G Q L V f t v k e a a a q q q
4OR2_A p r f s a w A Q V I I A S I L I S V Q L T L V V T L I I e p p p i l s y p s ~~~~~i k e v y l i c n t s n L G V V A P L G Y N G L L I M S C T Y Y A F K T R n v p ~~~~~
4009_A p r f s a x A Q L V I A F I L I C I Q L G I I V A L F I e p p d i h d y p s ~~~~~i r e v y l i c n t t n L G V V A P L G Y N G L L I L A C T F Y A F K T R n v p ~~~~~
4Z9G_A a y l t d r I R A W M F I C I G W G V P F P I I V A W A I g k l y d n e k ~~~~~c w a g k r p g v Y T D Y I Y Q G P M A L V L L I N F I F L F N I V r i l t k l r a s ~

```

TM4

TM5

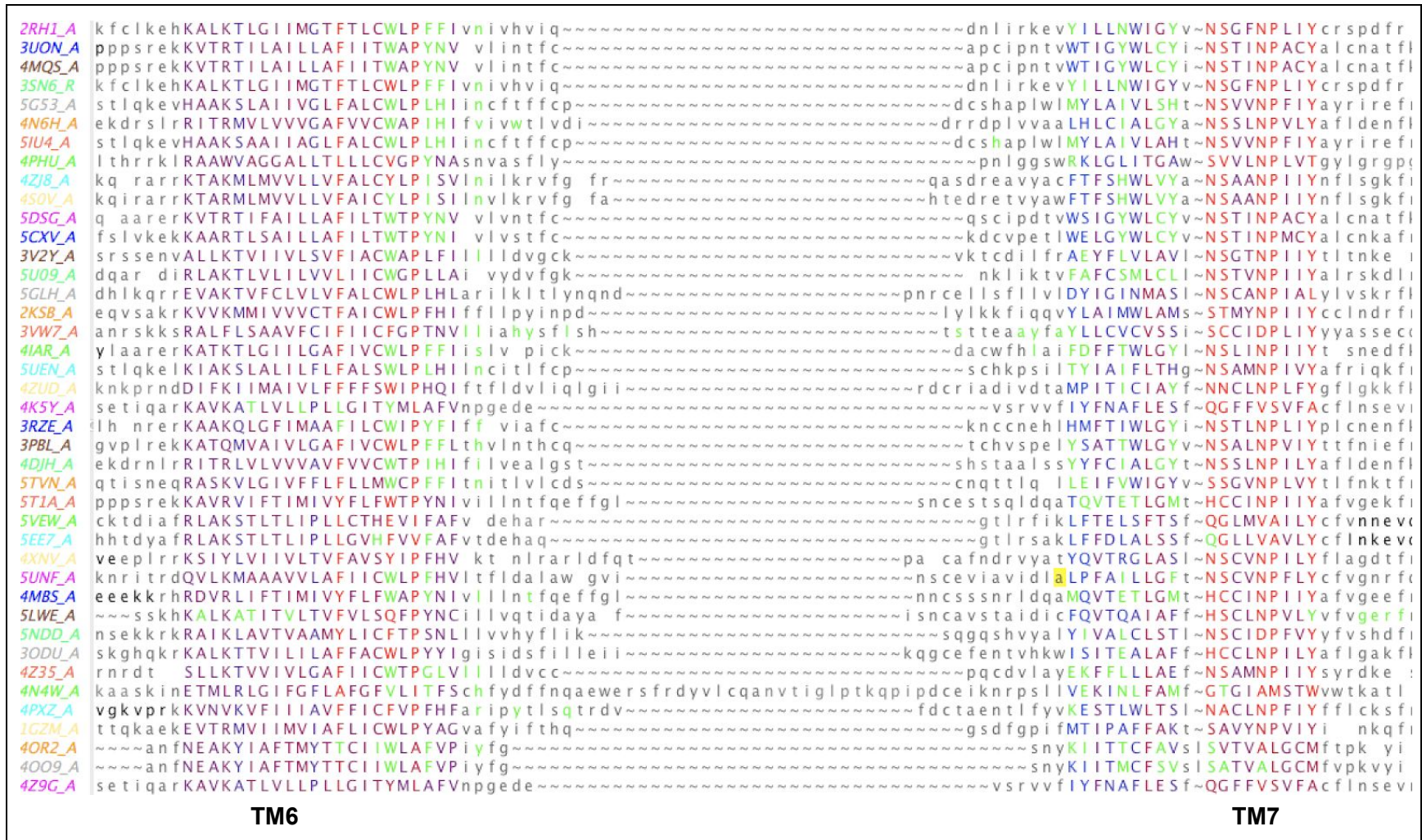


Figure S3.B - GPCR domains TM1-TM7 multiple alignment - (mostly Class A). Red = conserved, Green = ligand binding

A:1/2/3 A:5/6/7 BR:A/B/C BR:E/F/G	<b>1GZM_A</b> epwQFSMLAAYMFLLIMLg~FPINFLtlyvtvqhk~klrtpInyILNLAVALDFMVFGGFTTTLYTSLHgyfvf~gptgCNLEGGFFATLGGEIALWLSLV <b>1GZM_A</b> eetNNE SFVIYMFVVFHFIipLIVIFFcygqlvftvkeaaaqqqsattqkaeketrVIIMVIAFLICWLPYAGVAFYIFTHqgsdf~gpiFMTIPAFFAKTSAVYNPVIYI <b>1UAZ_A</b> etIWLIGITLLMLIGTFYf~IVKGGWvtd~keAREYYSITILVPGIASAAYLSMFFGigltevqvgseIdiyyaRYADWLFITPLLLLDLALLa <b>1UAZ_A</b> rrytWWLFFSTICMIVVLYFI~ATSLRAaaker~gpevASTFNTLTALVVLVLTAYPIILWIIIGtegagvvg~IgielIFMVLDVTAKVGGFGLLRSr
A:1/2/3 A:5/6/7 A:A/B/C A:E/F/G	<b>2Z1Y_A</b> yySLGIFIG~ICGIIICGGNGIIVYLFtkts ~qtpaNMFILNLA SDFTFSLVNGFPltiscfikkwiFGFAACKVYGFIGGIFGFMSIMTMAMISId <b>2Z1Y_A</b> trSNILCMFILGFFGPIILIFFCYFNIVsvsnhekeaaakrlnakelrkaqaganaerlAKISIVIVSQFLLSWSPYAVvallaq~fgplewVTPYAAQLPVMFAKASAIHNPMIYVSHpk <b>1GZM_A</b> fsMLAAYMF~LLIMLGFIPINFLTLTYVTvqhk ~rtpINyILNLAVALDFMVFGGFTTTLyts~lhgyfvFGPTGCNLEGGFFATLGGEIALWLSLVLAie <b>1GZM_A</b> neSFVIYMFvVHFIIPLIVIFFCYGQLVftvkeaaaqqe~sattqkaeketrMVIIMVIAFLICWLPYAGvafiyf~thqgsdFGPIFMTIPAFFAKTSAVYNPVIYIMMnkq
A :1-7 BR:A-G	<b>1GZM_A</b> wqfs LAAYMFLLIMLGFIPINFLtlyvtvqhkklrtpInyILNLAVALDFMVFGGFTTTLytslhgyfv~fgptgcnLEGGFFATLGGEIALWLSLVlaieryyvvcvkpsnfrfgenhai gVAFTWVMALACAAPplvgwsryiepeg <b>1UAZ_A</b>  lwigiGTLMLLIGTFYFIVKGGWvtdke~areyYSITILVPGIASAAYLSMffgigltevqvgseIdiyyaryADWLFITPLLLLDLALLak~vDRVSIIGTLVGVDAI ivtgIvgals~ <b>1GZM_A</b> ryiepeg qcs cgl dyypheetneSFVIYMFVVFHFIIPLIviffcygqlvftvkeaaaqqqsattqkaeketrMVIIMVIAFLICWLPYAgvafiyfth~qgsdfgpiFMTIPAFFAKTSAVYNPVIYI nqfrnc vttlccgkn <b>1UAZ_A</b> vgalss~htplarytwlFSTICMIVVLYFLATS raaa~kergeVASTFNTLTALVVLVLTAYPIILWIIIGtegagvvgIgielIFMVLDVTAKVGGFGLLRSrailgdteapepsagaeasa

**Figure S3.C - 1- Protodomains alignment of Rhodopsin (TM123 vs. TM567) vs. Bacteriorhodopsin (BR) (TM ABC vs. EFG).** (Class A) Bovine Rhodopsin (1GZM) protodomain 1 vs. protodomain 2: 3.11 Å; vs. Halobacterium salinarum bacteriorhodopsin (1UAZ) protodomain 1 2.58 Å, protodomain 2 3.69 Å. The sequence symmetry pattern is unique to each, yet a Methionine Identity match in TM1 and TM5 with the TM5 Methionine in contact with retinal in addition to the Schiff base linkage in TM7 xxAK. Similarly to Rhodopsin, BR shows a T in TM3 residue in a symmetrically equivalent position to K in TM7, as if that pair had been inherited from an ancestor in a divergent scenario, or coevolved in a similar way in a pseudo-symmetric coevolution scenario. The TM3/TM7 pseudo symmetric match in structure and partially in sequence is common to both with some of the key binding residues, but most noticeable key binding positions. The orientation of the trans-retinal is different, due to a larger number of aromatic residues, especially W residues at various positions, shifted in position. Most noticeable the (Fxxx)WxxY motif Rhodopsin TM6, binding retinal in both is shifted along the TM6 helix with an . **2- Protodomains alignment (TM123 vs. TM567) of Bovine Rhodopsin vs. Squid Rhodopsin (BR) TM ABC vs. EFG.** (RMSD 3.22/1.28/3.22A resp.) Here the (FFA)K/T motif in TM7/TM3 has evolved dissymmetrically as (MFA)K vs. (FIG)G in the rhodopsin family itself, where trans-retinal ligand and the full rhodopsin structures are superimposable within 1.43Å over 259 residues. Both have the protodomain “canonical” symmetry related pair match D in TM3 vs FxxxW in TM7, and most of the same retinal binding residues (highlighted in green) **3- Domain alignment** for reference Bovine Rhodopsin (1GZM) vs. Halobacterium salinarum bacteriorhodopsin (1UAZ) RMSD = 3.13Å while Bovine and Quid Rhodopsins will match with an overall RMSD = 1.43Å, and possess all the canonical GPCR Class A conserved residue.

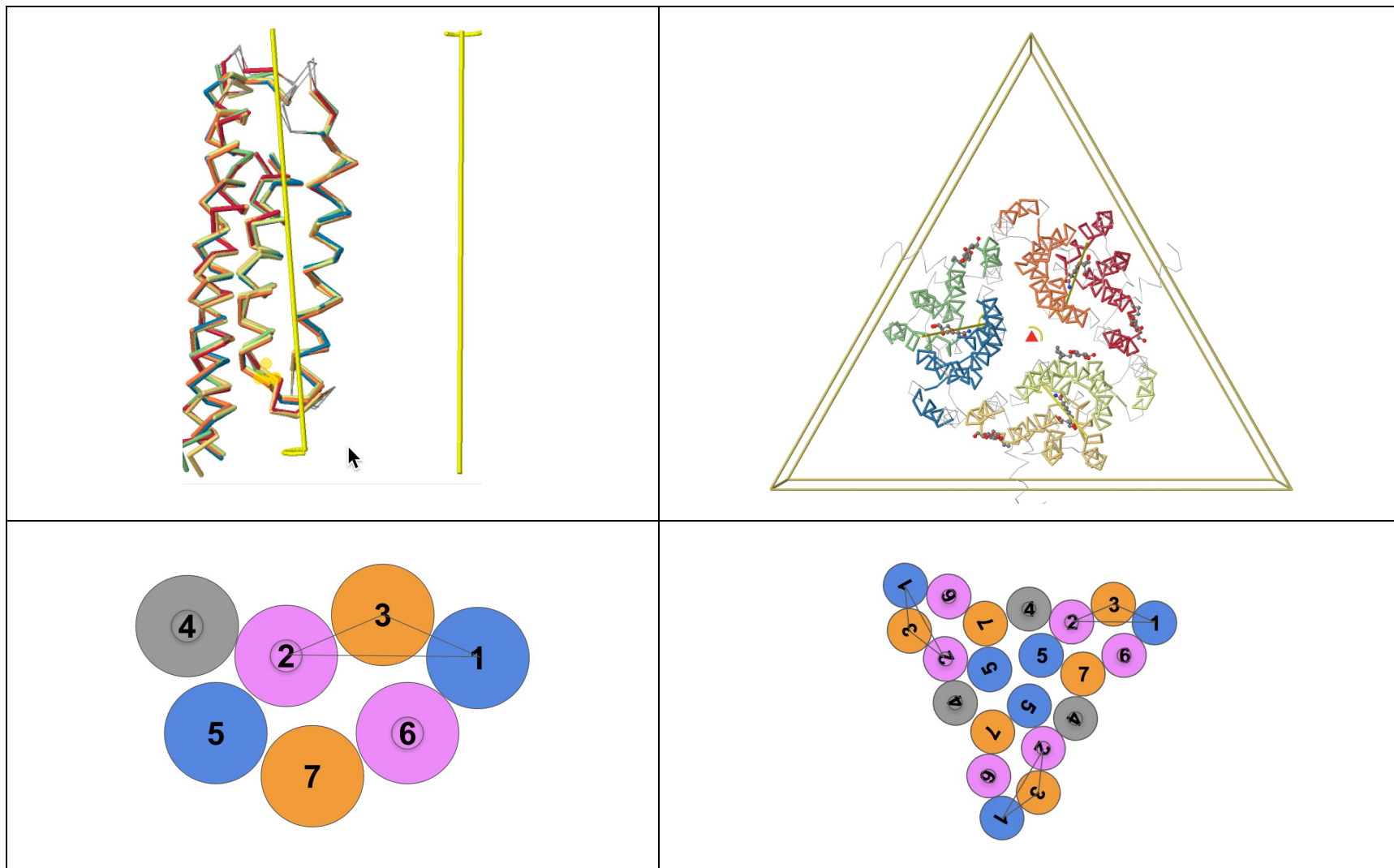


Figure S4 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry - SWEET Protodomain topology parallel C2 132c/s21a

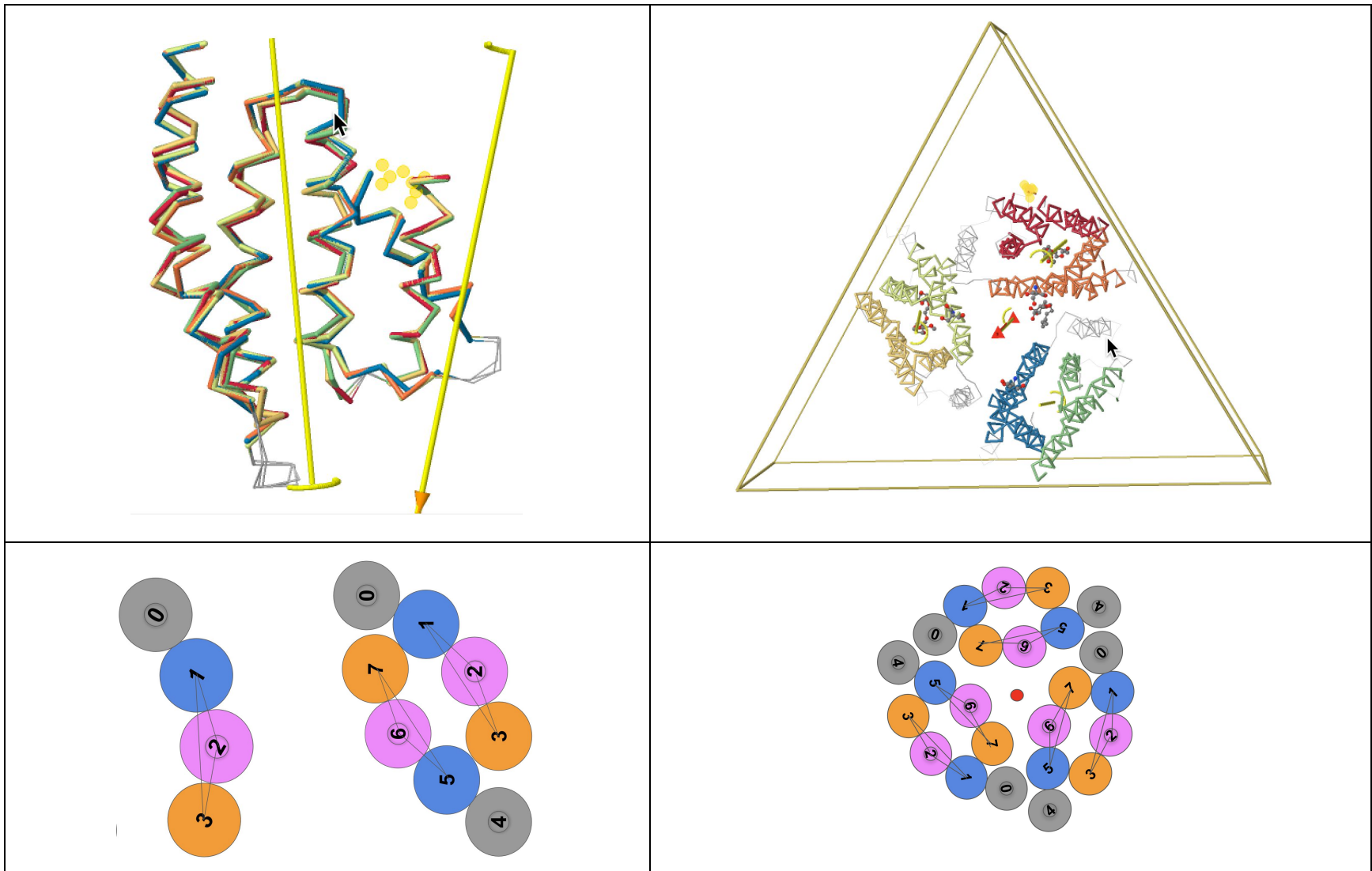


Figure S5 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry - PnuC Protodomain topology parallel C2 123c/s31p

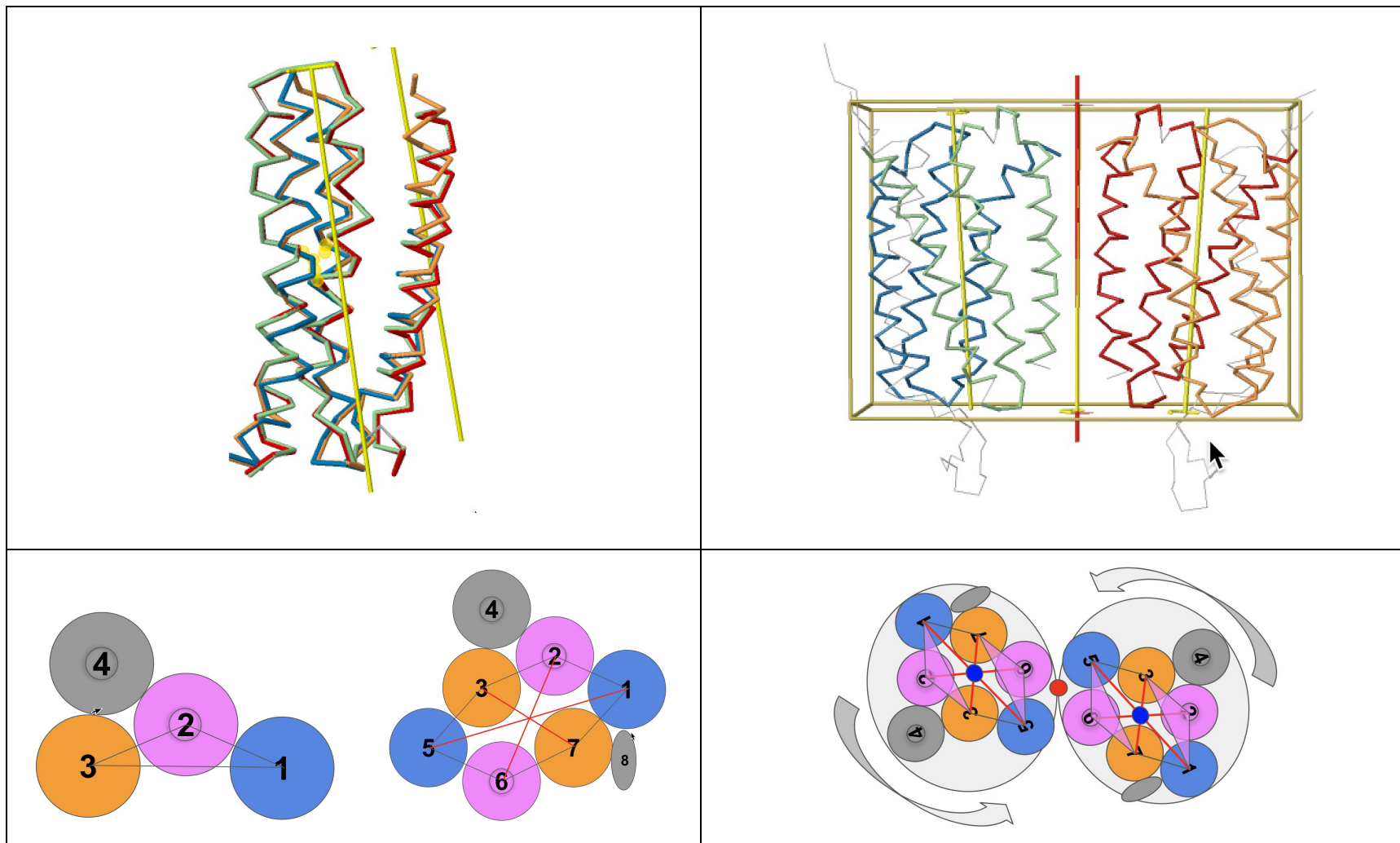


Figure S6 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry - GPCR Protodomain topology parallel C2 123cc/s31p.



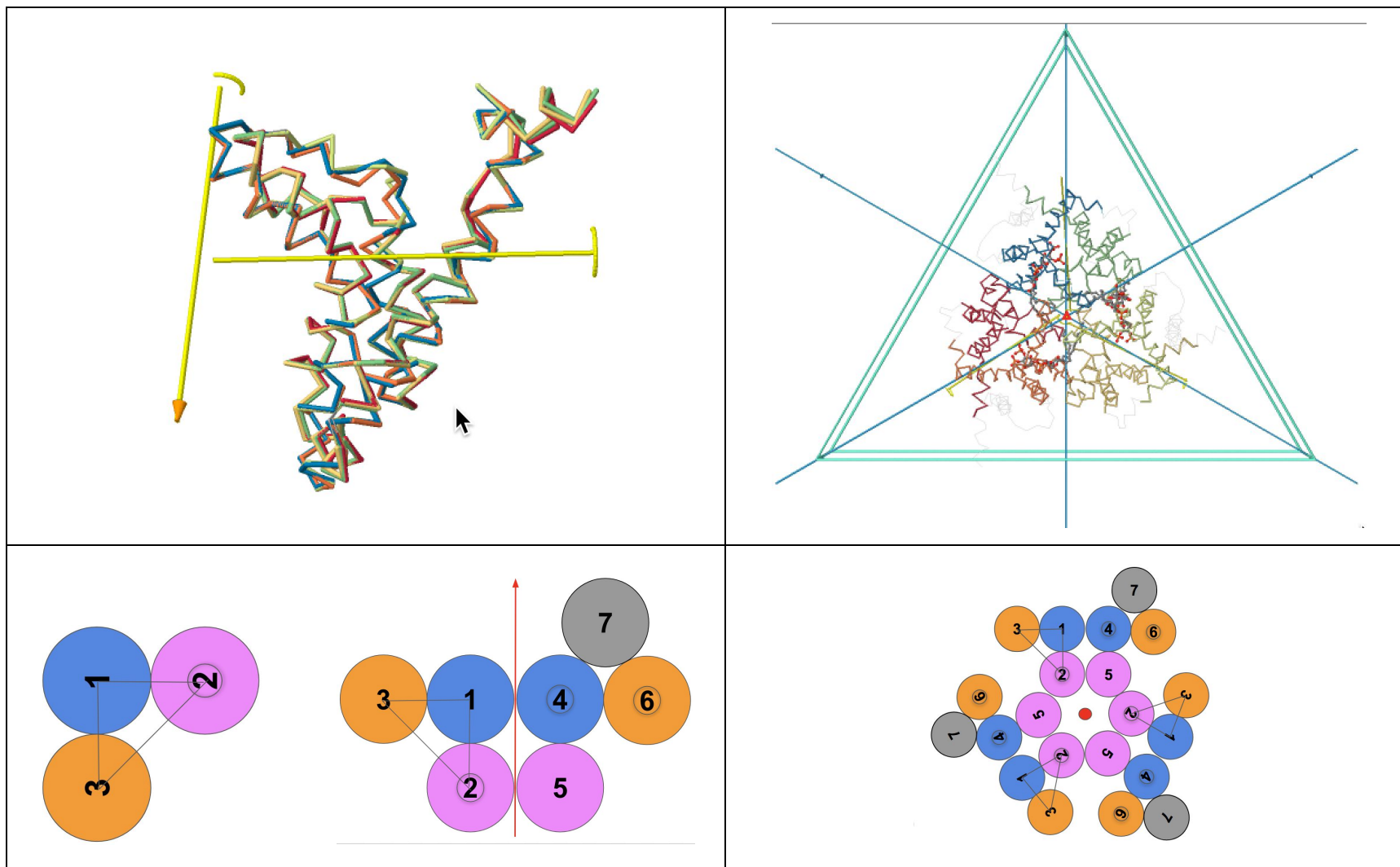
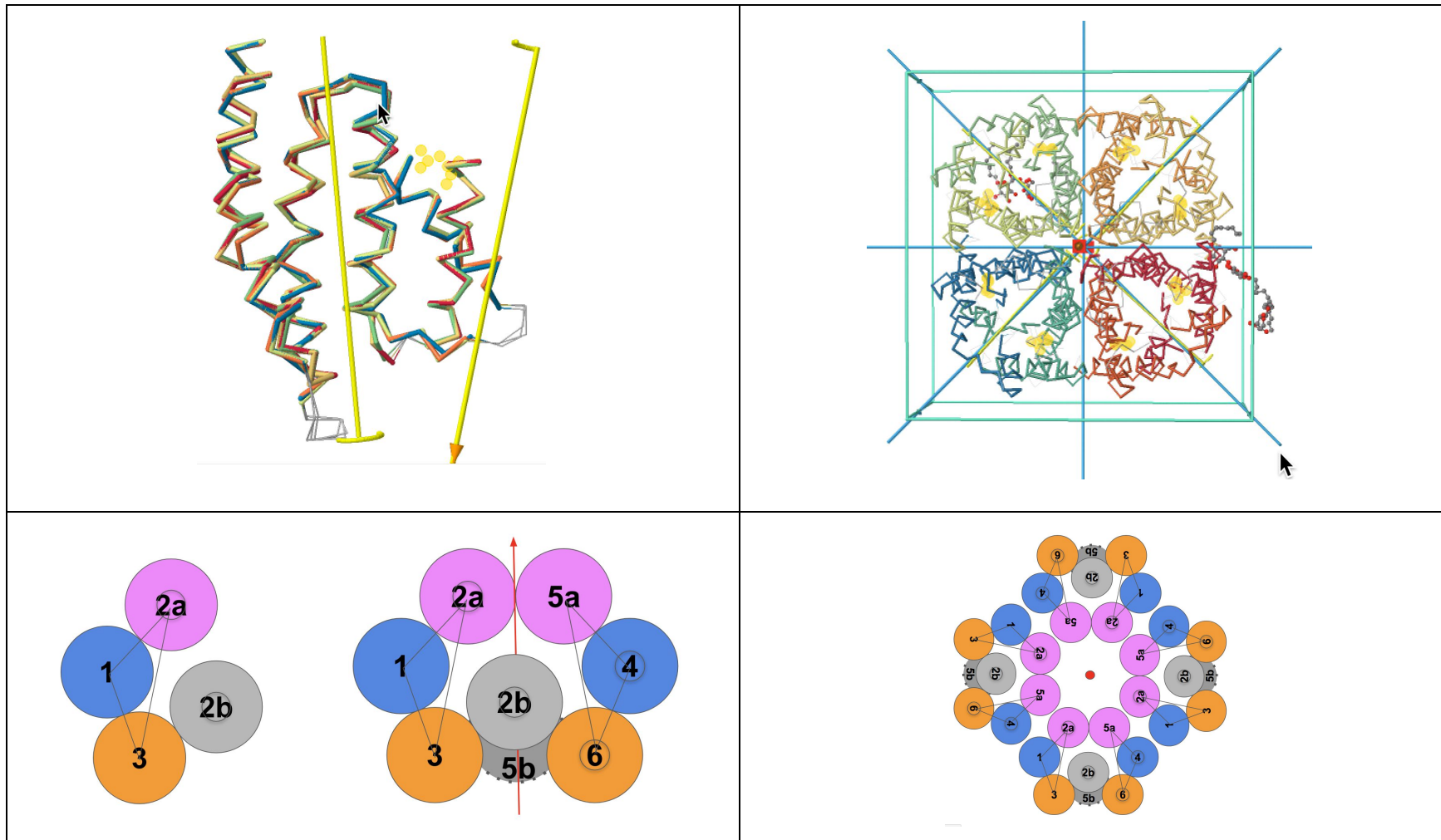
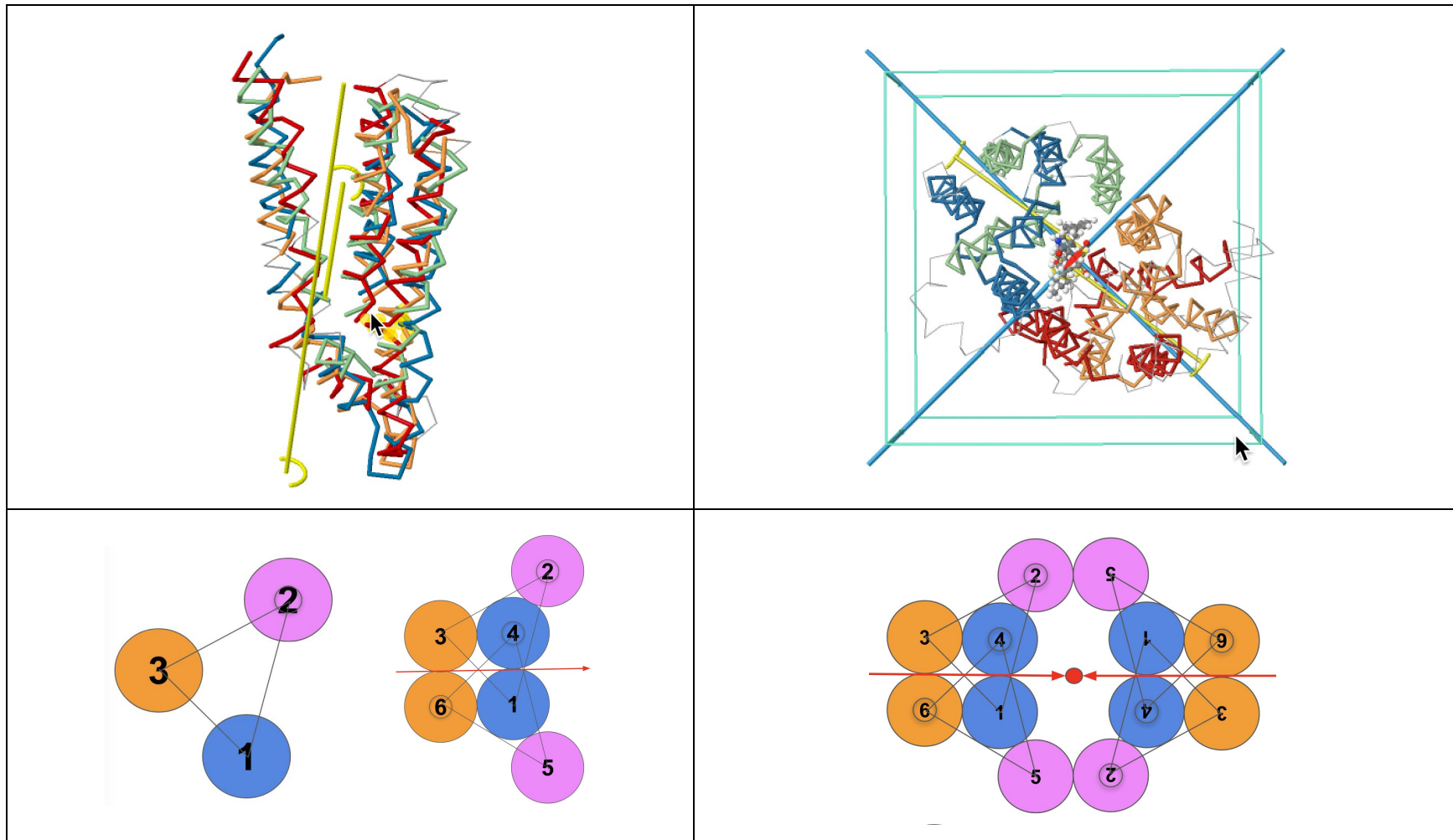


Figure S7 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry - TRIC Protodomain topology inverted 312c/s11a22a



**Figure S8 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry** - Aquaporin Protodomain topology inverted C2 312cc/a22a [forms a tetramer quaternary structure, FocA uses the same protodomain and forms a pentamer]. This is the only example with an asymmetric (a) 22 interface, of a very peculiar and idiosyncratic nature: helix 2 splits in 2 forming 2a and 2b to with a helix 2a-2a antiparallel interface from each protodomain on one side and 2b-2b on the other where the latter stack on top of each other one going up the other down - [iCn3D 3D visualization](#)

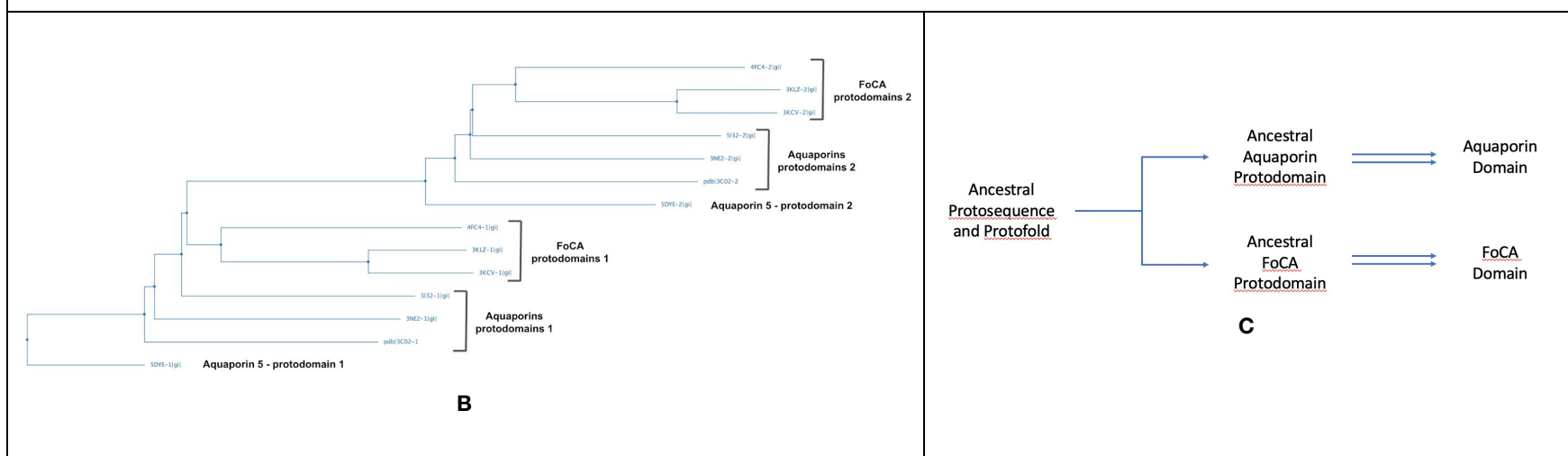


**Figure S9 - Protodomains, Domain Pseudo-symmetry and Quaternary symmetry - MFS** Protodomain topology inverted C2 interdigitated 132cc/s11a33a MFS - **Lower left**) Inverted Interdigitated Protodomains forming a C2 symmetric domain. It possesses two locally symmetric interfaces TM1-TM1 and TM3-TM3 (s11a33a) **Lower right**) C2 symmetric domain packing through a double (symmetry related) TM2-TM2 interface (2\*s22a) formed by the second helix of each of the 4 protodomains. The 2-domains/4-protodomains MFS protein has an overall D2 symmetry. (PDB: [5EQI](#)) since the internal symmetry axes of each domain colinearize and are orthogonal to the central axis (perpendicular to the membrane planes).

```

4FC4_A | gFWVSSAMAGAYVGLGII LIFTI gn lld ~~~~~psvrp lVMGATFGI ALTLVI IAGs ~~~~~eLFTGHTMFLT LGvkagtis ~~~~~hgg wai LPQTWLG NLVG SV~FVALLY SWGGs
4FC4_A | tvLFFK GAlCNWLVCLAIWMAIRte ~~~~~g tAKFLAIWCLLAF IASGy ~~~~~eHSVANMTL FALSwfghsday ~~~~~t lagighnLLWVT LGNTLSGVVFMGLGYWYATpk
3KLZ_A | ykSFL LAI SAGIQIGIAFVFYTVrttgahd ~~~~~pygvtk lLGGLAFSLGLI LVVITGg ~~~~~eLFTSSVLI LVAKAsgk i sw ~~~~~kelvrnWTVVY FGNLCGSI~ILVFI MLATRqf
3KCV_A | lqAFALGLMNCN I LVCLAVWMTFSar s ~~~~~l tDKVMVLI LPVAMVSSGf ~~~~~eHCIANMFQVPM Aigiky fapesfwa tgan iaqyadInfnv fivnnLI PVTLGN I VGGG~VFVGMWYWIyl
5132_A | lktFY LAITAGVFSI AFVFI Tattgtg ~~~~~pfg ak lVGGI CFSLGLI LCVVCGA ~~~~~dLFTSTVLI VVAKAsgr i tw ~~~~~gqlaknWLN VYFGN LVGAL~LFVLLMWLSGey
3KCV_A | leAVCLG I LANLMVCLAVWMSY Sgr s ~~~~~l DKAFIMVLPVAMVVASGf ~~~~~eHSIANMFIMPMG i vir dfa spe fwtavgsapen fshltv nfi tdnLI PVTLGN I VGGG~LLVGLTYWYIyl
5132_A | lrAYLAE FISTLLFV FAGVGSAlayak ltsda ~~~~~al dtpglvAI AVCHGFALFVAVAIGani sgg hvNP AVT FGLAVGGq i t ~~~~~v i tGVFYWIAQL LGST~AACFLKYYVTgg
5132_A | leGVVME I IITFALVYTVYATAAdpkkg ~~~~~s lgtIAPLAIGL I VGAN I LAAGp fsggs NPARSFGPAVAAGdf ~~~~~sghWVYVWGP LIGGG~LAGLI YGNVf g
3NE2_A | akRFTA EVVGTFI LVFFGPGA AVit l langadkpnefnigalggdwfAIGMAFALAI AAVIYSLgr i sgh iNP AVTIALVSI G r f ~~~~~greVVPYI VAQF I GAA~LGSLLFLACVgp
3NE2_A | gqAl LTEAIGTFLMLVIMGVAVdera ~~~~~ppgFAGLVIGLTVGGI I ITT I gn i tgs s INPARTFGPYLGDs l gi ~~~~~nlwqyFP IYVIGP I V GAV~AAAWLYNYLak
5DYE_A | lkAVFAE FLATLI FVFFGLGSA lkwpsa ~~~~~lpti IQIALAFGLAIGTLAQA Lgpvs gghi NPAITLALLVGNq i s ~~~~~l rAFFYVAAQLVGA I~AGAGI LYGVApI
5DYE_A | gqAMVVE LIITFQLALCI FASTDsrr t ~~~~~epvgSPALSI GLSVTLGHLVGI yftgcs MNPARSFGPAVVMnr f ~~~~~spaHWVFWGP I V GAV~LAAI LYFLLfp
3C02_A | vrE F IGEFLGT FVLMFLGEGATAnfhttg ~~~~~lsgdwyKLC LGWGLAVFFG I LVSAk i sgh iNLAVS IGLSS I nkfd ~~~~~l k kI PVYFFAQLLGA F~VGTSTVYGLYhg
A 3C02_A | tgAFFNEL I LTG I LLLVI LVVVDen i cg ~~~~~k fhi I KLSVVGLI I LCIGIT Fggn t g fa INPSRDLGSRFLSI I aygkdt ~~~~~f tkdnfyWVWP LVAPCVG SV~VFCQFYDKVICp

```



**D - Most similar pair of FocA vs. Aquaporin protodomains. 5DYE-2 / 4FC4-1**

```

4FC4_A | nnp l gFWVSSAMAGAYVGLGII LIFTI gn lldpsvrp lVMGATFGI ALTLVI IAGs ~~~~~E LFTGHTMFLT LGvkagtishgg wai LPQTWLG NLVG SVFVALLY S~
5DYE_A | ttqqgAMVVELIITFQLALCI FASTDsrr t ~~~~~epvgSPALSI GLSVTLGHLVGI yftgcsMNPARSFGPAVVMnr f ~~~~~spaHWVFWGP I V GAVLAAI LYFy

```

**E - Most distant pair of FocA vs. Aquaporin protodomains. 5DYE-1 / 4FC4-2**

```

5DYE_A | cs vafLKAVFAEFLATLI FVFFG lgsal kwpsalpti lqiALAFGLAIGTLAQA Lgpvs gghi NPAITLALLVGNq i ~~~~~s l rAFFYVAAQLVGA IAGAG I LYgva
4FC4_A | ttapaTVLFFK GAlCNWLVCLAIw airt ~~~~~egtakFLAIWCLLAF IASGy ~~~~~eHSVANMTL FALSwfghsdayt lagighnLLWVT LGNTLSGVVFMGLGYwya

```

**Figure S10: FocA vs Aquaporin A) Structure based multiple sequence alignment of protodomains of FocA (4FC4, 3KLZ, 3KCV) and Aquaporins (5DYE, 5132, 3NE2, 3C02) - RMS relative to the 4FC4-1: 1.88 Å (4FC4-2), 1.18 Å/2.05 Å (3KLZ), 1.54 Å/2.00 Å (3KCV), 2.17 Å/2.82 Å (5132), 2.25 Å/2.43 Å (3NE2), 2.33 Å/2.94 Å (5DYE), 2.21 Å/2.71 Å (3C02) with a common G/AxxxG motif in TM3 B) Computed evolutionary tree (using Jalview/Neighbor-Joining/BLOSUM62) from the alignment C) Proposed parallel evolutionary mechanism (see text) D) 4FC4-1 (FocA's first protodomain) vs. 5DYE-2 (Aquaporin's second protodomain) optimized alignment, with surprising sequence similarity in the middle of a hypothetical common evolutionary tree TM1. The sequence identity is 21% between protodomains (35% in TM3) (see Table S2 for details) E) 4FC4-2 (FocA's second protodomain) vs. 5DYE-2 (Aquaporin's first protodomain) at the extremes of a hypothetical common evolutionary tree in C) -optimized pairwise structural alignment (RMS of 2.16 Å) with a sequence identity is 11% between protodomains (TM3 involving the G/AxxxG/A motif). The color red is used for residues conserved between protodomains within a domain, as well as conserved across domains. Note: The alignment of hundreds of sequences for Aquaporin and FocA families do not unveil more common or different patterns than in our set. We consider our dataset as representative]**

Foca	TM1	TM2 a	TM2 b	TM3	Protodomain TM1-TM2-TM3	TM4	TM5 a	TM5 b	TM6	Protodomain TM4-TM5-TM6
EC	28	38	-	36	34	39	34	39	33	36
IC	40	52	41	43	44	44	42	-	52	46
Full	34	44	41	40	40	42	38	39	43	41

PnuC	TM0	TM1	TM2	TM3	Protodomain TM1-TM2-TM3	TM4	TM5	TM6	TM7	Protodomain TM5-TM6-TM7
EC	28	50	48	34	44	32	46	46	33	42
IC	28	43	58	46	49	32	57	61	31	50
Full	28	46	51	39	46	32	50	51	32	44

Tric	TM1	TM2	TM3	Protodomain TM1-TM2-TM3	TM4	TM5	TM6	TM7	Protodomain TM4-TM5-TM6
EC	38	57	40	45	41	47	36	28	41
IC	41	48	68	52	40	42	44	27	42
Full	40	53	54	49	41	44	39	28	41

MFS	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM8	TM9	TM10	TM11	TM12
EC	24	25	28	35	29	31	25	18	23	21	24	20
IC	22	41	35	34	28	30	22	26	21	16	21	23
Full	23	33	31	35	29	30	24	22	22	18	23	22
	Proto 1	29	Proto 2	31	Proto 3	23	Proto 4	21				

**Figure S11: Sequence Divergence of TMH Families** - Sequence similarity score (see Methods section for details) of the aligned EC half, IC half, and Full TM sequences for each of the TMs **TMH families Aquaporin, Foca, PnuC, Tric, and MFS**. Protodomain 1 and 2 scores are given along with those for EC-facing and IC-facing halves. Higher numbers mean high sequence similarity (or higher conservation), where a maximum score of 100 would mean identical sequences or two sequences with similar residues at each position in the sequence alignment. Protodomains specific to each protein family are discussed in the main text and shown in **Figure 1** and Supplementary **Figures S1** through **S9**. The list of proteins and PDB ids used for each family is provided in the Supplement File SF1. We use here a common 3TMH protodomain decomposition, including for PnuC (in the main text we used 4TMH)

**6CMO\_R**: chimera protein of Soluble cytochrome b562 and Rhodopsin  
**1GZM\_A**: RHODOPSIN

```

          10      20      30      40      50      60
6CMO_R 142 CgtegnfyvvpFSNAggvvrspfeypqyYLAEPWQFSmlaaymflilivlgfpinfltlyv 201
1GZM_A   3 NgtegnfyvvpFSNKggvvrspfeapqyYLAEPWQFSmlaaymflimlglfpinfltlyv 62

          70      80      90      100     110     120
6CMO_R 202 tvqhkklrtplnyILLNLAVADLFMVgGGFTSTLYTSlhgyfvfgptgcnlgGFFATLGG 261
1GZM_A  63 tvqhkklrtplnyILLNLAVADLFMVgGGFTTgTLYTSlhgyfvfgptgcnleGFFATLGG 122

          130     140     150     160     170     180
6CMO_R 262 EIALWgSLVVLAIERYVgVvckpmsnfrfGENHAIMGVAFgTWvmalacaapplagwsryipe 321
1GZM_A 123 EIALWgSLVVLAIERYVgVvckpmsnfrfGENHAIMGVAFgTWvmalacaapplvgwsryipe 182

          190     200     210     220     230     240
6CMO_R 322 glqcscgidyytlkpevnnesfviymfvvhftIPMIgIIFFCYQLVFTVKEgaaaqqgesa 381
1GZM_A 183 gmqcscgidyytphetnnesfviymfvvhfiIPLgLIVgIIFFCYQLVFTVKEgaaaqqgesa 242

          250     260     270     280     290     300
6CMO_R 382 ttqkaekvtrmviiyviaflicwvpyasvafyifthqgsCFgpifmTIPAFFAKgsaaiy 441
1GZM_A 243 ttqkaekvtrmviigmviaflicwlpgpyagvafyifthqgsDFgpifmTIPAFFAKgtsavy 302

          .....
6CMO_R 442 npviyimmN 450
1GZM_A 303 npviyimmN 311

```

**Figure S12** - VAST+ [Invariant substructure](#) alignment between the active (6CMO, human) vs, inactive (1GZM bovine) conformations of Rhodopsin - RMSD = 1.70 Å for 106 residues aligned (86% identity) . 3D visualization link: [iCn3D](#)