

Supplementary file: Table S3. Summary of selected false-negatives sequence hits with respect to problematic domains (Global-mode search)

Domain Name	Type, predicted region of alignment	Validated TM helices/ SP of model, reference	Sequence accession no. (No. of AA)	Sequence Description/ Taxonomy	Raw score/ E-value of FN hits with HMMER2	Raw score/ E-value of FN hits less SP/TM segments with HMMER2
PF08510.4 : PIG-P (phosphatidylinositol N-acetylglucosaminyl transferase subunit P) Gathering score : -11.4 Alignment length: 208 HMM length: 153	TM,1-91	8-24, 44-67 ref.[1]	1. EDL76721.1 (55 AA, fragment)	phosphatidylinositol glycan, class P, isoform CRA_b, <i>Rattus norvegicus</i>	-25.9/3.4e-4	67.9/4.1e-9
PF01569.13 : PAP2 (type 2 phosphatidic acid phosphatase) Gathering score : 8.3 Alignment length: 261 HMM length: 177	TM,200-261	129-143, 156-172 ref.[2]	2. NP_001096030.1 (176 AA, complete)	phosphatidate phosphatase PPAPDC1B isoform 3, <i>Homo sapiens</i>	-1.0/6.7e-3	52.1/2.2e-12
PF04387.6 : PTPLA (protein tyrosine phosphatase-like protein) Gathering score : 25 Alignment length: 177 HMM length: 168	TM,98-177	89-106, 138-155 refs.[3,4]	3. NP_001012396.2 (165 AA, complete) 4. CAI46276.1 (136AA, complete)	protein-tyrosine phosphatase-like member A isoform 2, <i>Mus musculus</i> protein tyrosine phosphatase-like, member a, splice variant PTPLAd5, <i>Canis familiaris</i>	-30.4/1.3e-4 -65.3/5.9e-2	32.6/2.0e-09 28.0/4.5e-09
F00672.17 : HAMP (cytoplasmic helical linker domain) Gathering score : 19.8 Alignment length: 106 HMM length: 79	TM,1-23	1-15 ref.[5]	5. NP_661555.1 (399AA, complete)	sensor histidine kinase, <i>Chlorobium tepidum</i> TLS	9.7/2.2e-1	45.3/2.4e-10
PF00690.18 : Cation_ATPase_N (Cation transporter/ATPase, N-terminus) Gathering score : 18.9 Alignment length: 107 HMM length: 87	TM, 85-105	TM,66-87 ref.[6]	6. XP_001868895.1 (814AA, complete)	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type, <i>Culex quinquefasciatus</i>	7.1/8.1e-4	26.9/7.9e-06
PF01544.10 : CorA (CorA-like Mg ²⁺ transporter protein) Gathering score : -61.3 Alignment length: 550 HMM length: 407	TM, 482-544	TM,341-407 ref.[7]	7. AAO72700.1 (158AA, fragment)	putative Mg transporter, <i>Oryza sativa Japonica Group</i>	-61.8/3.8e-3	17.3/4.8e-07

PF00558.11 : Vpu (Vpu protein) Gathering score : -9.5 Alignment length: 101 HMM length: 96	TM, 12-37	TM,6-28 ref.[8]	8. P08808 (34AA, fragment)	Protein Vpu, <i>Human immunodeficiency virus type 1 (WMJ2 ISOLATE)</i>	-16.0/1.3e-2	36.6/9.8e-08
PF07365.4 : Toxin_8 (Alpha conotoxin precursor) Gathering score : 25 Alignment length: 71 HMM length: 66	SP, 1-21	SP,1-21 ref.[9]	9. AAZ85382.1 (38AA, fragment) 10. Q1L777.1 (38AA, fragment) 11. ABD33869.1 (38AA, fragment) 12. ABD33862.1 (41AA, fragment) 13. ABD48795.1 (45AA, fragment) 14. ABD33857.1 (41AA, fragment) 15. P69658.1 (40AA, fragment) 16. AAZ85381.1 17. ABD33863.1 (38AA, fragment) 18. ABD33860.1 (40AA, fragment) 19. ABD33867.1 (53AA, fragment) 20. ABD33865.1 (41AA, fragment) 21. ABD33870.1 (41AA, fragment) 22. AAZ85380.1 (40AA, fragment)	TeA21P, <i>Conus textile</i> Alpha-conotoxin PeIA, <i>Conus pergrandis</i> alpha conotoxin Mr1.1, <i>Conus marmoreus</i> alpha conotoxin Qc1.5, <i>Conus quercinus</i> conotoxin Lp1.10, <i>Conus leopardus</i> alpha conotoxin Qca-L-1, <i>Conus quercinus</i> Alpha-conotoxin PIA, <i>Conus purpurascens</i> LeD2P, <i>Conus litteratus</i> alpha conotoxin Qc1.6, <i>Conus quercinus</i> alpha conotoxin Qc1.4a, <i>Conus quercinus</i> alpha conotoxin Ac4.3b, <i>Conus achatinus</i> alpha conotoxin Ac4.2, <i>Conus achatinus</i> alpha conotoxin Pu1.2, <i>Conus pulicarius</i> LiC22P, <i>Conus lividus</i>	16.1/1.5e-2 14.0/2.6e-2 12.8/3.6e-2 10.7/6.1e-2 15.0/2.1e-2 8.1/0.12 4.6/0.3 1.5/0.67 0.8/0.79 0.2/0.93 -0.8/1.2 -1.3/1.4 -1.9/1.6 -3.0/2.1	45.0/2.9e-10 42.9/1.2e-09 41.7/2.8e-09 39.6/1.2e-08 37.2/6.5e-08 37.0/7.3e-08 33.5/8.3e-07 30.4/7.1e-06 29.7/1.2e-05 29.1/1.8e-05 28.1/3.5e-05 27.6/5.0e-05 27.0/7.5e-05 25.9/1.6e-4
PF00482.11 : GSPII_F (Bacterial type II secretion system protein F domain) Gathering score : 25 Alignment length: 185 HMM length: 136	TM, 166-184	TM,118-136 ref. [10]	23. ZP_03270751.1 (163AA, fragment)	type II secretion system, subunit F/type IV pilus assembly protein TapC/PilC, <i>Burkholderia sp. H160</i>	0.1/7.6e-1	26.9/8.5e-05
PF04901.5 : RAMP (Receptor activity	TM, 85-106	TM, 85-106 ref. [11]	24. NP_848488.2 (103AA, complete)	receptor activity-modifying protein 1 isoform 2, <i>Mus</i>	-24.9/3.4e-2	5.4/7.3e-05

modifying family)				<i>musculus</i>		
Gathering score : 25 Alignment length: 115 HMM length: 115						

In the first column, we list selected Pfam domains with their accession, identifier, description and their gathering score (as in Pfam release 23) that have TM and/or SP regions included into the model. We also provide alignment length and the HMM length. The latter might be considerably shorter than the former as a result of hmmbuild defaults in HMMER2.

The region in the domain alignment that includes the predicted SP/TM segments (together with interlinking loops as described in Methods) is provided in the second column. We searched for experimental proof of these predictions in the literature and the corresponding references and the positional ranges for the respective SP/TM segments (with respect to the HMM but not the alignment) are given in the third column.

The next two columns provide running number, accession, sequence length, description and taxonomic origin of sequences that were found as false-negative hits of the respective HMMs when using HMMER2 in the global-mode search. The penultimate column provides their scores and E-values based on HMMER2. The last column gives the recomputed scores and E-values of the false-negative hits without their respective SP/TM segments.

Additional material such as hmmpfam outputs and alignments are available at the associated BII WWW site for this work.

References

1. Watanabe R, Murakami Y, Marmor MD, Inoue N, Maeda Y, Hino J, Kangawa K, Julius M, Kinoshita T (2000) Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. EMBO J 19: 4402-4411.
2. Sun L, Gu S, Sun Y, Zheng D, Wu Q, Li X, Dai J, Dai J, Ji C, Xie Y, Mao Y (2005) Cloning and characterization of a novel human phosphatidic acid phosphatase type 2, PAP2d, with two different transcripts PAP2d_v1 and PAP2d_v2. Mol Cell Biochem 272: 91-96.
3. Kihara A, Sakuraba H, Ikeda M, Denpoh A, Igarashi Y (2008) Membrane topology and essential amino acid residues of Phs1, a 3-hydroxyacyl-CoA dehydratase involved in very long-chain fatty acid elongation. J Biol Chem 283: 11199-11209.

4. Uwanogho DA, Hardcastle Z, Balogh P, Mirza G, Thornburg KL, Ragoussis J, Sharpe PT (1999) Molecular cloning, chromosomal mapping, and developmental expression of a novel protein tyrosine phosphatase-like gene. *Genomics* 62: 406-416.
5. Aravind L, Ponting CP (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett* 176: 111-116.
6. Xu C, Rice WJ, He W, Stokes DL (2002) A structural model for the catalytic cycle of Ca(2+)-ATPase. *J Mol Biol* 316: 201-211.
7. Lunin VV, Dobrovetsky E, Khutoreskaya G, Zhang R, Joachimiak A, Doyle DA, Bochkarev A, Maguire ME, Edwards AM, Koth CM (2006) Crystal structure of the CorA Mg²⁺ transporter. *Nature* 440: 833-837.
8. Montal M (2003) Structure-function correlates of Vpu, a membrane protein of HIV-1. *FEBS Lett* 552: 47-53.
9. Chi SW, Kim DH, Olivera BM, McIntosh JM, Han KH (2006) Solution conformation of a neuronal nicotinic acetylcholine receptor antagonist alpha-conotoxin Om1A that discriminates alpha3 vs. alpha6 nAChR subtypes. *Biochem Biophys Res Commun* 345: 248-254.
10. Abendroth J, Mitchell DD, Korotkov KV, Johnson TL, Kreger A, Sandkvist M, Hol WG (2009) The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *J Struct Biol* 166: 303-315.
11. Kusano S, Kukimoto-Niino M, Akasaka R, Toyama M, Terada T, Shirouzu M, Shindo T, Yokoyama S (2008) Crystal structure of the human receptor activity-modifying protein 1 extracellular domain. *Protein Sci* 17: 1907-1914.