Lab 3: KCNT1_CHICK analysis

Zackery Field

February 10, 2014

REMOVE THESE

Part 3: Choose BLAST
Choose search set: Database P.D.B.p (pdb)

Part 1

Pfam MDA prediction

Retrieve the KCNT1_CHICK protein from UniProt in FASTA format. Submit KCNT1_CHICK to Pfam and examine the predicted multi-domain architecture (MDA).

Family	Type	Clan	Start	End	E-value
Pfam-A					
Ion_trans_2	Domain	CL0030	257	325	$1.7E-12$
BK_channel_a	Family	n/a	474	578	$9.9E-36$
Pfam-B					
Pfam-B_4535	n/a	n/a	585	613	0.00016
Pfam-B_9815	n/a	n/a	687	746	6.9e-35
Pfam-B_1301	n/a	n/a	714	1022	1.2e-136
Pfam-B_15008	n/a	n/a	960	1026	1.2e-37
Pfam-B_12084	n/a	n/a	1062	1175	2e-56

Describe the MDA (as a series of Pfam domains); if a significant fraction of a sequence does not have any detectable Pfam domains, you may want to use Pfam-B. Please describe as "unknown fold" any long regions (>50aa) with no detectable Pfam-A domains, or for which Pfam-B domains give no actual functional or structural information. What fraction of the primary sequence is "covered" by informative Pfam domains? (If a Pfam-B domain provides no actual information, you should list that region as "unknown fold".)

Region				
Start	End	Domain	Database	Function
1	256	unknown fold		
257	325	Ion_trans_2	Pfam-A	transmembrane ion channel family
326	473	unknown fold		
474	578	BK_channel_a	Pfam-A	Potassium channel with high conductance
585	613	Pfam-B_4535	Pfam-B	unknown
614	713	unknown fold		
714	1022	Pfam-B_1301	Pfam-B	Automatically generated. Domain found in proteins with ion channel architecture
1062	1175	Pfam-B_12084	Pfam-B	Automatically generated. Domain found in proteins with BK channel and ion channel architectures.

sequence coverage

$$\frac{68[aa] + 104[aa] + 28[aa] + 308[aa] + 113[aa]}{1201[aa]} \approx 52\%$$

The Pfam-B domains were included in the sequence coverage because they reveal some information about the function of the protein. Each of the Pfam-B domains of note had some close relationship with ion channels, but none of them had any specific structural or clan classification. Of the domains that were identified only the Pfam-A domains revealed enough information for further analysis.

Ion_trans_2 (PF07885)

This domain has 158 domain architectures listed under the "Domain organisation" tab in Pfam. Since this domain has much more than a few dozen domains it is most likely a promiscuous domain.

Crystal Structure

There are two structures for this ion channel. One structure for the eukaryotic version of the protein, and one structure for the bacterial version of the protein. The both versions have the same SCOP classification 1bl8 and PDB ID: 1LNQ.

Clan

The clan for this domain contains 8 members. Clan: Ion_channel.

1. Domain GLTSCR1 has the same classifier as Ion trans 2. SCOP: 1bl8. This implies that this domain is related to Ion trans 2 down to and including the family level.

2. Domain Ion_trans also shares the same SCOP classification as Ion_trans_2 SCOP: 1bl8. This implies that the Ion trans domain and the Ion trans 2 domain are related down to and including the family level.

3. Domain IRK is the first domain in the clan to have a different SCOP classification than the original Ion trans 2 domain. SCOP: 1n9p. Surprisingly, SCOP places IRK in an entirely different class of proteins than Ion trans 2. SCOP superfamily and Pfam clan are commonly seen as equivalent, so this domain within the Ion_channel clan is an exception.

4. Domain KdpA has no structural information.

5. Domain Lig_chan is also an exception to the superfamily-clan equivalence rule. SCOP classifies this protein as being in a different class than Ion_trans_2. SCOP: 1gr2.

6. Domain PKD_channel has no structural information.

7. Domain TrkH has PDB entry number: 3PJZ. There is no reference to a SCOP classification and search of the SCOP database did not return any classification.

Taxonomic Distribution

The Pfam interactive species distribution map reveals that this domain is found in all 3 domains of life.

BK_channel_a (PF03493)

Crystal structure

This domain does not have a SCOP classification associated with it. Its PDB entry is 3U6N. The representative structure on the summary page of the Pfam entry contains 7 transmembrane helices as well as two cytoplasmic domains named RCK1 and RCK2. This structure is a voltage gated ion channel with the 4th (zero-labeled) transmembrane helix being the voltage controlled helix.

Clan

This domain also does not have a clan associated with it.

Taxonomic Distribution

According to the Pfam species sunburst diagram, this domain is restricted to the eukaryotes.

Part 2. Transmembrane helix prediction

After searching through exPASY looking for a hit on the accession number in UniProt in the SwissProt database, I downloaded the entire Uni-Swiss FASTA file in search of the SwissProt record for this sequence. There appears to be no SwissProt entry containing

information on the transmembrane helix prediction of this protein. Instead, I will use the transmembrane helix prediction provided on the UniProt site for this protein's entry.

TMHMM

sp|Q8QFV0|KCNT1_CHICK Length: 1201 sp|Q8QFV0|KCNT1_CHICK Number of predicted TMHs: 7 sp|Q8QFV0|KCNT1_CHICK Exp number of AAs in TMHs: 140.17303 sp|Q8QFV0|KCNT1_CHICK Exp number, first 60 AAs: 0 sp|Q8QFV0|KCNT1_CHICK Total prob of N-in: 0.98299

1. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 inside 1 95
2. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 96 115
3. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 outside 116 154
4. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 155 177
5. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 inside 178 185
6. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 186 208
7. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 outside 209 248
8. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 249 271
9. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 inside 272 277
10. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 278 300
11. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 outside 301 309
12. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 310 332
13. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 inside 333 352
14. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 TMhelix 353 375
15. sp|Q8QFV0|KCNT1_CHICK TMHMM2.0 outside 376 1201

Comparison of TMHMM and UniProt Transmembrane helix prediction

TMHMM predicts the existence of 7 transmembrane helices. Uniprot also lists, 7 transmembrane helices. The first 6 regions (3 TMHs) align between Uniprot and TMHMM. There is a short extracellular loop (aa:207-211) followed by another TMH at aa:212-224 in Uniprot, but the TMHMM analysis reveals one large extracellular loop between 209-248. The next two helices (249-271 and 278-300) align well, but they have opposed directionality.

The uniprot prediction at region 278-300 is not labeled as a TMH, but rather is labeled as a pore-forming region, and therefore does not change the handedness (intra vs. extra) of the adjacent loops. The TMHMM prediction then inserts a TMH at region 353-375 where Uniprot does not. Most importantly, this slight mismatching in TM helices causes a major C-terminal discrepancy between the two predictions. Region 375-1200 is labeled as being extracellular according to TMHMM, and it is labeled as being cytoplasmic according to Uniprot.

Part 3. Homologous PDB structure prediction

Description	Qry S-End	Subj S-End	Cover	E	Ident	Accession
Chain A, Structure Of The Human Slo3 Gating Ring >pdb|4HPF|B Chain B, Structure Of The Human Slo3 Gating Ring	350-969	3-578	51%	1e-24	23%	4HPF_A
Chain A, Crystal Structure Of The Human Bk Gating Apparatus	357-975	4-608	52%	1e-19	22%	3MT5_A
Chain A, Structure Of The Intracellular Gating Ring From The Human High-Conductance Ca2+ Gated K+ Channel (Bk Channel)	331-619	36-325	39%	2e-18	27%	3NAF_A
Chain A, Open Structure Of The Bk Channel Gating Ring >pdb|3U6N|B Chain B, Open Structure Of The Bk Channel Gating Ring	351-619	4-279	38%	9e-17	27%	3U6N_A
Chain A, Crystal Structure Of Mthk At 3.3 A >pdb|1LNQ|B Chain B, Crystal Structure Of Mthk At 3.3 A >pdb|1LNQ|C Chain C	231-356	4-121	10%	0.001	25%	1LNQ_A
Chain A, Mthk Channel, Ca2+-Bound >pdb|3RBZ|B Chain B, Mthk Channel, Ca2+-Bound >pdb|3RBZ|C Chain C, Mthk Channel,	231-356	4-121	10%	0.003	25%	3RBZ_A

The BLAST results are consistent with both the SCOP classifications and the Pfam domain classifications in Part 1. BLAST correctly identifies the Ion_trans_2 domain as well as the BK_Channel_a domain. In addition to the SCOP and Pfam domain classifications BLAST also identifies a provisional multi-domain unit labeled PRK10537 which is described as a voltage gated ion channel. This does not provide any new information, but does support the previous functional classification that this is some kind of potassium ion channel.

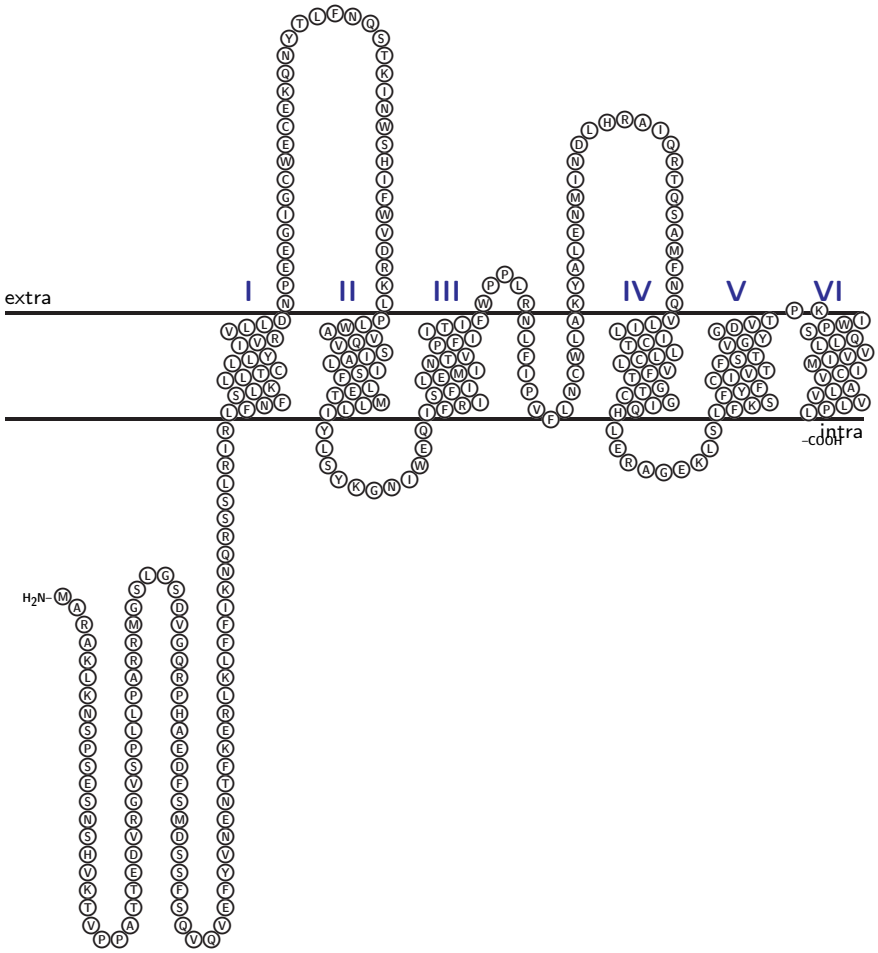

Figure 1: A transmembrane helix prediction from UniProt. The N-terminus is in the left-most cytoplasmic loop. Their are two adjacent large cytoplasmic domains at the C-terminus that are not pictured.

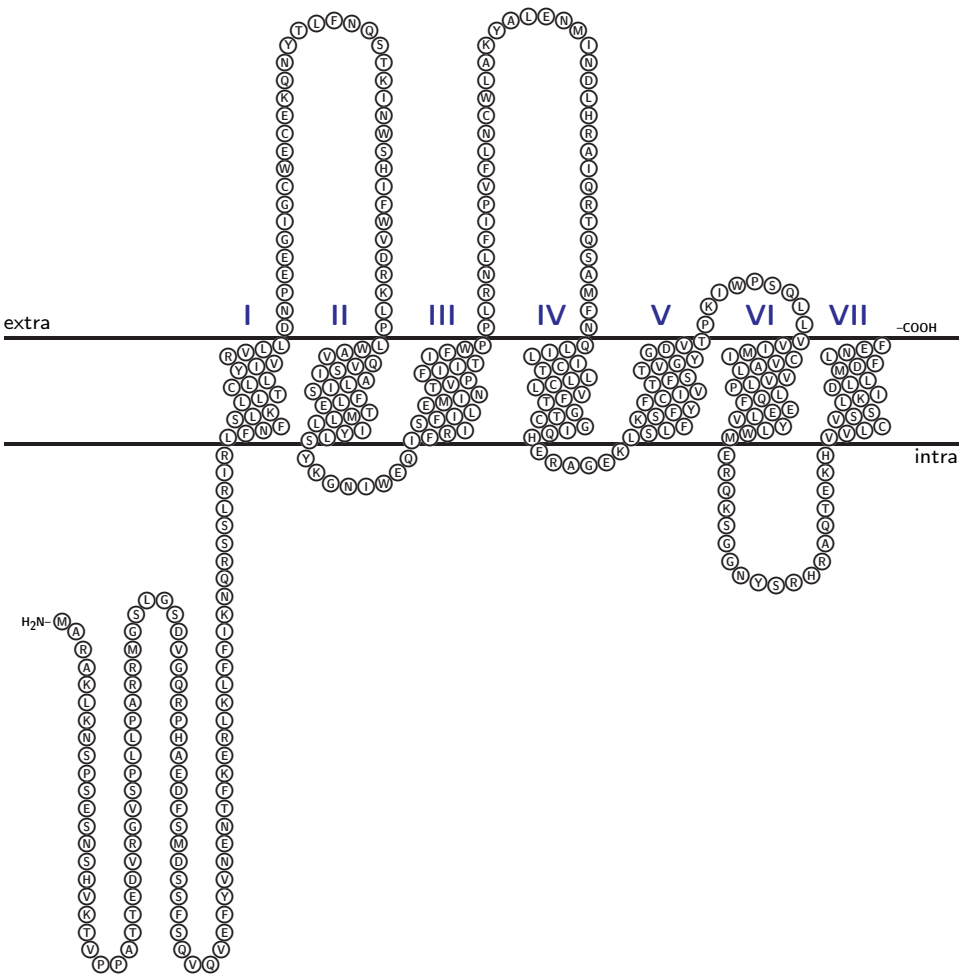

Figure 2: A transmembrane helix prediction from TMHMM. The N-terminus is in the left-most cytoplasmic loop. Their are two adjacent large extracellular domains at the C-terminus that are not pictured.