EXTENDED DATA: for Structural bioinformatics predicts that the Retinitis Pigmentosa-28 protein of unknown function FAM161A is a homologue of the microtubule nucleation factor Tpx2

Supplementary Figure S1

A. Hit 289 to titin in the second iteration of PSI-BLAST into the nr50 database.

XP_032812767.1 titin isoform X2 [Petromyzon marinus] Length=33831. Score = 59.6 bits (142), Expect = 6e-06. Identities = 99/604 (16%), Positives = 198/604 (33%), Gaps = 65/604 (11%)

```
Query  92    NEEYFKKVEELKAAHIETMAKLEKMYQDKLHLKEVQPVVIREDSLSDSSRSVSEKNSYHP  151
            +EE++++  E+K    +  AK E++Y++K+ +++V      E+   ++   V  K
Sbjct  9403  HEEHYEEKVEVKKVPPK--AKREELYEEKVEIRKVPAKPKPEEHYEET---VEIKKVPPK  9457

Query  152   VSLMTSFSEPDLGQSSSLYVSSSEEELPNLE-KEYPRKNRMMTYAKELINNMWTDFCVED  210
                  + E    +          E     +E K+ P K       +E +          +
Sbjct  9458  TKREELYEEKVEIRKVPAKPKPEELYEETVEIKKVPPKTEREELYEEKVEIRKVPAKPKP  9517

Query  211   YIRCKDTGFHAAEKRRKKRKEWVPTITVPEPFQMMIREQKKKEESMKSKSDIEMVHKALK  270
               ++T        + KR+E                  ++K E  K      +  KA +
Sbjct  9518  EEHYEETVEIKKVPPKAKREEHH---------------EEKVEIKK------VPPKAKR  9555

Query  271   KQEEDPEYKKKFRANPVPASVFLPLYHDLVKQKEERRRSLKEKSKEALLASQKPF---KF  327
            +  E+    +K  A P P   +           + RR  L E+  E      KP     +
Sbjct  9556  ELYEEKVEIRKVPAKPQPEEHYEETVEIKKALPKARREELYEEKVEIRKVPAKPKPEEHY  9615

Query  328   IAREEQKRAARAEKQLRDFLKYKKKTNRFKARPIPRSTYGSTTNDKLKEEELYRNLRTQLR  387
               E K+   + +  +  + K +  +  A+P P   Y  T   K    +  R    + +
Sbjct  9616  EETVEIKKVPPKARQEELYEEKVEIRKVPAKPKPEEHYEETVEIKKVPPKARREELYEEK  9675

Query  388   AQEHLQNSSPLPCRSACGC----------RNPRCPEQAVKLKCKHKVRCPTPDFEDL---  434
             +     + P P                 R     E+ V+++       P   +E+
Sbjct  9676  VEIRKVPAKPKPEEHYEETVEIKKVPPKARREELYEEKVEIRKVPAKPKPEEHYEETVEI  9735

Query  435   ----PERYQKHLSEHKSP--KLLTVCKPFDLHASPHASIKREKILADIEADEENLKETRW  488
                 P+  ++ L E K    K+    KP + +       K         E  EE ++  +
Sbjct  9736  KKVPPKARREELYEEKVEIRKVPAKPKPEEHYEETVEIKKVPPKARREELYEEKVEIRKV  9795

Query  489   PYLSPRRKSPVRCAGVNPVPCNCNPPVPTVSSRGREQAVRKSEKERMREYQRELEEREEK  548
            P      +      +  V                +   K + E   E   E+++   K
Sbjct  9796  PAKPKPEEHYEETVEIKKVLPKARREELYEEKVEIRKVPAKPKPEEHYEETVEIKKVPPK  9855

Query  549   LKKRPLLFERVA--QKNARMAAEKHYSNTL---------KALGISDEFVSKKGQSGKVLE  597
              ++  L  E+V   +  A+   E+HY  T+         +   + +E V  +    K
Sbjct  9856  ARREELYEEKVEIRKVPAKPKPEEHYEETVEIKKVLPKARREELYEEKVEIRKVPAKPKP  9915

Query  598   YFNNQETKSVTEDKESFNEEEKIEERENGEENYFIDTNSQDSYKEKDEANE---ESEEEK  654
              + +ET  + +        EE  EE+    +        ++ Y+E  E  +   ++  E+
Sbjct  9916  EEHYEETVEIKKVPPKARREELYEEKVEIRKV-PAKPKPEEHYEETVEIKKVLPKARREE  9974

Query  655   SVEE  658
              EE
Sbjct  9975  LYEE  9978
```

FAM161A (660 aa) aligns with Titin (33,800 aa) at 198 columns in 120 sites separated fairly evenly across 560 residues. Aligned residues in FAM161A are enriched for charged side-chains (DEKR, 52% included), but exclude small, polar amino acids (ACGST, 8% included).

B. Low complexity regions in FAM161A.

MATSHRVAKLVASSLQTPVNPITGARVAQYEREDP**lkalaaaeailedeeeekva**QPAGASADLNTSFSGVDEHAPIS
YEDFVNFPDIHHSNEEYFKKVEELKAAHIETMAKLEKMYQDKLHLKEVQPVVI**redslsdssrsvse**KNSYHPVSLMT
SFSEPDLGQ**ssslyvssseeel**PNLEKEYPRKNRMMTYAKELINNMWTDFCVEDYIRCKDTGFHAAE**krrkkrk**EWVP
TITVPEPFQMMIR**eqkkkeesmksks**DIEMVHKALKKQEEDPEYKKKFRANPVPASVFLPLYHDLVKQKEERRRSLKE
KSKEALLASQKPFKFIAREEQKRAAREKQLRDFLKYKKKTNRFKARPIPRSTYGSTTNDKLKEEELYRNLRTQLRAQE
HLQNSSPLPCRSACGCRNPRCPEQAVKLKCKHKVRCPTPDFEDLPERYQKHLSEHKSPKLLTVCKPFDLHASPHASIK
REKILADIEADEENLKETRWPYLSPRRKSPVRCAG**vnpvpcncnppvp**TVSSRGREQAVRKSEK**ermreyqreleere
eklkkrpllfer**VAQKNARMAAEKHYSNTLKALGISDEFVSKKGQSGKVLEYFNNQETKSVT**edkesfneeekieere
ngeen**YFIDTNSQDSY**kekdeaneeseeeksveesh**

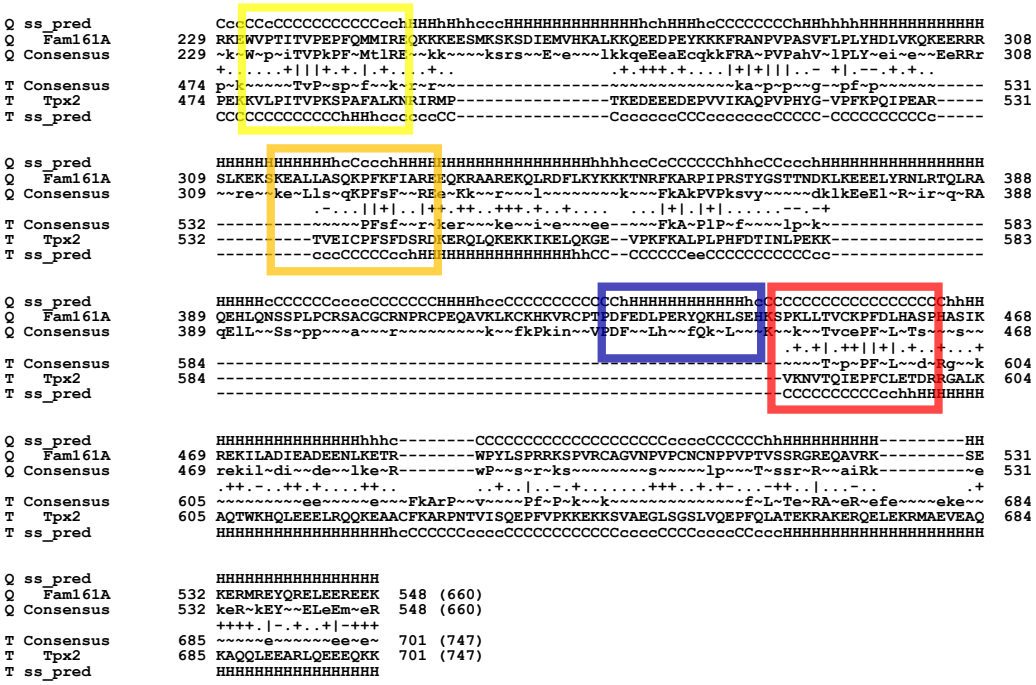
The low complexity filter in PSI-BLAST at NCBI identified nine low complexity blocks (bold, underlined, lower case) containing 147 residues (mainly charged).

Supplementary Figure S2

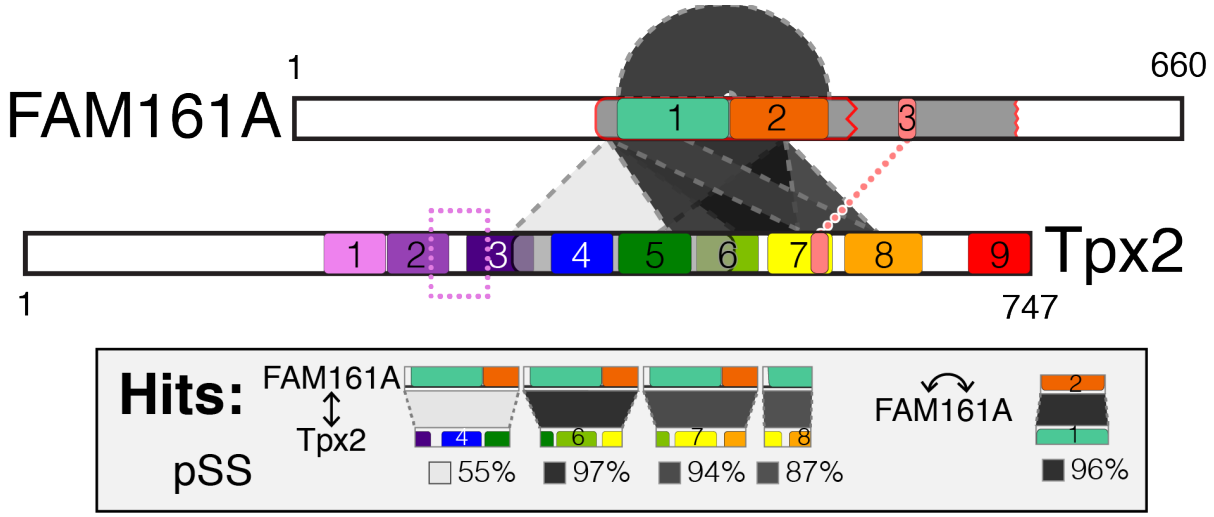
A. HHpred result of symmetrical pairwise alignment of Fam161A and Tpx2

No.	Hit	Prob	E-value	P-value	SS	Cols	Query HMM	Template HMM
1	Tpx2	96.7	1.7E-07	1.7E-07	10.5	211	229-548	474-701
2	Tpx2	93.1	2.5E-06	2.5E-06	5.5	112	228-367	524-637
3	Tpx2	83.9	0.00016	0.00016	11.2	103	233-363	584-721

No 1: Prob=96.67 E-value=1.7e-07 Aligned_cols=211 Identities=20% Similarity=0.288



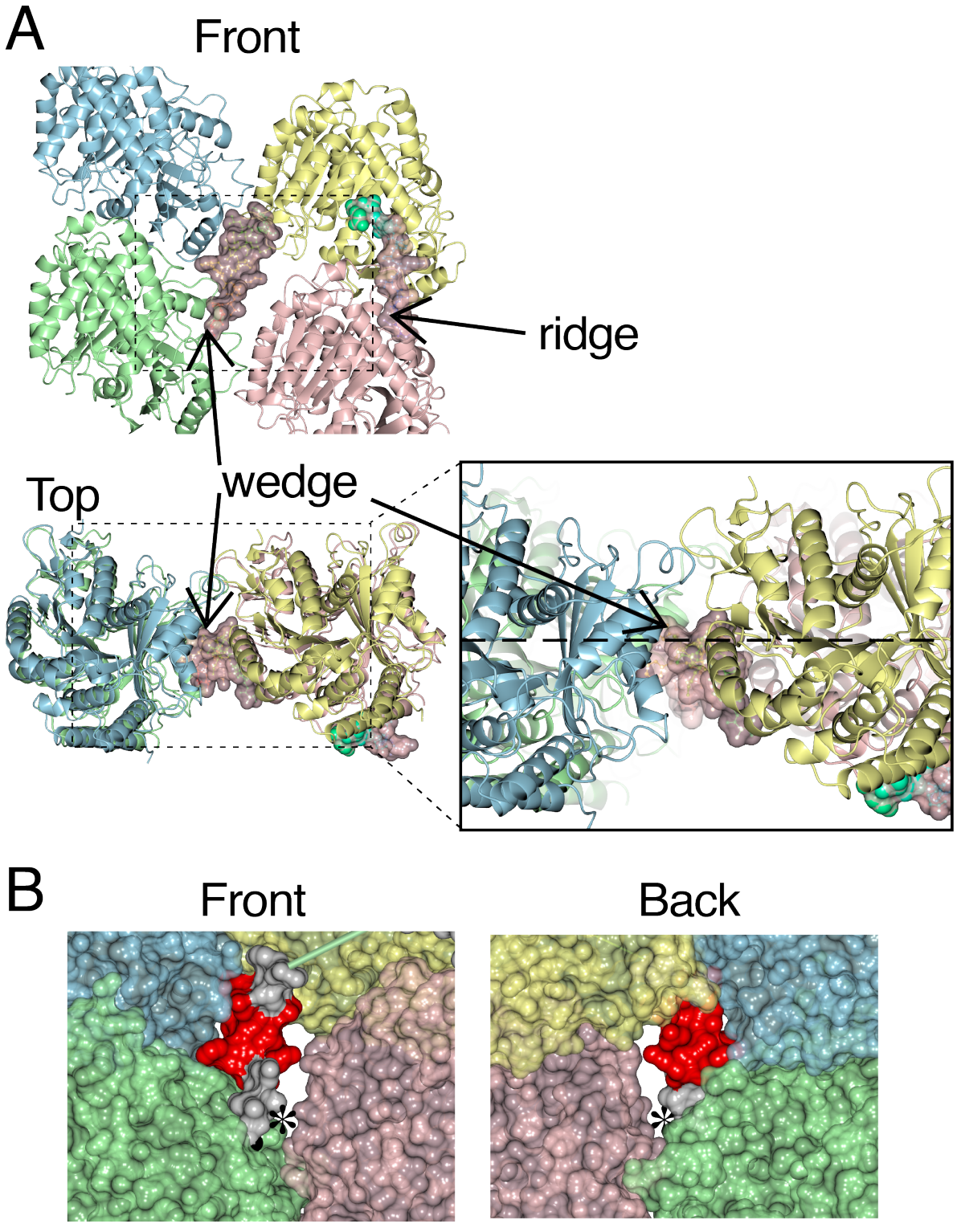
B. HHpred search with alignment realigned with Maximum Accuracy algorithm



A. HHpred output from pairwise profile-profile search of FAM161A and Tpx2, using two converged profiles, showing detail of the top hit. Three PF±Ø motifs in FAM161A are shown in yellow/orange/red boxes. An amphipathic helix in FAM161A residues 431-444 is in a blue box. B. Using the Maximum Accuracy algorithm, regions of FAM161A within residues 233-365 align with four different regions of Tpx2. The four hits are indicated by overlapping shading that varies on a grey scale (see key) with probability of shared structure (pSS) 55–97%. Homology between a partial third repeat in FAM161A (449-461, pink) and the C-terminus of repeat 7 in Tpx2 is indicated by a dotted line. All other aspects of homology are as in Figure 2B.

Supplementary Figure S3

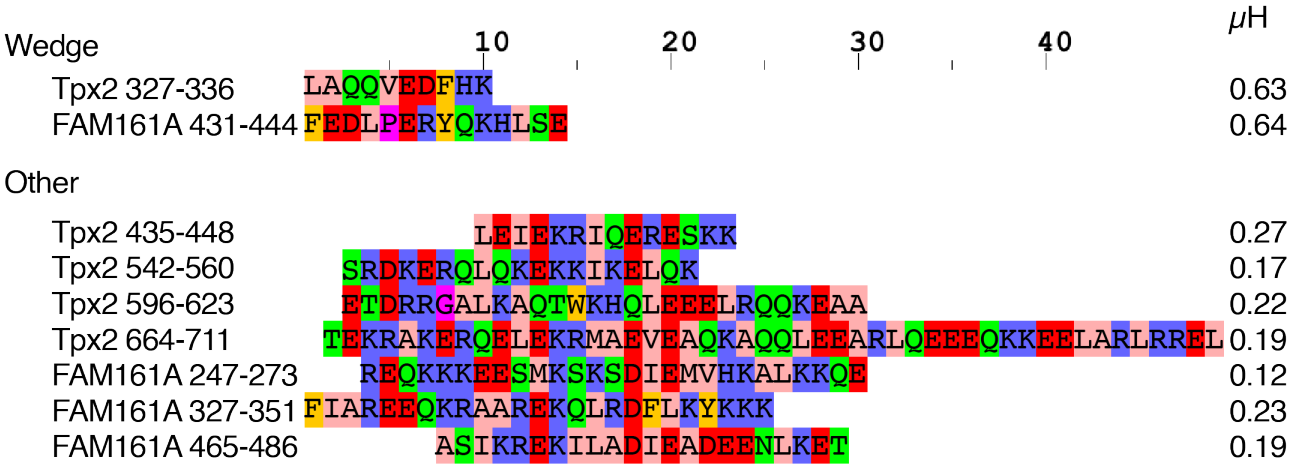
The wedge helix of Tpx2 is buried deeply in the pocket formed by four tubulin monomers



Views of Tpx2 bound to four tubulins (blue, yellow, green and pink), showing: A. the wedge binding into the pocket on the front of the tubulin polymer with such penetration that it partly crosses the mid-line between the front and back of the polymer (dotted line, right hand panel). B: Close up of region in box in part A, showing surface of all chains, with wedge coloured: red for helix and grey for extended loop. Space at its C-terminus is indicated by asterisks.

Supplementary Figure S4

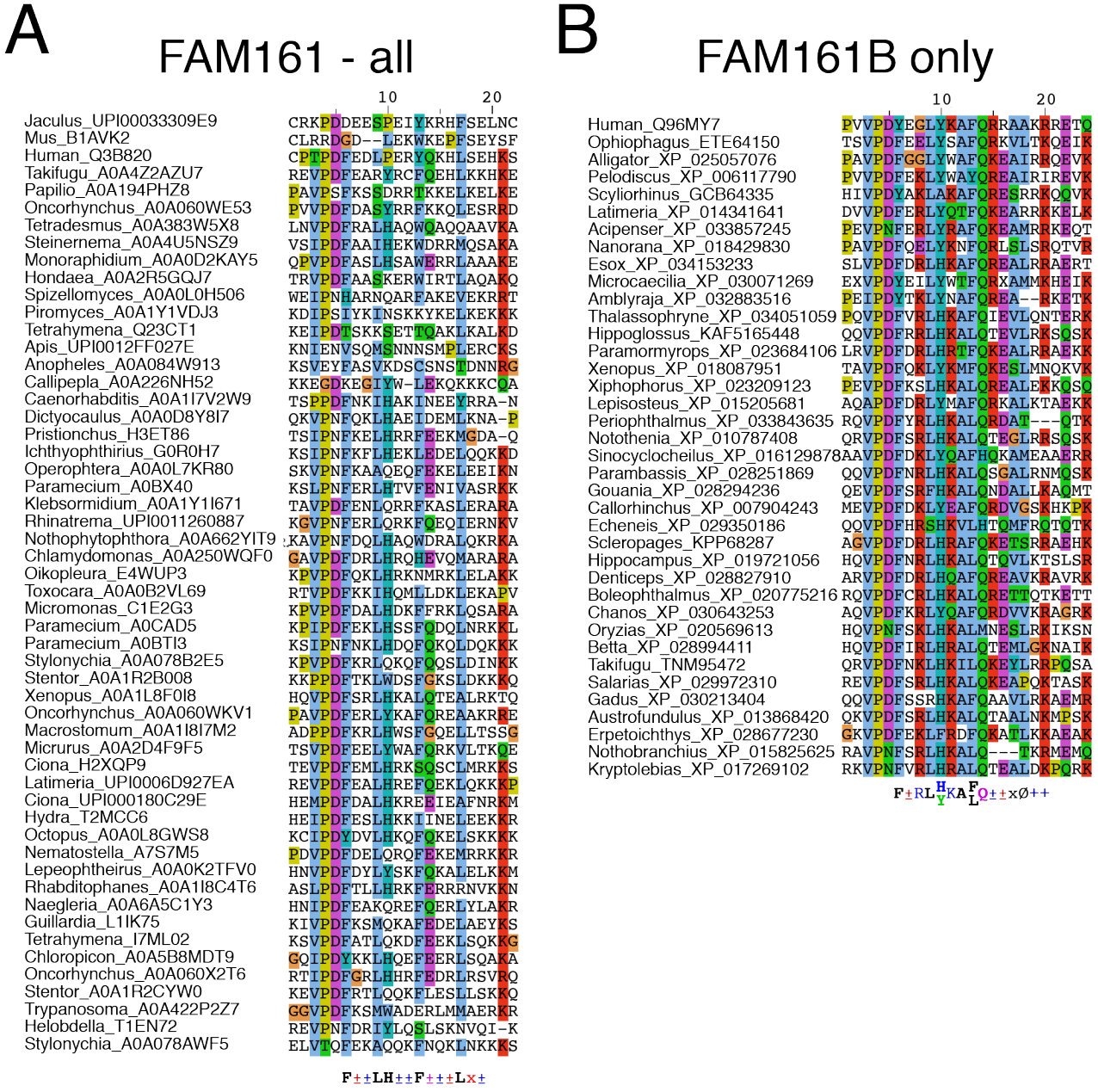
Properties of helices following ridge sequences in Tpx2 and FAM161A



Sequences predicted to be helical were aligned by MAFFT, and coloured to show residue properties (Zappo scheme). Hydrophobic moments (μH), indicating degree of amphipathicity were obtained for each full helix from Heliquest (Gautier *et al.*, 2008).

Supplementary Figure S5

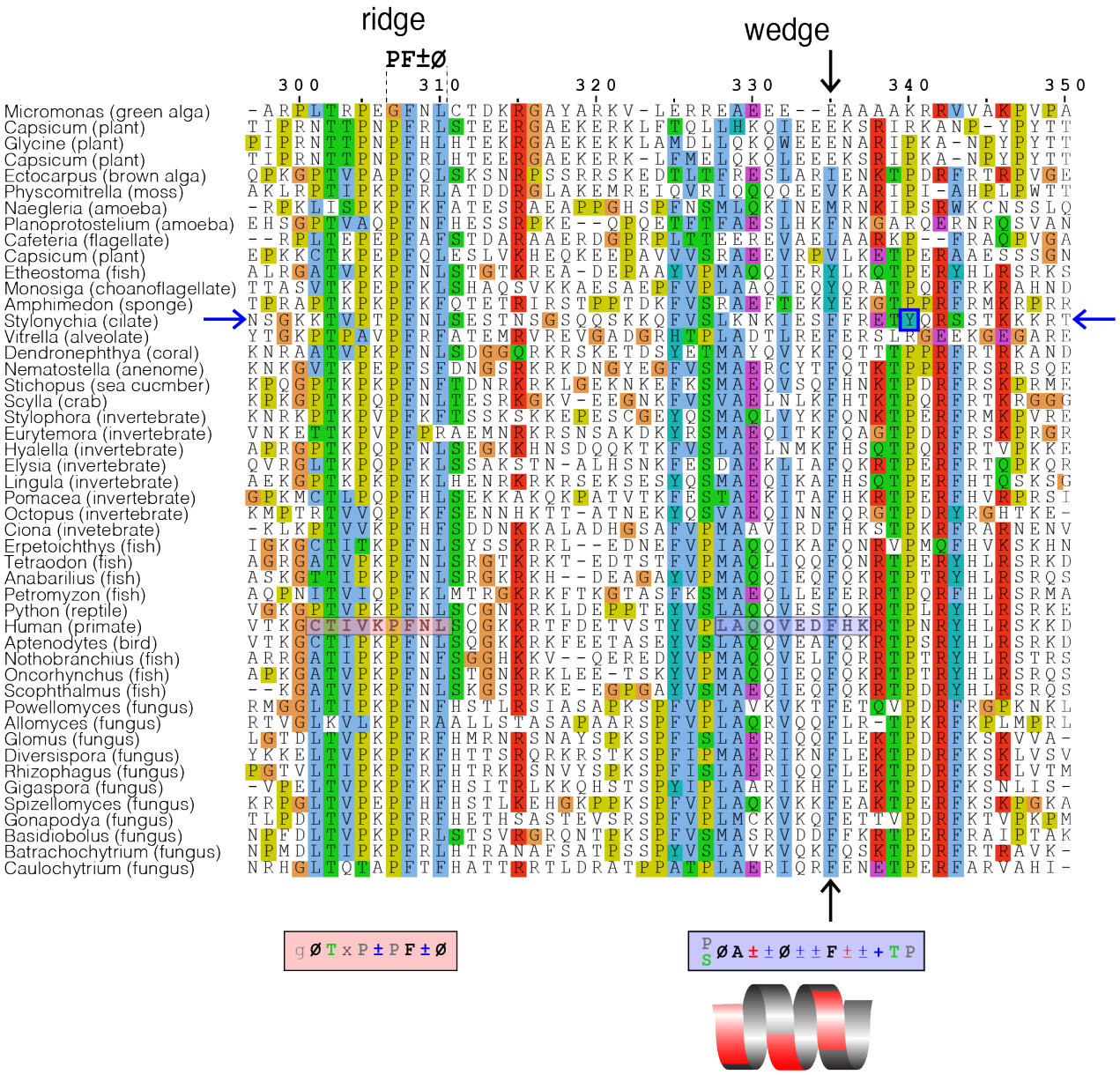
Sequences in amphipathic helices in the FAM161 family



Sequences of predicted amphipathic helices in FAM161 proteins were extracted from multiple sequence alignments, either (A) with all FAM161s (from the 9th iteration of HHblits), or (B) with FAM161B only (from BLAST *i.e.* no iterations and e-value cut-off set to exclude any proteins annotated as FAM161A). Residues are coloured by the Clustal scheme. Consensuses of the predicted helices are summarised at the bottom, and are coloured as in Figure S3B.

Supplementary Figure S6

Variation of Tpx2 ridge and wedge sequences across species.



A multiple sequence alignment of >1200 Tpx2 homologues created in HHblits was clipped down to the region aligning with human 297-375, and reduced to 48 sequences by excluding redundant and partial sequences (with >10 gaps across residues). The alignment was then reduced further to residues 297-350. The pink and blue boxes in the alignment indicate the human ridge and wedge sequences. Motifs below shows consensus sequences across the ridge (10 aa) and wedge (14 aa, 10 of which are predicted to be in a helix), with greater conservation indicated by capitals and bold; colouring: hydrophobic (Ø) – black; A/P/G: grey; K/R – blue; D/E – red; ± mixed KRDEQN, predominant charge either red or blue; S/T–green. The spacing of hydrophobic residues in the wedge (positions 1/5/8) is indicated b pink positions along a helix. Black arrows indicate the conserved F (334 in human) in the Tpx2 wedge. Blue arrows indicate a protist Tpx2 where the wedge lacks a terminal proline (blue box). Alignment coloured with Clustal scheme.