AIGR Deep-Research Report — F6LAX4 (WHEAT, LOC123103357)

Focus: computational_prediction · slug `prediction-animal-specific-terms` **Primary term in focus:** GO:0043005 *neuron projection* (CC) **Question:** Are the ProtNLM2 animal-specific predictions for wheat F6LAX4 appropriate, or are they cross-kingdom / taxon-inappropriate misassignments? Independently determine whether F6LAX4 is the PP2A scaffold (A/PR65) subunit.

Executive Judgment

Verdict: REFUTED (over-annotated) for the animal-context terms; the underlying protein identity in the seed hypothesis is SUPPORTED.

F6LAX4 is, with high confidence, the **wheat orthologue of the protein phosphatase 2A (PP2A) 65 kDa scaffold / structural "A" subunit (PR65 / PPP2R1)**. This is established independently by (i) UniProt naming it "Protein phosphatase 2A structural subunit", (ii) its ARM/HEAT-repeat α-α solenoid domain architecture (InterPro IPR051023 *PP2A_Regulatory_Subunit_A*, PANTHER PTHR10648:SF4, Pfam PF02985/PF22646), and (iii) computed orthology — **85–90 % global identity to the three Arabidopsis PP2A A subunits** (RCN1/A1, A2, A3) and **56–59 % to human PR65 α/β**, the exact conservation signature of this deeply conserved scaffold.

Given this identity, the ProtNLM2 predictions divide into two groups:

- **Taxon-inappropriate (REFUTE):** *neuron projection* (GO:0043005), *neuronal cell body* (GO:0043025) — plants have **no neurons**; *protein antigen binding* (GO:1990405) — plants have **no adaptive immune system / antibodies**. These are text-transfer artifacts from the extensive mammalian PP2A/PR65 neuronal literature. They must not be asserted on a wheat protein.
- **Conserved but indirect / imprecise (QUALIFY, do not assert on this subunit):** *chromosome centromeric region* (GO:0000775) and *chromosome segregation* (GO:0007059) describe a genuine, eukaryote-conserved Shugoshin–PP2A cohesion-protection function, but that role is mediated by the **B56 regulatory subunit** binding Shugoshin — not by the A

scaffold intrinsically — so for the A subunit these are holoenzyme-context, non-core terms with no direct wheat evidence. *Protein heterodimerization activity* (GO:0046982) mis-describes the scaffold, which nucleates an **A–B–C heterotrimer**; the informative MF is protein-phosphatase regulator/PP2A-complex scaffolding, already captured by the existing IBA annotations.

Most important caveat: F6LAX4 is a TrEMBL (unreviewed) entry with no wheat-specific experimental characterization; the identity call rests on strong sequence/domain orthology (computational + database), and the "conserved-but-indirect" mitotic terms cannot be positively excluded as PP2A-complex functions — they are simply not supportable for the *scaffold subunit* on current evidence.

Evidence Matrix

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding	Context
UniProt F6LAX4 (database)	review/ database	Supports	Protein identity = PP2A A subunit	Submission name "Protein phosphatase 2A structural subunit"; 587 aa; existing IBA GO: PP2A complex GO:0000159, phosphatase regulator GO:0019888, cytoplasm/cytosol/nucleus	Triticum aestivum, gene LOC123103
InterPro/Pfam/ PANTHER (database)	structural/ evolutionary	Supports	Domain architecture = ARM/HEAT solenoid scaffold	IPR051023 PP2A_Regulatory_Subunit_A; PTHR10648:SF4; PF02985; PROSITE PS50077 HEAT ×12; Gene3D ARM-like solenoid	Sequence-based
This report — global NW identity (computational)	computational	Supports	F6LAX4 is the plant PP2A A ortholog	85.0 % vs Ath RCN1/A1, 89.8 % vs A2, 88.6 % vs A3; 58.9 %/56.0 % vs human PR65 α/β; human vs Ath baseline 59.0 %	6 UniProt sequences
P 20133745	structural	Supports	PR65 is a HEAT-repeat α-α solenoid PP2A scaffold	"PR65 is the two-layered (alpha-alpha solenoid) HEAT-repeat … scaffold of protein phosphatase PP2A"	Biophysics human PR6
P 38820156	structural/ computational	Supports	PR65 scaffolds the heterotrimeric PP2A	"PR65 is the HEAT repeat scaffold subunit of the heterotrimeric protein phosphatase 2A (PP2A)"	MD + optica tweezers, human PR6
P 37290287	mutant phenotype / review	Supports	Plant PP2A A subunit (RCN1) is a real, growth/ development regulator	RCN1 = "a regulatory A1 subunit isoform of Arabidopsis PP2A"; rcn1 mutants alter UPR/ER-stress sensitivity	Arabidopsi plant
					Arabidopsi

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding	Context
P 26888284 / 28165500	mutant phenotype	Supports (function class)	Plant PP2A-A scaffolds developmental signaling	RCN1/PP2A regulates PIN auxin-transporter polarity & recycling with PINOID antagonism	
P 33973335	interaction/ structural	Qualifies	Centromere/ segregation role is via B56, not A scaffold	"shugoshin proteins are universal protectors of centromeric cohesin"; "a highly conserved pocket on the B56 regulatory subunit is required for hSgo1 binding"	Human somatic cel
Botanical fact (plants lack neurons & adaptive immunity)	review/first-principles	Refutes	Animal-context terms invalid in wheat	No neurons → GO:0043005/0043025 impossible; no antibodies → GO:1990405 impossible	Triticum aestivum
This report — AlphaFold + Kabsch superposition (computational)	structural/ computational	Supports	F6LAX4 3D fold = PR65 HEAT solenoid	F6LAX4 AF pLDDT=90.4; local 60-res-window Kabsch CA-RMSD to Ath RCN1 median **0.29 Å**, to human PR65 median **0.92 Å**; global RMSD (2.8/4.5 Å) inflated only by solenoid hinge-bending	AlphaFold v6 models
UniProt F6LAX4 GO set (database)	review/ database	Qualifies	Are ProtNLM terms already curated?	Curated GO = only 5 IBA terms; intersection with 6 ProtNLM terms = ∅ → action is "reject prediction," not "remove annotation"	Triticum aestivum
This report — PPP catalytic-motif scan (computational)	computational	Supports	F6LAX4 is A scaffold, not catalytic C subunit	F6LAX4 lacks all PPP metal-binding motifs (GDxHG, GDxVDRG, GNHE, RGNHE); positive control PPP2CA (P67775) has all four	Sequence-based

Structural Confirmation (added Iteration 2)

AlphaFold DB models were retrieved for F6LAX4 (global pLDDT 90.4, an elongated, confidently modeled HEAT-repeat solenoid), Arabidopsis RCN1 (Q38845, pLDDT 93.0) and human PR65α (P30153, pLDDT 94.9). A Kabsch superposition of Cα atoms over identical-residue anchors gives an elevated *global* RMSD (F6LAX4↔RCN1 2.82 Å; F6LAX4↔human PR65 4.46 Å), but a sliding 60-residue **local** window RMSD is near-zero (median **0.29 Å** vs RCN1, **0.92 Å** vs human PR65). This pattern — individually superimposable repeats but hinge-bent ends — is the structural signature of the elastic PR65 α-α solenoid (P 20133745) and independently confirms the scaffold identity. Provenance: `/tmp/aigr/AF-*.pdb`, superposition code in the executed analysis log.

Family Disambiguation & Annotation-Status Check (added Iteration 3)

Two curator-relevant confirmations:

1. **The ProtNLM2 terms are NOT curated GO annotations.** UniProt's curated GO set for F6LAX4 is exactly five IBA/GO_Central terms — cytoplasm (GO:0005737), cytosol (GO:0005829), nucleus (GO:0005634), PP2A complex (GO:0000159), protein phosphatase regulator activity (GO:0019888). The intersection with the six ProtNLM2 predictions is **empty**. The correct curation action is therefore **"reject / do not import the prediction,"** not "remove an existing annotation."
2. **F6LAX4 is the non-catalytic A scaffold, not the catalytic C subunit.** A motif scan finds **none** of the PPP-phosphatase catalytic metal-binding signatures (GDxHG, GDxVDRG, GNHE, RGNHE) in F6LAX4 (587 aa), whereas the genuine PP2A catalytic subunit human PPP2CA (P67775, 309 aa) contains **all four**. This excludes the main competing family assignment (catalytic subunit) and is consistent with the scaffold role.

GO Curation Implications (leads — require curator verification)

Term	Aspect	ProtNLM2 predicted	Recommended curator action
GO:0043005 neuron projection	CC	yes	**Do not annotate / reject.** Taxon-inappropriate (no neurons in plants). Flag as ProtNLM cross-kingdom error.
GO:0043025 neuronal cell body	CC	yes	**Do not annotate / reject.** Same rationale.
GO:1990405 protein antigen binding	MF	yes	**Do not annotate / reject.** No adaptive immunity/antibodies in plants.
GO:0000775 chromosome centromeric region	CC	yes	**Do not assert on this subunit.** Conserved Sgo–PP2A function is B56-mediated; non-core for the A scaffold; no wheat evidence.
GO:0007059 chromosome segregation	BP	yes	**Do not assert on this subunit.** Holoenzyme/B56-context; indirect for the scaffold.
GO:0046982 protein heterodimerization activity	MF	yes	**Replace with more informative scaffold term** (e.g., protein phosphatase 2A binding / structural constituent of PP2A holoenzyme). Heterodimerization mis-describes an A–B–C heterotrimer scaffold.
GO:0000159 protein phosphatase type 2A complex	CC	(existing IBA)	**Retain.** Correct and plant-appropriate.
GO:0019888 protein phosphatase regulator activity	MF	(existing IBA)	**Retain.** Correct core MF for the A scaffold.
GO:0005737/0005829/0005634 cyto/cytosol/nucleus	CC	(existing IBA)	**Retain.** Consistent with plant PP2A.

Net: the existing IBA annotations already capture the correct core function; the ProtNLM2 additions add nothing valid and three are actively wrong.

Mechanistic Scope

Immediate molecular function of F6LAX4: it is the **structural (A) scaffold subunit of the PP2A holoenzyme** — a rigid-yet-elastic α-α solenoid of ~15 HEAT repeats that simultaneously binds the catalytic (C) subunit and a variable regulatory (B/B56/B''/B''') subunit, thereby determining substrate specificity of a Ser/Thr phosphatase. It has **no catalytic activity itself** and no intrinsic DNA-, antigen-, or neuron-related activity. Downstream/holoenzyme-context phenomena (auxin-transport polarity, ethylene signaling, ER-stress/UPR in plants; Tau dephosphorylation, cohesion protection in animals) are properties of *specific PP2A holoenzymes*, not of the isolated scaffold, and must be separated from the scaffold's direct GO annotation.

Conflicts and Alternatives

- **Paralog/ortholog confusion (the actual error source):** ProtNLM2 transferred descriptors from mammalian PR65/PP2A entries whose partner holoenzymes act in neurons and immune cells. The wheat sequence is a bona fide ortholog (56–59 % to human), so a naive language model reuses mammalian tissue/organelle vocabulary.
- **Subunit confusion:** the centromere/segregation terms belong to **PP2A-B56 + Shugoshin**, i.e., the *regulatory* subunit, not the A scaffold. Assigning them to the A subunit conflates holoenzyme function with scaffold function.
- **No competing identity hypothesis survives:** 85–90 % identity to three independent Arabidopsis A subunits rules out any non-PP2A-A assignment. The alternative that this is a generic ARM/HEAT protein (e.g., importin, karyopherin) is excluded by the PP2A-A-specific InterPro/PANTHER family and by orthology.

Knowledge Gaps

1. **No wheat experimental data.** Checked: UniProt (TrEMBL, PE=3 "inferred from homology"), literature. Matters because all functional claims are homology-based. Resolved by: wheat PP2A pulldown/complex-MS, or a wheat rcn1-like mutant.
2. **Which wheat PP2A A homoeolog/paralog is this?** Hexaploid wheat has A/B/D homoeologs and multiple A-subunit genes; LOC123103357 is one copy. Checked: single UniProt entry only. Matters for gene-level annotation propagation. Resolved by: wheat genome synteny / homoeolog mapping.
3. **Does wheat PP2A participate in Sgo-mediated cohesion (GO:0000775/0007059)?** Checked: no wheat-specific paper found; conserved in animals/fungi and reported in plants via SGO1. Matters only if a curator considers those terms. Resolved by: wheat/cereal meiosis cohesion studies (SGO1–PP2A-B56).

Discriminating Tests

- **Orthology/phylogeny (done, decisive):** global identity to Arabidopsis A subunits (85–90 %) vs catalytic subunits (would be <25 %) confirms scaffold, not catalytic, identity.
- **AlphaFold geometry:** predict/inspect the AlphaFold model of F6LAX4 for the characteristic hooked HEAT-repeat solenoid and superpose on human PR65 (PDB 1B3U) — expected low RMSD over the solenoid.
- **Holoenzyme co-IP MS in wheat:** recovery of PP2A C and B/B56 subunits (not neuronal or immune proteins) would positively confirm scaffold function and simultaneously refute the animal-context terms.
- **Motif check for B56/Sgo pocket:** the Sgo-binding pocket is on B56, absent from the A subunit — confirming the centromere terms do not map to F6LAX4.

Curation Leads (require curator verification)

- **Reject** GO:0043005, GO:0043025, GO:1990405 on F6LAX4 as ProtNLM2 cross-kingdom errors (organism = plant; no neurons, no adaptive immunity).

- **Do not propagate** GO:0000775 / GO:0007059 to the A scaffold subunit; if any PP2A-complex mitotic annotation is desired, it belongs to the B56 subunit with a wheat/plant reference.
- **Replace** GO:0046982 with an informative scaffold MF (protein phosphatase 2A binding / structural constituent of PP2A holoenzyme) rather than generic heterodimerization.
- **Retain** existing IBA terms GO:0000159, GO:0019888, GO:0005737/0005829/0005634.
- **Candidate references & exact snippets to verify:**
- P 20133745
  — "PR65 is the two-layered (alpha-alpha solenoid) HEAT-repeat ... scaffold of protein phosphatase PP2A."
- P 38820156
  — "PR65 is the HEAT repeat scaffold subunit of the heterotrimeric protein phosphatase 2A (PP2A)."
- P 37290287
  — RCN1 "a regulatory A1 subunit isoform of Arabidopsis PP2A" (plant ortholog exists; function class = growth/development/signaling).
- P 33973335
  — "a highly conserved pocket on the B56 regulatory subunit is required for hSgo1 binding and cohesion protection" (centromere terms map to B56, not A).
- **Suggested question for curators:** should AIGR add a taxon-constraint QC rule flagging animal-only CC/MF terms (neuron*, antibody/antigen) predicted on Viridiplantae proteins?
- **Suggested experiment:** wheat PP2A affinity-MS to positively identify holoenzyme partners and close the "no wheat data" gap.

Provenance / Artifacts

- Orthology identity matrix (global Needleman–Wunsch, computed this run): `/tmp/aigr/orthology_identity.csv`
- Per-term verdict table: `/tmp/aigr/protnlm_term_verdicts.csv`
- Sequences retrieved from UniProt REST: F6LAX4, P30153, P30154, Q38845, Q38950, Q38951.

Computational note: identities were computed with a simple match/mismatch (+1/−1, gap −2) global alignment rather than a BLOSUM-scored aligner; the plant-vs-plant (85–90 %) versus cross-kingdom (56–59 %) separation is large and robust to scoring scheme, so the ortholog assignment is not sensitive to alignment parameters.
