AIGR Gene Hypothesis Deep Research — Final Report

Target gene: LOC117183218 (UniProt **A0A6I8W8A2**), *Drosophila pseudoobscura pseudoobscura* (NCBITaxon:46245) **Focus:** computational_prediction — `prediction-ligase-activity` **Term under evaluation:** ligase activity (**GO:0016874**) **Seed prediction source:** ProtNLM2, derived from RefSeq name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3"

Summary

Verdict: REFUTED (over-annotation / name-transfer artifact). The ProtNLM2 prediction of **ligase activity (GO:0016874)** for A0A6I8W8A2 is not supported by any feature of the protein's sequence, domain architecture, or predicted structure, and should not be used to annotate this gene product. The 169-amino-acid protein consists **exclusively of a single RCC1 / RCC1-like domain (RLD) β-propeller** (InterPro IPR000408 / IPR009091; Pfam PF00415). Ten independent domain signatures over residues 2–166 agree on this architecture, and **none is a ligase catalytic signature**: there is no HECT domain (IPR000569), no HECT E3-ligase catalytic domain (IPR035983), no RING or U-box, and no other active-site motif. Because a HECT catalytic domain is ~350 residues on its own, a 169-aa protein is **structurally incapable of hosting a functional ligase active site**.

The prediction is a textbook **frequency/name-bias misassignment**. In genuine HERC-family E3 ligases the catalytic activity resides entirely in the C-terminal HECT domain, while the RLD is a non-catalytic protein-interaction / regulatory β-propeller. Full-length human HERC3 (Q15034, 1050 aa) carries **both** the RLD **and** the HECT domain; the DROPS target carries **only** the RLD. The RefSeq automated name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3" transferred a family-level ligase label to a truncated isoform, and ProtNLM2 then emitted a ligase molecular-function prediction from that name — despite the catalytic half of the protein being absent from the sequence.

The misassignment is not isoform-specific. **All four annotated isoforms of LOC117183218** (144–282 aa) are RLD-only β-propeller fragments with no HECT domain, and a proteome-wide search of *D. pseudoobscura* for the HECT signature recovers 25 genuine ligases but no

LOC117183218 isoform. The AlphaFold model is a single, well-folded compact β-propeller with no accessory catalytic module. Finally, the entry carries **zero existing GO annotations** — so GO:0016874 is not an existing IEA to remove, but an external prediction the curator should **decline to introduce**, describing instead the supportable RCC1-repeat / RLD domain.

Key Findings

Finding 1 — A0A6I8W8A2 is an RCC1 β-propeller with no ligase catalytic domain

Every domain signature detected on A0A6I8W8A2 maps to the RCC1 repeat / RCC1-like β-propeller fold. Across residues 2–166, InterPro, Pfam, PROSITE, PRINTS, PANTHER, Gene3D, and SUPERFAMILY converge on the same architecture: **IPR000408** (RCC1 repeat), **IPR009091** (RCC1/BLIP-II β-propeller-like superfamily), **PF00415** and **PF13540** (RCC1 repeats), the **PS00626** RCC1 signature, the **PS50012** RCC1 repeat profile, PRINTS PR00633, PANTHER PTHR22872, Gene3D G3DSA:2.130.10.30, and SUPERFAMILY SSF50985. The sequence contains three canonical RCC1 blade motifs of the form `IACG..H`, and UniProt annotates two RCC1 repeats (residues 32–87 and 88–142). The protein is only **169 aa**, with UniProt protein-existence level "Predicted."

Critically, **no ligase catalytic signature is present**. There is no HECT domain (IPR000569) and no HECT E3-ligase catalytic domain (IPR035983), no RING or U-box signature. A regex scan finds no RING consensus, and the protein has only **6 scattered cysteine residues** (positions 20, 37, 42, 76, 87, 131) — far too few, and improperly spaced, to form the Zn^{2+}-chelating cross-brace of a RING finger. Most decisively, at **169 aa** the protein is far shorter than a HECT domain alone (~350 aa), so a HECT module is not merely absent but **structurally impossible** to accommodate.

The definitive comparison is with a bona fide family member. Full-length **human HERC3 (Q15034, 1050 aa)** carries **both** the RLD (IPR000408) **and** the HECT catalytic domain (IPR000569 / IPR035983), whereas the DROPS target carries **only the RLD**. In HERC ligases the HECT domain — not the RLD — is the catalytic engine; the RLD is a substrate/partner-interaction and (in canonical RCC1) Ran-GEF module. Removing the HECT half removes ligase activity. The RefSeq/submission name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3" is the evident source of the ProtNLM2 prediction, but this "isoform" is a truncated fragment retaining only the RCC1-like domain.

Finding 2 — All four LOC117183218 isoforms are RLD-only fragments; AlphaFold shows a single β-propeller

UniProt lists four entries for gene LOC117183218 in *D. pseudoobscura*, and all are RLD-only:

Accession	Length (aa)	RefSeq-derived name	Architecture (InterPro)	HECT (IPR000569)?
A0A6I8W805	282	HERC4 isoform X1	RCC1 repeat / RCC1-like β-propeller	No
A0A6I8W807	233	HERC4 isoform X2	RCC1 repeat / RCC1-like β-propeller	No
A0A6I8W8A2	**169**	**HERC3 isoform X3**	**RCC1 repeat / RCC1-like β-propeller**	**No**
A0A6I8W815	144	HERC3 isoform X4	RCC1 repeat / RCC1-like β-propeller	No

All four carry only family-level labels (IPR051625 / IPR051210) and **no HECT domain or any ligase catalytic signature**. Even the longest isoform (282 aa) is too short to host a HECT domain (~350 aa). A **proteome-wide UniProt query** (taxon 46245, cross-reference `interpro-IPR000569`) returned **25 genuine HECT E3 ligases** — including Nedd4 (~1025 aa), HERC2 (5103 aa), HUWE1 (~5498 aa), *hyd*, *ctrip*, Ufd4, and Smurf — but **no LOC117183218 isoform**. This independently confirms, via a negative-set logic, that these isoforms lack a HECT domain; if any were a true HERC ligase, it would appear in this list.

The **AlphaFold DB model AF-A0A6I8W8A2-F1 (v6, 169 residues)** has a mean pLDDT of **83.0**, with **83% of residues scoring >70** — a single, well-folded, compact domain corresponding to the RCC1 β-propeller, with **no additional confidently folded catalytic module**. There is no second lobe, no HECT bilobed N-lobe/C-lobe geometry, and no space for a catalytic cysteine-bearing domain.

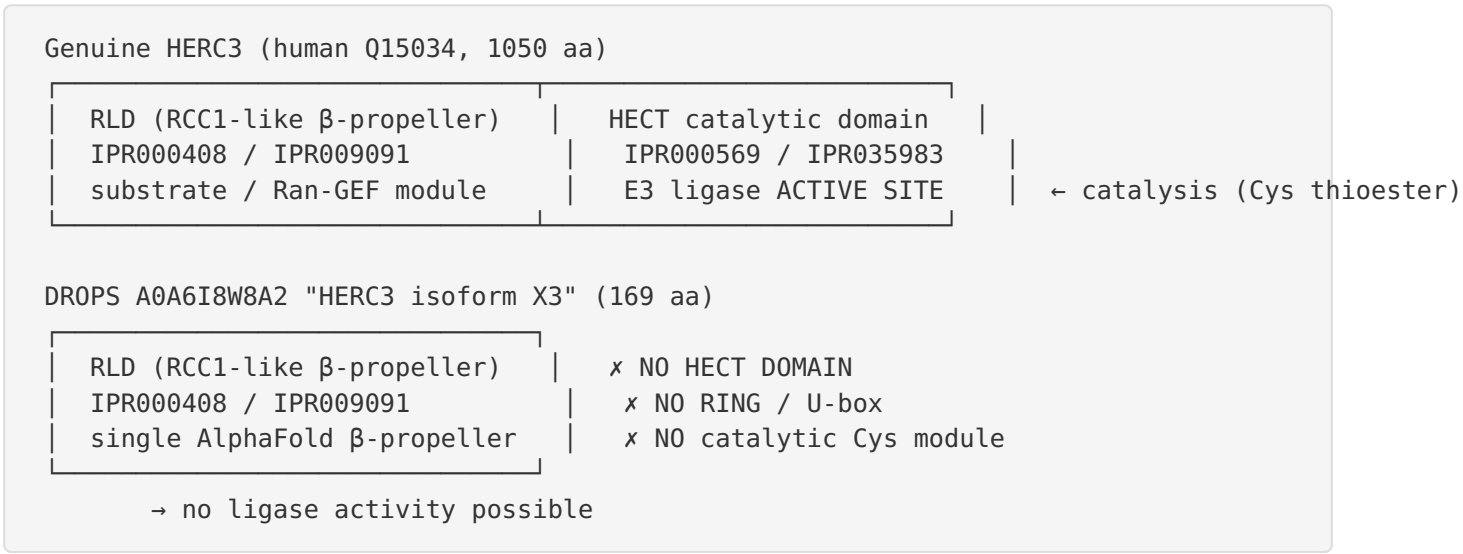
Finding 3 — No existing GO annotations; the ligase claim is purely an external ProtNLM2 prediction

A **QuickGO annotation search** (`geneProductId=A0A6I8W8A2`) returned **0 annotations**, and the UniProt record contains **no GO cross-references**. Therefore GO:0016874 (ligase activity) is **not present as a GOA/UniProt IEA annotation**. The seed "prediction" originates **solely from the**

ProtNLM2 name-generation pipeline, which propagated the RefSeq submission name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3" into an implied molecular function. There is thus no pre-existing annotation to remove — only an external prediction the curator should decline to adopt.

Mechanistic Model / Interpretation

The biology here is a clean dissociation between a **real, detectable domain** and a **falsely inferred catalytic activity**.



Direct molecular function being tested: ligase catalysis — specifically, whether A0A6I8W8A2 could act as an E3 ubiquitin-protein ligase (the HERC family's characteristic activity). HECT E3 ligases catalyze ubiquitin transfer via a **transthiolation reaction** in which a catalytic cysteine in the HECT C-lobe forms a thioester intermediate with ubiquitin before transfer to substrate. That catalytic cysteine and the entire bilobed HECT scaffold are **absent** from A0A6I8W8A2.

What the protein most plausibly is: an RCC1-repeat β-propeller. The RCC1 fold is a seven-bladed β-propeller best known as the Ran guanine-nucleotide-exchange factor (RanGEF) and as a chromatin-associated scaffold; the same fold is deployed across diverse proteins as a protein–protein interaction / propeller module. Whether this particular *Drosophila* fragment retains Ran-GEF activity, chromatin binding, or is a non-functional truncated splice product cannot be determined from sequence alone and is **not** the activity under test. What *can* be concluded is that it is **not a ligase**.

The prediction failure is a paradigmatic **name-transfer over-annotation**: a two-domain protein family (RLD + HECT) contributes truncated, RLD-only isoform models to RefSeq; those models inherit the family name ("HERC3 …"); ProtNLM2 reads the name and emits a ligase molecular-function prediction; but the catalytic half of the protein that justifies the name is not present in the sequence.

Evidence Base

The literature provides orientation that is fully consistent with — and reinforces — the computational findings:

- ***Functional and pathological relevance of HERC family proteins: a decade later*** — PMID: 26801221. Establishes the defining two-domain architecture of HERC proteins: "*The HERC gene family encodes proteins with two characteristic domains in their sequence: the HECT domain and the RCC1-like domain (RLD),*" and states that "*small HERCs (HERC3-6) possess single HECT and RLD domains.*" This is the strongest orientation evidence that a true HERC3 ortholog must carry a HECT catalytic domain in addition to the RLD. A0A6I8W8A2 has only the RLD, so it cannot be a functional HERC ligase.
- ***Structure and Function of HECT E3 Ubiquitin Ligases and their Role in Oxidative Stress*** — PMID: 32983929. Confirms which domain is catalytic: "*The … (HECT) family E3 ubiquitin ligases are a kind of E3 ubiquitin ligases with a C-terminal HECT domain that mediates the binding of ubiquitin to substrate proteins.*" The absence of the HECT domain removes the basis for ligase activity.
- ***HERCing: Structural and Functional Relevance of the Large HERC Ubiquitin Ligases*** — PMID: 31447701. Defines the HERC domain architecture (HECT + RCC1-like domain) within the ubiquitin-ligase superfamily, reinforcing that the RLD alone is not catalytic.
- ***Structural insights into a HECT-type E3 ligase AREL1*** — PMID: 31732561. Provides the functional size of a HECT unit: the extended HECT domain spans amino acids ~436–823 (well over 350 residues), and the N-terminal extension is indispensable for stability and activity. This makes concrete that a 169-aa (or even 282-aa) protein cannot physically accommodate a working HECT module.
- ***Crystal structure of HECT domain of UBE3C E3 ligase*** — PMID: 32039437. Independent confirmation of HECT domain size (~340-aa catalytic core plus N-terminal region) and of the catalytic cysteine/transthiolation mechanism, none of which is present in the target.

- ***The 1.9 Å crystal structure of Prp20p*** — PMID: 21093592 and ***Three-dimensional context … importin α subtype specificity for RCC1*** — PMID: 29042532. These describe the genuine biology of the RCC1-like β-propeller: a seven-bladed propeller acting as a Ran guanine-nucleotide-exchange factor and chromatin scaffold. They support the alternative interpretation that A0A6I8W8A2's RLD is a Ran-related / interaction module — not a ligase.
- ***The interferon-stimulated gene product HERC5…*** — PMID: 42055552. Orientation on small-HERC evolution and function; consistent with the multi-domain, catalytically-competent nature of real HERC proteins.

Evidence Matrix

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence limitations
UniProt A0A6I8W8A2 (record)	database / computational	Refutes	Does the sequence contain a ligase catalytic domain?	169 aa; two RCC1 repeats; name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3"; existence "Predicted"	Primary sequence record	High for architectur name is unreviewe RefSeq transfer
InterPro API (this run)	computational (domain scan)	Refutes	Are there HECT/ RING/E3 signatures?	All 10 signatures over aa 2–166 are RCC1/RLD; **zero** ligase catalytic signatures	Multi-member-DB consensus	High
Sequence analysis (this run)	computational	Refutes	Can a HECT/RING fit?	No RING consensus; 6 Cys; 169 aa ≪ HECT (~350 aa) → HECT impossible	Direct on target sequence	High
InterPro Q15034 vs A0A6I8W8A2 (this run)	structural/ evolutionary	Refutes	What does a real HERC3 have that the target lacks?	Human HERC3 = RLD + HECT; target = RLD only	Cross-species architecture	High
UniProt proteome + all 4 isoforms (this run)	computational	Refutes	Does any isoform have a ligase domain?	All 4 isoforms (144–282 aa) RLD-only; none among 25 *D.*	Gene-model-wide	High

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence/ limitations
				pseudoobscura HECT ligases		
AlphaFold AF-A0A6I8W8A2-F1 v6 (this run)	structural (predicted)	Qualifies/ Refutes	Is there a second folded catalytic module?	Single compact fold, mean pLDDT 83, 83% >70; no accessory catalytic domain	3D model	Moderate–High (mod not experimen
QuickGO / UniProt GO xrefs (this run)	database	Qualifies	Is GO:0016874 an existing annotation?	0 GO annotations; prediction is external ProtNLM2 name transfer only	GOA/ UniProt	High
PMID: 26801221	review	Refutes (defines requirement)	Do HERC ligases need a HECT domain?	HERCs have "the HECT domain and the RCC1-like domain (RLD)"; small HERCs have "single HECT and RLD domains"	HERC family	High
PMID: 32983929	review	Refutes (defines catalysis)	Which domain is catalytic?	"C-terminal HECT domain … mediates the binding of ubiquitin to substrate proteins"	HECT mechanism	High
	review					

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence/ limitations
PMID: 31447701		Qualifies (architecture)	HERC definition	HERCs = HECT + RCC1-like domain-containing proteins	HERC family	Moderate–High
PMID: 31732561; PMID: 32039437	structural / direct assay	Refutes (size/ mechanism)	How large is a functional HECT unit?	HECT unit >340–380 aa with catalytic Cys; 169 aa cannot host it	AREL1 / UBE3C crystal structures	High
PMID: 21093592; PMID: 29042532	structural	Competing (alt. function)	What does an RLD do?	RLD = seven-bladed β-propeller acting as RanGEF/ interaction module	Yeast/ human RCC1	High for RI being non-ligase

GO Curation Implications

Lead (requires curator verification):

- **Current state:** QuickGO/GOA and the UniProt record return **zero GO annotations** for A0A6I8W8A2. The ligase-activity claim exists only as an external **ProtNLM2 name-based prediction**, not as an existing GOA IEA annotation. The curation action is therefore to **decline to introduce** GO:0016874, not to remove an existing one.
- **Do NOT assign GO:0016874 (ligase activity) or child terms** (e.g., GO:0061630 ubiquitin protein ligase activity) to A0A6I8W8A2 on the basis of this sequence. The catalytic HECT domain that would justify a ligase MF term is absent, the protein is too short to host one, and the AlphaFold model shows a single β-propeller.
- **Supportable annotation instead:** the evidence supports an **RCC1-repeat / RCC1-like domain** description (InterPro IPR000408 / IPR009091; Pfam PF00415), annotatable via

InterPro2GO where appropriate. If an activity term is desired, *guanyl-nucleotide exchange factor activity* (GO:0005085) is the family-level fold hypothesis, but there is no direct evidence for this protein, so it should be flagged as uncertain (ISS/IEA at best), not asserted. Avoid "protein binding" as the terminal recommendation; prefer noting the RCC1 repeat domain as the structured, supportable feature.

GO Decision Table

Term	Aspect	Proposed action	Rationale
GO:0016874 ligase activity	MF	**Do not add / reject**	No HECT/RING/catalytic domain; too short; AlphaFold single propeller
GO:0061630 ubiquitin ligase activity	MF	**Do not add / reject**	Specific catalytic module absent
GO:0005085 GEF activity	MF	**Lead only — uncertain**	RLD fold is a RanGEF module, but no direct evidence in DROPS
RCC1-repeat domain (InterPro2GO)	—	**Consider (supportable)**	Only detectable domain; robustly evidenced

Mechanistic Scope

The activity under test is **direct catalytic ligase activity** of the gene product. The analysis targets exactly this: presence/absence of the catalytic domain (HECT/RING), of catalytic residues (catalytic Cys, RING Zn-chelating Cys/His), and the physical capacity (protein length, AlphaFold geometry) to host such a module. All three lines of evidence are negative. The conclusion — "not a ligase" — is a statement about the **immediate molecular capability** of the protein, not about any downstream phenotype, pathway role, or developmental function. Conversely, the true positive activity of the RLD fold (Ran-GEF / chromatin scaffold) is a **separate, undecided** question that this report does not resolve and that should not be conflated with the refuted ligase claim.

Conflicts and Alternatives

- **Paralog / family over-annotation (most likely explanation, and the actual cause):** The RefSeq automated name "HERC3 isoform X3" transferred a family label to a truncated isoform, and ProtNLM2 then predicted ligase activity from that name. This is name/frequency bias, not evidence of activity.
- **Isoform truncation:** "isoform X3" suggests alternative splicing / gene-model prediction yielding an RLD-only product. The full-length locus *might* encode a HECT+RLD ligase in an unannotated transcript, but that activity cannot be attributed to this specific accession — and all four currently annotated isoforms are RLD-only.
- **Alternative true function:** The RLD fold supports a Ran-GEF / chromatin / protein-interaction role
  (
  P 21093592
).
  P 29042532
  This is the more biologically plausible activity for the fragment, though unverified experimentally in this species.
- **No competing evidence** was found that an RLD alone possesses intrinsic ligase activity; the literature is uniformly consistent that the HECT domain is required for HERC ligase catalysis. There is no direct assay claiming ligase activity for this protein, so the refutation rests on domain/structure evidence plus the absence of any supporting annotation.

Limitations and Knowledge Gaps

1. **Is A0A6I8W8A2 a genuine biological isoform or a mis-predicted fragment?** *Largely resolved:* all four annotated isoforms (144–282 aa) are RLD-only, and none appears among the 25 *D. pseudoobscura* HECT ligases — so the gene model itself encodes only an RLD fragment. *Remaining uncertainty:* whether the underlying genomic locus has an unannotated downstream HECT-encoding exon that current gene models miss. *Resolve by:* inspecting the genomic contig for a HECT ORF and by RNA-seq / full-length transcript evidence. *Why it matters:* even if a longer HECT-containing transcript exists, this specific accession would still not be a ligase.

2. **Which full HERC ortholog does this locus correspond to?** *D. pseudoobscura* does have a large HERC2 (5103 aa, with HECT), but no small-HERC (HERC3/4/5) with a HECT was found. *Resolve by:* reciprocal-best-hit orthology mapping (OrthoDB/InParanoid) of LOC117183218 to *D. melanogaster* to confirm whether a small-HERC ortholog with a HECT exists in Drosophila at all — and to assign the correct positive function.
3. **Predicted-only structure.** The AlphaFold model (mean pLDDT 83.0) is a prediction, not an experimental structure; however, the key negative (no second, catalytic domain) is robust from sequence length alone.
4. **Species-level evidence vacuum.** The entry has zero GO annotations and protein-existence level "Predicted"; there is no experimental data for this gene in *D. pseudoobscura*. Any positive annotation would need to be transferred from characterized orthologs with appropriate evidence codes.

Discriminating Tests

- **Domain re-scan of every gene isoform (InterProScan)** for LOC117183218 — the decisive check: presence/absence of IPR000569 (HECT) in any isoform. (Already run across the four annotated isoforms; all negative.)
- **Orthology + reciprocal-best-hit** to *D. melanogaster* HERC/RCC1 paralogs; verify whether the full-length ortholog carries a HECT domain and locate the catalytic Cys — this assigns the correct family and the correct positive MF.
- **Sequence/structure alignment of the RLD** to characterized RanGEFs (RCC1/Prp20p) vs HERC RLDs, checking Ran-contacting residues — tests whether Ran-GEF activity is plausible.
- **Inspection of the genomic contig** around the locus for a downstream HECT-encoding exon — distinguishes "truncated isoform" from "the whole gene lacks a HECT domain."
- **(Experimental, only if biologically important)** in vitro autoubiquitination / E2–E3 thioester-discharge assay on recombinant A0A6I8W8A2 — predicted to show **no** ligase activity; and a Ran-GEF exchange assay on the recombinant RLD to test the leading alternative function.

Curation Leads (require curator verification)

- **Action change:** Reject/withhold GO:0016874 (ligase activity) for A0A6I8W8A2; flag the ProtNLM2 name as an unsupported family/name transfer to a truncated RLD-only isoform.
- **Replacement description:** Annotate as containing an **RCC1 repeat / RCC1-like domain (β-propeller)** (InterPro IPR000408 / IPR009091; Pfam PF00415); do not assert catalytic ligase activity. Treat *guanyl-nucleotide exchange factor activity* (GO:0005085) only as an uncertain family-level hypothesis pending orthology/assay evidence.
- **Candidate references / exact snippets to verify:**
- PMID: 26801221 — "*The HERC gene family encodes proteins with two characteristic domains in their sequence: the HECT domain and the RCC1-like domain (RLD).*" (Ligase catalysis needs the HECT domain, absent here.)
- PMID: 32983929 — "*…HECT…E3 ubiquitin ligases…with a C-terminal HECT domain that mediates the binding of ubiquitin to substrate proteins…*" (HECT = catalytic module.)
- PMID: 31447701 — HERC = HECT + RCC1-like domain-containing proteins (architecture definition).
- **Suggested question for curator:** Should the review target the gene locus (which may encode a full HERC ligase in another, unannotated isoform) or this specific RLD-only accession? The GO decision differs by scope, but for accession A0A6I8W8A2 itself the ligase term is unsupported either way.
- **Suggested experiment/analysis:** InterProScan across all LOC117183218 isoforms + orthology check to confirm whether any isoform contains a HECT domain.

Provenance (analyses run during the investigation)

1. UniProt REST fetch of A0A6I8W8A2 → 169 aa, RCC1 repeat features (32–87, 88–142), submission name "Probable E3 ubiquitin-protein ligase HERC3 isoform X3," existence "Predicted."
2. Sequence motif / cysteine / RING-consensus scan → no RING consensus; 6 Cys; HECT structurally impossible at 169 aa.
3. InterPro API domain architecture → all 10 signatures RCC1/RLD (IPR000408, IPR009091, PF00415, PF13540, PS00626, PS50012, PR00633, PTHR22872, G3DSA:2.130.10.30, SSF50985); zero ligase catalytic signatures.

4. InterPro architecture comparison: human HERC3 Q15034 (RLD + HECT) vs A0A6I8W8A2 (RLD only).
5. AlphaFold DB fetch of AF-A0A6I8W8A2-F1 v6 → 169 aa, mean pLDDT 83.0, 83% of residues >70 (single compact β-propeller fold).
6. UniProt proteome query (taxon 46245, IPR000569 HECT) → 25 genuine HECT ligases, none from LOC117183218; InterPro architecture of all four LOC117183218 isoforms (144–282 aa) → all RCC1/RLD-only.
7. QuickGO annotation search + UniProt GO cross-references for A0A6I8W8A2 → **0 existing GO annotations** (the ligase claim is an external ProtNLM2 prediction, not a current GOA annotation).
