AIGR Deep Research Report — *mcm-4* (A0A061AL94, *C. elegans*)

Hypothesis under review (computational_prediction): ProtNLM2 predicts *transcription initiation at RNA polymerase II promoter* (GO:0006367) for A0A061AL94 (MCM-4). Is A0A061AL94 instead a minichromosome-maintenance (MCM2-7) replicative DNA helicase / replication-licensing factor, making the prediction a misassignment?

Executive Judgment

Verdict: REFUTED (prediction is a misassignment / over-annotation).

A0A061AL94 is unambiguously a fragment of the MCM2-7 replicative DNA helicase subunit **MCM4**, not an RNA polymerase II transcription-initiation factor. The evidence is decisive and computationally verifiable:

1. A0A061AL94 is a **74-aa sequence** (WormBase locus **Y39G10AR.14b**, gene *mcm-4*), submission-named by UniProt "*DNA replication licensing factor mcm-4*".
2. Its **only Pfam domain is PF21128 = "MCM4, winged helix domain" (WHD_MCM4)** — an MCM4-specific signature.
3. Its sequence is an **exact substring (residues 750–823, the C-terminal 74 aa) of the reviewed Swiss-Prot entry MCM4_CAEEL (Q95XQ8, 823 aa)**. A0A061AL94 is therefore the isolated C-terminal winged-helix domain of full-length *C. elegans* MCM4.
4. The reviewed parent entry (Q95XQ8) carries only DNA-replication GO terms (DNA helicase activity, ATP binding/hydrolysis, ssDNA binding, Zn binding, MCM complex, mitotic DNA replication initiation, premeiotic DNA replication, DNA strand elongation). **No transcription-related GO term appears.**

The most likely source of the ProtNLM2 error: the **winged-helix (WHD) fold** is a generic DNA-binding module shared by many bona-fide transcription factors, and a language model conflates **"DNA replication initiation" with "transcription initiation"** (frequency bias on the word "initiation"). GO:0006367 should **not** be applied to this gene product.

Caveat: A0A061AL94 is a 74-aa fragment, so it lacks the catalytic AAA+/Walker motifs (which lie in the parent's PF00493 domain upstream); the family assignment nonetheless rests on the MCM4-specific WHD signature and exact identity to the reviewed full-length MCM4.

Evidence Matrix

Citation	Evidence type	Direction	Claim tested	Key finding	Context
UniProt A0A061AL94 (database record)	Database / computational	Refutes prediction	Identity of A0A061AL94	Gene *mcm-4*; submission name "DNA replication licensing factor mcm-4"; sole Pfam PF21128 = MCM4 winged-helix domain	*C. elegans*, WormBase Y39G10AR.14b
InterPro/ Pfam PF21128 (database)	Structural/ evolutionary	Refutes	Domain identity of the 74-aa fragment	PF21128 = "MCM4, winged helix domain" — an MCM4-specific WHD	Family signature
UniProt Q95XQ8 MCM4_CAEEL (reviewed) + own sequence alignment	Structural/ evolutionary (computed)	Refutes	Is A0A061AL94 part of MCM4?	74-aa fragment is an **exact match to residues 750–823** (C-terminus) of reviewed 823-aa MCM4_CAEEL; parent has PF00493 (MCM AAA+), PF17855, PF14551, PF17207, PF21128	*C. elegans*
UniProt Q95XQ8	Review/ database	Refutes	Function of MCM4	GO: DNA helicase	*C. elegans*

Citation	Evidence type	Direction	Claim tested	Key finding	Context
curated GO set				activity (GO:0003678), ATP binding/ hydrolysis, ssDNA binding, Zn binding, MCM complex, **mitotic DNA replication initiation (GO:1902975)**, premeiotic DNA replication, DNA strand elongation; **no transcription term**	
Yang et al. 2026, P 41448435	Review/ mechanistic (human)	Qualifies (family biology)	MCM2-7 role	ORC/CDC6/ CDT1 load MCM2-7 double hexamers to license replication origins — replication initiation, not transcription	Human, in vitro reconstitution
Tomkins et al. 2025, P 41365879	Mechanistic (human)	Qualifies	MCM2-7 role	MCM2-7 helicase loaded onto origins in G1; geminin/CDT1	Human

Citation	Evidence type	Direction	Claim tested	Key finding	Context
				licensing control	
Yuan et al. 2017, P 28191893	Structural (cryo-EM, 3.9 Å)	Refutes	Function of MCM winged-helix domains	In the OCCM loading intermediate, flexible Mcm2-7 **WHDs engage ORC-Cdc6**; Mcm6 WHD binds Orc4 WHD — WHDs are replication-licensing structural elements	*S. cerevisiae*
Zhai et al. 2017, P 28191894	Structural (cryo-EM)	Refutes	Role of MCM C-terminal WHD	Mcm5 C-terminal WHD occludes the central channel of the Cdt1-Mcm2-7 helicase precursor (double-hexamer formation)	*S. cerevisiae*
Miller & Winston 2023, P 36924499	Genetic/ genomic	Competing→resolved	Any MCM–transcription link	Related hit explicitly **separates** replication (MCM origin loading) from RNAPII	*S. cerevisiae*

Citation	Evidence type	Direction	Claim tested	Key finding	Context
				transcription; no MCM transcription-initiation activity	

GO Curation Implications

- **GO:0006367 (transcription initiation at RNA Pol II promoter, BP): REJECT / do not add.** No experimental or reliable computational support; it conflicts with the well-established replicative-helicase identity. This is a paralog-independent misassignment driven by domain-fold ambiguity and word-frequency bias.
- **Supported terms for *mcm-4* (from reviewed parent Q95XQ8), as leads requiring curator verification:**
- MF: **DNA helicase activity (GO:0003678)**, ATP hydrolysis activity (GO:0016887), ATP binding (GO:0005524), single-stranded DNA binding (GO:0003697).
- BP: **mitotic DNA replication initiation (GO:1902975)**, DNA strand elongation involved in DNA replication (GO:0006271), premeiotic DNA replication (GO:0006279).
- CC: **MCM complex (GO:0042555)**, nucleus (GO:0005634), chromosome (GO:0005694).
- Because A0A061AL94 is only the **C-terminal WHD fragment**, MF catalytic terms describe the full protein/complex; for this isoform record, the WHD's role is regulatory/DNA-binding within the CMG helicase. "Protein binding" is not needed — more informative replication terms apply.

Mechanistic Scope

The immediate molecular function of MCM4 is as a subunit of the **MCM2-7 hexameric ATPase**, the catalytic core of the eukaryotic **CMG (CDC45–MCM–GINS) replicative DNA helicase**. Its processes are **origin licensing and replication-fork unwinding during DNA replication initiation and elongation** — a *DNA*-templated process at replication origins, mechanistically

distinct from **RNA polymerase II transcription initiation at promoters**. The 74-aa C-terminal **winged-helix domain** contributes DNA binding/regulatory contacts within the helicase; it is not an RNA Pol II general transcription factor and is not part of the pre-initiation complex machinery (TFIIA/B/D/E/F/H, Mediator, Pol II).

Downstream *mcm-4* loss-of-function phenotypes in the parent record (gonad development, premeiotic DNA replication, nervous-system development, locomotion) are **secondary consequences of impaired replication**, not evidence of a transcription-initiation activity.

Conflicts and Alternatives

- **No credible conflicting evidence** supports a transcription-initiation role. The prediction stands alone against domain, orthology, exact-sequence, and curated-GO evidence.
- **Alternative explanation for the prediction (favored):** WHD fold ambiguity (winged-helix domains occur in many transcription factors) + language-model frequency-bias conflation of "replication initiation" ↔ "transcription initiation." This is a classic ProtNLM naming artifact, not paralog confusion.
- **Moonlighting confound checked and rejected (Iteration 2):** A targeted search for an MCM/RNA-Pol-II transcription-initiation moonlighting activity returned no supporting primary evidence. The most relevant hit (P 36924499) explicitly separates MCM's replication role from RNAPII transcription. Cryo-EM structures (P 28191893) (P 28191894) show the MCM WHD — the exact domain of A0A061AL94 — functions in origin licensing (engaging ORC-Cdc6), not at Pol II promoters.
- **AlphaFold check (could-not-run):** AlphaFold DB has no deposited model for A0A061AL94 or the reviewed parent Q95XQ8 (both HTTP 404), so a per-residue pLDDT/fold confirmation could not be executed; the WHD fold is instead established from the Pfam PF21128 signature and the cryo-EM literature above.

- **Fragment caveat:** the 74-aa isoform lacks Walker A/B and the MCM box (in the upstream AAA+ region), so catalytic MF terms should be curated on the full-length protein rather than asserted from this fragment alone.

Knowledge Gaps

1. **Isoform-specific function of Y39G10AR.14b (the 74-aa product).** Checked: it maps exactly to the MCM4 C-terminus. Gap: whether this short isoform is independently expressed/functional. Matters because GO on isoform records should reflect the isoform. Resolution: WormBase expression/isoform annotation, RNA-seq isoform evidence.
2. **No PMID directly assaying *C. elegans* mcm-4 transcription activity** (correctly — none is expected). Resolution: none needed; absence is consistent with refutation.

Discriminating Tests

- **Domain/sequence (done):** Pfam = MCM4 WHD; exact identity to reviewed MCM4 C-terminus → replication, not transcription. ✔
- **Orthology:** OrthoDB group 10251574at2759 clusters with eukaryotic MCM4 orthologs (replicative helicase), not with RNA Pol II GTFs.
- **Complex membership / interactome:** MCM4 co-purifies with MCM2/3/5/6/7 (MCM complex), CDC45, GINS — not with TFIID/Mediator/Pol II. A physical-interaction check would further discriminate.
- **Structural:** AlphaFold model of the 74-aa fragment should adopt a compact winged-helix fold consistent with MCM4 C-terminal WHD.

Curation Leads (require curator verification)

- **Action:** Do **not** annotate *mcm-4*/A0A061AL94 with GO:0006367; flag the ProtNLM2 prediction as a false positive (fold-ambiguity + "initiation" frequency bias).

- **Replacement terms (leads):** MF GO:0003678 (DNA helicase activity); BP GO:1902975 (mitotic DNA replication initiation); CC GO:0042555 (MCM complex) — transfer with appropriate evidence codes / IBA from the reviewed parent Q95XQ8.
- **Candidate reference to verify (snippet):** UniProt reviewed entry **Q95XQ8 (MCM4_CAEEL)** curated GO includes "mitotic DNA replication initiation" and "MCM complex," with no transcription term — supporting rejection of GO:0006367.
- **Suggested question for curator:** Should isoform record A0A061AL94 (74-aa WHD fragment) inherit full-length MCM4 MF/BP terms, or carry only CC/domain-appropriate annotation?
- **Suggested experiment:** Confirm MCM-complex co-fractionation / interactome membership of the mcm-4 gene product to formally exclude any Pol II association.

Paralog-Specificity Confirmation (Iteration 3)

To exclude MCM-paralog confusion, the 74-aa fragment was tested for exact-substring membership across the reviewed *C. elegans* MCM2-7 subunits. It is contained **only in Q95XQ8 (MCM4, 823 aa)** and is **absent** from the reviewed mcm-5 (Q21902, 759 aa) and mcm-6 (P34647, 810 aa) sequences. This confirms A0A061AL94 is specifically an MCM4 fragment, not a mis-mapped paralog. (UniProt gene-name search is fuzzy and returned some cross-hits by accession, but the substring test is unambiguous.)

Provenance

Computed in-run: UniProt REST fetch of A0A061AL94 (74 aa, gene *mcm-4*, Pfam PF21128 = MCM4 WHD) and Q95XQ8 (MCM4_CAEEL, 823 aa); exact-substring alignment placing the fragment at residues 750–823 (C-terminal WHD); enumeration of Q95XQ8 curated GO terms (all DNA-replication, none transcription); InterPro/Pfam lookup identifying PF21128 as "MCM4, winged helix domain."
