AIGR Gene Hypothesis Deep Research — Final Report

Target gene: A0A8B6GS20 (*Mytilus galloprovincialis*, MYTGA, NCBITaxon:29158) **UniProt automated name:** Myotubularin-related protein 9 (ECO:0000313 / EMBL VDI68180.1) **Length:** 607 aa **Hypothesis under test:** ProtNLM2 predicts *phosphatidylinositol-3-phosphate phosphatase activity* (GO:0004438). **Focus type:** computational_prediction

Verdict: REFUTED (over-annotation)

Summary

The ProtNLM2 prediction that *Mytilus galloprovincialis* protein A0A8B6GS20 possesses phosphatidylinositol-3-phosphate phosphatase activity (GO:0004438) is **refuted** by direct inspection of the catalytic machinery required for that activity. Myotubularins are members of the protein-tyrosine-phosphatase (PTP) superfamily, and every catalytically active myotubularin dephosphorylates PtdIns3P (and PtdIns(3,5)P_2) using a strictly conserved P-loop of the form **CX_5R(T/S)**, in which a nucleophilic cysteine forms the covalent phospho-enzyme intermediate and a downstream arginine stabilizes the transition-state oxyanion. Loss of the catalytic cysteine is the defining, diagnostic lesion that separates active myotubularins from the "pseudophosphatase" members of the family.

In A0A8B6GS20 the P-loop is present in fold but degenerate in substance. Anchored on the invariant myotubularin histidine scaffold (`V.VH-[Cx5R]`), active human orthologs read **VLVH-CSDGWDRT** (MTM1; catalytic Cys375/Arg381) and **VVVH-CSDGWDRT** (MTMR2; catalytic Cys417). The aligned segment in the *Mytilus* target reads **VLVH-GSEGFDTT**: the catalytic nucleophilic cysteine is replaced by glycine (C→G) and the catalytic arginine by threonine (R→T). A whole-protein scan for any `CX5R` motif found only a single incidental match (`CNNEKER` at Cys485) that lacks the obligatory downstream Thr/Ser, is not at the P-loop position, and sits in a charged surface loop — i.e., there is no intact, correctly positioned CX_5R(T/S) catalytic motif anywhere in the 607-residue sequence. Because there is no active-site thiolate to attack the 3-phosphate, intrinsic PI3P dephosphorylation is mechanistically impossible.

This assignment is fully consistent with the biology of the protein's named human ortholog, **MTMR9**, a textbook catalytically dead pseudophosphatase that functions as a regulatory adaptor for active myotubularin partners (MTMR6, MTMR7, MTMR8) rather than as an enzyme in its own right. The recommended curation action is therefore to **not** assign GO:0004438 as a direct molecular function; the biologically supported alternative is a **non-catalytic pseudophosphatase** / **adaptor** role. Two caveats temper the peripheral details without changing the core verdict: (1) k-mer similarity was too weak to confirm the exact human paralog, so the specific "MTMR9" name is provisional; and (2) the target carries an atypical N-terminal architecture (zona-pellucida/D8C_UMOD-like repeats, IPR057774/PF23283) absent from vertebrate MTMR9, raising questions about the gene model or a divergent mollusc architecture. Neither caveat rescues the missing catalytic residues.

Key Findings

Finding 1 — The catalytic P-loop is degenerate: the target is a pseudophosphatase (refutes the prediction)

The decisive computational test is whether the residues required for PI3P phosphatase catalysis are present and correctly spaced. Myotubularin *fold* membership is not in dispute — A0A8B6GS20 is annotated through multiple orthogonal signatures (InterPro **IPR030564** Myotubularin; **IPR010569** Myotubularin-like phosphatase domain; **IPR029021** PTP-like; PROSITE **PS51339** PPASE_MYOTUBULARIN; Pfam **PF06602** Myotub-related). The question is purely catalytic competence within that fold.

Anchoring on the conserved myotubularin P-loop scaffold `V.VH-[Cx5R]`, the alignment is unambiguous:

Protein	Aligned P-loop segment	Catalytic Cys	Catalytic Arg	Active?
MTM1 (human, active)	`VLVH` **`C`** `SDGWDR` **`T`**	**C**375	**R**381	Yes
MTMR2 (human, active)	`VVVH` **`C`** `SDGWDR` **`T`**	**C**417	**R**	Yes
A0A8B6GS20 (target)	`VLVH` **`G`** `SEGFD` **`T`** `T`	**G** (Gly ← C)	**T** (Thr ← R)	**No**

Position-by-position, the P-loop reads `C→G, S→S, D→E, G→G, W→F, D→D, R→T, T→T` — **both** catalytic residues are lost. A full-protein regular-expression scan for `C.{5}R` returned exactly one match, `CNNEKER` at Cys485, which (a) lacks the required downstream Thr/Ser, (b) is not at the structural P-loop position, and (c) lies in a charged loop. No functional $CX_5R(T/S)$ P-loop exists anywhere in the protein. This is the signature of a **catalytically dead myotubularin pseudophosphatase** and directly refutes the GO:0004438 prediction.

The interpretation is anchored to primary literature. PMID: 22647598 states that within the family, "*Seven members are inactive because they lack the conserved cysteine residue in the CX(5)R motif required for activity*" — precisely the lesion observed here (C→G). The same study confirms the named ortholog is inactive: active MTMRs are "*all of which dimerize with the catalytically inactive MTMR9*." Independent work (PMID: 31704058) reinforces this, describing that "*the inactive phosphatase MTMR9 localizes to the intermediate compartment and to the Golgi apparatus and is able to recruit its active phosphatase partners MTMR6 and MTMR8 to these locations*" — i.e., MTMR9 acts as an adaptor, not an enzyme.

Finding 2 — Family-wide catalytic-loop panel places the target unambiguously in the inactive subfamily

To confirm the diagnostic is not an artifact of comparing against only two enzymes, the catalytic loop was compared across 15 human myotubularins, anchored on the invariant histidine scaffold. The result is a clean binary split:

Class	Member(s)	Catalytic loop (H…)	Catalytic Cys present?
Active	MTM1, MTMR1, MTMR2, MTMR7, MTMR8	`HCSDGWDRT`	Yes
Active	MTMR3, MTMR4	`HCSDGWDRTP`	Yes
Active	MTMR6	`HCSDGWDRTS`	Yes
Active (variant)	MTMR14	`HCISGWDRT`	Yes
Inactive	MTMR5/SBF1	`H-PEPVIRFH`	No
Inactive	MTMR9	`H-RPFYPAVE`	No
Inactive	MTMR11	`H-GRDFRLLR`	No
Inactive	MTMR12	`H-LLPGEQLL`	No
Inactive	MTMR13/SBF2	`H-SVFKTDVH`	No
Target	**A0A8B6GS20**	`VLVH-G-SEGFDTT`	**No (Gly)**

All nine catalytically active members carry the nucleophilic cysteine immediately after the anchoring histidine; all five known pseudophosphatases lack it. The *Mytilus* target has glycine at that position, grouping it unambiguously with the inactive subfamily and excluding it from the active-enzyme class. Presence/absence of the catalytic cysteine is the accepted family-level discriminator (PMID: 22647598).

Finding 3 — Exact human paralog unresolved; AlphaFold model exists but cannot rescue absent residues

A 4-mer Jaccard comparison of the target's C-terminal phosphatase-domain region against all 15 human myotubularins produced uniformly near-noise similarity (maximum Jaccard 0.026 — 14 shared 4-mers, vs MTM1; all others ≤0.022). The ranking does not cleanly favor MTMR9, so the automated "Myotubularin-related protein 9" name is **not corroborated at the paralog level** and should be treated as provisional. This is a naming/subfamily caveat only: the catalytic verdict rests on absence of the catalytic cysteine/arginine, which is independent of which specific human paralog is closest, since **all** candidate inactive paralogs equally lack the catalytic cysteine.

AlphaFold DB does host a model for this accession (**AF-A0A8B6GS20-F1**, Monomer v2.0, 2022-06-01). Structural geometry cannot restore catalytic function: the nucleophilic cysteine and catalytic arginine are simply **absent from the sequence** (P-loop = VLVH-GSEGFDTT), so there is no thiolate to position in an active site regardless of how well-folded the domain is. A predicted 3D model of a pseudophosphatase fold is not evidence of enzymatic activity.

Separately, the target has an atypical N-terminal architecture — three zona-pellucida / D8C_UMOD-like repeats (IPR057774, PF23283, MatchStatus 2) — that is absent from vertebrate MTMR9 (which carries an N-terminal PH-GRAM domain). This may indicate a divergent mollusc architecture or a mis-joined gene model. Either way, the C-terminal myotubularin domain is unambiguously catalytically degenerate.

Mechanistic Model / Interpretation

Myotubularin catalysis follows the canonical two-step PTP mechanism, which requires the intact $CX_5R(T/S)$ P-loop:

```
ACTIVE MYOTUBULARIN (e.g., MTM1, MTMR2)
---------------------------------------
     PtdIns3P (3-phosphate)
           |
           v
  [ Cys-S(-) ]  nucleophilic attack --> covalent phospho-Cys intermediate
     Arg(+)     stabilizes transition-state oxyanion
     Asp        general acid/base
           |
           v
     PtdIns  +  Pi     (PI3P dephosphorylated)


TARGET A0A8B6GS20  (P-loop = VLVH-GSEGFDTT)
-----------------------------------------
     PtdIns3P
           |
           X   no Cys thiolate  (C -> G)
           X   no catalytic Arg (R -> T)
           |
     No catalysis possible  ==>  PSEUDOPHOSPHATASE
```

The biological role of catalytically dead myotubularins is not "nothing." Across the family, pseudophosphatase members act as **obligate regulatory partners** of active members: catalytically inactive MTMR5 binds MTMR2, increases its activity, and dictates its localization (PMID: 12668758); MTMR13/SBF2 partners MTMR2 and, when mutated, causes CMT4B2 despite being catalytically inactive (PMID: 12687498); and MTMR9 heterodimerizes with, and dictates the activity/substrate specificity/localization of, MTMR6, MTMR7, and MTMR8 (PMID: 22647598), while also regulating ER-to-Golgi trafficking and WNT3A secretion (PMID: 31704058).

The most defensible model for A0A8B6GS20 is therefore: a myotubularin-fold protein that has **lost direct PI3P phosphatase activity** and most likely functions as a **non-catalytic adaptor/regulator** within a myotubularin heterodimer, plausibly contributing to phosphoinositide/membrane-trafficking pathways *indirectly* through an active partner rather than by dephosphorylating PtdIns3P itself.

Evidence Base

Citation	Evidence type	Direction	Claim tested	Key finding	Context
This run (UniProt A0A8B6GS20 + alignment)	Computational / structural-evolutionary	**Refutes**	Does the target retain the $CX_5R(T/S)$ catalytic P-loop?	Catalytic Cys→Gly and Arg→Thr; canonical CSDGWDRT → GSEGFDTT; no intact $CX_5R(T/S)$ anywhere in 607 aa	*M. galloprovinci* vs human MTM1/MTMI
This run (15-member family panel)	Computational (comparative)	**Refutes**	Active vs inactive clade membership	Catalytic-Cys position = Gly → clusters with 5 pseudophosphatases, excluded from 9 active enzymes	Human MTM family vs tar
PMID: 22647598 (Zou et al., 2012, *J Biol Chem*)	Direct assay / family biochemistry	**Supports refutation**	Are motif-lacking MTMRs inactive; is MTMR9 inactive?	"*Seven members are inactive because they lack the conserved cysteine residue in the CX(5)R motif required for activity*"; MTMR6/7/8 dimerize with "*the catalytically inactive MTMR9*"	Human myotubulari in vitro assay
PMID: 31704058 (Doubravská et al., 2020)	Localization / functional	**Supports refutation**	What does MTMR9 do if not a phosphatase?	"*the inactive phosphatase MTMR9…is able to recruit its active phosphatase partners MTMR6 and MTMR8*"; regulates ER-Golgi trafficking / WNT3A secretion	Human cells
	Interaction / biochemistry	Qualifies		Inactive MTMR5 binds MTMR2,	Human

Citation	Evidence type	Direction	Claim tested	Key finding	Context
PMID: 12668758 (2003)			Can an inactive MTMR regulate an active one?	increases activity, dictates localization	
PMID: 12687498 (2003)	Mutant phenotype / genetics	Qualifies	Are pseudophosphatases biologically essential?	MTMR13/SBF2, a pseudophosphatase, causes CMT4B2 when mutated	Human patie
PMID: 17917119 (Bolis et al., 2007)	Review	Orientation	Family active/ inactive split	Of 14 human MTMRs, 8 active / 6 catalytically inactive	Human PNS
PMID: 23630283 (2013)	Loss-of-function (shRNA)	Qualifies	Does MTMR9 influence PI signaling?	MTMR9 silencing alters T-cell differentiation & AKT/PIP3 signaling — indirect regulatory effect, not direct PI3P catalysis	Mouse T cells
AlphaFold AF-A0A8B6GS20-F1 (v2.0)	Structural (prediction)	Neutral / qualifies	Can structure rescue activity?	Model exists but cannot restore absent Cys/Arg	in silico
4-mer Jaccard paralog scan (this run)	Computational (orthology)	Qualifies (naming only)	Is "MTMR9" the correct paralog?	Max Jaccard 0.026 — too divergent to assign a specific paralog	target vs 15 human MTM

GO Curation Implications (leads — require curator verification)

- **GO:0004438 (PI3P phosphatase activity, MF): do NOT assign as a direct molecular function.** The catalytic residues required for this activity (nucleophilic Cys and transition-state Arg of the $CX_5R(T/S)$ P-loop) are absent (C→G, R→T). The ProtNLM2 prediction is an **over-annotation** driven by whole-domain/fold homology (the protein *is* a myotubularin by fold) without checking active-site integrity — the classic pseudophosphatase misassignment. If any myotubularin MF term is used, it should carry a **NOT** qualifier or be omitted.
- **Preferred alternative annotation:** capture the non-catalytic role rather than the uninformative "protein binding." The informative direction is **regulation/potentiation of a phosphoinositide phosphatase via heterodimerization** (a phosphatase-activator/regulator MF, or an adaptor-in-complex annotation) — flagged as a lead based on human MTMR9 data (PMID: 22647598, PMID: 31704058), contingent on confirming ortholog identity and partner conservation in *Mytilus*.
- **BP leads (by orthology, non-core):** phosphatidylinositol-mediated signaling, ER-to-Golgi vesicle-mediated transport, regulation of protein secretion — plausible but downstream and organism-unverified; mark inferred-from-ortholog, low confidence.
- **Provenance note:** IPR030564/PS51339 correctly place the protein in the myotubularin family, but family membership alone does **not** justify the catalytic MF because this member is a pseudophosphatase.
- **Name caveat:** the automated "Myotubularin-related protein 9" is **not corroborated** at the paralog level (k-mer similarity near noise). Recommend a more conservative "myotubularin-family pseudophosphatase (subfamily unresolved)."

Mechanistic Scope

- **Immediate molecular function tested:** intrinsic 3-phosphatase hydrolysis of PtdIns3P at the D-3 position via the PTP-like CX_5R P-loop (Cys nucleophile + Arg oxyanion stabilization). This specific activity **fails** — the active site is dead.
- **Adaptor/regulatory activity:** heterodimerization with and modulation of active MTMRs — *plausible* by orthology, not directly tested in *Mytilus*.

- **Pathway consequences (downstream, not direct):** altered PI(3,4,5)P_3/AKT signaling and T-cell differentiation on MTMR9 knockdown (PMID: 23630283); ER-Golgi trafficking and autophagy modulation — these are consequences of the adaptor role and must not be recorded as this gene's direct catalytic function.
- **Disease associations** of family pseudophosphatases (e.g., CMT4B2 for MTMR13) — paralog- and organism-specific, and not evidence of intrinsic catalysis.

A curator should treat only the "direct activity" line as bearing on the GO:0004438 MF decision.

Conflicts and Alternatives

1. **Fold-based homology vs active-site reality.** The strongest apparent "support" for the prediction is that A0A8B6GS20 genuinely is a myotubularin by every domain signature. This is exactly the trap that produces pseudophosphatase over-annotation: automated pipelines project the family's modal function (active phosphatase) onto every member. The active-site check breaks the tie decisively against catalysis.
2. **Paralog confusion.** The k-mer analysis could not confirm MTMR9 specifically (Jaccard ≤0.026). The target could be closer to a different inactive myotubularin (MTMR11/12, or an invertebrate-specific subfamily). This changes the *name* and specific BP leads, not the catalytic verdict.
3. **Atypical N-terminal architecture.** The ZP/D8C_UMOD repeats (IPR057774, PF23283) are not part of vertebrate MTMR9, which has an N-terminal PH-GRAM domain. This raises the possibility of a divergent mollusc architecture or a mis-joined gene model; either way the C-terminal phosphatase domain is degenerate.
4. **Indirect PI-pathway effects mistaken for direct activity.** MTMR9 loss-of-function alters PIP3/AKT signaling (PMID: 23630283); a reader could misread this as "MTMR9 regulates phosphoinositides → therefore a PI phosphatase." The mechanism is regulatory (via active partners), not enzymatic.
5. **No competing evidence for retained activity.** Even human MTMR9 lacks the P-loop, so there is no basis to expect the *Mytilus* protein to be active.

Limitations and Knowledge Gaps

Gap	What was checked	Why it matters	What would resolve it
No direct enzyme assay	Motif/active-site analysis only (in silico)	GO MF ideally rests on assay (IDA) or well-supported ortholog inference	In vitro malachite-green / BODIPY-PI3P phosphatase assay on recombinant protein
Exact human paralog unresolved	4-mer Jaccard vs 15 human MTMRs (max 0.026)	Determines gene name and BP leads	ML phylogenetic tree (MSA + bootstrap); reciprocal best-hit BLAST
N-terminal ZP/D8C domain reality	Pfam PF23283 D8C_UMOD ×2, ZP-like repeats	May indicate chimeric/ mis-joined gene model inflating length/ architecture	Genomic/transcript evidence; AlphaFold multi-domain geometry
Mytilus-specific partners unknown	Ortholog inference only	Adaptor annotation needs a documented partner	Co-IP / Y2H / AlphaFold-Multimer against candidate active myotubularins
PH-GRAM lipid-binding module not assessed	Not analyzed here	Relevant to residual membrane targeting even without catalysis	Domain analysis / structure inspection

Discriminating Tests

The following would most efficiently confirm or overturn the verdict:

1. **Recombinant phosphatase assay.** Express the phosphatase domain and test PtdIns3P / PtdIns(3,5)P_2 dephosphorylation (malachite green / BODIPY-PI3P), with an active MTMR (e.g., MTMR2) as positive control. Prediction: **no activity.** The single most decisive experiment.

2. **Active-site rescue control.** Introduce G→C and T→R substitutions to restore the $CX_5R(T/S)$ P-loop and re-assay; recovery of activity would confirm these residues are the sole reason for inactivity.
3. **Phylogenetic placement.** Build an MSA of the target with all human myotubularins plus other bivalve/invertebrate myotubularins and infer an ML tree to identify the true subfamily (active vs inactive) and settle the "MTMR9" name.
4. **Partner interaction screen.** AlphaFold-Multimer and/or co-IP of the target against *Mytilus* active myotubularins to test the adaptor hypothesis; assay whether it potentiates a partner's activity (as MTMR9 does).
5. **hmmscan / active-site profile & AlphaFold active-site inspection.** Quantify Pfam PF06602 / PROSITE PS51339 hits, confirm P-loop coordinates, and superpose the AF model on MTM1 to confirm no cryptic Cys/Arg occupies the catalytic pocket.

Proposed Follow-up Actions (Curation Leads — verify before applying)

- **Action change:** Reject/withhold GO:0004438 as a direct MF for A0A8B6GS20; do not carry the ProtNLM2 prediction into the review as asserted function. Consider a NOT-qualified or omitted phosphatase MF, and reclassify as a pseudophosphatase.
- **Candidate references + exact snippets to verify:**
- PMID **22647598** — "*Seven members are inactive because they lack the conserved cysteine residue in the CX(5)R motif required for activity.*"
- PMID **22647598** — "*all of which dimerize with the catalytically inactive MTMR9*"
- PMID **31704058** — "*the inactive phosphatase MTMR9…is able to recruit its active phosphatase partners MTMR6 and MTMR8 to these locations*"
- **Candidate replacement/new terms (leads):** pseudophosphatase adaptor MF (phosphatase regulator/activator); phosphoinositide-metabolism / ER-to-Golgi transport BP; Golgi / ER-Golgi-intermediate-compartment CC — all by-orthology, contingent on confirming ortholog identity in *Mytilus*.
- **Suggested questions for curator:** Is the automated "MTMR9" call supported by phylogeny? Is the ZP/D8C N-terminus a real fusion or a gene-model artifact? Are *Mytilus* MTMR6/7/8 partners present?

- **Suggested experiments:** in vitro PI3P phosphatase assay (expect negative) ± active-site rescue; partner co-IP; AlphaFold active-site inspection.

Bottom Line

The myotubularin catalytic $CX_5R(T/S)$ P-loop is degenerate in A0A8B6GS20 (Cys→Gly, Arg→Thr; CSDGWDRT → GSEGFDTT), with no intact $CX_5R(T/S)$ motif anywhere in the sequence. Combined with the well-documented catalytic death of its named ortholog MTMR9, the ProtNLM2 PI3P-phosphatase prediction (GO:0004438) is **refuted** — this is a pseudophosphatase, and the term should not be assigned as a direct molecular function. A non-catalytic adaptor/pseudophosphatase role is the biologically supported alternative, and the "MTMR9" name should be treated as provisional pending phylogenetic confirmation.
