# AIGR Gene Hypothesis Deep Research — Final Report

## Target: *S. pombe* gaa1 (UniProt Q9US48) · GO:0004197 (cysteine-type endopeptidase activity)

**Focus type:** computational_prediction · **Hypothesis slug:** prediction-cysteine-endopeptidase **Prediction source:** GO-GPT (via BioReason-Pro) · **Reference context:** doi:10.64898/2026.03.19.712954

## Summary

The GO-GPT/BioReason-Pro pipeline predicts **cysteine-type endopeptidase activity (GO:0004197)** for *Schizosaccharomyces pombe* **gaa1** (Q9US48), a subunit of the GPI transamidase (GPI-T) complex that attaches glycosylphosphatidylinositol (GPI) anchors to the C-termini of substrate proteins. This report evaluates, independently, whether gaa1 is itself the catalytic cysteine protease of the complex, or whether the catalytic transamidation/cysteine-protease chemistry resides in a different subunit — **Gpi8 (mammalian PIG-K)** — with gaa1 acting as a non-catalytic recognition/scaffolding component.

**The prediction is REFUTED.** Every tier of evidence — classical yeast genetics and biochemistry, active-site mutagenesis, human and fungal cryo-EM structures, and direct protein-domain/annotation queries — places the catalytic cysteine-protease/transamidase activity in **Gpi8/PIG-K**, a caspase/legumain-related Peptidase C13 enzyme with a defined Cys–His(–Asn) catalytic triad. gaa1 carries **no** peptidase domain, **no** annotated catalytic residues, and **no** EC number; its role is to recognize/bind the GPI anchor and stabilize the pentameric complex. The prediction is therefore a textbook **subunit-level misassignment**: it attributes the catalytic chemistry of one subunit (Gpi8) to a different, non-catalytic subunit (gaa1).

Two practical points sharpen the curation guidance. First, gaa1 does **not** currently carry GO:0004197, so the prediction would be a **new, incorrect molecular-function addition** rather than a correction of an existing annotation — the correct action is simply to reject it. Second,

even the most generous alternative reading (that Gaa1/GPAA1 assists the amide-bond-forming second half-reaction) does not rescue the prediction: the only peptidase-like fold anywhere in the Gaa1 family is a **Zn/M28 metallopeptidase** superfamily hit (seen in human GPAA1, absent even from *S. pombe* gaa1), which is mechanistically **incompatible** with cysteine-type catalysis.

---

## Key Findings

### Finding 1 — gaa1 is a non-catalytic GPI-T subunit; the catalytic cysteine protease is Gpi8/PIG-K

The GPI transamidase reaction proceeds through a covalent **thioester (carbonyl) intermediate**: an active-site cysteine attacks the carbonyl of the substrate's ω-site residue after cleavage of the C-terminal GPI-attachment signal peptide, then the GPI anchor's ethanolamine amino group displaces the enzyme to form the amide-linked anchor. The seed hypothesis asks which subunit supplies that catalytic cysteine.

Classical genetics/biochemistry answers this directly. Ohishi et al. (2000) showed that *GAA1*-knockout cells fail to accumulate the carbonyl intermediate, but the essential cysteine and histidine residues that generate the intermediate belong to **Gpi8p**, not Gaa1p: "*cysteine and histidine residues of Gpi8p, which are conserved in members of a cysteine protease family, are essential for generation of a carbonyl intermediate... Gpi8p is a catalytic component that cleaves the GPI attachment signal peptide*" (PMID:10793132). The GAA1 requirement is therefore **indirect** — it reflects the need for an intact, functional complex, not a catalytic activity intrinsic to Gaa1.

Meyer et al. (2000) independently mapped Gpi8p's active site and identified it as a **caspase-related cysteine protease** whose catalytic Cys199 and His157 (S. cerevisiae numbering) are essential: "*gpi8 alleles mutated at Cys199 or His157 are nonfunctional*" (PMID:10727241). This directly localizes the catalytic cysteine/histidine to Gpi8p.

Structural biology confirms this across kingdoms. The human GPI-T cryo-EM structure places catalysis in PIGK (the Gpi8 ortholog): "*The PIGK subunit functions as the catalytic component, in which we identified a C206-H164-N58 triad that is critical for the transamination reaction*" (PMID:35165458); GPAA1 — the direct ortholog of *S. pombe* gaa1 — is non-catalytic. The fungal

GPI-T cryo-EM structure reinforces this, describing "*the dynamic accommodation of catalytic subunit Gpi8*" and casting Gaa1/Gab1 as GPI-anchor-binding/scaffold components (PMID:41085069).

Finally, direct UniProt provenance shows the domain asymmetry cleanly. *S. pombe* gaa1 (Q9US48) carries **only** Pfam PF04114 "Gaa1" (IPR007246) — no Peptidase_C13 domain, no annotated active site, no EC number, no cysteine-endopeptidase MF term — whereas *S. pombe* gpi8 (Q9USP5) carries Pfam PF01650 "Peptidase_C13" (IPR001096), annotated ACT_SITE His145/Cys187, EC 3.-.-.-, and GO:0003923. Genetics, mutagenesis, structure, and annotation all converge: **the cysteine-protease catalysis is Gpi8's, not gaa1's.**

## Finding 2 — Domain analysis: gaa1 has zero cysteine-protease signatures; the only peptidase-like fold in the Gaa1 family is a Zn/M28 metallopeptidase, not cysteine-type

To exclude a "hidden" or divergent protease activity in gaa1, direct InterPro API queries were run against the target and its orthologs. *S. pombe* gaa1 (Q9US48) returns **only** Gaa1-component family signatures across all member databases — InterPro IPR007246, Pfam PF04114, PIRSF036762, PANTHER PTHR13304 — and **no** peptidase/protease domain and **no** catalytic-triad signature of any kind. By contrast, *S. pombe* gpi8 (Q9USP5) returns the full catalytic signature set: InterPro IPR001096 "Peptidase C13, legumain," Pfam PF01650 "Peptidase C13 family," PIRSF019663 "Transamidase/peptidase, legumain type," PIRSF500138 "catalytic/GPI8 component," and CATH G3DSA:3.40.50.1460 (the caspase/legumain-like fold).

The human ortholog GPAA1 (O43292) carries the same Gaa1-family entries **plus** a CATH-Gene3D structural superfamily hit G3DSA:3.40.630.10, "Zn peptidases" — an **M28 metallopeptidase-like fold**. This matters for two reasons. First, a metallo/Zn-dependent peptidase fold is **mechanistically incompatible** with a cysteine-type endopeptidase (GO:0004197): one uses a metal ion and an activated-water nucleophile, the other uses a catalytic cysteine thiolate. Second, *S. pombe* gaa1 did **not** even return this metallopeptidase CATH hit, so there is no protease-like structural signal in the target at all. The domain analysis thereby closes the loophole: if any enzymatic role for Gaa1 exists (some models propose Gaa1/GPAA1 contributes to the amide-bond-forming second half-reaction), it would be **metallo/synthetase** in character, not cysteine-type. Either way, GO:0004197 is contradicted at the domain level.

### Finding 3 — Current annotation status: gaa1 has no GO:0004197; its sole MF term is GO:0003923 (TAS) — the prediction would be a novel, incorrect addition

A QuickGO annotation query for Q9US48 confirms that **GO:0004197 is not present** in gaa1's current record. Its only molecular-function annotation is **GO:0003923 (GPI-anchor transamidase activity)**, evidence code TAS, reference

P 26563290

Its other annotations are biological-process terms (GO:0016255 attachment of GPI anchor to protein, IBA/NAS; GO:0006506 GPI anchor biosynthetic process, IEA; GO:0031505 fungal-type cell wall organization, NAS) and cellular-component terms (GO:0042765 GPI-anchor transamidase complex; GO:0005789 ER membrane; GO:0005783 ER; GO:0098553 lumenal side of ER membrane, IC). No cysteine-protease/endopeptidase term exists anywhere in the record.

The curation implication is concrete: the prediction is **not** a duplicate of an existing annotation — it is a proposal to **add a new** molecular-function term. Because that term is refuted by all available evidence, the correct action is to **reject the addition**. Separately (and not as justification for the prediction), the existing GO:0003923 TAS annotation on gaa1 attributes whole-complex catalytic activity to a non-catalytic subunit; by the same subunit-assignment logic it is a candidate for curator review, though annotating whole-complex activity to each subunit is an accepted GO convention. The bottom line for this hypothesis is unchanged: **cysteine-type endopeptidase activity is not supported for gaa1.**

---

## Mechanistic Model / Interpretation

The GPI transamidase (GPI-T) is a pentameric, ER-membrane–embedded enzyme that catalyzes the second major post-translational step in GPI anchoring: it cleaves the C-terminal GPI-attachment signal peptide of a nascent substrate at the ω-site and, in the same catalytic cycle, replaces it with a preassembled GPI glycolipid via an amide (transamidation) bond. The five subunits and their roles are summarized below.

```
      GPI Transamidase (GPI-T) complex — subunit roles
-----------------------------------------------------------------
Subunit (S. pombe / human)   Role
-----------------------------------------------------------------
Gpi8  / PIG-K                CATALYTIC cysteine protease (Peptidase
                             C13 / legumain-caspase-like). Active-site
                             Cys-His(-Asn) triad forms the thioester
                             carbonyl intermediate.  <-- GO:0004197 HERE
Gaa1  / GPAA1  (TARGET)      Non-catalytic. Binds/recognizes the GPI
                             anchor; stabilizes the complex; proposed
                             role in the amide-bond (second) half-
                             reaction. NO cysteine-protease domain.
Gpi16 / PIG-T                Scaffold; disulfide-links to Gpi8.
Gpi17 / PIG-S                Scaffold / substrate handling.
Cdc91 / PIG-U                Accessory.
-----------------------------------------------------------------

Reaction (catalysis at Gpi8/PIG-K, NOT Gaa1):

 Substrate-omega-C(=O)-signal-peptide
            |  Gpi8 active-site Cys thiolate attacks
            v
 Substrate-omega-C(=O)-S-Gpi8   (thioester "carbonyl intermediate")
            |  GPI-ethanolamine-NH2 nucleophile (Gaa1 assists binding)
            v
 Substrate-omega-C(=O)-NH-GPI   +   Gpi8-SH (regenerated)
```

The prediction under review takes a property that is real for the *complex* and for *one specific subunit (Gpi8/PIG-K)* and mis-attributes it to a *different subunit (Gaa1)*. All four evidence tiers place the catalytic cysteine in Gpi8: (1) *GAA1*-null cells lose intermediate formation only because the complex is non-functional, while the actual catalytic Cys/His are Gpi8's ( P 10793132 ); (2) direct mutagenesis pinpoints Gpi8's Cys199/His157 ( P 10727241 ); (3) human and fungal cryo-EM place the Cys–His–Asn triad in PIGK/Gpi8 and label Gaa1/GPAA1 non-catalytic ( P 35165458 ),

);

P 41085069

and (4) domain databases show gaa1 has no protease signature while gpi8 carries the full Peptidase C13/legumain catalytic signature.

A subtle but useful refinement: even the *possibility* that Gaa1 contributes catalytically to transamidation is constrained to be **non-cysteine** in nature. The only peptidase-like structural fold detected anywhere in the Gaa1 family is a **Zn/M28 metallopeptidase** superfamily hit (in human GPAA1, and absent even from *S. pombe* gaa1). A metallopeptidase-like fold cannot support GO:0004197, which is defined by cysteine-thiolate catalysis. So both the "Gaa1 is purely structural/recognition" model and the "Gaa1 assists the amide-bond second half-reaction" model are consistent with **refuting** cysteine-type endopeptidase activity.

---

# Evidence Base / Evidence Matrix

| Citation | Evidence type | Direction | Claim tested | Key finding | Context | Confidence & limitations |
|---|---|---|---|---|---|---|
| PMID:10793132 (Ohishi 2000) | Mutant phenotype + biochemistry | **Refutes** | Does gaa1 supply the catalytic Cys/His? | *GAA1*-null cells fail to form the carbonyl intermediate, but the essential Cys/His belong to Gpi8p ("Gpi8p is a catalytic component that cleaves the GPI attachment signal peptide"). | Mouse F9 cells / yeast; in vitro GPI-anchoring | High. GAA1 requirement is indirect (complex integrity), not catalytic. |
| PMID:10727241 (Meyer 2000) | Direct assay / mutagenesis | **Refutes** | Where is the GPI-T active site? | Gpi8p is caspase-related; Cys199/His157 mutants are nonfunctional. | *S. cerevisiae* | High. Directly identifies catalytic residues in Gpi8, not gaa1. |
| PMID:35165458 (human GPI-T cryo-EM) | Structural | **Refutes** | Which subunit holds the catalytic triad? | PIGK holds the Cys206–His164–Asn58 triad; GPAA1 (gaa1 ortholog) non-catalytic. | *Homo sapiens*, cryo-EM ~3.1 Å | High. Direct structural localization of catalysis to PIGK/Gpi8 |
| PMID:41085069 (fungal GPI-T cryo-EM) | Structural | **Refutes** | Subunit roles in fungal GPI-T | Gpi8 explicitly the "catalytic subunit"; Gaa1/Gab1 | Fungal, cryo-EM | High. Kingdom-matched to *S. pombe*. |

| Citation | Evidence type | Direction | Claim tested | Key finding | Context | Confidence & limitations |
|---|---|---|---|---|---|---|
| | | | | GPI-binding/ scaffold. | | |
| UniProt Q9US48 (gaa1) — direct fetch | Sequence/ domain (database) | **Refutes** | Does gaa1 carry a protease domain/ active site? | Only PF04114 "Gaa1" (IPR007246); no Peptidase_C13, no ACT_SITE, no EC. | *S. pombe* | High. Direct provenance; database-level. |
| UniProt Q9USP5 (gpi8) — direct fetch | Sequence/ domain (database) | **Contrast** | Where do catalytic annotations map? | PF01650 Peptidase_C13, ACT_SITE His145/ Cys187, EC 3.-.-.-, GO:0003923. | *S. pombe* | High. Shows catalytic annotation belongs to gpi8. |
| InterPro API (executed) | Structural/ evolutionary (database) | **Refutes** | Any cysteine-protease signature in gaa1? | gaa1: zero peptidase signatures. gpi8: IPR001096 legumain, CATH caspase/ legumain fold (3.40.50.1460). | *S. pombe* | High. Direct, reproducible |
| InterPro API (executed) | Structural (database) | **Qualifies** | Is the "Gaa1 is peptidase-like" idea cysteine-type? | Human GPAA1 (O43292) shows CATH "Zn peptidases" (M28-like) fold — metallo, not cysteine; *S.* | Human vs *S. pombe* | Moderate; fold-level only, argues against GO:0004197. |

| Citation | Evidence type | Direction | Claim tested | Key finding | Context | Confidence & limitations |
|---|---|---|---|---|---|---|
| | | | | *pombe* gaa1 lacks even this hit. | | |
| QuickGO (executed) | Review/ database | **Qualifies** | Does gaa1 already carry GO:0004197? | No. Sole MF = GO:0003923 (transamidase activity, TAS, P 26563290). | *S. pombe* | High. Prediction would be a new, incorrect MF addition. |
| PMID:27208238 (Arabidopsis AtGPI8) | Mutant phenotype | **Refutes** (context) | Which subunit is the Cys protease across kingdoms? | AtGPI8/PIG-K described as "a Cys protease that transfers an assembled GPI anchor to proteins." | *Arabidopsis thaliana* | Medium-high. Confirms Cys-protease identity is GPI8's in plants too. |
| PMID:12958211 (Nagamune 2003) | Interaction/ composition | **Qualifies** | Is GAA1 a conserved subunit distinct from GPI8? | GAA1 and GPI8 are distinct, separately conserved core subunits across eukaryotes. | *Trypanosoma brucei* | Medium. Reinforces GAA1 ≠ GPI8 |

## GO Curation Implications

**Current live annotation set for gaa1 (Q9US48):** the only MF term is **GO:0003923 (GPI-anchor transamidase activity), evidence TAS, ref** .

P 26563290

**GO:0004197 is NOT currently annotated.** BP: GO:0016255, GO:0006506, GO:0031505. CC: GO:0042765, GO:0005789, GO:0005783, GO:0098553.

| GO term | Aspect | Action for gaa1 | Rationale |
|---|---|---|---|
| GO:0004197 cysteine-type endopeptidase activity | MF | **Do not add / reject** | Catalysis is Gpi8's; gaa1 has no protease domain/active site |
| GO:0003923 GPI-anchor transamidase activity | MF | Retain (flag as whole-complex convention) | Currently annotated (TAS); belongs most directly to Gpi8 |
| GO:0016255 / GO:0006506 (BP) | BP | Retain | Well-supported role in GPI attachment |
| GO:0042765 GPI-anchor transamidase complex (CC) | CC | Retain | gaa1 is a bona fide complex subunit |

**Leads (curator verification required):**

- **Reject** the addition of **GO:0004197** to gaa1 (Q9US48). This would be a new, incorrect MF term — a subunit-level misassignment; the catalytic cysteine-protease/transamidase activity belongs to Gpi8/PIG-K.
- **Review (lower priority)** the existing GO:0003923 (TAS) on gaa1, which attributes whole-complex catalytic activity to a non-catalytic subunit. This is a judgment call (whole-complex activity is sometimes annotated to each subunit by convention), not a clear error.
- The catalytic MF term (GO:0003923; mechanistically GO:0004197) correctly belongs to **gpi8 (Q9USP5)**, where GO:0003923 is already annotated.
- **MF for gaa1:** No catalytic MF is supported by direct evidence. If a molecular-function statement is desired, the defensible level is complex-membership/recognition, not enzymatic activity. Avoid "protein binding" as the terminal recommendation; the informative characterization is *non-catalytic structural/recognition subunit of the GPI-T complex*.

## Mechanistic Scope

The immediate molecular function under test is **cysteine-type endopeptidase (peptide-bond hydrolysis via a catalytic cysteine)**. The relevant chemistry — nucleophilic attack by an active-site Cys thiolate on the substrate ω-site carbonyl to form a covalent thioester intermediate, followed by aminolysis by the GPI ethanolamine to form the amide-linked anchor — is a **direct gene-product activity of Gpi8/PIG-K**. For **gaa1**, the tested activity is **absent**: gaa1's direct contributions are (i) recognition/binding of the GPI-attachment signal and/or the GPI glycolipid, and (ii) structural stabilization of the pentameric complex. The observed loss of intermediate formation in *GAA1*-null cells ( P 10793132 ) is a **downstream, loss-of-function consequence** of complex dysfunction — not evidence of a catalytic activity intrinsic to gaa1. Distinguishing these is the crux: mutant phenotypes of a non-catalytic subunit can mimic loss of catalysis without the subunit being catalytic.

## Conflicts and Alternatives

- **Apparent conflict (resolved):** GAA1 knockout abolishes carbonyl-intermediate formation ( P 10793132 ), which could naively suggest Gaa1 is catalytic. The same paper resolves this: the essential catalytic Cys/His are in Gpi8p; Gaa1 is a required partner, not the enzyme.
- **Paralog/subunit confusion (likely source of the prediction):** gaa1 is a *subunit of a cysteine-protease-containing complex*; association/frequency bias can propagate the complex's catalytic term onto any member. This is precisely the misassignment the seed hypothesis anticipates.
- **Alternative activity for GAA1 (M28-like/second half-reaction):** Some literature proposes GPAA1 assists amide-bond formation and resembles M28 metallopeptidases; this is corroborated here at the fold level (human GPAA1 carries a CATH "Zn peptidases"/M28-like hit). Even if catalytically relevant, this is a metallo/amide-synthetase character, **mechanistically incompatible with GO:0004197**. *S. pombe* gaa1 returns no protease fold signature at all, so the prediction is unsupported even under the most generous reading.

- **Organism-specific differences:** None that rescue the hypothesis. The Cys-protease identity of GPI8/PIG-K is conserved from yeast to human to plant (P 10727241 P 35165458 ), P 27208238 and the distinctness of GAA1 vs GPI8 is conserved even in trypanosomes ( P 12958211 ).
- **Database carry-over:** No existing GO:0004197 annotation on gaa1 to carry over; the current record's MF is GO:0003923 only.

## Limitations and Knowledge Gaps

1. **Does *S. pombe* gaa1 have any weak intrinsic proteolytic/amide-forming activity?** Checked: no domain, active site, or assay supports this. Matters because a positive result could justify a *different* (metallo/synthetase) MF, but never cysteine-endopeptidase. Resolution: reconstituted GPI-T activity assays with catalytically dead gpi8 vs gaa1 point mutants.
2. **Structural triad absence for *S. pombe* gaa1.** Orthologs (human PIGK, yeast) were assessed structurally; pombe gaa1 has not been individually solved. Resolution: map Q9US48 onto the human/yeast GPI-T cryo-EM structures (or AlphaFold model) to confirm no Cys–His–Asn triad in the Gaa1 fold.
3. **Provenance of the GO-GPT prediction.** The reference DOI (2026.03.19.712954) was not programmatically accessible in this run; its internal reasoning could not be inspected. This does not change the verdict.
4. **Positive MF for gaa1 remains under-specified.** The evidence robustly refutes GO:0004197 and supports a non-catalytic recognition/stabilization role, but a precise, primary-evidence-backed "GPI-attachment-signal/GPI-anchor binding" MF term was not pinned to a specific GO ID here.

## Discriminating Tests

1. **Active-site swap/mutagenesis:** Reconstitute GPI-T with catalytically dead gpi8 (Cys→Ala/Ser) vs wild-type gaa1, and vice versa; only gpi8 active-site mutation should abolish transamidation — already demonstrated in orthologs ( P 10727241 ).
2. **Thioester-intermediate trapping / activity-based probes:** Cysteine-protease-directed covalent probes on isolated GPI-T, with subunit identification by MS. Prediction: label maps to Gpi8/PIG-K, not gaa1.
3. **Structural triad mapping:** Superpose an AlphaFold model of Q9US48/gaa1 onto the human PIGK catalytic pocket; confirm the absence of a nucleophilic cysteine positioned for catalysis.
4. **Domain/HMM scan (done at annotation level):** Confirm gaa1 lacks Peptidase_C13 (PF01650) and any MEROPS clan CD cysteine-protease signature — only PF04114 is present.

---

## Proposed Follow-up Actions / Curation Leads (require curator verification)

- **Action:** Reject the GO:0004197 prediction for gaa1 (Q9US48). Do not add cysteine-type endopeptidase activity.
- **Candidate references to attach:**
  - P 10793132 ,
  - P 10727241 ,
  - P 35165458 ,
  - P 41085069 .
- Verify snippet ( P 10793132 ):
  *"cysteine and histidine residues of Gpi8p, which are conserved in members of a cysteine*

*protease family, are essential for generation of a carbonyl intermediate ... Gpi8p is a catalytic component that cleaves the GPI attachment signal peptide."*

- Verify snippet ( ): P 10727241
  *"gpi8 alleles mutated at Cys199 or His157 are nonfunctional."*
- Verify snippet ( ): P 35165458
  *"The PIGK subunit functions as the catalytic component, in which we identified a C206-H164-N58 triad that is critical for the transamination reaction."*
- Verify snippet ( ): P 41085069
  *"the dynamic accommodation of catalytic subunit Gpi8."*
- **Retain existing gaa1 annotations:** GO:0042765 (CC), GO:0016255 (BP), GO:0006506 (BP), GO:0005789 (CC).
- **Route the catalytic term correctly:** the cysteine-protease/transamidase MF belongs to gpi8 (Q9USP5), where GO:0003923 is already annotated.
- **Suggested curator question:** Should gaa1 carry any positive MF term beyond complex membership? Current evidence supports only a non-catalytic recognition/scaffolding role.

---

## Provenance / Artifacts

- Direct UniProt fetches (executed): Q9US48 (gaa1) and Q9USP5 (gpi8) domain/active-site/GO lines.
- InterPro API queries (executed): gaa1 (Q9US48), gpi8 (Q9USP5), human GPAA1 (O43292).
- QuickGO annotation snapshot (executed): Q9US48 MF/BP/CC terms.
- Literature: PMIDs 10793132, 10727241, 35165458, 41085069, 27208238, 12958211.

**Bottom line:** The cysteine-type endopeptidase (GO:0004197) prediction for gaa1 is **REFUTED** — a subunit-level misassignment of Gpi8/PIG-K's catalytic activity. gaa1 is a non-catalytic GPI-transamidase subunit that binds the GPI anchor and stabilizes the complex; it should not receive GO:0004197.