# AIGR Deep Research Report — Q2U1U6 (ASPOR): Glycoside Hydrolase vs Polysaccharide Lyase

**Gene:** Q2U1U6 (Q2U1U6_ASPOR), *Aspergillus oryzae* RIB40, ORF AO090138000091 · 134 aa · PE=4 (Predicted) **Hypothesis under test (computational_prediction):** ProtNLM2 predicts **GO:0004553** "hydrolase activity, hydrolyzing O-glycosyl compounds" (i.e. glycoside hydrolase, GH) for Q2U1U6, whose only signature is **IPR008929 / SUPFAM SSF48230** (Chondroitin AC/alginate lyase fold). Is Q2U1U6 a GH (hydrolysis) or a polysaccharide lyase (PL, β-elimination), and is the prediction supported or a class misassignment?

---

## Executive Judgment

**Verdict: REFUTED (class misassignment), with the residual signal being at most a weak, fold-level, non-catalytic resemblance.**

The ProtNLM2 GO:0004553 (O-glycosyl **hydrolase**) prediction is not supported and is best read as a CAZy **class misassignment**:

1. **Wrong mechanistic class.** The single piece of evidence behind any enzymatic call is a fold-level match to **SSF48230 / IPR008929**, the (α/α)-toroid scaffold of CAZy **polysaccharide lyases** (PL5/PL8/PL15/PL35/PL40). These enzymes cleave glycosidic bonds by **β-elimination** (a lyase, EC 4.2.2.-), producing Δ4,5-unsaturated uronate — mechanistically distinct from **hydrolysis** (EC 3.2.1.-, GO:0004553). If any activity were implied by the fold, it would be lyase, not hydrolase. Predicting a **hydrolase** from a **lyase** fold is a category error.
2. **Too small and only a partial match.** Q2U1U6 is **134 aa**; the SSF48230 match covers only **residues 6–97 (92 aa)** with a weak SUPFAM E-value (**4.03e-06**). Quantitatively decisive: **all 53 reviewed IPR008929 members are 331–1372 aa (median 682); 0% are shorter than 134 aa**, and the smallest members are annotated alginate/polysaccharide lyases. Q2U1U6 is ~2.5× smaller than the smallest family member and ~5× below the median — it cannot reconstitute the $(\alpha/\alpha)_6$ toroid, its substrate groove, or the His/Tyr(/Arg/Asn) catalytic constellation, nor does it provide the Asp/Glu acid–base pair of a hydrolase (geometry check

found no canonical GH di-carboxylate pair; only a weak, incomplete His124–Tyr17–Asn85 cluster).

3. **No positive evidence for catalysis of any kind.** UniProt has **no EC and no GO** for this entry; PE=4 (Predicted); name is a placeholder ("DNA, SC138"). No primary literature characterizes this ORF. The AlphaFold model is confidently folded (global pLDDT 92.3) but is a compact **all-α** domain (~58% helix, no β-sheet, Rg 15.1 Å) — consistent with a fragment of the toroid superfamily, not a functional enzyme.

**Most important caveat:** "Refuted" applies specifically to the **GH / GO:0004553** claim. The data do not license a positive PL annotation either; the honest state is *uncharacterized protein with a weak all-α fold resemblance to the chondroitin/alginate lyase superfamily*. This should not be annotated with any molecular-function enzyme term without experimental evidence.

---

# Evidence Matrix

| Citation | Evidence type | Supports/ Refutes/ Qualifies | Claim tested | Key finding | Context | C li |
|---|---|---|---|---|---|---|
| UniProt Q2U1U6 (record) | database | Refutes | Is GO:0004553 a curated fact? | No EC, **no GO**, PE=4 "Predicted", 134 aa, sole signature SSF48230/ IPR008929 | *A. oryzae* RIB40 genome ORF | H r G o P p |
| InterPro API (this run) | computational | Refutes/ Qualifies | Extent & strength of the lyase-fold match | SSF48230 match = res **6–97** only, E-value **4.03e-06**; homologous-superfamily (fold) level, not a functional family | Sequence signature | H a a |
| AlphaFold AF-Q2U1U6-F1 v6 + this run | computational/ structural | Qualifies | Does it fold into a full toroid? | Confident model (pLDDT 92.3), but **all-α** compact domain (~58% helix, no β-sheet, Rg 15.1 Å); only fragment-sized | In silico model | H in a fr H |
| P 20507980 (Atu3025, PL15) | structural, direct assay | Supports (misassignment) | Mechanism/ size of toroid PLs | α/α-barrel enzyme degrades alginate by **β-elimination**; catalytic **His311/Tyr365**; His531 present (≥531 aa) | *Agrobacterium* alginate lyase | H to e |

| Citation | Evidence type | Supports/ Refutes/ Qualifies | Claim tested | Key finding | Context | C li |
|---|---|---|---|---|---|---|
| P 34392362 (chondroitinase ABC I) | structural, direct assay | Supports (misassignment) | GH vs PL mechanism for this fold family | "cleaves the β-1,4 glycosidic linkage… by **β-elimination**" | Chondroitin sulfate lyase | H ly h |
| P 41532755 (Uly1040, PL40) | structural, direct assay | Supports (misassignment) | Catalytic chemistry of the superfamily | His/Tyr catalytic dyad; Tyr catalytic acid — lyase β-elimination on glycosidic bonds | Marine ulvan lyase | H H G |
| P 22297996 (ODV-E66) | structural | Supports (size argument) | Size of a fungal-adjacent chondroitin lyase | Functional chondroitin lyase spans res 67–704 (~640 aa) | Baculovirus | H u a s |
| P 18500999 | review/ database | Qualifies | Base rate of unknowns in *A. oryzae* | >50% of annotated *A. oryzae* genes were hypothetical proteins | *A. oryzae* genome annotation | M s "t C |
| InterPro API — IPR008929 members (this run) | computational/ evolutionary | Refutes | Is 134 aa plausible for this family? | All 53 reviewed members are **331–1372 aa** (median 682); **0%** shorter than 134 aa; smallest members are annotated alginate/ polysaccharide **lyases** | Reviewed-protein length distribution | H ~ th m c fu |

| Citation | Evidence type | Supports/ Refutes/ Qualifies | Claim tested | Key finding | Context | C li |
|---|---|---|---|---|---|---|
| AF-Q2U1U6-F1 active-site geometry (this run) | computational/ structural | Refutes/ Qualifies | Coherent GH vs PL active site? | No canonical GH di-carboxylate pair; only a single His124–Tyr17 (5.3 Å) + Asn85 cluster loosely echoing (incompletely) the lyase His/ Tyr motif | AlphaFold model geometry | M s d f a m |
| UniProt — A. oryzae IPR008929 paralogs (this run) | computational/ evolutionary | Refutes/ Qualifies | Paralog context of the mis-prediction | A. oryzae has 4 IPR008929 proteins: Q2TXR4 (406), Q2USR3 (453), Q2UB60 (395), Q2U1U6 (**134**); Q2U1U6 is the lone truncated outlier | *A. oryzae* proteome | H tr m p e |
| UniProt — IPR008929 members ≤200 aa (this run) | computational/ database | Refutes | Is GH ever the function of small members? | Short members are "Alginate lyase", "Poly(β-D-mannuronate) lyase", "Dermatan sulfate epimerase" (EC 5.1.3.19) or uncharacterized — **none are glycoside hydrolases** | All organisms | H s β e c h |

## GO Curation Implications (leads requiring curator verification)

- **GO:0004553 (MF, hydrolase / O-glycosyl): Do NOT add.** Not supported; a mechanistic class misassignment (lyase fold → hydrolase term). If it appears in any prior action, recommend **removal / rejection** of the ProtNLM-derived prediction.
- **Do NOT substitute a PL/lyase MF term either** (e.g., GO:0016837 carbon-oxygen lyase, or GO:0030246-type). The evidence is fold-level and partial only; there is no experimental or full-domain basis for a positive enzyme annotation.
- **Recommended stance:** leave molecular function **uncharacterized** (no enzyme MF term). At most, an internal note that the protein shows a weak, partial structural resemblance to the chondroitin/alginate lyase (α/α-toroid) superfamily — orientation only, not an annotation.
- Avoid "protein binding" as a fallback; it is uninformative and equally unsupported.

## Mechanistic Scope

The term tested is a **direct molecular-function** claim (catalysis of O-glycosidic bond hydrolysis). The analysis addresses that directly: (a) the fold family is catalytically a **lyase**, not a hydrolase; (b) the protein lacks the size and complete active-site architecture for either mechanism. No downstream phenotype, pathway, or developmental inference is involved — this is purely a molecular-activity assignment, and it fails at the molecular level.

## Conflicts and Alternatives

- **Paralog/label-propagation bias (now supported):** *A. oryzae* has four IPR008929 paralogs — Q2TXR4 (406 aa), Q2USR3 (453 aa), Q2UB60 (395 aa) and Q2U1U6 (134 aa). Q2U1U6 is the lone truncated outlier; ProtNLM2 (a name/label language model) likely over-generalized a carbohydrate-active label from the larger paralogs, collapsing the mechanistically distinct PL vs GH distinction onto a fragment.
- **No GH anywhere in the superfamily:** Across all organisms, small (≤200 aa) IPR008929 members are annotated as alginate/mannuronate **lyases**, **dermatan sulfate epimerase** (EC

5.1.3.19), or uncharacterized — i.e. β-elimination/epimerase (carbanion) chemistry. None are glycoside hydrolases, confirming GO:0004553 is off-class for this fold.

- **Alternative benign interpretation:** Q2U1U6 may be a non-catalytic small all-α protein (structural/binding module) or a degenerate/truncated remnant of a lyase-fold ancestor. The confident AlphaFold fold plus small size is consistent with a stable mini-domain (or a truncated gene model) rather than a complete enzyme.
- **Database carry-over risk:** because the entry is TrEMBL/PE=4 with no GO, any downstream GH annotation would propagate an unverified, off-class prediction.

## Knowledge Gaps

1. **No biochemical assay** for Q2U1U6 — checked: no primary literature on this ORF. Matters because activity (GH vs PL vs none) is undecidable from sequence alone. Resolve with recombinant expression + activity screen (see below).
2. **No structural superposition to a characterized PL** was run (no Foldseek/DALI available here). Checked fold class computationally (all-α, partial); a structural search against PDB PL/GH representatives would sharpen the "fragment vs domain" call.
3. **Catalytic-residue geometry** not confirmed in 3D: candidate His7/His51/His124 and multiple Tyr exist in sequence, but whether any form a lyase-like His/Tyr dyad in the model was not measured. Matters for judging even a lyase hypothesis.

## Discriminating Tests

- **Foldseek/DALI** of AF-Q2U1U6-F1 against PDB: does it match a full toroid PL catalytic domain or only an isolated α-hairpin cluster? (Fast, decisive on the "partial fold" claim.)
- **Active-site geometry check** in the model: measure spatial proximity/orientation of His/Tyr candidates and any Asp/Glu pair; absence of a coherent catalytic center supports "non-catalytic."
- **HMMER/CAZy dbCAN** scan: does Q2U1U6 hit any PL or GH family HMM above threshold, or only the SUPFAM fold? Lack of a family-level PL/GH hit corroborates misassignment.
- **Experimental (definitive):** recombinant expression and (i) thiobarbituric-acid / A232 assay for β-eliminative unsaturated-product formation (lyase) vs (ii) reducing-sugar/pNP-glycoside assays (hydrolase), on chondroitin/alginate/ulvan and generic glycosides.

## Curation Leads (verify before acting)

- **Action:** Reject/remove the ProtNLM2 GO:0004553 prediction as a **PL→GH class misassignment**; do not add any enzyme MF term.
- **Candidate snippets to verify** (exact quotes):

  P 34392362

  — *"cleaves the β-1,4 glycosidic linkage of chondroitin sulfate (CS) by β-elimination"*;

  P 20507980

  — *"degraded by alginate lyases through a beta-elimination reaction"* and *"His(311) and Tyr(365) as the catalytic base and acid."*
- **Suggested reviewer question:** Should the entry carry a fold-level "chondroitin/alginate lyase-like superfamily" note (orientation only) while remaining unannotated for MF?
- **Suggested experiment:** heterologous expression + paired lyase (A232/TBA) and hydrolase (reducing-sugar/pNP) assays to establish whether Q2U1U6 has any glycan-cleaving activity at all.

---

### Provenance

Analyses executed (code + outputs retained in the iteration log): - **Iteration 1:** UniProt text fetch (134 aa, no GO/EC, PE=4); InterPro API (SSF48230 match res 6–97, E=4.03e-06); AlphaFold API + PDB parse (global pLDDT 92.3; region 6–97 pLDDT 94.3; ~58% helix; no β-sheet; Rg 15.1 Å; catalytic-candidate residue inventory). - **Iteration 2:** IPR008929 reviewed-member length distribution (n=53: 331–1372 aa, median 682, 0% < 134 aa; smallest members = alginate/PL lyases); AlphaFold active-site geometry test (no canonical GH di-carboxylate acid/base pair; single His124–Tyr17 5.3 Å dyad + Asn85, incomplete lyase-type cluster). - **Iteration 3:** UniProt paralog query (4 A. oryzae IPR008929 proteins: 406/453/395/**134** aa — Q2U1U6 is the truncated outlier); UniProt query of ≤200-aa IPR008929 members across all organisms (annotated as alginate/mannuronate lyases, dermatan sulfate epimerase, or uncharacterized — **no glycoside hydrolases**). - Literature via PubMed (PMIDs 20507980, 34392362, 41532755, 22297996, 18500999).

---