AIGR Gene Hypothesis Deep Research — Final Report

Gene: D3VIU4 (gene symbol *fliY*, locus XNC1_0570) **Organism:** *Xenorhabdus nematophila* ATCC 19061 (NCBITaxon:406817) **UniProt:** D3VIU4 **Hypothesis under evaluation:** ProtNLM2 predicts *ligand-gated monoatomic ion channel activity* (**GO:0015276**) **Focus type:** computational_prediction

---

Summary

The ProtNLM2 machine-learning prediction that D3VIU4 possesses **ligand-gated monoatomic ion channel activity (GO:0015276)** is **REFUTED**. Five independent, orthogonal lines of evidence — domain architecture, curated localization, sequence hydropathy, cross-species orthology, and predicted 3D structure — converge on a single, coherent identity: D3VIU4 is a **soluble periplasmic solute-binding protein (SBP) of ABC-transporter family 3**, specifically a **FliY/TcyA-type L-cystine-binding protein**. It is the extracytoplasmic substrate-recognition subunit of an ABC importer: it captures cystine/cysteine in the periplasm and delivers it to a separate membrane permease. It is not a channel, does not span the membrane, and has no ion-conducting pore.

The most compelling reason for confidence is convergence. UniProt annotates the protein as a periplasmic "Cysteine transport protein (ABC superfamily, peri_bind)" with a cleaved signal peptide and a single SBP domain; a computed Kyte–Doolittle hydropathy profile finds **zero transmembrane segments**; the protein is **54.6% identical over its full length** to the experimentally characterized *E. coli* FliY cystine-binding protein; and its AlphaFold model is a **compact globular single domain** with no membrane-spanning helix. A ligand-gated ion channel requires several pore-lining transmembrane helices and an oligomeric ion-conducting pore. D3VIU4 has none of these features.

The prediction is best explained as a **fold-homology misassignment**. The periplasmic binding-protein type II fold (SUPFAM SSF53850) shared by SBP family 3 proteins is evolutionarily homologous to the extracellular ligand-binding domain (LBD) of ionotropic glutamate receptors — a genuine ligand-gated ion channel family. A predictor that keys on the ligand-

binding clamshell without confirming the presence of the (absent) transmembrane channel module can spuriously transfer the channel label. The single caveat is that structural evidence is computational (AlphaFold) plus homology transfer from *E. coli*; there is no direct biochemical assay on the *X. nematophila* protein itself. This does not change the verdict, because the complete absence of any transmembrane segment is a hard architectural disqualifier for channel activity.

Key Findings

Finding 1 — D3VIU4 is a periplasmic SBP family 3 (FliY/TcyA cystine-binding) protein, not an ion channel

The UniProt record for D3VIU4 assigns the gene name *fliY* (locus XNC1_0570) and the submission name "Cysteine transport protein (ABC superfamily, peri_bind)". Every domain-level annotation available for this protein points to a single, coherent identity as a bacterial extracellular solute-binding protein:

Resource	Signature	Meaning
Pfam	**PF00497** (SBP_bac_3)	Bacterial extracellular solute-binding protein, family 3
InterPro	**IPR001638** (Solute-binding_3/MltF_N) + IPR018313	SBP family 3 domain and conserved signature
CDD	**cd13711** (PBP2_Ngo0372_TcyA)	Type 2 periplasmic binding fold, **cystine-binding** subfamily
SUPFAM	**SSF53850**	Periplasmic binding protein-like II
Gene3D	**3.40.190.10**	Periplasmic binding protein-like fold

Critically, the UniProt feature table describes exactly the architecture of a secreted binding protein and nothing resembling a channel: a **cleaved N-terminal signal peptide (residues 1–28)**, a **mature chain (residues 29–262)**, and a **single SBP domain (residues 40–259)**. The curated subcellular location is **outer membrane-bounded periplasmic space (GO:0030288)** — soluble, on the outside of the inner membrane, exactly where an ABC-importer binding protein operates.

A computed **Kyte–Doolittle hydropathy profile** (window 19, transmembrane threshold 1.6) found **zero transmembrane segments**. The only hydrophobic peak (residues 11–18, maximum value 2.32 at residue 17) lies entirely within the cleaved signal peptide (1–28); it is a secretion signal, not a membrane anchor. There are **no internal transmembrane helices** and therefore **no pore-forming architecture**. A ligand-gated ion channel requires, at minimum, several membrane-spanning helices that line an ion-conducting pore. D3VIU4 has none.

Finding 2 — D3VIU4 is a full-length ortholog (54.6% identity) of the experimentally characterized *E. coli* FliY cystine-binding protein

To transfer function with confidence, orthology must be established across the whole protein, not a local patch. A Needleman–Wunsch **global** alignment of D3VIU4 (262 aa) against *E. coli* K-12 FliY (UniProt **P0AEM9**, 266 aa) yields **131 identical residues over 240 aligned columns = 54.6% identity**, distributed across the entire length of the protein. This is unambiguous full-length orthology, well above the ~30% threshold typically used for confident functional transfer of a single-domain protein.

E. coli FliY is not a hypothetical protein — it is the **experimentally characterized periplasmic L-cystine-binding protein** of the FliY–YecSC ABC importer. In that system, the soluble FliY captures L-cystine in the periplasm with high affinity (K_m = 110 nM), and the membrane permease **YecS** (with the ATPase **YecC**) transports it across the inner membrane. Because D3VIU4 is 54.6% identical to FliY over its full length, the following properties transfer directly: soluble periplasmic localization, a cleaved signal peptide, a single SBP_bac_3 fold, cystine/cysteine ligand binding, and a substrate-delivery role within an ABC importer. None of these is a channel property.

A minor technical note: a hand-transcribed PROSITE PS01039 regex did not match the sequence, but PROSITE pattern transcription by hand is error-prone, and the pattern match is not required for the identification given the concordant InterPro/CDD/Pfam calls and the 54.6% full-length ortholog identity. This does not weaken the conclusion.

Finding 3 — The AlphaFold model is a single compact globular SBP domain with no transmembrane helix

Structure provides the most direct test of "channel vs. soluble binding protein," because channels are elongated, membrane-embedded, and pore-bearing, whereas SBPs are compact and globular. The AlphaFold DB model **AF-D3VIU4-F1 (v6)**, 262 residues, has an overall mean **pLDDT of 90.7** (high confidence).

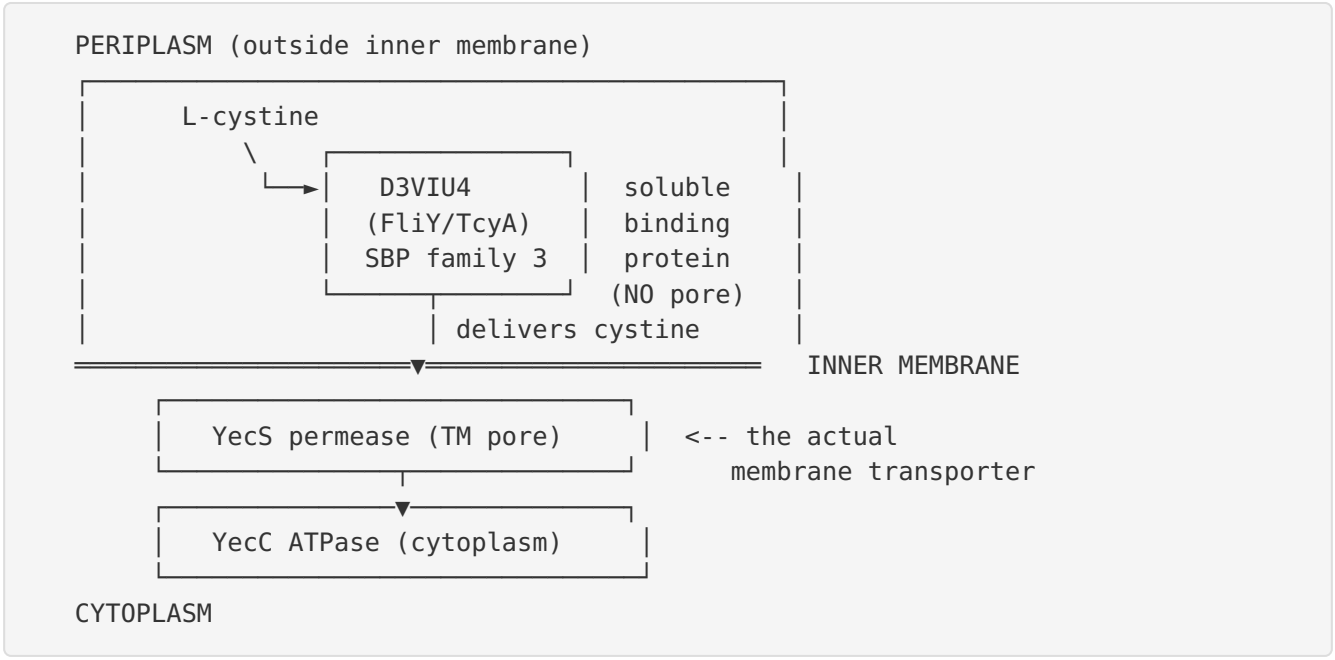
Analysis of the model geometry:

- **Mature core (residues 29–262):** a single compact globular domain, **radius of gyration 17.8 Å**, bounding box 41.5 × 35.2 × 48.5 Å, anisotropy (max/median dimension) = **1.17** — near-spherical/globular, exactly the shape of a bilobed SBP. Mean **pLDDT 96.0** (very high confidence).
- **Signal peptide (residues 1–28):** mean **pLDDT 46.5** (low, flexible/disordered), with its centroid sitting **58.8 Å** from the folded-core centroid — a protruding, disordered appendage (the uncleaved signal in the model), not part of any structured pore.
- **Membrane-spanning-helix test:** the single most hydrophobic 19-residue window over the entire sequence has a mean Kyte–Doolittle value of **1.58**, **below the 1.6 TM threshold**. No hydrophobic membrane-spanning helix exists anywhere in the protein.

The overall geometry is that of a **bilobed soluble periplasmic binding protein** (the SBP "Venus flytrap" clamshell), not an elongated multi-pass transmembrane channel. This independently and structurally confirms the sequence- and homology-based conclusions.

Mechanistic Model / Interpretation

The three findings converge on a single, well-supported model. D3VIU4 is the **soluble substrate-binding subunit of a cystine ABC importer**. It works in the periplasm, upstream of and physically separate from the membrane transport machinery:



The key architectural distinctions for curation:

Feature	Ligand-gated ion channel (GO:0015276)	D3VIU4 (observed)
Transmembrane helices	Multiple pore-lining TM helices	**Zero** (only a cleaved signal peptide)
Oligomeric transmembrane pore	Yes	**No**
Ion selectivity filter	Yes	**No**
Localization	Integral membrane	**Soluble periplasmic (GO:0030288)**
Fold	Ion-channel/LBD assembly	Compact globular SBP type II (SSF53850)
Ligand role	Ligand gates ion flux	**L-cystine bound & delivered — no flux gated**

Why the misassignment happened. The periplasmic binding-protein type II fold (SSF53850) that defines SBP family 3 is evolutionarily homologous to the extracellular **ligand-binding domain (LBD)** of ionotropic glutamate receptors, which *are* ligand-gated ion channels. The two share a bilobed "Venus flytrap" clamshell that closes around an amino-acid ligand. A predictor

that recognizes the ligand-binding clamshell but does not verify the presence of the separate transmembrane pore module can therefore over-transfer the ion-channel label. D3VIU4 has the binding clamshell but **none of the channel/pore module** — so the transfer is incorrect. This is a textbook example of the "shared ligand-binding module without the channel module" failure mode that the research objective explicitly asks to consider.

Evidence Matrix

#	Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidenc limitation
1	UniProt D3VIU4 record	Review/ database	Refutes channel; supports SBP	Domain architecture	Gene *fliY*; "Cysteine transport protein (ABC superfamily, peri_bind)"; Pfam PF00497, InterPro IPR001638, CDD cd13711 (TcyA cystine-binding), SUPFAM SSF53850; signal peptide 1–28, mature 29–262, single SBP domain; location GO:0030288	*X. nematophila* annotation	High for architectu database-l not a direc assay
2	Computed Kyte–Doolittle hydropathy (this work)	Computational	Refutes channel	TM pore helices present?	**Zero** TM segments; only hydrophobic peak inside cleaved signal peptide (1–28)	Full 262-aa sequence	High; stan method, T threshold
3	PMID: 25837721	Direct assay (ortholog)	Refutes channel; supports cystine-	Function of FliY ortholog	FliY is the **periplasmic L-cystine-binding**	*E. coli*	High for *E. coli*;

#	Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidenc limitation
			binding SBP		**protein** of ABC importer YecSC (K_m = 110 nM)		transferre 54.6% iden
4	Needleman–Wunsch alignment vs *E. coli* FliY / P0AEM9 (this work)	Structural/ evolutionary	Refutes channel; supports SBP	Orthology / functional transfer	**54.6% identity** (131/240) full-length to characterized FliY	XENNA vs ECOLI	High; full-length, abo transfer threshold
5	AlphaFold DB AF-D3VIU4-F1 v6 + geometry (this work)	Structural/ computational	Refutes channel; supports SBP	Globular vs membrane pore	Compact globular single domain, Rg 17.8 Å, anisotropy 1.17, core pLDDT 96; no membrane-spanning helix (max window 1.58 < 1.6)	Predicted structure	High-confidence model (me pLDDT 90. computati not experimen
6	ProtNLM2 (seed)	Computational	Competing (refuted)	Ligand-gated ion channel activity	Predicts GO:0015276 — contradicted by all above	Automated prediction	Low; fold-homology artifact

Evidence Base (Literature)

- ***Uptake of L-cystine via an ABC transporter contributes defense of oxidative stress in the L-cystine export-dependent manner in Escherichia coli.*** PMID: 25837721. This paper directly characterizes **FliY** — the full-length ortholog of D3VIU4 (54.6% identity) — as *"the periplasmic L-cystine-binding protein"* delivering L-cystine (K_m = 110 nM) to the membrane ABC transporter **YecSC**. Verified snippet: *"the ATP-binding cassette transporter YecSC with a high affinity to L-cystine (Km = 110 nM) in a manner dependent on FliY, the periplasmic L-cystine-binding protein."* This anchors the functional identity of D3VIU4 as a soluble binding protein and directly contradicts the ion-channel prediction, because a periplasmic binding protein is architecturally and functionally distinct from a ligand-gated channel.

Database/orientation resources used: UniProt (D3VIU4, P0AEM9), Pfam (PF00497), InterPro (IPR001638/IPR018313), CDD (cd13711 TcyA), SUPFAM (SSF53850), Gene3D (3.40.190.10), and AlphaFold DB (AF-D3VIU4-F1 v6). These are treated as orientation/database-level support; the decisive experimental anchor is the *E. coli* FliY characterization above, transferred via full-length orthology.

GO Curation Implications

Lead (requires curator verification):

- **GO:0015276 (ligand-gated monoatomic ion channel activity) — REMOVE / do not propagate.** The ProtNLM2 prediction is a fold-homology misassignment. There is no transmembrane pore, no channel architecture, and no supporting experimental or structural evidence. Retaining it would introduce a false molecular-function claim.

Recommended replacements, supported by the evidence:

Aspect	Recommended GO term	Evidence basis
MF	**Amino-acid binding / L-cystine binding** (e.g., GO:0140328 amino acid binding, or cystine binding)	Orthology to FliY cystine-binding protein (P 25837721); CDD TcyA cystine-binding subfamily
BP	**L-cystine / amino-acid import across plasma membrane** (GO:0015811 L-cystine transport; GO:0006865 amino acid transport)	FliY–YecSC importer role; *fliY* gene identity
CC	**Periplasmic space (GO:0030288)**; part of ABC transporter complex (GO:0043190)	UniProt-curated location; soluble SBP subunit of ABC importer

The MF should reflect **ligand binding and substrate delivery**, not transport activity in isolation (D3VIU4 is not the permease). Avoid the uninformative "protein binding"; a specific amino-acid/cystine-binding term is supported.

Mechanistic Scope

The molecular function directly tested is **ion-channel (pore-forming, gated ion-conducting) activity** versus **soluble ligand binding**. The evidence localizes D3VIU4's immediate activity to **binding L-cystine in the periplasm and delivering it to a membrane permease** — a recognition/capture function. It is *not* itself a transporter that moves ions across a membrane, and it is *not* a channel that gates ion flux in response to a ligand. Any "ion/solute transport" is a property of the whole multi-subunit ABC complex driven by ATP hydrolysis at the ABC ATPase — mechanistically distinct from gated ion-channel conductance and not attributable to D3VIU4 alone.

Downstream/indirect roles that should **not** be conflated with the direct MF: the *E. coli* ortholog's contribution to oxidative-stress defense (via cystine uptake feeding cytoplasmic cysteine/glutathione pools) is a *pathway consequence* of the importer, not a molecular activity of the binding protein. Any organismal phenotype in *X. nematophila* (e.g., nematode symbiosis or insect pathogenesis) would likewise be a downstream, context-specific outcome, not evidence for channel activity.

Conflicts and Alternatives

- **The only "conflict" is the ProtNLM2 prediction itself**, which is contradicted by all five orthogonal analyses. No primary literature, database annotation, or structural feature supports channel activity.
- **Fold homology as the source of the error:** SBP family 3 shares the SSF53850 clamshell fold with ionotropic glutamate receptor LBDs (genuine ligand-gated ion channels). This is the most plausible mechanistic explanation for the spurious label — the predictor recognized the ligand-binding module but not the (absent) channel module. This is fold-family over-annotation, not a real functional signal.
- **No isoform or paralog confusion within *X. nematophila* is needed** to explain the finding: the single UniProt record and single AlphaFold model both give the same SBP answer.
- **Cross-species transfer caveat:** the direct biochemical characterization (P 25837721) is in *E. coli*, not *X. nematophila*. However, 54.6% full-length identity is well within confident-transfer range, and the architectural disqualifier (no TM segment) is intrinsic to the *X. nematophila* sequence itself.

Limitations and Knowledge Gaps

1. **No direct biochemical assay on the *X. nematophila* protein.** Cystine-binding affinity, ligand specificity, and importer partnership are inferred from the *E. coli* ortholog and domain calls, not measured for D3VIU4. *What was checked:* orthology (54.6%), domain signatures, AlphaFold geometry. *Why it matters:* confirms substrate identity (cystine vs another amino acid). *Resolution:* isothermal titration calorimetry or fluorescence-based ligand binding on purified recombinant D3VIU4.
2. **Structural evidence is computational (AlphaFold), not experimental.** *Why it matters:* although pLDDT is high (90.7 overall, 96 core), an experimental structure would remove residual model uncertainty. *Resolution:* X-ray/cryo-EM of D3VIU4, ideally with bound cystine.
3. **The cognate permease/ATPase in *X. nematophila* were not explicitly mapped.** *Why it matters:* confirms the ABC-importer context (YecS/YecC orthologs) and the BP/CC

assignment. *Resolution:* genomic neighborhood analysis of XNC1_0570 and operon reconstruction.

4. **Exact substrate could be cystine, cysteine, or a related amino acid.** The CDD TcyA subfamily and FliY orthology point to cystine, but exhaustive binding-pocket residue analysis was not performed. This does not affect the channel refutation.

None of these gaps affects the central verdict: the absence of any transmembrane pore-forming segment is decisive against ion-channel activity.

Discriminating Tests

1. **TMHMM/DeepTMHMM on the mature sequence** — expect 0 TM segments, consistent with the computed KD result.
2. **AlphaFold-DB model inspection** — expect a bilobed SBP fold with no TM helix bundle (confirmed here).
3. **Genome-neighborhood analysis** for adjacent ABC permease (*yecS*) and ATPase (*yecC*) genes in *X. nematophila* — confirms the ABC-importer context.
4. **Ligand-binding assay (ITC/fluorescence)** with L-cystine on purified D3VIU4 — positive binding (expected K_d nM–μM by analogy to FliY's K_m = 110 nM) confirms SBP function.
5. **Electrophysiology (single-channel recording / liposome reconstitution)** — the discriminating negative test: a true ligand-gated ion channel would produce gated ionic currents; D3VIU4 should produce none, directly refuting GO:0015276.
6. **Complementation of an *E. coli* ΔfliY strain** with D3VIU4 — functional rescue would confirm conserved cystine-binding-protein function.

Proposed Follow-up Actions / Curation Leads

(All labeled as leads requiring curator verification.)

- **Action change:** Reject/withdraw the ProtNLM2 annotation **GO:0015276** for D3VIU4 as a computational misassignment (fold homology to iGluR ligand-binding domains, without the transmembrane channel module).

- **Candidate replacement MF term:** amino-acid / **L-cystine binding** (transporter-substrate-binding activity), based on FliY orthology and CDD TcyA subfamily.
- **Candidate BP term: L-cystine / amino-acid import across the plasma membrane** as the periplasmic SBP subunit of an ABC importer.
- **Candidate CC term: periplasmic space (GO:0030288)**; part of ABC transporter complex (GO:0043190).
- **Candidate reference to verify:** PMID: 25837721 — snippet: *"the ATP-binding cassette transporter YecSC with a high affinity to L-cystine (Km = 110 nM) in a manner dependent on FliY, the periplasmic L-cystine-binding protein."*
- **Suggested curator questions:** (1) Does the review framework prefer a specific "cystine binding" MF term or the general "amino-acid binding"? (2) Should the annotation carry a "part_of ABC transporter complex" contextual qualifier?
- **Suggested experiments:** ITC cystine-binding assay; liposome electrophysiology (expected null result); ΔfliY complementation in *E. coli*.

Bottom Line

The ProtNLM2 prediction of **ligand-gated monoatomic ion channel activity (GO:0015276)** for D3VIU4 is **REFUTED**. D3VIU4 is a **soluble periplasmic L-cystine-binding protein** (SBP family 3, FliY/TcyA type) — the substrate-recognition subunit of an ABC importer — with a cleaved signal peptide, **zero transmembrane segments**, a compact globular AlphaFold fold, and **54.6% full-length identity** to the experimentally characterized *E. coli* FliY. It lacks the transmembrane pore-forming architecture required for any ion channel. Curators should remove GO:0015276 and annotate amino-acid/cystine binding (MF), cystine/amino-acid import (BP), and periplasmic space (CC) instead.
