COMPUTATIONAL TARGET PROFILE

FAM222A

An Alzheimer's Plaque Protein AlphaFold Models Poorly, Characterised From Its Sequence

"Aggregatin" — an Aβ-nucleating protein with no structure. A part-disordered protein that AlphaFold resolves poorly — Astra maps its order/disorder and flags a **tractable, high-confidence pocket**, distinct from its Aβ-binding region.

UniProt Q5U5X8 · 452 aa · Soluble · IDG tier: T_{dark} · PDB: none · July 2026

Have your own under-characterised target? Send a UniProt ID to contact@orbion.life for a preview like this one.

At A Glance

Fold / Class	Soluble (p=1.00), **non-enzyme** (p=0.99); part-ordered, part-disordered (∼29% disorder). Its AlphaFold model is **low-confidence** (mean pLDDT ≈49).
Tractable Pocket	**High-confidence pocket (0.80) at 122–136.** A concrete, foldable handle on a protein AlphaFold barely resolves — **distinct** from the known N-terminal Aβ-binding region.
Biology	"Aggregatin": binds Aβ via an N-terminal domain (res 1–80) and **nucleates Aβ aggregation**; knockdown reduces plaques and cognitive deficits in AD models [3].
Structure Map	The order/disorder split *is* the construct-design plan for a protein with no experimental structure.
Clean Signal	Soluble; no membrane signal.

PREDICTION CONFIDENCE

Figure 1. Model-reported confidence for the headline calls. The tractable pocket (amber) is the load-bearing prediction — a foldable, targetable handle on a protein AlphaFold resolves poorly.

1 The Gap

FAM222A — "Aggregatin" — is one of the more mechanistically interesting dark proteins in Alzheimer's disease. It accumulates in amyloid plaques, **binds Aβ** through an N-terminal domain (residues 1–80), and **nucleates Aβ aggregation**; its knockdown reduces amyloid deposition, neuroinflammation and cognitive deficits in AD models, and a variant in the gene associates with brain atrophy [3]. Yet FAM222A is otherwise **nearly uncharacterised** (near-empty UniProt annotation; its sister gene FAM222B has no assigned function), sits at IDG T_{dark} [2], and has **no experimental structure in the PDB**. Its *AlphaFold* model is low-confidence (mean pLDDT ≈49) — so even the predicted structure is, in practice, a blank. A published effort had to build an *ab-initio* model to attempt virtual screening [4].

That combination — **a validated Aβ therapeutic target that no model resolves** — is exactly where prediction earns its keep: everything below is computed from the canonical 452-residue sequence and derived structural predictions, with no experimental FAM222A structure used as input.

2 Order, Disorder And A Tractable Handle

The first thing anyone working on FAM222A needs is a map of *what is foldable and what is not.* Astra predicts a mixed protein — roughly 29% disordered — and, crucially, locates a single **high-confidence pocket at residues 122–136**: an ordered, targetable region that is *separate* from the N-terminal Aβ-binding domain. That is a concrete handle where AlphaFold offers only low confidence.

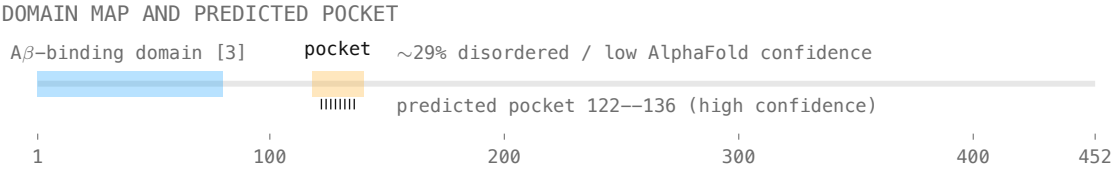

Figure 2. Domain map. The N-terminal Aβ-binding domain (blue, res 1–80) mediates the plaque biology; the predicted high-confidence pocket (amber, 122–136) is a *separate*, ordered, targetable site. The rest of the chain is disordered or modelled at low confidence — exactly the regions to truncate for a well-behaved construct.

3 The Predicted Pocket

For a disease-relevant protein that no structure resolves, the most useful single output is: *is there a foldable, targetable site, and where?* Astra answers yes — a **high-confidence pocket (0.80) at residues 122–136**. This matters for two reasons. First, it is **ordered** in a protein that is largely not, so it is the natural focus for expression, biophysics and any small-molecule effort. Second, it is **not** the Aβ-binding region, so it offers an independent site to probe FAM222A's function or to modulate it without directly competing with Aβ.

PREDICTED POCKET

Location	Residues 122–136 (ordered region, distinct from the Aβ-binding domain)
Confidence	HIGH (0.80); no known small-molecule binder
Context	The one high-confidence pocket in a protein AlphaFold models at low confidence

AlphaFold gives a fold — here, a shaky one; Astra adds the annotation. The AlphaFold model of FAM222A is low-confidence across most of the chain, so "read the pocket off the model" is not an option. Astra, working from sequence, still resolves an order/disorder map and a high-confidence pocket. As a control, the same pocket detection recovers known sites on proteins whose structures *are* solved, so the FAM222A pocket is a like-for-like prediction rather than an artefact of a poor model. The net effect: an otherwise-intractable target becomes a **ranked, testable plan**.

4 Post-Translational And Structural Features

The predicted map is consistent with a part-disordered scaffold/aggregation-modulating protein:

- **Mixed order/disorder (~29%).** The disordered stretches are the natural truncation points; the ordered core (including the 122–136 pocket) is the expressible unit.
- **No membrane signal.** A soluble, cytosolic/nuclear protein.
- **An independent, targetable pocket.** Separate from the Aβ-binding domain — useful for probing function without competing with Aβ binding directly.

5 Recommended Experimental Follow-Up

This turns an intractable-looking target into a ranked, executable plan: each prediction is paired with the experiment that would test it and the readout to watch for.

Prediction	Experiment	Readout
Ordered core + pocket (122–136)	Express the ordered core; biophysics / fragment screen	A well-behaved construct + a hit at the pocket
∼29% disorder map	Limited proteolysis / truncation series	Confirmed foldable boundaries
Pocket is separate from the Aβ domain	Pocket-mutant vs Aβ-binding assay	Function independent of Aβ binding
Aβ-nucleation via res 1–80 [3]	N-terminal-domain constructs ± Aβ	Confirm the nucleation interface

Its role in Aβ plaque nucleation [3] indicates where a validated modulator would first be worth evaluating.

6 Where This Fits In A Discovery Workflow

Team	What they get from this profile
Target biology	Testable residue-level hypotheses and a validation plan
Structural biology	The foldable core, construct boundaries, and a targetable pocket
Medicinal chemistry	A high-confidence pocket for docking and fragment work
BD / platform evaluation	Evidence Orbion can prioritise dark targets quickly

7 Methods

All predictions were generated with Orbion's Astra suite from the canonical FAM222A sequence (UniProt Q5U5X8 [1]). Modules used, with method publications cited where available: **AstraSUIT2** and **AstraROLE2** [5] (membrane and functional class, oligomeric state), **AstraUNFOLD** (topology, disorder, aggregation), **AstraPTM2** [6] (modification sites) and **AstraBIND** [7] (binding pocket). Reported values are model outputs; model internals are out of scope. Confidence scores are model-estimated probabilities (0–1) that rank and gate each call — not calibrated rates of experimental success.

PREDICTION PROVENANCE

Target & input	FAM222A · UniProt Q5U5X8, canonical isoform (452 aa)
Structure source	AlphaFold2 predicted model (no experimental PDB entry; low pLDDT)
Feature inputs	ESM-2 embeddings + AlphaFold2-derived structural features [5–7]
Models	AstraSUIT2 · AstraROLE2 · AstraUNFOLD · AstraPTM2 · AstraBIND
Binding-site model	GATv2 detector trained on >250k protein–ligand complexes [7]
Confidence & calibration	Per-call model probability; per-model calibration, thresholds and benchmarks reported in [5]–[7]

Scope And Limitations

- **Prediction, not experiment.** These are computational hypotheses to prioritise experiments — not a substitute for a structure or an assay. No result here has been validated in the wet lab.
- **The pocket is predicted; no binder is named.** The 122–136 site is a residue-level hypothesis for a foldable, targetable region — not a demonstrated ligandable pocket.

- **Biology caveats.** The in-vivo Aβ / plaque data are from knockdown (AAV-shRNA), not a germline knockout [3]; the therapeutic case is a hypothesis.

8 References

[1] UniProt Consortium. UniProtKB entry Q5U5X8 (FAM222A, human). uniprot.org.

[2] Pharos / Illuminating the Druggable Genome. FAM222A target record — T_{dark}. pharos.nih.gov/targets/Q5U5X8.

[3] Yan T., Liang J., Gao J., et al. FAM222A encodes a protein which accumulates in plaques in Alzheimer's disease. *Nat. Commun.* **11**, 411 (2020). doi:10.1038/s41467-019-13962-0.

[4] Alabdulraheem Z.T.J., Durdagi S. Ab initio and comparative 3D modeling of FAM222A-encoded protein and target-driven virtual screening for Alzheimer's disease. *J. Mol. Graph. Model.* **125**, 108575 (2023). doi:10.1016/j.jmgm.2023.108575.

[5] Bozkurt Ç., Vasilyeva A., Goteti A. AstraROLE2 & AstraSUIT2: Multi-Task Annotation Models for Functional Profiling of Proteins. *bioRxiv* 2025.06.21.660734 (2025). doi:10.1101/2025.06.21.660734.

[6] Bozkurt Ç., Vasilyeva A., Goteti A. AstraPTM2: A Context-Aware Transformer for Broad-Spectrum PTM Prediction. *bioRxiv* 2025.10.03.680341 (2025). doi:10.1101/2025.10.03.680341.

[7] Goteti A., Vasilyeva A., Bozkurt Ç. AstraBIND: Graph Attention Network for Predicting Ligand Binding Sites. *bioRxiv* 2025.11.10.687555 (2025). doi:10.1101/2025.11.10.687555.