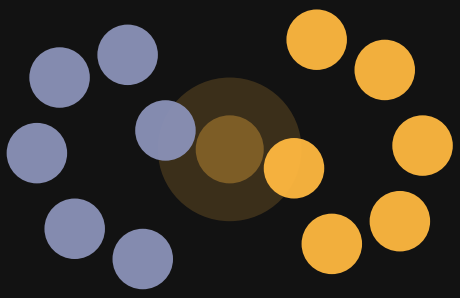
orb::ion



Model Performance Series

Astra AI on Transporters

How Orbion's six AI models perform across 6,203 membrane transport proteins.

AUTHORS

Çağlar Bozkurt[1,*], Aniruddh Goteti[1]

[1]Orbion GmbH, Berlin, Germany

[*]Correspondence: caglar.bozkurt@orbion.life

June 9, 2026 · orbion.life

Executive Summary

Membrane transport proteins move ions, nutrients, neurotransmitters, and drugs across the lipid bilayer. They are among the most important — and most under-exploited — drug-target classes: the solute-carrier (SLC) superfamily alone spans roughly 450 human genes and is the second-largest family of membrane proteins in the human genome [8], the ABC transporters govern multidrug resistance in oncology, and individual carriers such as the serotonin transporter (SERT) are the direct targets of entire drug classes (the SSRIs). Yet transporters are exceptionally hard to characterize computationally: they span a wide range of folds and transmembrane-helix counts, they cycle through multiple conformational states by the *alternating-access* mechanism, and a substantial subset are themselves enzymes (the ATP-driven primary active transporters).

This document reports how the six AI models in Orbion's Astra AI Suite perform on the transporter class. We ran every Astra AI model on **6,203 reviewed transport-associated proteins** from UniProt Swiss-Prot and compared each output against the strongest publicly available experimental reference: Swiss-Prot literature annotation for sequence-level features [4], PDB co-crystal contacts at 4 Å resolution for ligand binding [5], and **197 mutations on the serotonin transporter** with experimentally measured thermal-shift data [10] for stability prediction.

Headline Numbers Across the Six AI Models on Transporters

- **Protein Function and Family Classification.** Canonical transporter families are recognised near-perfectly: SLC carriers 97 %, ABC transporters 100 %, aquaporins 100 %. The model correctly flags the ∼15 % of the cohort that are enzymatic (P-type ATPases, ABC ATPase domains) via the EC head.
- **Topology and Membrane Class.** Transporters span 1–14 transmembrane helices; the model resolves this diversity rather than assuming a fixed architecture, and assigns membrane class (multi-pass / single-pass / peripheral / soluble) across the heterogeneous cohort.
- **Residue-Level Transmembrane Topology.** Per-residue prediction agrees with UniProt-annotated transmembrane segments at AUROC 0.95, AUPRC 0.89, F1 0.86 (n=3,524), uniformly across folds.
- **Post-Translational Modification (PTM) Site Prediction.** On the strongest classes: F1 0.88 for disulfide bonds, 0.83 for N-linked glycosylation, 0.79 for myristoylation. All 39 modification classes covered; two operating points (high-precision, high-recall) reported per class.
- **Ligand Binding Pocket Prediction.** 56 % recall on PDB-observed ligand identities across 833 transporters with co-crystal data; strongest on the nucleotide-binding trafficking GTPases (RAN, Rab, SAR1B; residue F1 up to 0.81), reflecting the model's nucleotide-pocket strength.
- **Thermostability Prediction (ΔT_m).** On 197 mutations of the serotonin transporter, scored against *independent* experimental thermal-shift data: MAE 2.13 °C, Spearman ρ 0.59.

Classification, topology and PTM numbers measure production performance on reviewed real-world transporter targets (their training corpus included); ΔT_m is the independent experimental-validation benchmark (§6).

The transporter class is defined by its heterogeneity, and the results reflect that honestly. Where a family is structurally well-defined — SLC carriers, ABC transporters, aquaporins — recognition and topology prediction are excellent. The broad Swiss-Prot transport keyword also captures enzymatic ATPases and trafficking GTPases whose dominant function the model assigns elsewhere; we report this directly rather than averaging it away. Binding-pocket prediction is strongest where the ligand is a nucleotide, and thermostability prediction is conservative on extreme stabilizers. Each of these is documented in §4.

The centrepiece of this paper is §3, where we walk every Astra AI model through one transporter — the serotonin transporter (SLC6A4, UniProt P31645), the target of the SSRI antidepressants — and show what a program team would do with the integrated output.

What This Document Is. A performance report on Orbion's Astra AI Suite on the transporter class. Every number traces to a checked-in source-data artifact. The classification, topology and PTM metrics characterise how the deployed models behave on this class in production — including proteins drawn from their training corpora; held-out generalisation is reported in the per-model preprints. Thermostability (ΔT_m) is the exception, evaluated on independent experimental data (§6).

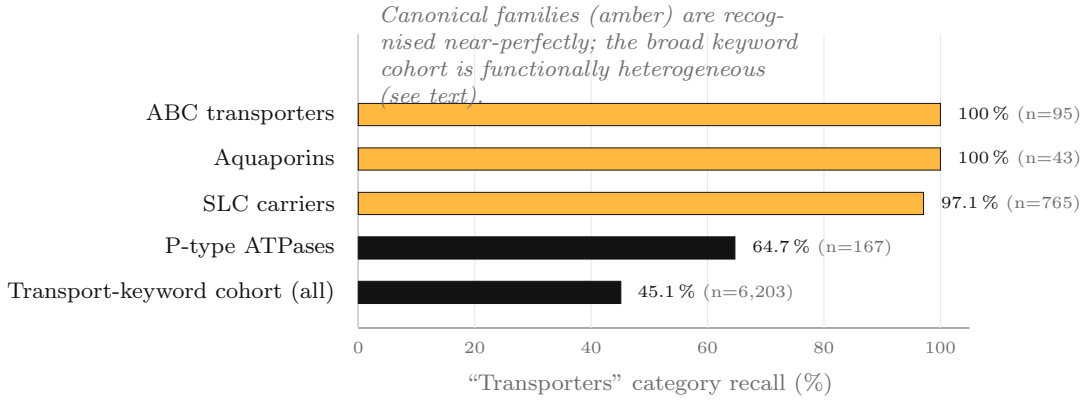


Figure 1: *Family recognition across the transporter cohort. Canonical families (amber) — SLC, ABC, aquaporins — are recognised as transporters near-perfectly; P-type ATPases score lower because the model correctly identifies many as enzymes; the broad transport-keyword cohort is functionally heterogeneous. Detail in §2.1.*

1 Why Transporters

Membrane transport proteins are the gatekeepers of the cell. They set ion gradients, import nutrients, clear neurotransmitters, extrude xenobiotics, and — inconveniently for drug developers — determine the pharmacokinetics and resistance profile of many therapeutics. The solute-carrier (SLC) superfamily is the second-largest family of membrane proteins in the human genome (∼450 genes) [8] and an increasingly prominent target class in drug discovery [9]; the ATP-binding-cassette (ABC) transporters drive multidrug resistance in oncology; and single carriers anchor whole therapeutic areas — SERT, DAT and NET for antidepressants and ADHD, the glucose transporters for metabolic disease, the bile-acid and urate carriers for liver and gout indications. Despite this, transporters remain comparatively under-represented in the structural and computational literature relative to their therapeutic importance [8].

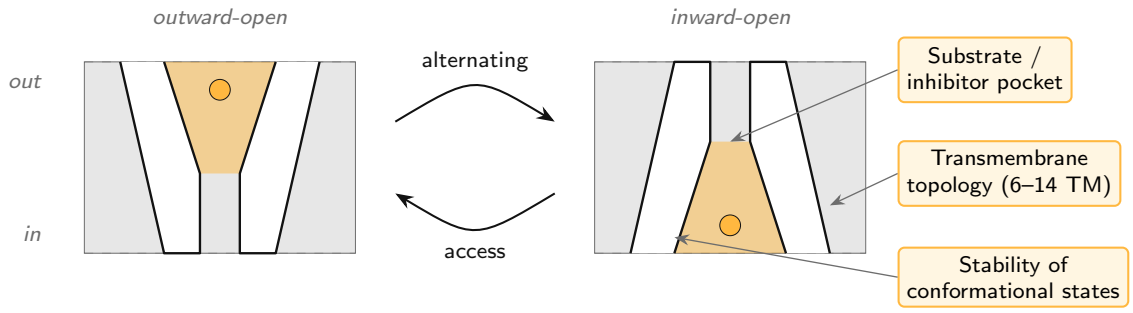


Figure 2: *The alternating-access mechanism that defines the transporter class. A carrier cycles between outward-open and inward-open conformations to move substrate across the membrane, exposing its binding site to one side at a time. The prediction tasks the Astra suite addresses — the substrate / inhibitor pocket, the transmembrane topology (which ranges from 6 to 14 helices across the class), and the stability of the conformational states — are annotated at right.*

Structural diversity. Unlike the receptor classes, transporters do not share a single architecture. SLC carriers fold into 6-, 10-, 12- or 14-transmembrane-helix bundles built from inverted structural repeats; ABC transporters pair transmembrane domains with cytoplasmic nucleotide-binding domains; P-type ATPases add a phosphorylation and actuator domain; aquaporins form compact six-helix channels. A model for this class cannot assume a fixed topology — it must resolve a wide and uneven range of folds.

The alternating-access mechanism. Transport is not a static binding event but a conformational cycle. A carrier alternates between outward-open and inward-open states, exposing its substrate site to one side of the membrane at a time (Figure 2). This conformational plasticity makes both pocket prediction and stability engineering harder: the relevant pocket and the relevant stability depend on which state the protein is in.

A functional spectrum from facilitator to enzyme. Secondary active and facilitative carriers move substrates down or coupled to electrochemical gradients and carry no catalytic activity. Primary active transporters — the P-type ATPases and the ATP-hydrolysing ABC pumps — *are* enzymes. Any honest functional-annotation model on this class must therefore place a substantial subset of "transporters" on the enzyme side of the line, and we show in §2.1 that the model does exactly that.

Experimental bottlenecks. Transporters are notoriously difficult to express, stabilise, and crystallise; thermostabilising mutations are routinely engineered to enable structural work, but the mutational landscape is far too large to screen exhaustively by thermal-shift assay. A computational prefilter that ranks candidate mutations is directly useful here — and is the basis of the serotonin-transporter worked example in §3.

Why a unified AI approach. The Astra AI Suite is built on a shared protein-language-model foundation — a common sequence representation feeding task-appropriate model architectures for each prediction problem. The suite was validated on this premise: one consistent input pipeline, evaluated transparently per protein family. The remainder of this document reports how the suite performs on the 6,203 reviewed transport-associated proteins in UniProt Swiss-Prot [4], one model at a time, with the experimental reference data and metrics for each.

2 The Astra AI Suite, Capability by Capability

The Astra AI Suite is six models that share a common protein-language-model foundation (a sequence representation augmented with structural and physicochemical features), each

built on the machine-learning technique best suited to its task — spanning transformer networks, graph neural networks, and regression models. Each output below was evaluated on the 6,203-protein transporter cohort against the strongest publicly available experimental reference: Swiss-Prot literature annotation [4], PDB co-crystal heavy-atom contacts at 4 Å resolution [5], Gene Ontology with ancestor-closure semantic matching [6, 7], and experimentally measured thermal shifts for stability [10]. Methodological details, metric definitions, evaluation scope, and per-model cohorts are in §6.

Across the suite, the models are calibrated to be conservative: where signal is weak they tend to omit a prediction rather than fabricate one, so the dominant failure mode is a missed call rather than a false positive — the safer error for wet-lab triage. For binding-site prediction this property is reinforced by a seven-gate anti-hallucination audit (§2.5).

2.1 Protein Function and Family Classification

What We Predict. From a protein sequence alone, the protein's broad functional category, its Enzyme Commission family, its molecular-function Gene Ontology terms, and its pathway memberships. Evaluated on all 6,203 proteins in the transport cohort.

Headline Numbers. Transporters are a functionally heterogeneous class, and the results are best read per family rather than in aggregate.

- **Canonical families recognised near-perfectly:** SLC carriers 97.1 %, ABC transporters 100 %, aquaporins 100 % assigned the Transporters category.
- **P-type ATPases 64.7 %** — the model assigns many to Enzymes, which is correct: they are ATP-hydrolysing pumps. The EC head predicts a non-zero enzyme probability for **∼15 %** of the cohort, concentrated in exactly these primary-active transporters.
- **GO molecular-function top-5 (ancestor closure): 84.6 %**; top-1: 67.7 %.

Strong and Weak. Across the whole transport-keyword cohort, only 45 % are assigned the Transporters category — but this reflects the breadth of the Swiss-Prot transport keyword, which captures enzymatic ATPases, trafficking GTPases, and accessory proteins whose dominant annotated function is something else, not a failure to recognise bona-fide carriers. On the structurally-defined families that a discovery team actually works on, recognition is at or near ceiling — SLC carriers 97 %, ABC transporters 100 %, aquaporins 100 % (Figure 1). Read the per-family numbers, not the 45 % aggregate.

Use as: production triage on canonical transporter families; the EC head additionally separates primary-active (enzymatic) from secondary/facilitative carriers.

2.2 Topology and Membrane Class

What We Predict. Membrane topology class (multi-pass / single-pass / peripheral / soluble), transmembrane-helix count, and auxiliary biochemical classifications (cofactor, subcellular localization, quaternary structure).

Headline Numbers. The transport cohort is topologically diverse, and the model's output reflects that spread rather than collapsing to a single class: **48 % multi-pass**, 24 % soluble, 19 % peripheral-membrane, 8 % single-pass. Predicted transmembrane-helix counts span the full range from 1–3 helices to >13, with no single dominant bucket —

consistent with the structural breadth of the class (6-TM aquaporins, 12-TM SLC carriers, larger ABC and P-type architectures).

Strong and Weak. This is a genuine point of difference from single-architecture classes: there is no "correct" helix count to score against. The value here is that the model recovers the diversity — it does not force every transporter into one topology — which is what allows the residue-level head below to perform uniformly across folds.

Use as: membrane-class confirmation and topology routing before structural modeling.

2.3 Residue-Level Topology

What We Predict. For each residue, the probability of lipid-bilayer embedding (and, from the same head, intrinsic disorder and amyloidogenicity). Evaluated on the 3,524 transporters with UniProt transmembrane annotations.

Headline Numbers.

- Per-protein **mean AUROC 0.951, AUPRC 0.893**, F1 at threshold-optimal cutoff 0.858, median AUROC 0.978.

Strong and Weak. Transmembrane prediction is strong and, importantly, *uniform across folds*: per-family mean AUROC is 0.94–0.98 across SLC carriers, ABC transporters, aquaporins, and P-type ATPases alike (Figure 4). This is the cleanest result in the suite for transporters — the model maps the membrane-embedded residues of a carrier reliably regardless of its helix count.

Use as: production triage — residue-level topology mapping ahead of construct design or structural work.

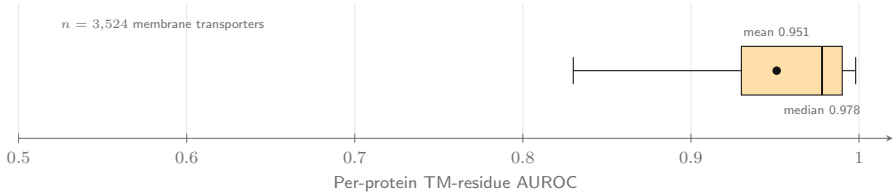


Figure 3: *Distribution of per-protein transmembrane-residue AUROC across the 3,524 transporters with UniProt transmembrane annotations. Mean 0.951, median 0.978.*

2.4 Post-Translational Modification (PTM) Site Prediction

What We Predict. Per-residue probabilities for 39 PTM classes, at two operating points — a high-precision call for confident wet-lab handoff and a high-recall call for hypothesis generation [1]. The dual-output design responds to the incomplete-annotation regime characteristic of PTM literature. Evaluated on the subset of the cohort with $\geq$1 Swiss-Prot-curated positive per modification class; the transporter cohort provides large per-class samples (phosphorylation n=2,636, N-linked glycosylation n=1,648, acetylation n=1,064, disulfide bond n=1,004).

Headline Numbers. Figure 5 reports F1 at the threshold-optimal operating point per modification class. The high-recall operating point dominates the high-precision operating point on every class with sufficient ground truth — most dramatically for the structural and lipidation PTMs: disulfide bond F1 0.88, N-linked glycosylation 0.83, myristoylation 0.79.

Strong and Weak. Disulfide bonds and N-linked glycosylation — the extracellular-loop modifications central to transporter folding and trafficking — are the strongest classes.

	n	Transporters recall	TM-residue AUROC
ABC transporters	95	**1.00**	0.954
Aquaporins (AQP)	43	**1.00**	0.964
SLC carriers	765	**0.97**	0.936
P-type ATPases	167	0.65	0.977
Other transport-keyword	5,018	0.36	0.954

Cell shade ∝ value. Canonical families (ABC, aquaporin, SLC) are recognised as transporters near-perfectly; P-type ATPases score lower because many are correctly identified as enzymes; the broad transport-keyword cohort is functionally heterogeneous. Residue-level transmembrane topology is strong across every family.

Figure 4: *Performance across transporter families. Canonical families (ABC, aquaporin, SLC) are recognised as transporters near-perfectly; P-type ATPases score lower on category because many are correctly identified as enzymes; residue-level transmembrane topology is strong (AUROC 0.94–0.98) across every family.*

	High-precision F1	High-recall F1	
Disulfide bond	0.63	**0.88**	*n=1,004*
N-linked glycosylation	0.44	**0.83**	*n=1,648*
Myristoylation	0.40	**0.79**	*n=81*
Phosphorylation	0.46	0.50	*n=2,636*
Succinylation	0.18	0.44	*n=158*
S-palmitoylation	0.35	0.43	*n=198*
Acetylation	0.24	0.37	*n=1,064*

Cell shade ∝ F1; bold values mark the higher of the two operating points.

Figure 5: *PTM site prediction F1 at the threshold-optimal operating point on transporter Swiss-Prot reference, per modification class, for the high-precision and high-recall operating points.*

Phosphorylation, concentrated in the cytoplasmic N- and C-terminal tails that regulate carrier internalisation, is moderate (F1 0.50 at the high-recall point) and, as in other classes, limited by the model's conservative per-protein output budget on heavily-phosphorylated tails.

Use as: production triage at the high-precision operating point (wet-lab handoff); hypothesis generation at the high-recall operating point.

2.5 Ligand Binding Pocket Prediction

What We Predict. Given a target sequence, a ranked panel of plausible ligand candidates with each ligand's predicted binding-residue set [3]. A seven-gate anti-hallucination audit verifies position validity, confidence bounds, ligand-ID realism, reproducibility, and biological plausibility on every batch. Evaluated on the 833 transporters with both

predictions and PDB co-crystal contacts at 4 Å.

Headline Numbers.

- **Ligand-identity recall: 56 %** — of PDB-observed ligand codes per protein, the model predicts 56 % of them.
- **Pocket-success rate: 49 %** — of model predictions, 49 % overlap at least one PDB-observed ligand-contact residue set.

Residue-level F1 on ligand-ID-matched predictions is 0.27 — the *strict* 4 Å atomic-contact metric. As in the other classes, the suite is built for pocket localisation and ligand-panel triage, not atomic-contact prediction; read the recall and pocket-success rates as the operative numbers.

Strong and Weak. The standout result is on the nucleotide-binding trafficking GTPases — RAN (residue F1 0.81), SAR1B (0.80), and the Rab family (RAB8A 0.79, RAB25 0.76) — which populate the transport cohort because they regulate vesicular trafficking (Figure 6). This is a coherent, architecture-driven strength: AstraBIND2's strongest ligand category is nucleotides, and these proteins present well-defined GTP/GDP pockets. The same logic favours the ATP-binding nucleotide-binding domains of ABC transporters. Conversely, secondary carriers whose substrate site is a transient feature of one conformational state in the alternating-access cycle are harder, and large soluble-domain binding sites are out of scope (§4).

Use as: ligand-panel triage and pocket-hypothesis generation — not accurate pocket mapping across all transporters — strongest for nucleotide-binding transport proteins (GTPases, ABC nucleotide-binding domains).

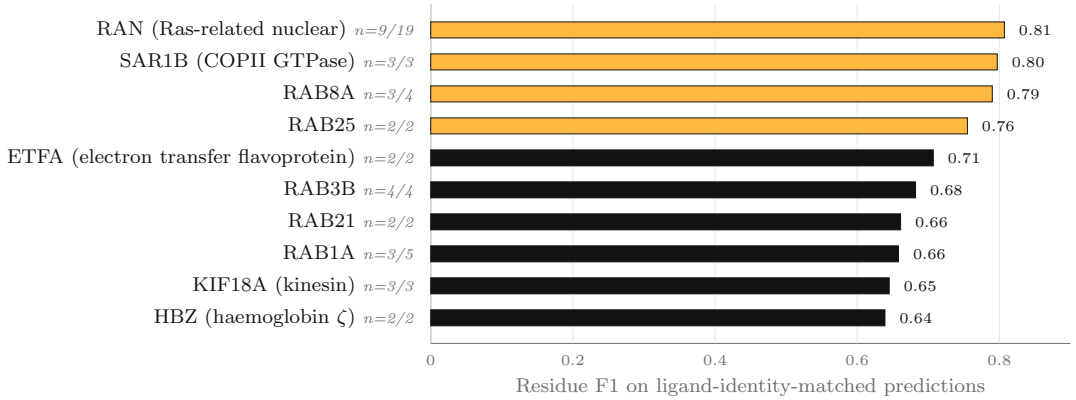


Figure 6: *Top 10 proteins in the transport-associated keyword cohort by AstraBIND2 residue F1 on ligand-identity-matched predictions. The ranking is led by nucleotide-binding trafficking GTPases (RAN, Rab family, SAR1B), reflecting the model's nucleotide-pocket strength. Because the cohort is the broad Swiss-Prot transport keyword (§2.1), it also surfaces nucleotide-binding proteins that are not themselves carriers (e.g. kinesin KIF18A, electron-transfer flavoprotein ETFA) — the same keyword breadth discussed in §4.*

2.6 Thermostability Prediction (ΔT_m)

What We Predict. For a target protein and a point mutation, the change in melting temperature (ΔT_m) of the mutant relative to wild-type, in degrees Celsius, together with a confidence interval. Evaluated on **197 mutations of the serotonin transporter (SLC6A4, P31645)**, the transporter with the most curated thermal-shift entries, using

experimentally measured thermal-shift values from the MPTherm membrane-protein thermodynamics database as ground truth [10]. Unlike the annotation-recall metrics elsewhere in this report, this is an *independent* experimental evaluation: the measured ΔT_m values are not part of the sequence models' training corpus (see §6).

Headline Numbers.

- **MAE 2.13 °C**; Pearson r 0.55; **Spearman ρ 0.59**.
- On the 114 strong-effect stabilizing mutations (experimental $\Delta T_m > 2\,°C$), sign accuracy 81 %; on the small number of strong destabilizers, the directional call is correct.

Note on the Experimental Measurement Floor. The aggregate MAE of 2.13 °C is within the between-replicate reproducibility envelope of the CPM thermal-shift assay against which it is compared (±1.5 °C between replicates [11]). Predictions in the small-effect band ($|\Delta T_m| < 2\,°C$, either direction) are being compared against a measurement whose own resolution is comparable. Performance on strong-effect mutations is markedly higher than the aggregate metrics suggest.

Strong and Weak. Unlike the receptor sets, SERT's experimental landscape is dominated by stabilizing mutations, and the model performs well on them (81 % directional accuracy on n=114). It is most accurate on small-to-moderate effects and least reliable on the extremes — the deep mutational scan includes outliers such as V512F (experimental +8.9 °C, a sign miss at predicted −0.3 °C) and H143A (experimental −18.5 °C, correct sign but compressed magnitude), the documented mean-regression behaviour on extreme effects (Figure 7).

Use as: production pre-screen for ranking thermostabilization mutation panels ahead of a thermal-shift assay; below actionable resolution for small effects ($|\Delta T_m| < 2\,°C$).

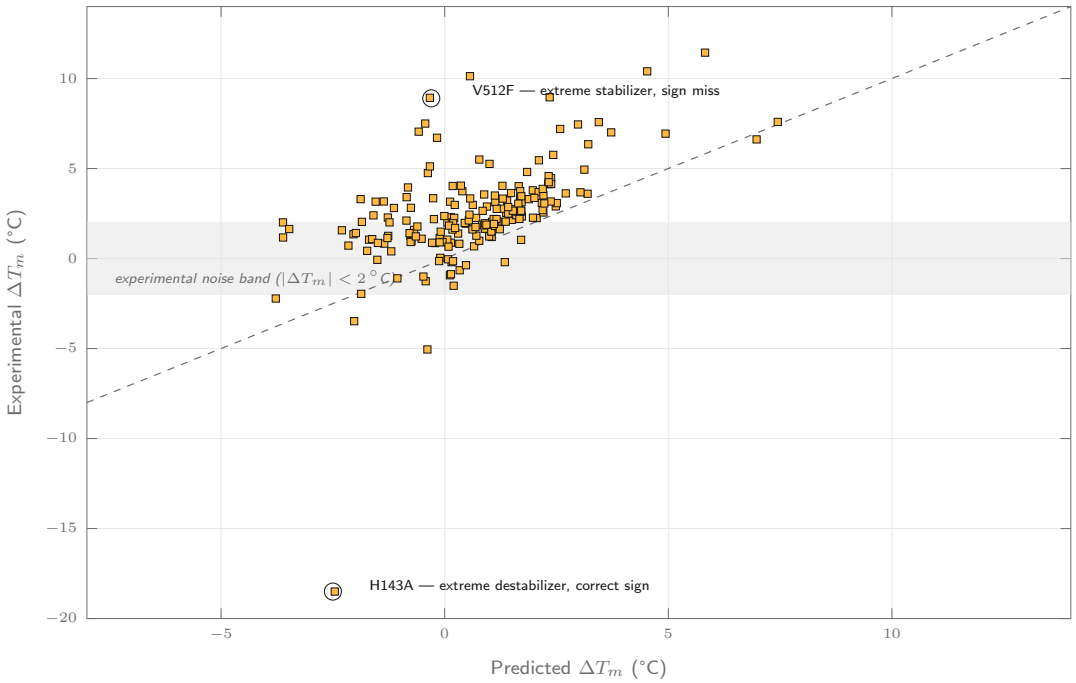


Figure 7: *Predicted vs. experimentally measured ΔT_m across 197 mutations of the serotonin transporter (SLC6A4). The shaded band marks the experimental measurement floor ($|\Delta T_m| < 2\,^\circ C$); the dashed diagonal is the line of perfect prediction. Two extreme-effect outliers are annotated.*

3 Worked Example: The Serotonin Transporter Across the Full Suite

The headline numbers in §2 aggregate across thousands of proteins. In practice, a discovery team works one transporter at a time. This section walks every Astra AI model through a single, exceptionally well-characterized target — the serotonin transporter (SERT; gene SLC6A4, UniProt P31645, 630 residues) — and shows what the integrated output looks like for a real program. SERT clears serotonin from the synapse and is the direct molecular target of the SSRI antidepressants (fluoxetine, sertraline, escitalopram); it adopts the 12-transmembrane LeuT/neurotransmitter-sodium-symporter fold and is one of very few transporters with a deep experimental mutational-stability dataset, which makes it an ideal end-to-end test.

Figure 8 overlays every per-area output on the receptor's structure. The function-and-family classification returns **Transporters** as the top category; the topology classification returns **multi-pass membrane at P = 0.98** and a transmembrane-helix count of **8–12 helices at P = 1.00**, correctly placing SERT in the 12-TM regime. The residue-level head resolves the transmembrane segments of the bundle directly from sequence.

The PTM site prediction returns a biologically coherent set concentrated exactly where SERT is known to be regulated and folded: a **cytoplasmic N-terminal phosphorylation cluster** (positions 2, 3, 7, 12, 32, 39, 43, 46) and **N-terminal ubiquitination sites** (9, 28, 36) — the tail modifications that govern SERT internalisation and turnover — together with the **extracellular-loop-2 disulfide bond** (Cys199–Cys208) and the **N-linked glycosylation site** (Asn216) required for trafficking to the cell surface.

The binding-pocket prediction returns a panel of twelve candidate ligands, including **SRO (serotonin itself)** — the endogenous substrate, recovered directly — alongside a

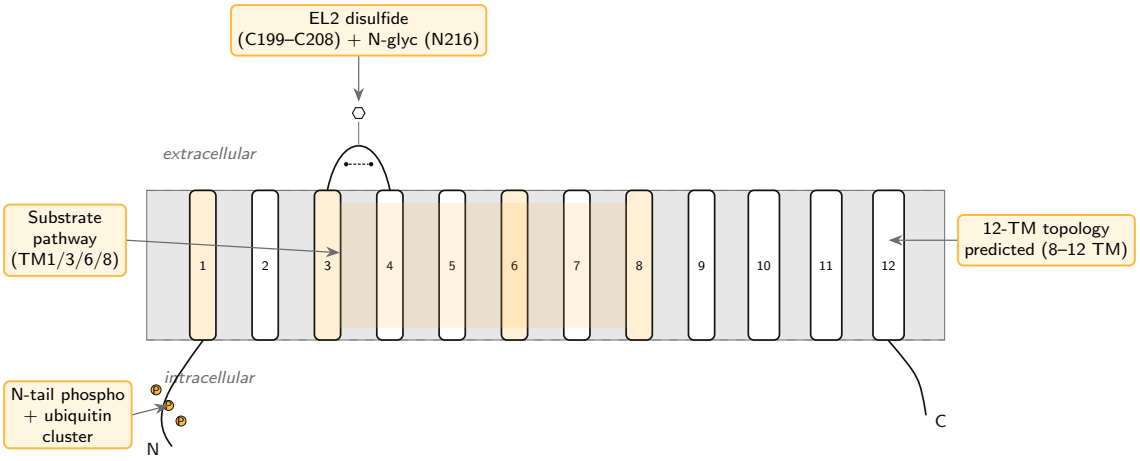


Figure 8: *Integrated view of all model predictions on the serotonin transporter (SLC6A4, P31645, 630 residues). The 12-TM LeuT fold is correctly predicted; core helices (amber) line the central substrate pathway recovered by the binding model; the EL2 disulfide (Cys199–Cys208) and N-glycosylation site (Asn216) are identified on the extracellular side; the cytoplasmic N-terminal tail carries the predicted phosphorylation and ubiquitination cluster that regulates internalisation. All predictions are produced from sequence alone in a single inference pass.*

set of small-molecule inhibitor-class candidates. The predicted contact residues localise to the central substrate pathway formed by the core helices, consistent with the orthosteric site occupied by both substrate and SSRIs.

The thermostability prediction scores SERT's deep experimental thermal-stability dataset (197 mutations) at **MAE 2.13 °C, Spearman ρ 0.59**. The model is most accurate on small-to-moderate effects (V367F: predicted −0.1 °C, experimental −0.1 °C; C473A: +0.7 vs +0.7; I266L: −1.1 vs −1.1) and least reliable on the extremes — V512F, an experimental +8.9 °C stabilizer, is a *sign miss* (predicted −0.3 °C), and H143A, an experimental −18.5 °C destabilizer, is called with the correct sign but compressed magnitude (Figure 7).

What a Program Team Would Do With This Output. A team working SERT — or any SLC carrier — from sequence alone can move directly to a defensible construct design and a wet-lab mutation list:

- Preserve the EL2 disulfide (Cys199–Cys208) and the N-glycosylation site (Asn216) during any engineering — both are required for folding and surface trafficking.
- Mutate the cytoplasmic N-terminal phosphorylation/ubiquitination cluster to probe (or suppress) regulated internalisation.
- Rank thermostabilization candidates by predicted ΔT_m before committing to a thermal-shift screen — the rank correlation (ρ 0.59) makes the prediction a useful triage prioritiser, with the caveat that extreme-magnitude effects are unreliable (the V512F sign miss above).
- Use the recovered substrate/inhibitor panel and its predicted pathway-lining contact residues to seed virtual screening or to focus mutagenesis on the orthosteric site.

The output yields a ∼20–40 variant panel and a construct design ready for wet-lab follow-up, starting from nothing but the sequence.

4 Limits and Operating Envelope

Three operating boundaries are worth making explicit for the transporter class. Each is a scope statement, not a recovery roadmap.

Functional-Category Aggregates Reflect Cohort Breadth. The 45 % whole-cohort "Transporters" category recall in §2.1 should not be read as a 45 % accuracy. The Swiss-Prot transport keyword is deliberately broad — it captures enzymatic P-type ATPases, vesicular-trafficking GTPases, and accessory proteins whose dominant annotated function is enzymatic, regulatory, or structural. The model assigns those to their dominant functional category (and the EC head correctly flags the enzymatic subset). On the structurally-defined carrier families a discovery team actually works on — SLC, ABC, aquaporins — recognition is 97–100 %. Read the per-family numbers, not the aggregate.

Binding-Pocket Prediction Is Strongest for Nucleotide Pockets. The 56 % ligand-identity recall is carried by well-defined, chemically-distinct pockets — above all the nucleotide sites of trafficking GTPases and the ATP-binding domains of ABC transporters (§2.5). The harder case is the secondary-active and facilitative carriers, whose substrate site is a transient feature of one state in the alternating-access cycle rather than a persistent pocket; here both recall and residue precision are lower. Large soluble-domain binding sites remain out of the current model's scope.

Thermostability Operating Envelope — The Experimental Measurement Floor. The SERT MAE of 2.13 °C is within the between-replicate reproducibility envelope of the CPM thermal-shift assay against which it is compared. Predictions in the small-effect band ($|\Delta T_m| < 2$ °C, either direction) are being compared against a measurement whose own resolution is comparable. This is a property of the evaluation, not the model; the rank correlation (ρ 0.59) and the strong-effect directional accuracy are the metrics to cite for production triage. As elsewhere, extreme stabilizers are subject to mean-regression (the SERT V512F example in §3). Finally, this evaluation rests on a single deeply-scanned transporter; broadening ΔT_m validation across additional carriers is the priority next step (§1 contact).

5 Decision Framework: Which Models for Which Question

The Astra AI Suite is most useful when used in combination. The table below maps five common transporter program questions to the prediction-area combinations that address them, the operating points to use, and the expected output for a discovery team. Numbers in parentheses reference the per-area evidence in §2.

Table 1: *Decision framework mapping common transporter program questions to combinations of Orbion's Astra AI Suite. The integrated suite is most valuable when used as a workflow; single-prediction-area use is appropriate for triage but generally leaves information on the table.*

Program Question	Recommended Workflow Across Orbion's Astra AI Suite
1. Thermostabilizing a Carrier for Structure	The **thermostability prediction** ranks candidate point mutations by predicted ΔT_m (rank correlation ρ 0.59 on the SERT benchmark); prioritise the top-ranked candidates for a thermal-shift screen. Cross-check positions against the **residue-level topology prediction** to keep mutations in the transmembrane bundle rather than flexible loops, and against **PTM site prediction** to avoid disrupting folding-critical disulfides or glycosylation sites. A 20–40 variant panel for CPM-assay validation is the typical output.
2. Substrate / Inhibitor Pocket Identification	The **ligand binding pocket prediction** returns a ranked candidate-ligand panel with predicted contact residues — strongest for nucleotide-binding transporters (GTPases, ABC nucleotide-binding domains). Cross-check the predicted pocket residues against the **residue-level topology prediction** to place the site within the substrate pathway. For secondary carriers whose pocket is state-dependent, treat the output as a starting hypothesis and confirm against any available structural state (see §4).
3. ABC / Multidrug-Resistance Transporter Characterisation	The **function and family classification** confirms the ABC assignment and (via the EC head) the ATPase activity; the **binding-pocket prediction** localises the nucleotide-binding site; the **residue-level topology prediction** maps the transmembrane domains. Together these scope the construct boundaries and the catalytic versus transport modules before expression.
4. SLC Disease-Variant Interpretation	Many SLC carriers are Mendelian disease genes. The **thermostability prediction** estimates the stability impact of each missense variant; combine with **PTM site prediction** to flag variants that disrupt trafficking-critical modification sites and the **residue-level topology prediction** to identify variants in functionally critical transmembrane positions. ClinVar or gnomAD variants can be re-ranked by a combined Astra-derived deleteriousness score.
5. Mapping Regulation & Trafficking (PTM Landscape)	The **PTM site prediction** (high-recall operating point) returns the candidate-site set across all 39 modification classes; for transporters, focus on the cytoplasmic-tail phosphorylation and ubiquitination that govern internalisation and the extracellular glycosylation that governs surface delivery. Intersect with the literature-curated set in Swiss-Prot to surface novel candidate sites for mass-spectrometry follow-up.

The recurring pattern across all five workflows is the same: a primary prediction area

produces a candidate set, and one or two complementary areas filter or contextualize the candidates against other protein properties. This is the integration story — Orbion's Astra AI Suite is built so that one sequence-input call returns enough complementary outputs to make multi-criteria decisions in a single pass.

About Orbion

Orbion is an AI-powered protein engineering platform that compresses the workflow from sequence to experiment-ready protocol. The Astra AI Suite described in this document is the prediction layer of the platform — functional annotation, post-translational modification site prediction, ligand binding-site identification, residue-level topology and disorder, and thermostability prediction across ΔT_m and $\Delta\Delta G$. Sitting alongside it on the same platform are AlphaFold2 / AlphaFold-Multimer structure prediction, a construct engineering workspace with composite scoring against an organization's expression vector library, automated experimental protocol generation across expression / purification / crystallization / cryo-EM / stabilization, and a mutation engine for thermostabilization campaigns. Orbion was founded in 2024, is headquartered in Berlin, and works with contract research organizations, pharmaceutical and biotech teams, and academic groups. Free access is available to academic users with a partnership agreement.

Platform. app.orbion.life **Web.** orbion.life **Contact.** `contact@orbion.life` **Academic access.** orbion.life/researcher-program

6 Methods

Cohort. 6,203 reviewed transport-associated proteins from UniProt Swiss-Prot [4] by the transport keyword (KW-0813), all organisms, de-duplicated against sequence-integrity checks. The keyword is deliberately broad and the cohort is functionally heterogeneous (see §2.1); per-area evaluation subsets are stated inline in each §2 subsection. Family assignments (SLC, ABC, P-type ATPase, aquaporin) are by gene-name pattern.

Experimental Reference. UniProt feature table for sequence-level features (transmembrane segments, modified residues, glycosylation sites, disulfide bonds, lipidation, regions). Gene Ontology molecular-function annotations expanded to the is_a / part_of ancestor closure via QuickGO [6, 7]. PDB ligand-contact reference (§2.5): heavy-atom contacts at 4.0 Å between each ligand and protein residue, aggregated across all deposited co-crystal structures per UniProt accession [5]. ΔT_m reference (§2.6): experimentally measured thermal-shift (ΔT_m) values for 197 serotonin-transporter point mutations, drawn from the MPTherm membrane-protein thermodynamics database [10].

Metrics. AUROC, AUPRC [12], F1 at threshold-optimal cutoff, recall at $k = \mathrm{n}_{\mathrm{positives}}$, Brier score, MAE in degrees Celsius, Pearson and Spearman correlations, sign accuracy (computed on the subset of mutations with $|\Delta T_m| > 0.5\,^{\circ}\mathrm{C}$). Per-protein metrics aggregated by mean; distributions reported where shape is informative. The 5 GB per-protein prediction cache was processed by streaming (one protein at a time) to keep memory bounded.

Evaluation Scope. The classification, topology, and PTM metrics in §2.1–§2.4 measure how the deployed Astra models perform on the transporter class as they are run in production. The cohort is drawn from the reviewed proteome, which overlaps the corpora these models were trained on; the figures therefore characterise within-distribution behaviour on real targets rather than held-out generalisation. Held-out

test-split performance for each model is reported in its respective preprint [1–3]. The contribution of this report is the per-family transporter performance breakdown and the integrated single-protein workflow (§3). The thermostability evaluation (§2.6) is the exception: it scores predictions against an independent experimental thermal-shift dataset that is not part of the sequence models' training corpus, and so reflects genuine predictive generalisation.

Reproducibility. Every aggregate number traces to a checked-in source-data artifact in Orbion's evaluation repository, versioned alongside the published methods [1–3].

References

[1] 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.

[2] 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.

[3] 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.

[4] The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, **51**(D1):D523–D531 (2023). doi:10.1093/nar/gkac1052.

[5] Burley, S. K. *et al.* RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, **47**(D1):D464–D474 (2019). doi:10.1093/nar/gky1004.

[6] Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, **25**(22):3045–3046 (2009). doi:10.1093/bioinformatics/btp536.

[7] Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**:25–29 (2000). doi:10.1038/75556.

[8] César-Razquin, A. *et al.* A Call for Systematic Research on Solute Carriers. *Cell*, **162**(3):478–487 (2015). doi:10.1016/j.cell.2015.07.022.

[9] Lin, L., Yee, S. W., Kim, R. B., Giacomini, K. M. SLC transporters as therapeutic targets: emerging opportunities. *Nature Reviews Drug Discovery*, **14**(8):543–560 (2015). doi:10.1038/nrd4626.

[10] Kulandaisamy, A., Sakthivel, R., Gromiha, M. M. MPTherm: database for membrane protein thermodynamics for understanding folding and stability. *Briefings in Bioinformatics*, **22**(2):2119–2125 (2021). doi:10.1093/bib/bbaa064.

[11] Alexandrov, A. I. *et al.* Microscale fluorescent thermal stability assay for membrane proteins. *Structure*, **16**(3):351–359 (2008). doi:10.1016/j.str.2008.02.004.

[12] Davis, J., Goadrich, M. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240 (2006). doi:10.1145/1143844.1143874.