ChiralCall Technical Validation Supplement

Eutomer Prediction Accuracy by Compound Class | Known-Compound Validation Dataset

49	1,197	98.5%	[97.6, 99.0]	1051
Compound Classes	Validated Compounds	Overall Accuracy	Wilson 95% CI	Known Compounds

1. Definitions & Scope

Eutomer. For a given chiral compound, the *eutomer* is the enantiomer exhibiting the greater pharmacological activity at the primary therapeutic target. ChiralCall predictions identify the historically favored eutomer reflecting dominant pharmacological activity from peer-reviewed literature. Where multiple targets exist, the eutomer designation follows the primary indication documented in the source literature.

Known-Compound Validation. Each compound in the validation dataset has a literature-confirmed eutomer assignment. ChiralCall's prediction for each compound is compared against this ground truth. A prediction is scored as correct when ChiralCall's predicted eutomer matches the published assignment.

Scope. The current validation dataset covers 49 compound classes: 25 validated (production-ready) and 24 exploratory (under evaluation). Classes span pharmaceutical (36 classes) and agricultural (13 classes) chemistry.

2. Dataset Provenance

Validation compounds are drawn from two primary sources:

- **ChEMBL (version 36)** — Stereo-defined bioactivity records from the European Molecular Biology Laboratory's curated database. Compounds were selected via substructure search, filtered for defined stereochemistry, and verified to have differential enantiomer bioactivity data.
- **Published literature** — Primary research articles, review papers, and pharmacology textbooks documenting stereoselective activity. Each compound's eutomer assignment is traceable to at least one peer-reviewed source.

CIP (Cahn-Ingold-Prelog) labels (R/S) are computed from SMILES representations using RDKit. All SMILES used in production are matched via InChIKey (stereo-stripped first layer) to ensure notation-independent compound identification.

3. Compound Class Assignment Methodology

ChiralCall employs a deterministic, first-principles scaffold classifier. The prediction engine is **not** a machine-learning model: it uses no trained weights, no neural networks, and no statistical fitting. Instead, predictions are derived from the molecular structure itself through a three-pass matching algorithm:

- **Pass 1 (Exact Match):** Query compound is matched against the known-compound database (1051 entries) via InChIKey. If found, the literature-confirmed eutomer is returned directly.
- **Pass 2 (Scaffold Classification):** If no exact match, the SMILES string is tested against scaffold keys for each compound class. When a match is found, the class-level prediction rule is applied to determine the eutomer.
- **Pass 3 (Out of Scope):** Compounds matching no scaffold are reported as out-of-scope rather than producing a low-confidence guess.

This deterministic approach ensures that identical inputs always produce identical outputs, independent of training data or model version. The methodology is fully reproducible.

4. Confidence Tier System

Each compound class is assigned a confidence tier based on validation performance. Wilson binomial confidence intervals are used to account for small-sample uncertainty:

Tier	Criteria	Count	Interpretation
T1	N ≥ 10, Acc ≥ 90%, Wilson CI lower > 80%	**25**	High confidence — production-ready
T2	N ≥ 6, Acc ≥ 90%, Wilson CI lower > 54%	**14**	Moderate confidence — expanding validation
T3	Below T2 thresholds	**10**	Early-stage — limited validation data

5. Per-Compound-Class Validation Summary

Table below shows all 49 compound classes sorted by validation sample size (N). Wilson 95% confidence intervals are computed for each class independently.

Compound Class	N	Correct	Wrong	Accuracy	Tier	CI Low	CI High
Amino-alcohol (other) — REDACTED modular	195	186	9	**95.4%**	T1	91.5%	97.6%
Catecholamine (3,4-dihydroxy)	84	78	6	**92.9%**	T1	85.3%	96.7%
Aryloxy-propanolamine β-blocker (N-tert-butyl)	83	83	0	**100.0%**	T1	95.6%	100.0%
Aryl-ethanolamine β-agonist (non-saligenin)	77	76	1	**98.7%**	T1	93.0%	99.8%
Tertiary carbinol triazole (spectral centre selection)	73	72	1	**98.6%**	T1	92.6%	99.8%
Benzamide antipsychotic	70	70	0	**100.0%**	T1	94.8%	100.0%
Morphinan opioid antagonist	69	69	0	**100.0%**	T1	94.7%	100.0%
Classical 4-aminoquinoline antimalarial	49	49	0	**100.0%**	T1	92.7%	100.0%
Arylpropionic acid (profen NSAID)	44	44	0	**100.0%**	T1	92.0%	100.0%
Pyrrolidinone SV2A ligand	25	25	0	**100.0%**	T1	86.7%	100.0%
ACE inhibitor (proline-scaffold angiotensin-converti...	25	25	0	**100.0%**	T1	86.7%	100.0%
JAK inhibitor (7H-pyrrolo[2,3-d]pyrimidine, R-active)	24	24	0	**100.0%**	T1	86.2%	100.0%
Axial biaryl atropisomer	21	21	0	**100.0%**	T1	84.5%	100.0%
8-Aminoquinoline antimalarial	20	20	0	**100.0%**	T1	83.9%	100.0%
Serotonin-norepinephrine reuptake inhibitor	19	19	0	**100.0%**	T1	83.2%	100.0%
Morphinan opioid agonist	18	18	0	**100.0%**	T1	82.4%	100.0%
Dihydropyridine CCB (C4 ring-chiral)	18	18	0	**100.0%**	T1	82.4%	100.0%
Pyrimidine nucleoside analog (uracil/thymine class)	18	18	0	**100.0%**	T1	82.4%	100.0%
ACCase herbicide (aryloxyphenoxy-propionate)	17	17	0	**100.0%**	T1	81.6%	100.0%
Proton pump inhibitor (sulfoxide-chiral benzimidazole)	17	17	0	**100.0%**	T1	81.6%	100.0%
Dihydropyridine CCB (chiral ester side chain)	16	16	0	**100.0%**	T1	80.6%	100.0%
Coumarin anticoagulant	16	16	0	**100.0%**	T1	80.6%	100.0%
PARP inhibitor (chiral amine class)	16	16	0	**100.0%**	T1	80.6%	100.0%
Propylamine antihistamine (pyridyl-aryl class)	16	16	0	**100.0%**	T1	80.6%	100.0%
PI3K inhibitor (purine/pyrazolopyrimidine, S-active)	16	16	0	**100.0%**	T1	80.6%	100.0%
ALK inhibitor (chiral amine class)	11	11	0	**100.0%**	T2	74.1%	100.0%
Arylcyclohexylamine dissociative (ketamine class)	11	11	0	**100.0%**	T2	74.1%	100.0%
Pipecoloxylidide anesthetic (N≥C3-alkyl sub-class)	10	10	0	**100.0%**	T2	72.2%	100.0%
Oxazolidinone antibiotic (C5-chiral)	10	10	0	**100.0%**	T2	72.2%	100.0%
Cephalosporin β-lactam antibiotic (6R,7R-active)	10	10	0	**100.0%**	T2	72.2%	100.0%
Pyrethroid insecticide (cyclopropane 1R-active)	9	9	0	**100.0%**	T2	70.1%	100.0%
Amino acid sulfoxide (methionine class)	8	8	0	**100.0%**	T2	67.6%	100.0%
Fluoroquinolone antibiotic (chiral quinolone)	8	8	0	**100.0%**	T2	67.6%	100.0%
Carbapenem β-lactam antibiotic (5R,6S-active)	8	8	0	**100.0%**	T2	67.6%	100.0%
Triazole fungicide (1,3-dioxolane core)	7	7	0	**100.0%**	T2	64.6%	100.0%
Triazole fungicide (open-chain carbinol)	7	7	0	**100.0%**	T2	64.6%	100.0%
Triazole fungicide (fused ring / epoxide)	6	6	0	**100.0%**	T2	61.0%	100.0%
α-Carbon triazole fungicide	6	6	0	**100.0%**	T2	61.0%	100.0%
3-Hydroxy-1,4-benzodiazepine (oxazepam class)	6	6	0	**100.0%**	T2	61.0%	100.0%
Methoxyacetyl acylalanine fungicide	4	4	0	**100.0%**	T3	51.0%	100.0%

Compound Class	N	Correct	Wrong	Accuracy	Tier	CI Low	CI High
Phenylacetyl acylalanine fungicide	4	4	0	**100.0%**	T3	51.0%	100.0%
Ryanodine receptor diamide insecticide	4	4	0	**100.0%**	T3	51.0%	100.0%
Neonicotinoid insecticide	4	4	0	**100.0%**	T3	51.0%	100.0%
Chloroacetanilide herbicide (metolachlor class)	4	4	0	**100.0%**	T3	51.0%	100.0%
Imidazolinone herbicide (imazethapyr class)	4	4	0	**100.0%**	T3	51.0%	100.0%
Aryloxyphenoxypropionic acid herbicide (FOP class)	4	4	0	**100.0%**	T3	51.0%	100.0%
Thienopyridine antiplatelet (clopidogrel class)	3	3	0	**100.0%**	T3	43.8%	100.0%
1,5-Benzothiazepine calcium channel blocker (diltiaz...	2	2	0	**100.0%**	T3	34.2%	100.0%
Pipecoloxylidide anesthetic (N-methyl sub-class)	1	0	1	**0.0%**	T3	0.0%	79.3%
TOTAL / OVERALL	**1197**	**1179**	**18**	**98.5%**		**97.6%**	**99.0%**

6. Wrong Prediction Analysis

Of 1,197 validated compounds, 18 predictions (1.5%) did not match the literature-confirmed eutomer. Wrong predictions are concentrated in 5 compound classes:

- **Amino-alcohol (other) — REDACTED modular**: 9 of 195 wrong (95.4% accuracy, Wilson CI [91.5%, 97.6%])
- **Catecholamine (3,4-dihydroxy)**: 6 of 84 wrong (92.9% accuracy, Wilson CI [85.3%, 96.7%])
- **Aryl-ethanolamine β-agonist (non-saligenin)**: 1 of 77 wrong (98.7% accuracy, Wilson CI [93.0%, 99.8%])
- **Tertiary carbinol triazole (spectral centre selection)**: 1 of 73 wrong (98.6% accuracy, Wilson CI [92.6%, 99.8%])
- **Pipecoloxylidide anesthetic (N-methyl sub-class)**: 1 of 1 wrong (0.0% accuracy, Wilson CI [0.0%, 79.3%])

Failure analysis indicates that wrong predictions cluster in compound classes with high structural diversity within the class. Amino-alcohol (other) accounts for 9 of the 18 errors, reflecting the broad structural heterogeneity of this catch-all class. Ongoing sub-class decomposition is expected to reduce these errors as more specific scaffold patterns are identified.

7. Statistical Methods

Wilson score interval. Confidence intervals are computed using the Wilson score method for binomial proportions (Wilson, 1927). This method is preferred over the Wald (normal approximation) interval because it provides accurate coverage even for small sample sizes and proportions near 0 or 1.

Tier boundaries. The Wilson CI lower bound thresholds (80% for T1, 54% for T2) are chosen to reflect practical confidence levels for decision support. T1 classes have sufficient evidence that even the pessimistic bound exceeds 80% accuracy. T2 classes exceed chance performance with statistical significance.

8. References

[1] Mendez, D. et al. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1), D930–D940.

[2] Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209–212.

[3] Landoni, M.F. & Soraci, A.L. (2001). Pharmacology of chiral compounds: 2-arylpropionic acid derivatives. *Current Drug Metabolism*, 2(1), 37–51.

[4] Hutt, A.J. & Tan, S.C. (1996). Drug chirality and its clinical significance. *Drugs*, 52(Suppl 5), 1–12.

[5] Cahn, R.S., Ingold, C. & Prelog, V. (1966). Specification of molecular chirality. *Angewandte Chemie International Edition*, 5(4), 385–415.

[6] Landrum, G. (2023). RDKit: Open-source cheminformatics. https://www.rdkit.org.

This document is provided for evaluation purposes under mutual NDA. Methodology details beyond what is described here are proprietary and trade-secret protected. For questions, contact: pilot@chiralcall.com