ChemAxon logP Predictor – Published Experimental Benchmarks

Why these benchmarks

The octanol/water partition coefficient (logP) of the neutral species is a foundational descriptor in drug discovery. It influences solubility, permeability, distribution, protein binding, and metabolic stability, and underpins models for logD, PAMPA, and CNS penetration.

A logP predictor is useful only if it generalizes beyond curated showcase sets. We evaluated the latest Chemaxon logP model on two independent, publicly available experimental references: one large-scale, multi-source aggregate, and one compact, pharma-curated drug set. Together they span two orders of magnitude in dataset size and reflect different curation approaches.

Data set processing

Environment

- cxcalc CLI
- v25.5.0

① Wang et al. (2023) aggregate

Public collection assembled for the RTlogD study, built from four independent sources, normalized and deduplicated to 13,688 unique molecules. The authors describe these values as experimental logP.

Statistical metrics

n = 13,688	Pearson	R^2	MAE	RMSE
v25.5.0	0.893	0.788	0.553	0.860

Ranged delta distribution

	Occurrence	
$\Delta < 0.25$	4,898	36%
$0.25 \leq \Delta < 0.5$	3,532	26%
$0.5 \leq \Delta < 1$	3,345	24%
$1 \leq \Delta \leq 2$	1,526	11%
$\Delta > 2$	387	3%

② Saaidpour (2014) drug set

Compact, pharma-focused benchmark of 150 experimental logP values assembled by Saaidpour for QSPR/ANN modelling. The set spans -1.85 to 6.03 logP units and covers a diverse drug-like and non-drug-like cross-section.

Statistical metrics

n = 150	Pearson	R^2	MAE	RMSE
v25.5.0	0.914	0.773	0.522	0.738

Ranged delta distribution

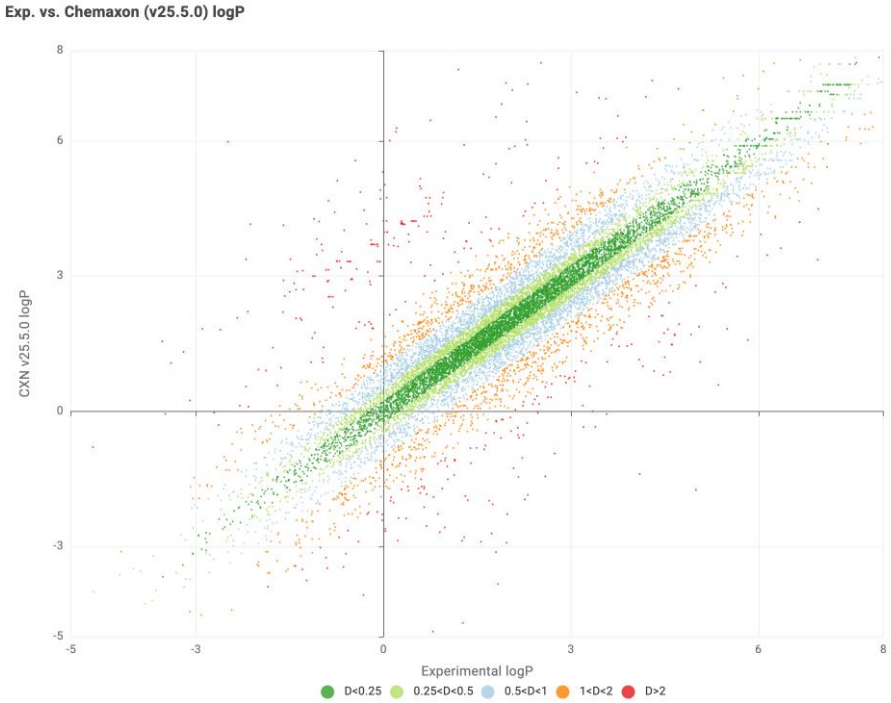
	Occurrence	
$\Delta < 0.25$	58	39%
$0.25 \leq \Delta < 0.5$	36	24%
$0.5 \leq \Delta < 1$	36	24%
$1 \leq \Delta \leq 2$	14	9%
$\Delta > 2$	6	4%

Bottom line

- **Consistent performance** across two reference sets with distinct curation of philosophies: a large multi-source aggregate and a compact pharma-curated drug compilation.
- **Strong correlation** in both sets (Pearson 0.89–0.91, $R^2 \approx 0.78$) and **sub-unit mean error** (MAE 0.52–0.55, RMSE 0.74–0.86).
- **Outlier-resistant:** Only 3–4% of predictions deviate by more than 2 logP units, even on the 13.7k-compound aggregate containing non-drug-like chemistry.
- **Fit for purpose** for ranking analogs, lipophilicity flagging in virtual libraries, ADME triage, and as an upstream descriptor for logD7.4 and solubility models.

Scatter Plots

Wang et al. (2023) aggregate



Saaidpour (2014) drug set

