Monroe

A Molecular Foundation Model for In-Context Probabilistic Inference

Błażej Banaszewski
Graphcore
Max Planck Institute of Biochemistry

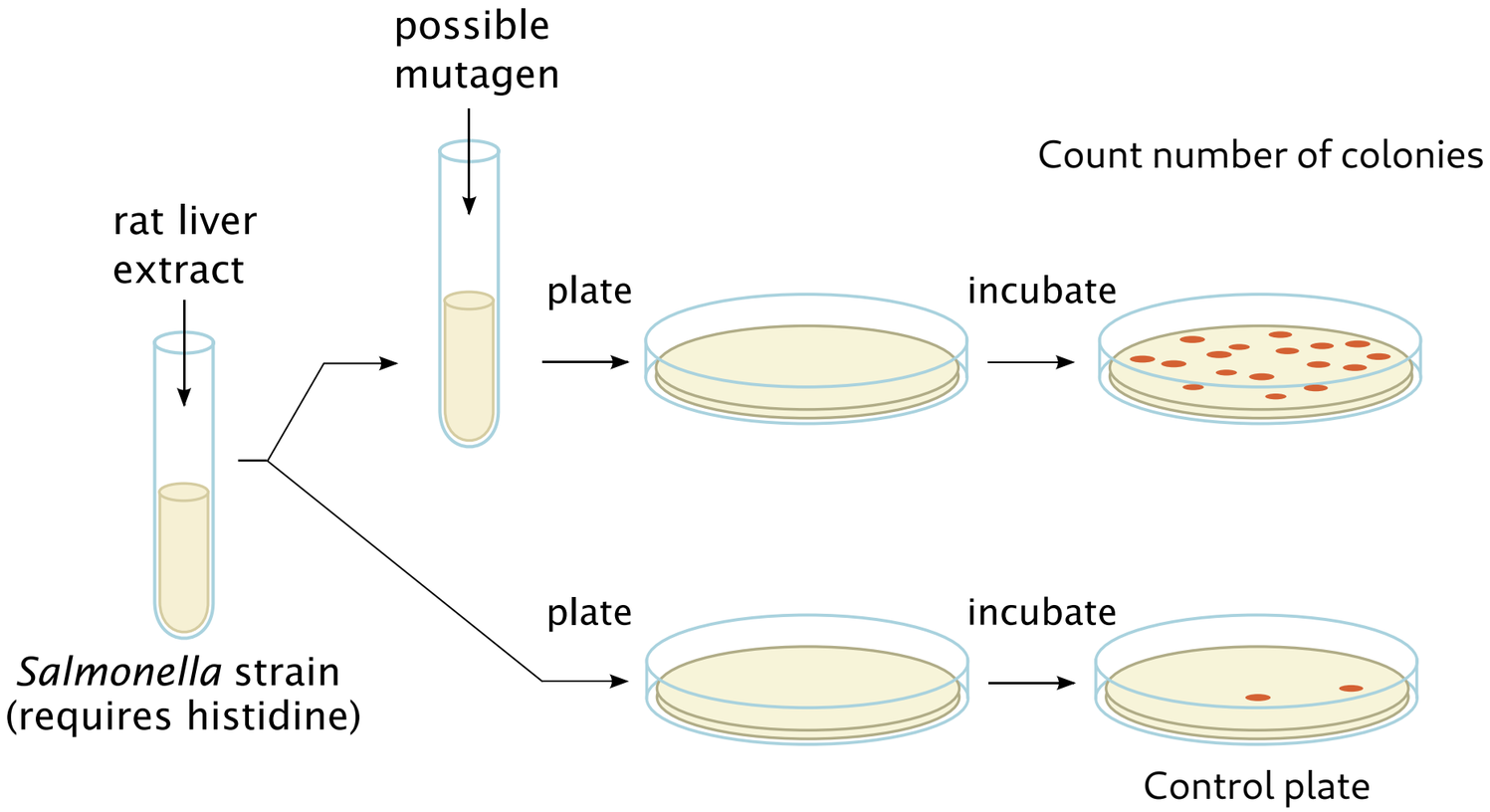
Andrew W. Fitzgibbon
Graphcore
University of Cambridge

What is molecular property prediction?

Does this compound cause mutations?

Compare bacterial colonies with and without the compound.

More colonies can indicate more reverse mutations.



Wikipedia: Bioassay

Why is molecular property prediction challenging?

A Datasets are small

Labelled molecules per task

Fang HPPB	**160**
DILI	**475**

B Search space is vast

166 billion

enumerated molecules in GDB-17

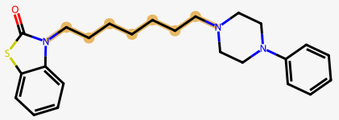
At most 17 heavy atoms

C Extrapolation is common

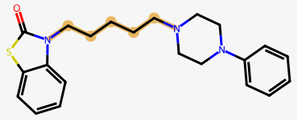
Caco-2 efflux ratio, ExpansionRx

Train max	**105.64**
Test max	**607.29**

D Activity landscape is rugged



0.27 nM



95.66 nM

354× affinity difference 5-HT1A, Ki

E No universal context

Same molecule, different question

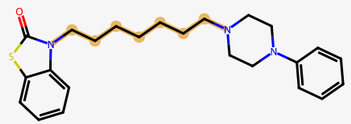
Assay	**Ames**	**Lipophilicity**
Species	**Human**	**Mouse**

Prediction = f(molecule, context)

A: Polaris datasets B: GDB-17, Ruddigkeit et al. (2012) C: OpenADMET ExpansionRx D: MoleculeACE

Learn the new task from examples

Molecules



Labelled examples
+ query molecule

Monroe
graph encoder
FROZEN

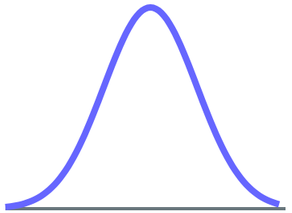
Embedding table

Features	Label
	y1
	y2
	y3
	?

720 features per molecule

TabPFN
in context

Prediction

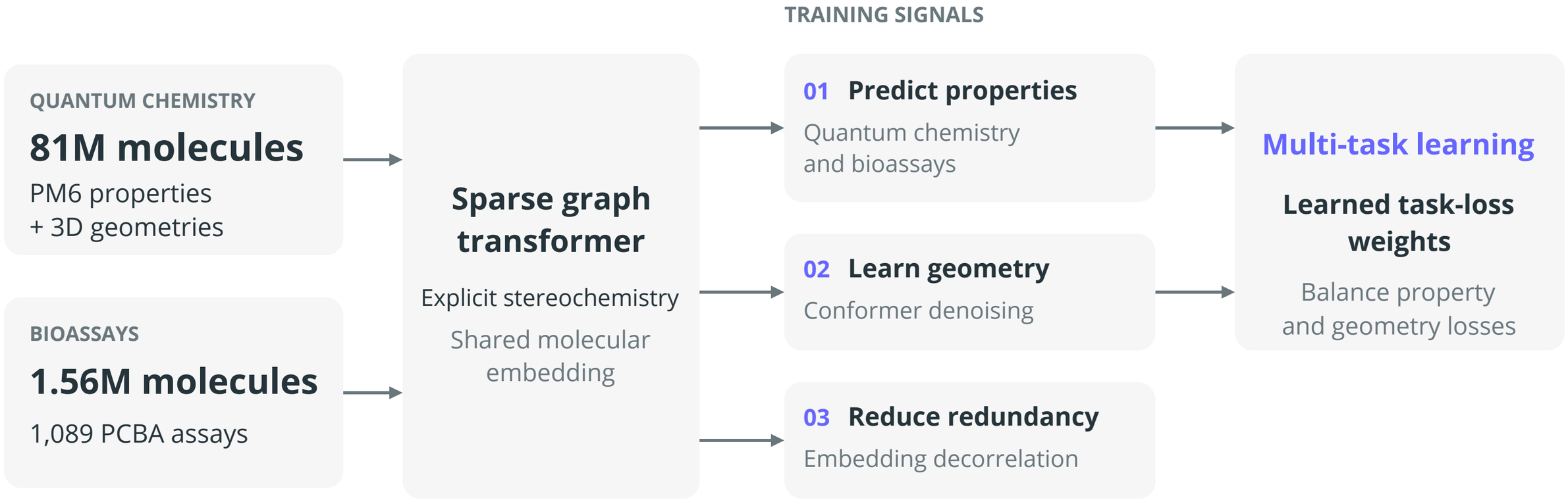


Predictive distribution

No task-specific gradient updates in Monroe's downstream pipeline

TabPFN: Hollmann et al., Nature 2025.

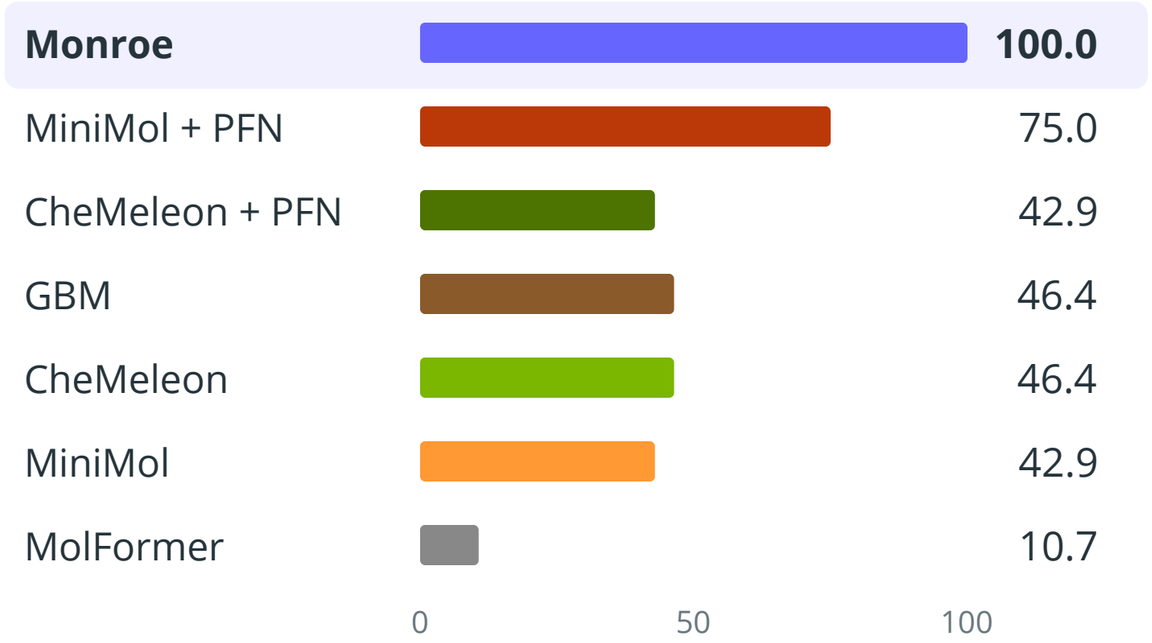
Teach the encoder chemistry before the new assay



Strong across assays, including activity cliffs

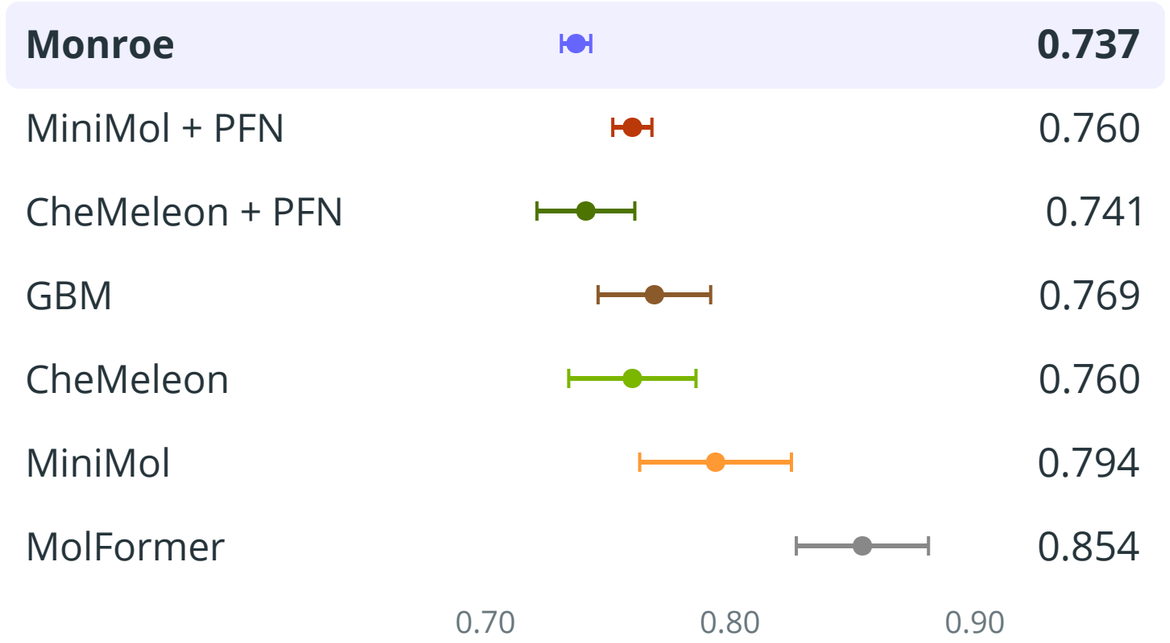
Polaris (28 tasks)

Win or Tie (%) ↑



MoleculeACE (30 tasks)

Activity Cliff RMSD ↓



Polaris: Tukey HSD + Benjamini–Hochberg correction, FDR 5%. Non-significance does not establish equivalence.
Five seeds per model and task. MoleculeACE bars show mean standard deviation across seeds.

The downstream predictor matters

MiniMol

Frozen pretrained encoder

Native **0.703** → + TabPFN **0.642**

8.7% lower mean RMSE

CheMeleon

Task-finetuned encoder for PFN

Native **0.662** → + TabPFN **0.641**

3.2% lower mean RMSE

MoleculeACE overall RMSE, averaged over 30 tasks. Lower is better.

Molecular pretraining + in-context prediction

01 **Learn a transferable molecular representation**

Quantum chemistry, bioassays and structural supervision

02 **Condition on the new assay's labelled molecules**

Frozen Monroe embeddings + in-context prediction with TabPFN

03 **Test both broad transfer and local activity changes**

Strong results on Polaris and MoleculeACE

github.com/blazejba/monroe