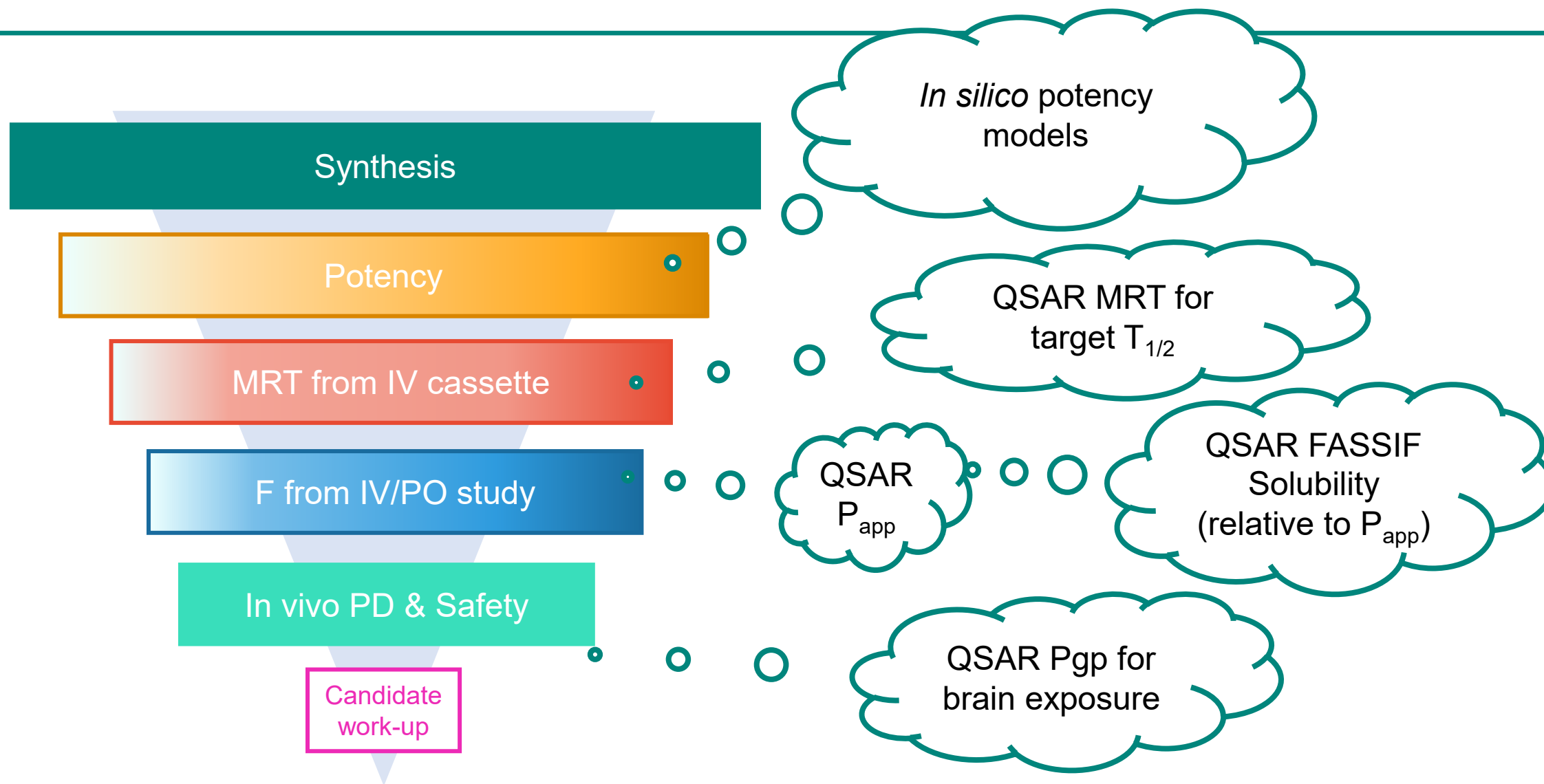# Applications of Machine Learning in Discovery and Development at Merck

IQ Machine Intelligence for Quantitative Modeling in Drug Discovery & Development Applications Workshop
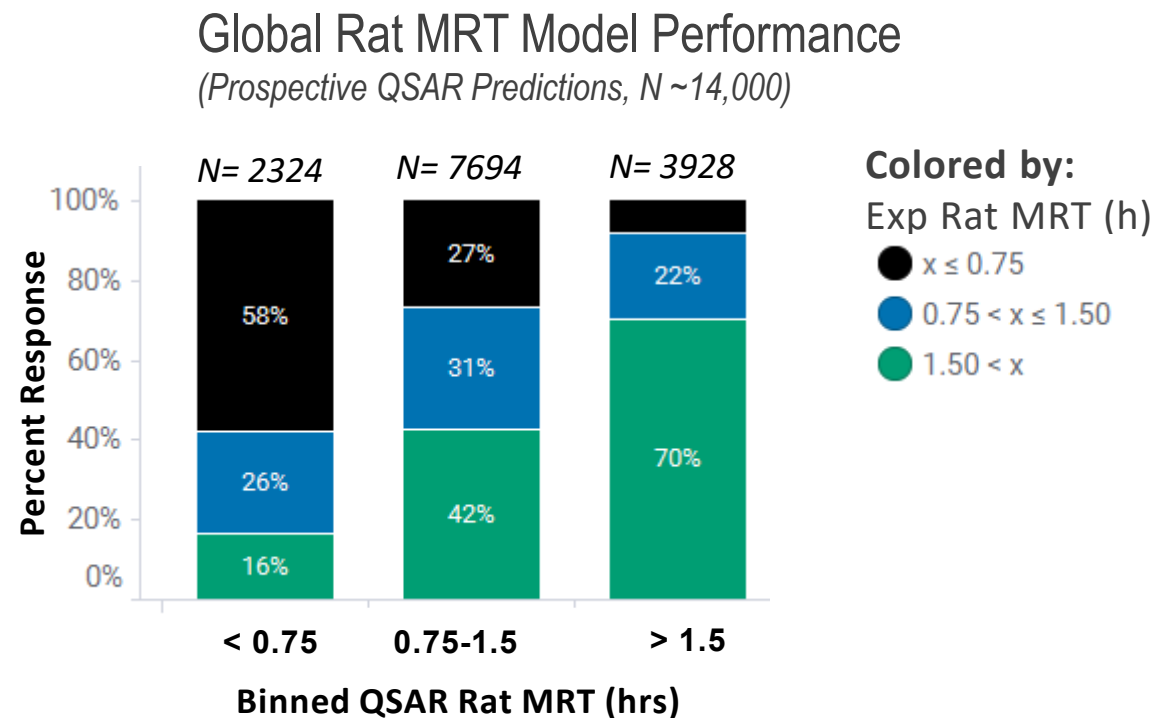
September 16, 2022

Use models relevant to human dose for molecule design and progression

Synthesis

Potency

MRT from IV cassette

F from IV/PO study

In vivo PD & Safety

Candidate work-up

In silico potency models

QSAR MRT for target $T_{1/2}$

QSAR P_{app}

QSAR FASSIF Solubility (relative to P_{app})

QSAR Pgp for brain exposure

Use machine learning models to inform decisions at every stage of the compound life cycle

How have we gotten better at using models for design and progression?

- Treat continuous models categorically
  - Emphasizes probability over r^2 and helps adoption

- Highlight opportunity for model uptake
  - Usage statistics and model performance

- Socialize benefits of improving compound quality
  - Speed and cost

- Focus synthesis/testing resources on designs/molecules compatible with program needs

Global Rat MRT Model Performance
(Prospective QSAR Predictions, N ~14,000)

Colored by:
Exp Rat MRT (h)
- ● $x \leq 0.75$
- ● $0.75 < x \leq 1.50$
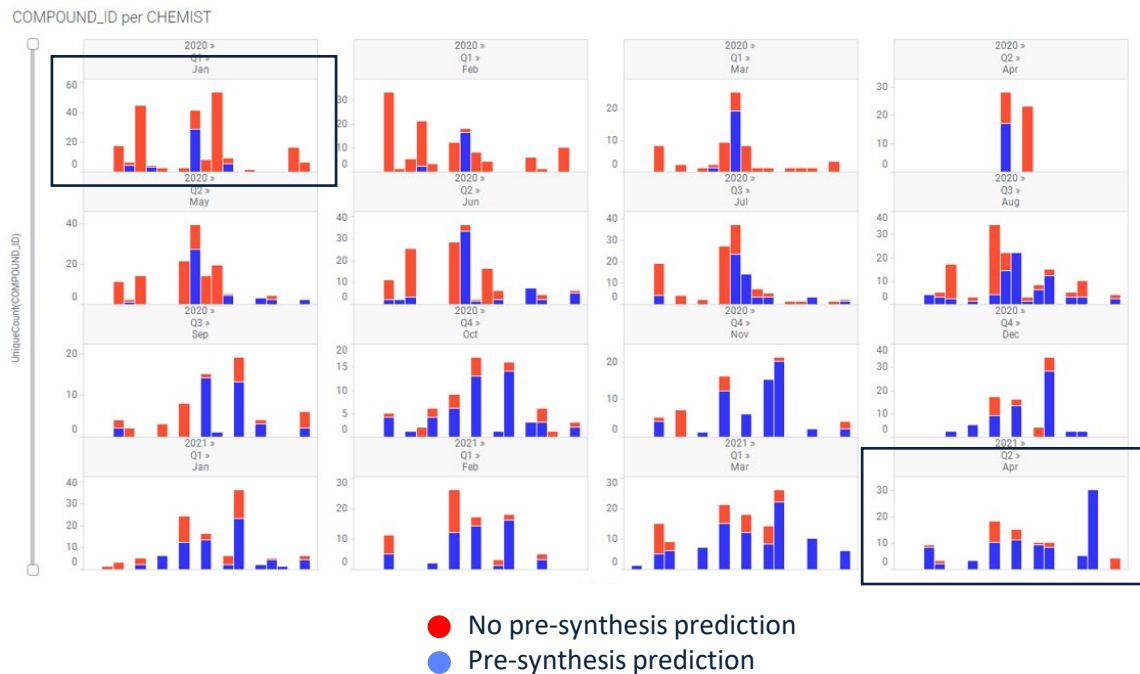- ● $1.50 < x$

Measuring the adoption of prospective model usage

Our virtual design database enables analysis of model usage

- Virtual and real registration <u>dates</u> indicates usage
- Property <u>prediction value</u> distributions suggest acceptance

Pre-synthesis utilization of predictive tools has steadily increased...

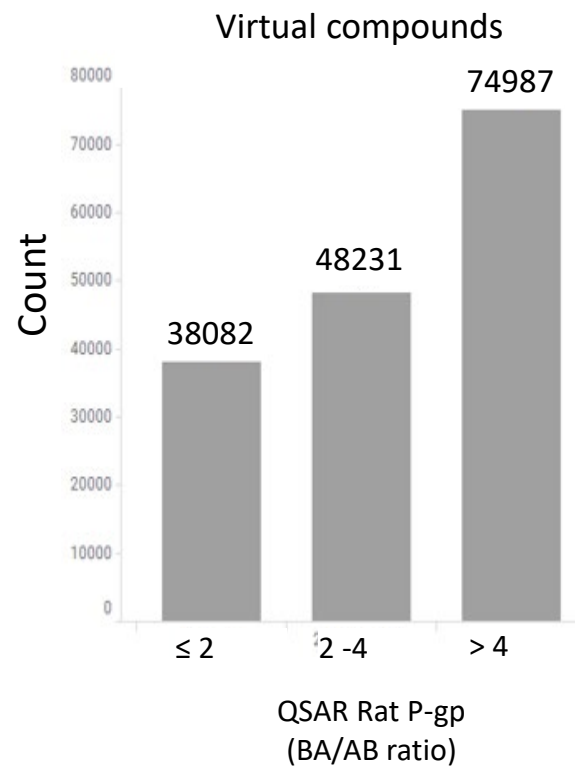*But are we using the predictions to make decisions?*

- 🔴 No pre-synthesis prediction
- 🔵 Pre-synthesis prediction

MERCK 272
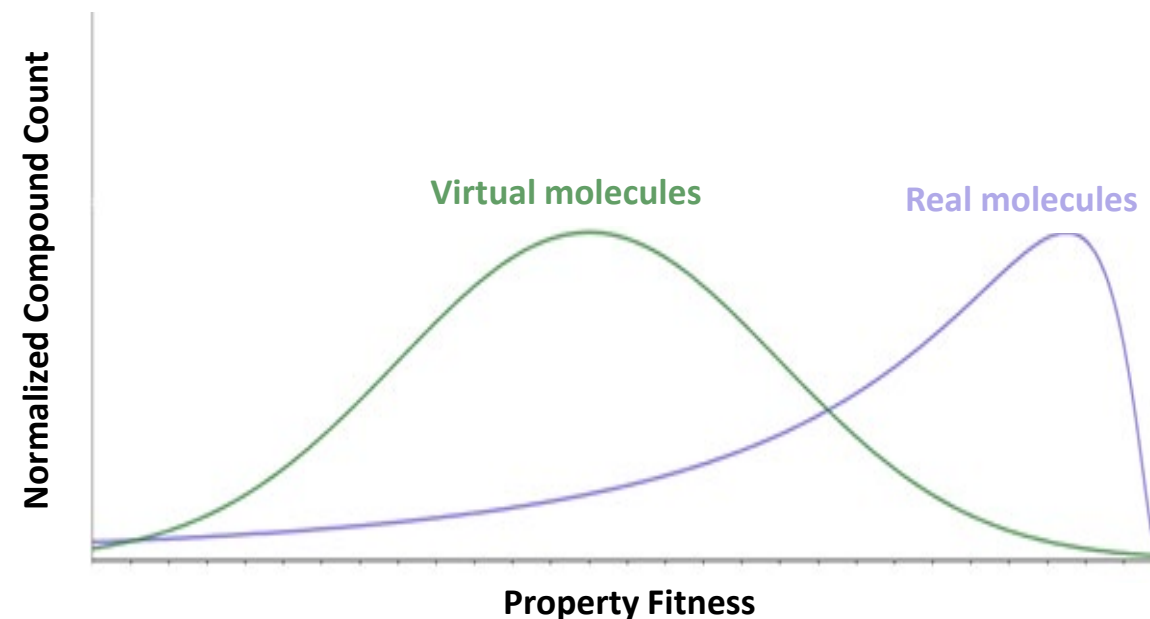
Project X: Using Rat P-gp predictions to prioritize chemistry

Design and synthesis prioritization based on QSAR Rat P-gp
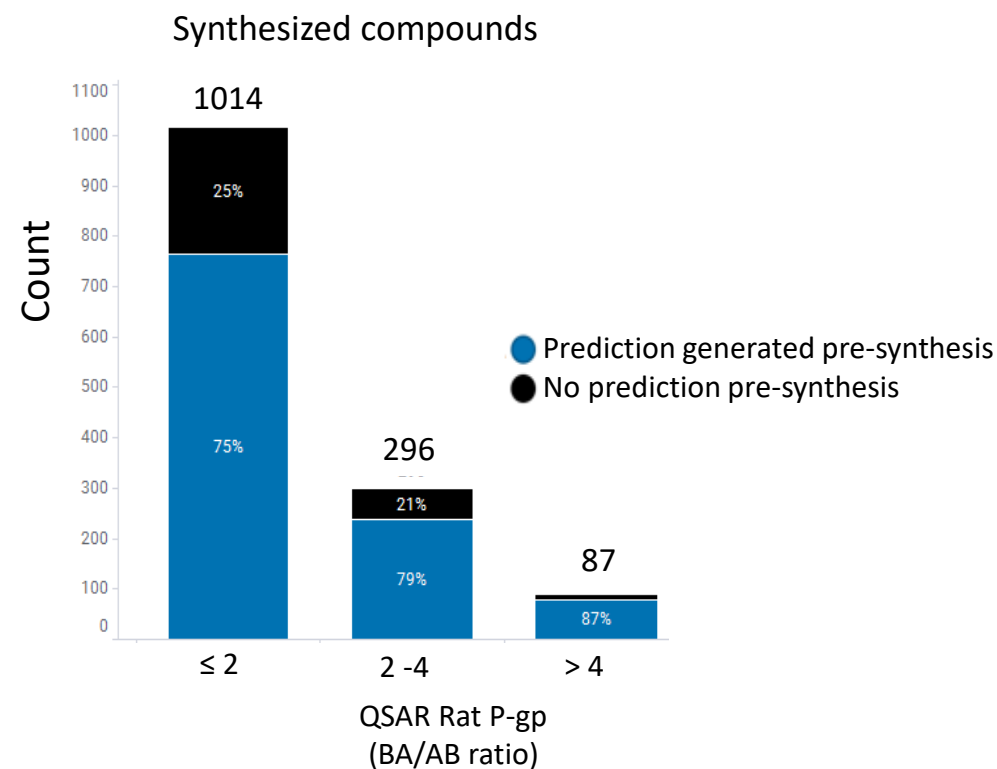
DESIGN → MAKE

Designs explored across the range of QSAR P-gp
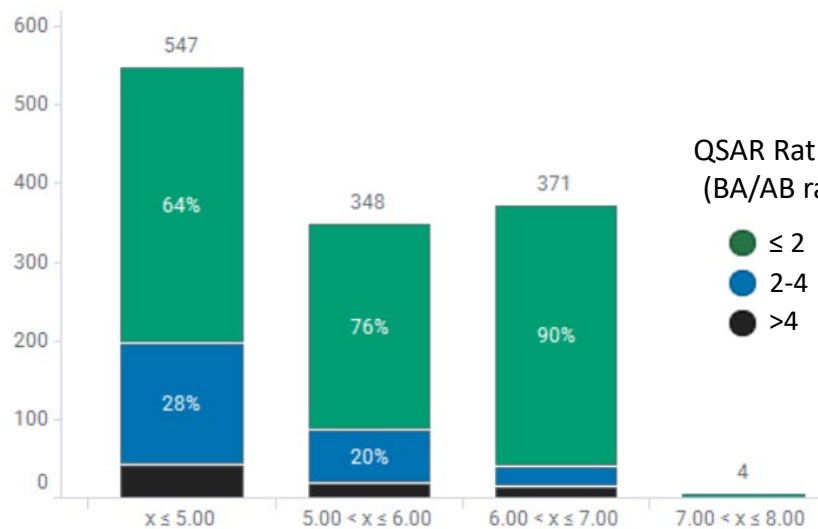
Synthesis skewed towards compounds with QSAR P-gp < 4

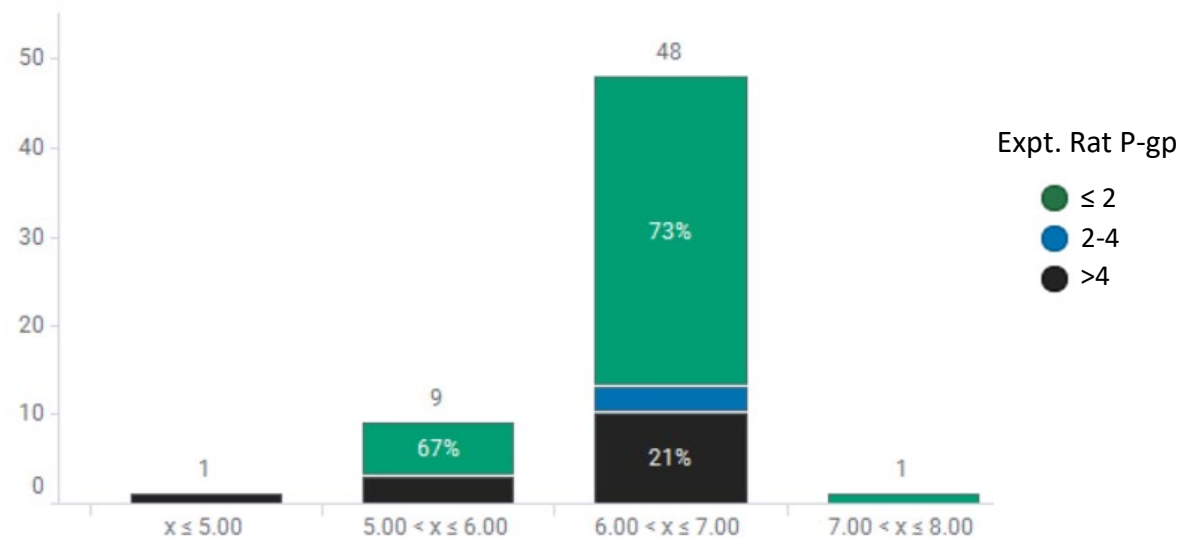# Project X: Using Rat P-gp predictions to improve measured properties

Rat P-gp Submission Prioritization based on QSAR Rat P-gp and corrected potency

MAKE → TEST

QSAR Rat P-gp (BA/AB ratio)
- ● ≤ 2
- ● 2-4
- ● >4

Expt. Rat P-gp
- ● ≤ 2
- ● 2-4
- ● >4

Corrected potency metric

Corrected potency metric

Most synthesized compounds have QSAR rat P-gp BA:AB < 4

Most compounds with measured P-gp have BA:AB <4

MERCK 274

Project Y: Using rat MRT predictions to prioritize chemistry

- Project team had shorter half-lives, so MRT predictions reviewed to set design guidelines

- Chemists began avoiding designs with QSAR rat MRT <0.75 hr

- <u>Predicted</u> MRT distributions for synthesized compounds shifted
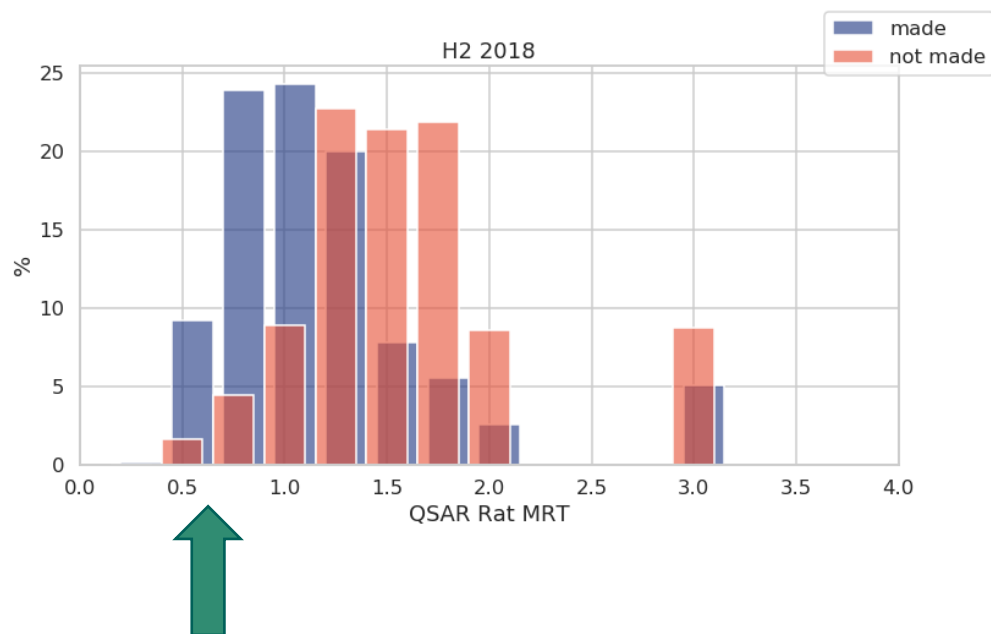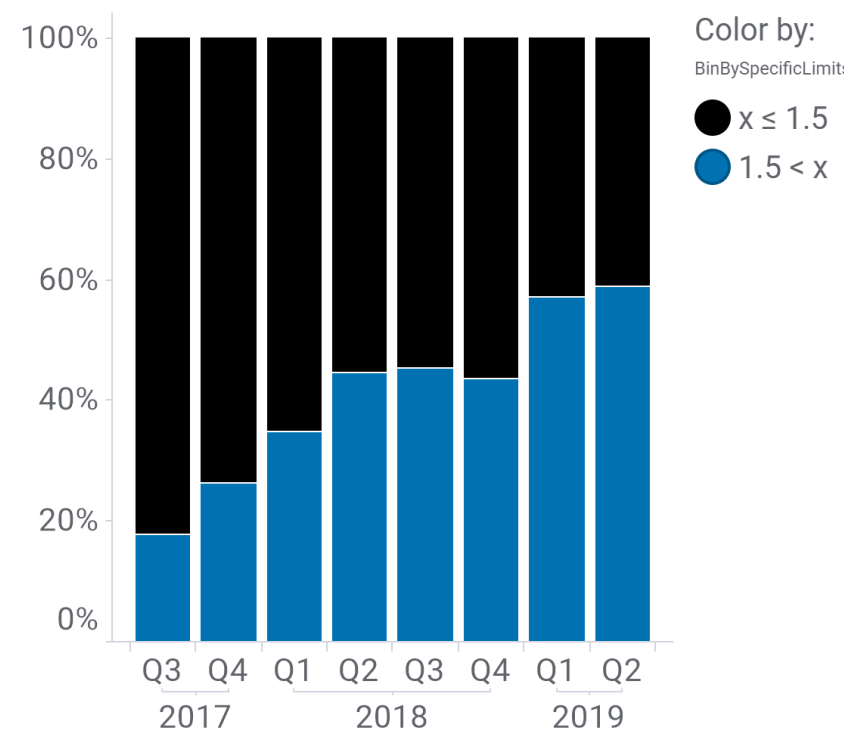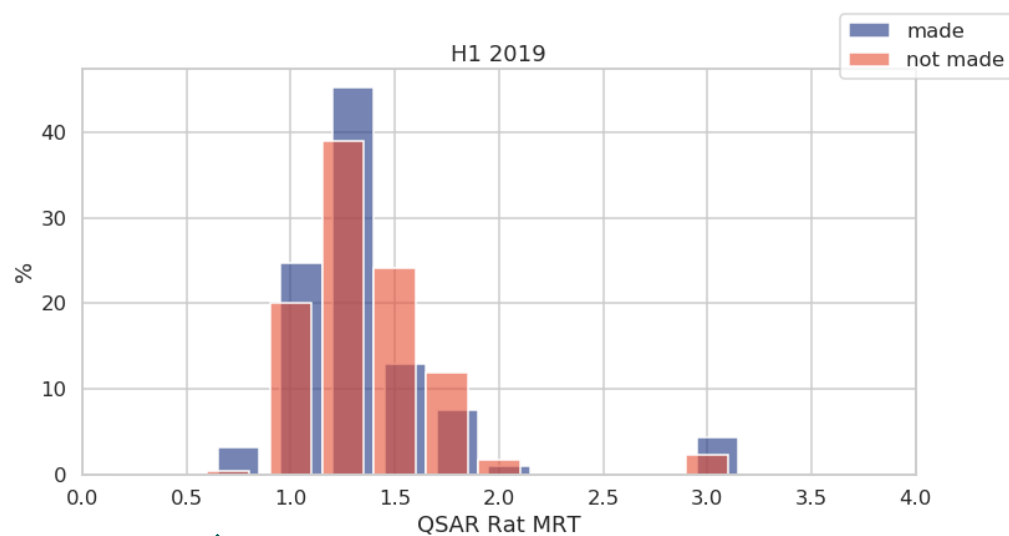

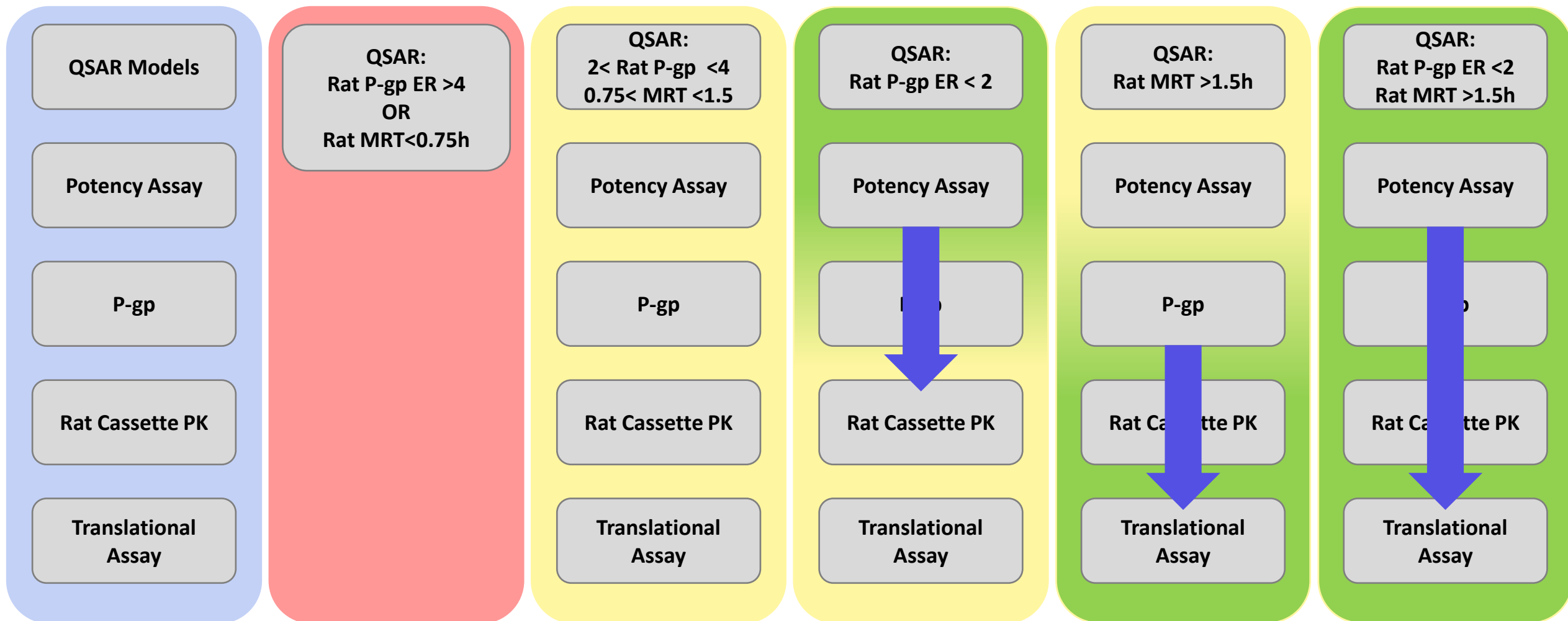Deprioritize designs with predicted rat MRT <0.75 hr
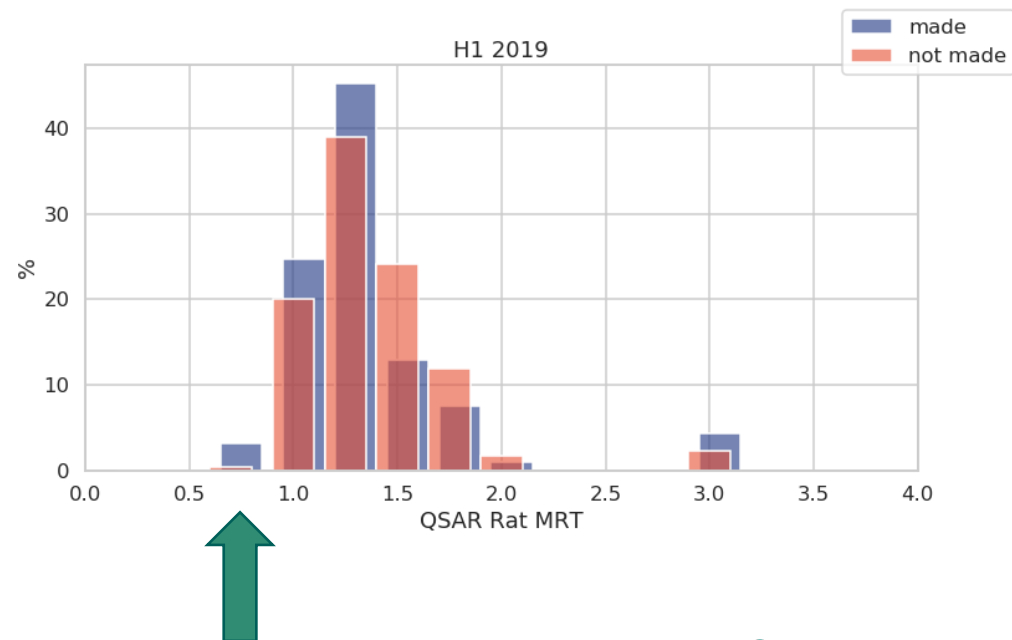
Project Y: Using rat MRT predictions to improve measured properties

- After adopting new QSAR MRT design criteria, more molecules had <u>measured</u> MRT values >1.5 hr

  - Improving the prediction distribution helped the measured distribution
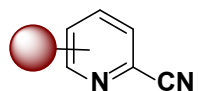
Adaptive compound testing – faster and leaner

QSAR Models	QSAR: Rat P-gp ER >4 OR Rat MRT<0.75h	QSAR: 2< Rat P-gp <4 0.75< MRT <1.5	QSAR: Rat P-gp ER < 2	QSAR: Rat MRT >1.5h	QSAR: Rat P-gp ER <2 Rat MRT >1.5h
Potency Assay		Potency Assay	Potency Assay	Potency Assay	Potency Assay
P-gp		P-gp	P-gp	P-gp	P-gp
Rat Cassette PK		Rat Cassette PK	Rat Cassette PK	Rat Cassette PK	Rat Cassette PK
Translational Assay		Translational Assay	Translational Assay	Translational Assay	Translational Assay

Progressing compounds with ML guidelines may get to key compound decision points faster

MERCK 277
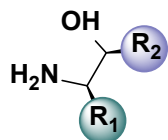
How do we apply machine learning for development projects?

MERCK
INVENTING FOR LIFE

Pyridine-Oxazoline (PyOx) ligands

- We use catalysis kits and automation to accelerate chemistry

<u>*PyOx ligands in Asymmetric Catalysis*</u>

building block 1   **building block 2**

*PyOx Ligand*

- *Reductions*
- *Heck-type reactions*
- *Difunctionalizations of olefins*
- *Additions of arylboron to electron deficient bonds*
- *Cross-couplings*
- *C-H functionalizations*

- *Easily prepared in two steps*

- *Access to natural chiral pool - aminoalcohols!*

- *Used with various metals: Pd, Ni, Ir, Cu*

Can we apply machine learning to high throughput synthesis data and iterate?

Yang, G., Zhang, W. *Chem. Soc. Rev.* **2018**, *47, 1783*

Ligand evolution and iteration with machine learning

Typical workflow searching for more selective catalysts/ligands
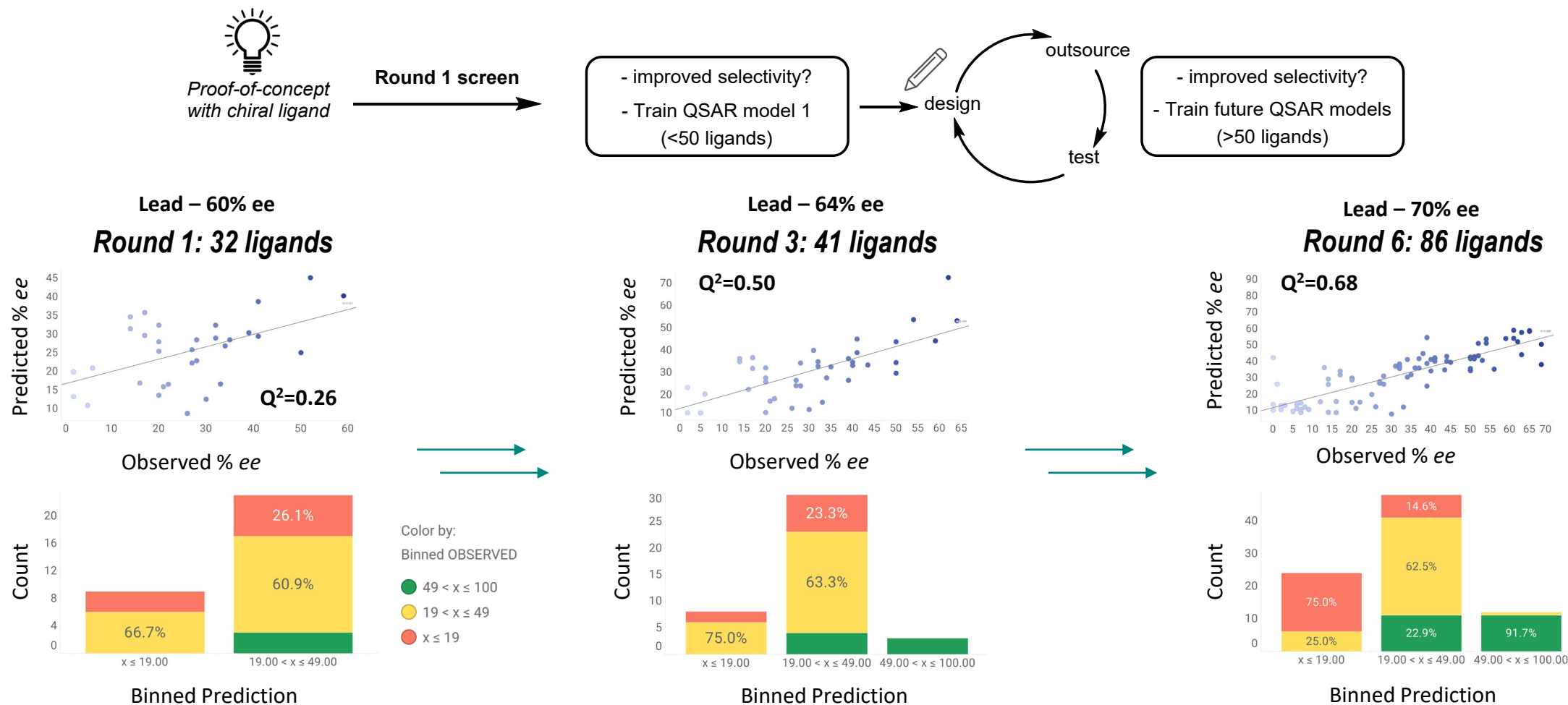
Creating 96-well plates for high throughput experimentation

- Use clustering to select diverse ligands to enable machine learning
- Create screening plate that enables ML and iterative ligand design in future projects

21,474 ligands

Chemist-Informed
Database
Literature

96 cluster K–means partitioning

96-clusters

Select ligand closest to centroid

96-selection

Parse selection

96-well HTE PyOx plate
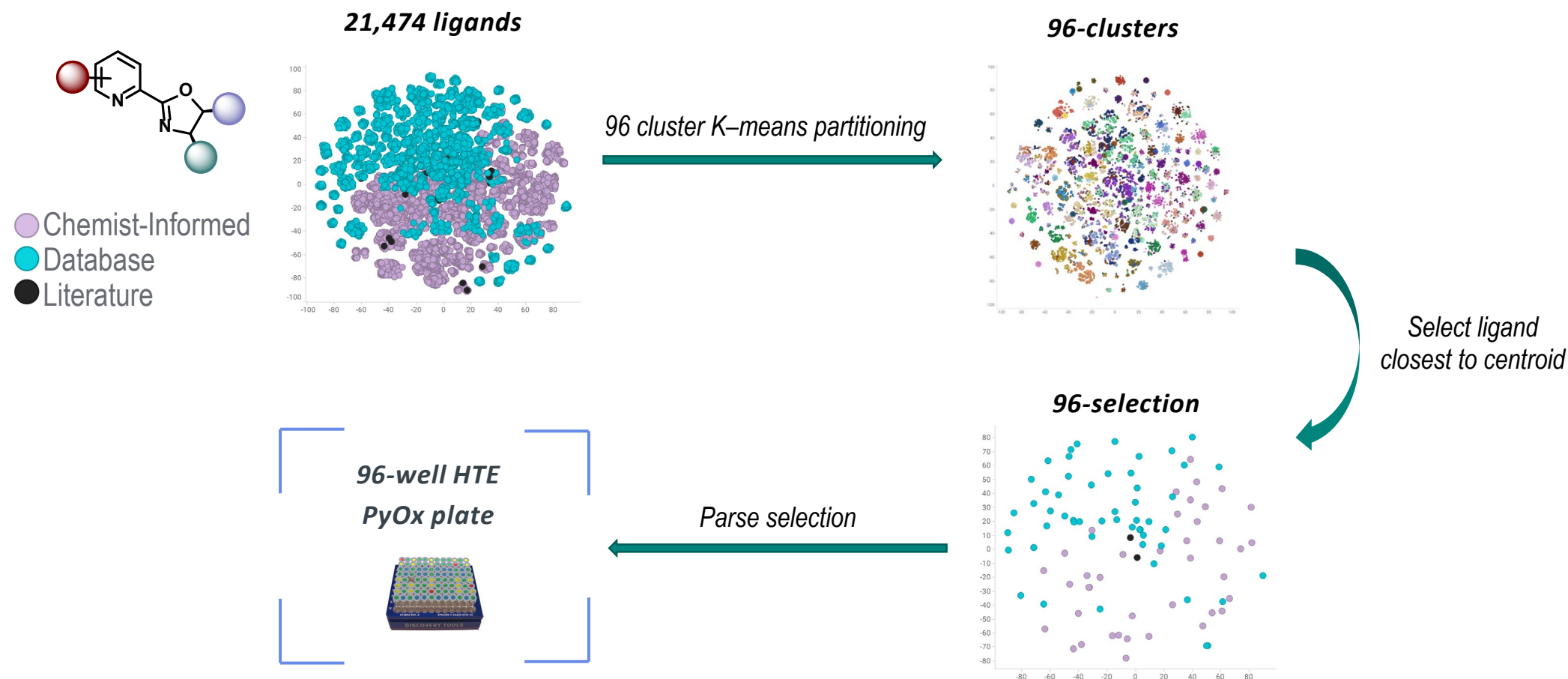
Conclusions and outlook for machine learning in pharma

- Machine learning being applied to select quality designs, advance molecules, and optimize chemistry

- We are realizing benefits and investing further!

  - Working to increase prediction capacity

  - Developing experience with generative models

  - Improving our retrosynthetic methods and learning how to maximize impact

  - Developing more models for toxicology questions

  - Gaining experience with learned potentials for physics-based modeling

- We anticipate increased organizational ML fluency and for traditional modeling to evolve

Acknowledgments

Juan Alvarez

Duane DeMong

Jesus Estrada

Cyndi He

Luca Iuzzolino

Beth Joshi

Prabha Karnachi

Dan Lehnherr

Maria Madeira

Karin Otte

Phieng Siliphaivanh

Thomas Struble