# Ranking, not predicting

## *Antibacterial lead optimization in 15 months with AI/ML*

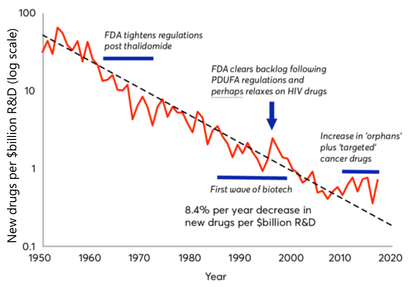
ARREPATH · NINE MARKETED ANTIBACTERIALS ACROSS THE TEAM · PRINCETON SPINOUT, 2021

Most drug discovery teams discard cross-assay data to avoid noise. We pooled it, learned to rank rather than predict, and built models that zero-shot generalize to new chemistry — rescuing an antibacterial program in 15 months with a team of eight.

### Productivity in drug discovery is declining

AI/ML has so far failed to improve drug discovery efficiency.

Conventional approaches work well within a series but often fail to generalize.



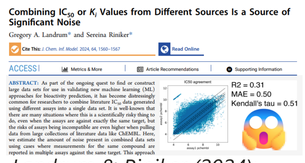

*Jones et al. 2018. 10.13140/RG.2.2.15613.0560.*

### Pooling the data everyone else discards

Drug discovery data is sparse, noisy, and scattered across species, strains, and assay types. Standard practice: throw most of it away and model a clean subset.

Absolute values don't reproduce across assays, but *rank orders* do — so we trained models to learn rankings instead, using scaffold-agnostic molecular representations.

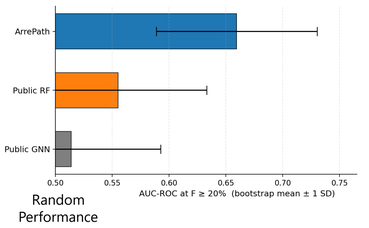
Using this heterogeneous data others discard allows our models to generalize by learning across more data.




*Landrum & Riniker (2024): even against the same target, combining $IC_{50}$ values across assays introduces substantial noise. $R^2$ = 0.31.*

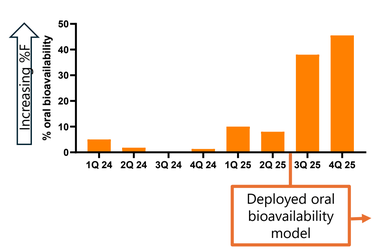
### Zero-shot generalization to new molecules

ArrePath's oral bioavailability model successfully ranked oral bioavailability of our lead program *despite having no training data from that program.*




### AI enabled oral dosing of our lead program

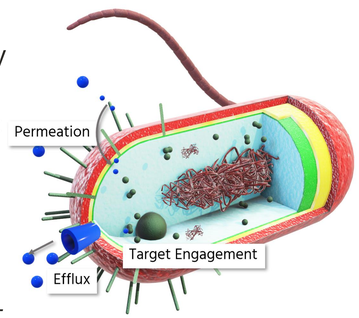
Scoring all ~1000 compounds to date identified a molecule with *37% oral bioavailability*; the best molecule before that was only 10% orally bioavailable




Current molecules are >80% orally bioavailable

### Antibiotics are a stress test for an AI platform

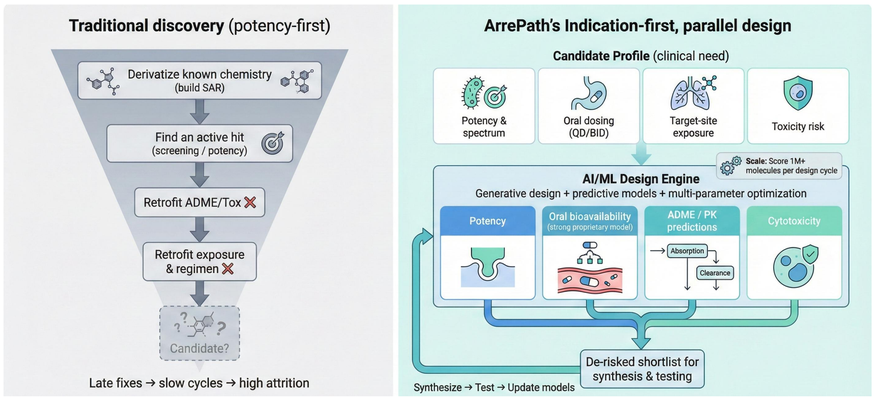
Antibacterials face all standard drug discovery constraints *plus* bacterial-specific ones: permeation, efflux, and resistance.




Discovering antibacterials against novel targets is *hard* — over two decades, GSK and AstraZeneca screened 100+ essential bacterial targets and no molecule ever reached the clinic

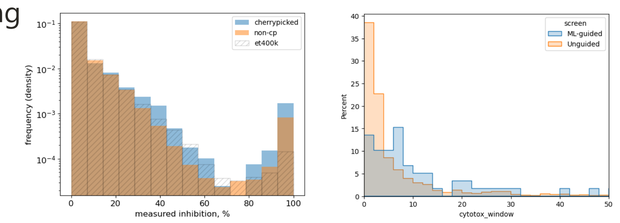
A platform that works here should work anywhere.

### Define criteria, then design the molecule




### 3x more actionable hits from AI-guided screening

Screening for novel antibiotics by selecting for predicted antibacterial and not cytotoxic compounds gave ~3x more progressible hits than unguided screening




**15**
months
lead optimization
*vs. 4-5 years typical*

**8**
people
*vs. 20-30 typical*

**6**
new antibacterial classes over 4.5 years

**Let's talk.** Partnerships, investment, and technical deep-dives. Scan or email kurt.thorn@arrepath.com

ARREPATH