ATOM

A Scalable, [illegible] Molecules:
Case Study in Discover[illegible] Inhibitors with Favorable
[illegible]

Can I ask a (super)computer to design a drug?

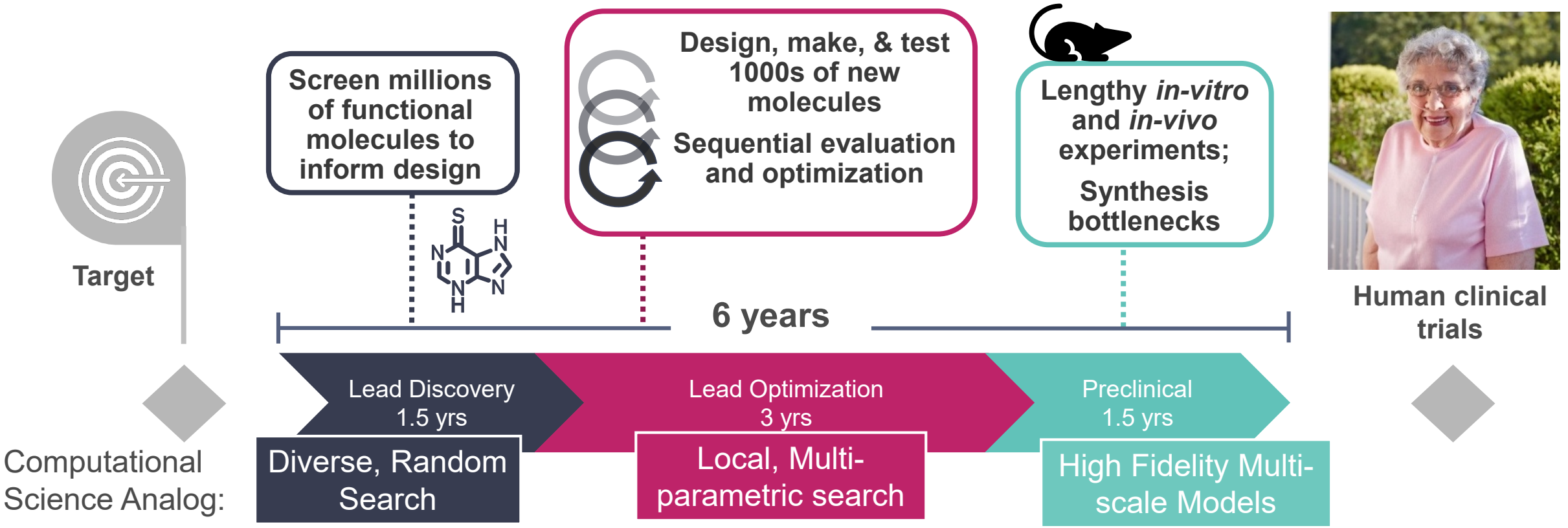
Andrew Weber

Research Director, ATOM Consortium

SC 2019 Computational Approaches for Cancer Workshop

Current drug discovery: long, costly, high failure

Is there a better way to get medicines to patients?

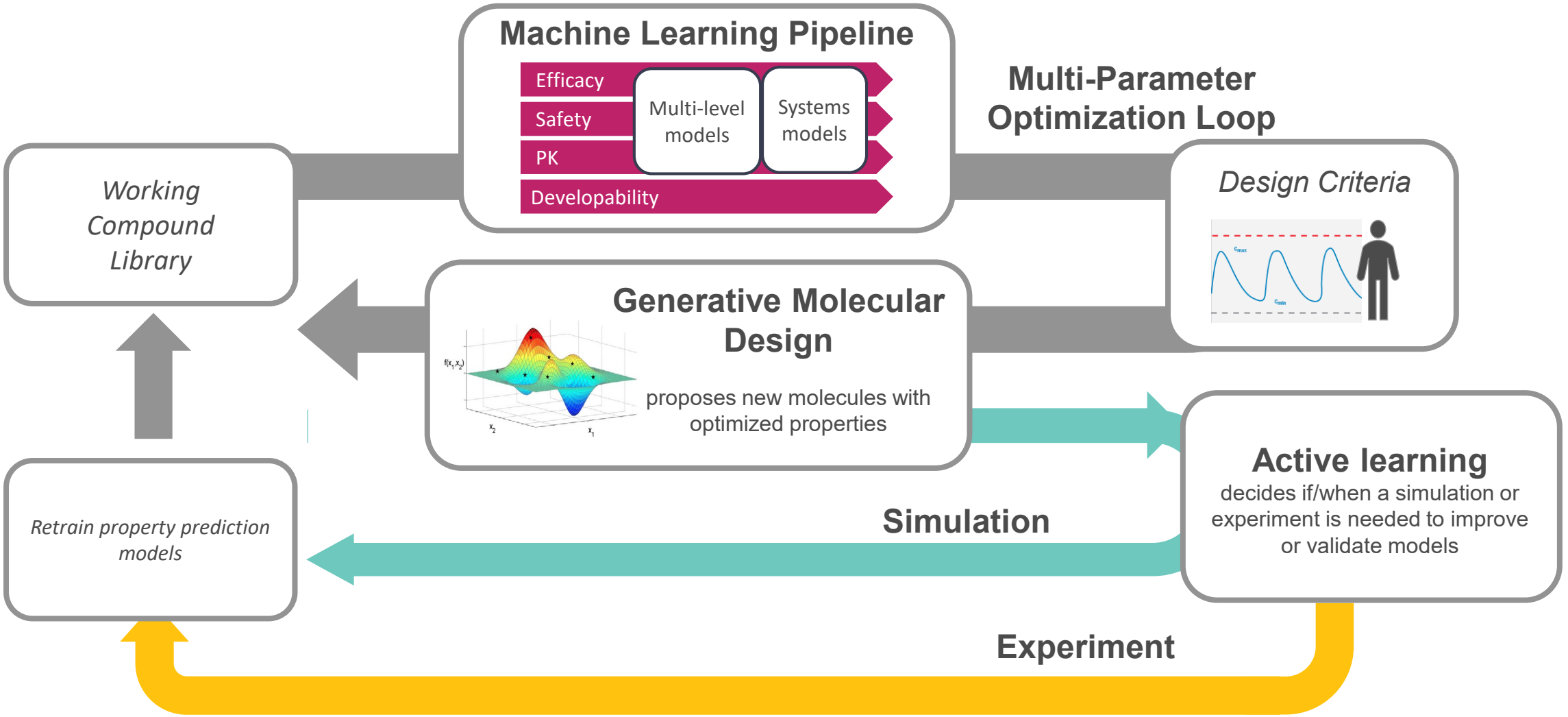


- **33% of total cost of medicine development**
- **Clinical success only ~12%, indicating poor translation in patients**

Source: http://www.nature.com/nrd/journal/v9/n3/pdf/nrd3078.pdf

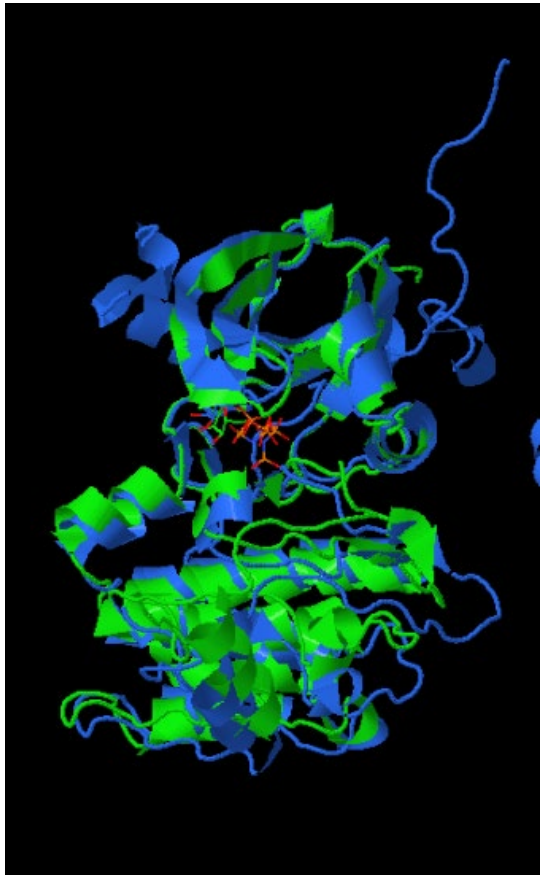
The ATOM Platform

Active Learning Drug Discovery Framework



Application Generative Molecular Design to Aurora Kinase Case Study

Case Study: Can the ATOM generative design framework develop a novel, potent, selective Aurora Kinase B inhibitor, while maintaining select developability properties (i.e. PK/safety)



Why Aurora Kinase?

Oncology Relevant

A family of serine/threonine kinases involved in cell division pathways

Validated Target

≥12 inhibitors of AURK A and or AURK B have progressed to clinical trials

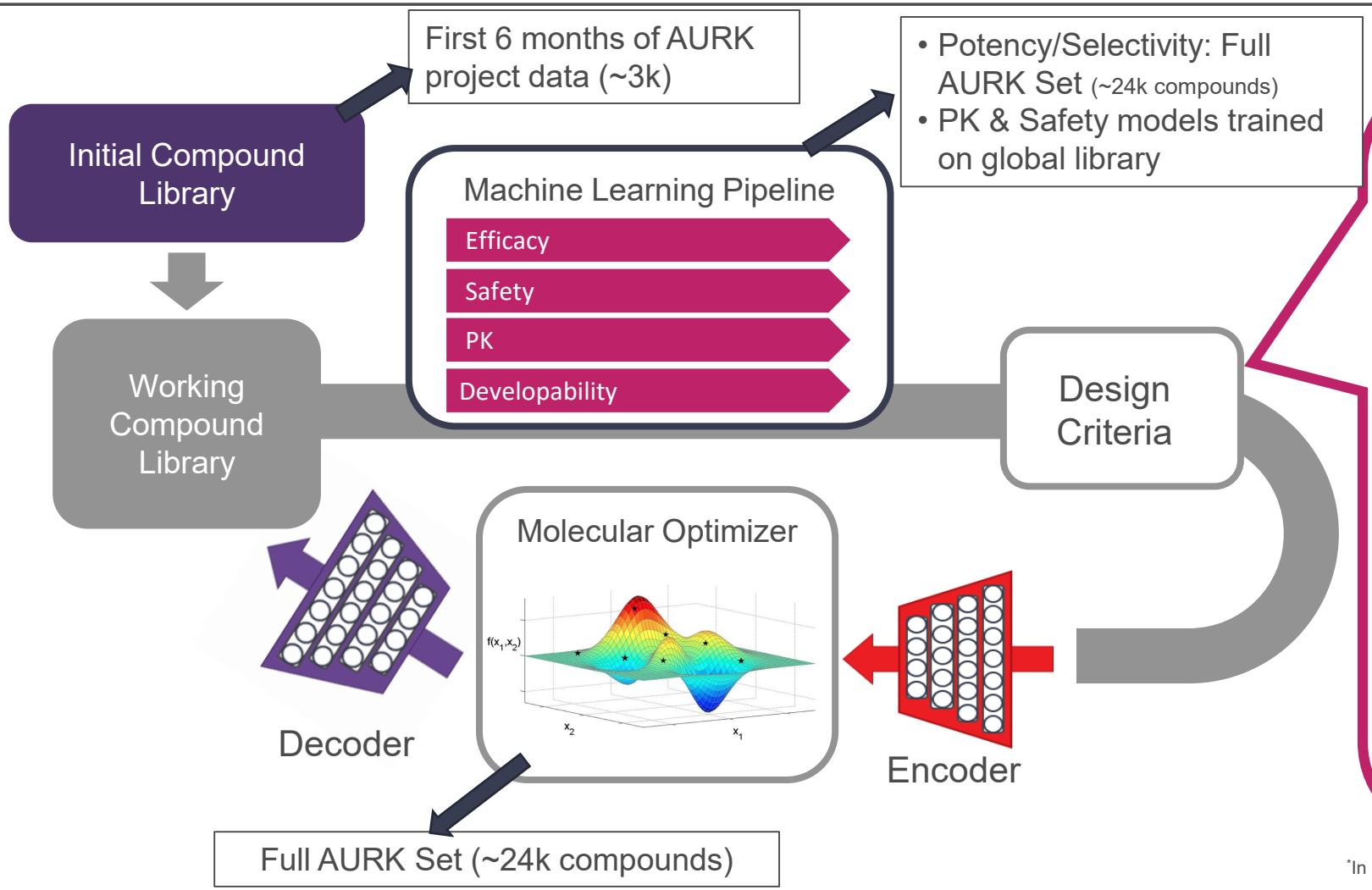
Representative Drug Discovery Problem

Selectivity between kinases is a challenging, but not intractable problem for drug discovery

Availability of Target-Specific Data

Potency data on ~24k compound available for AURK A and/or AURK B, which allows for assessment of problem at different phases of discovery

Case Study Design and Targets



Pilot Candidate Quality Panel

Efficacy
AURK B (pIC50 > 9)
AURK B/A Selectivity > 1000

Safety
BSEP (pIC_{50} < 4.5)*
hERG (pIC_{50} < 4.5)*

PK
Solubility (>10uM)
CL_{int} (<3 mL/min/g)

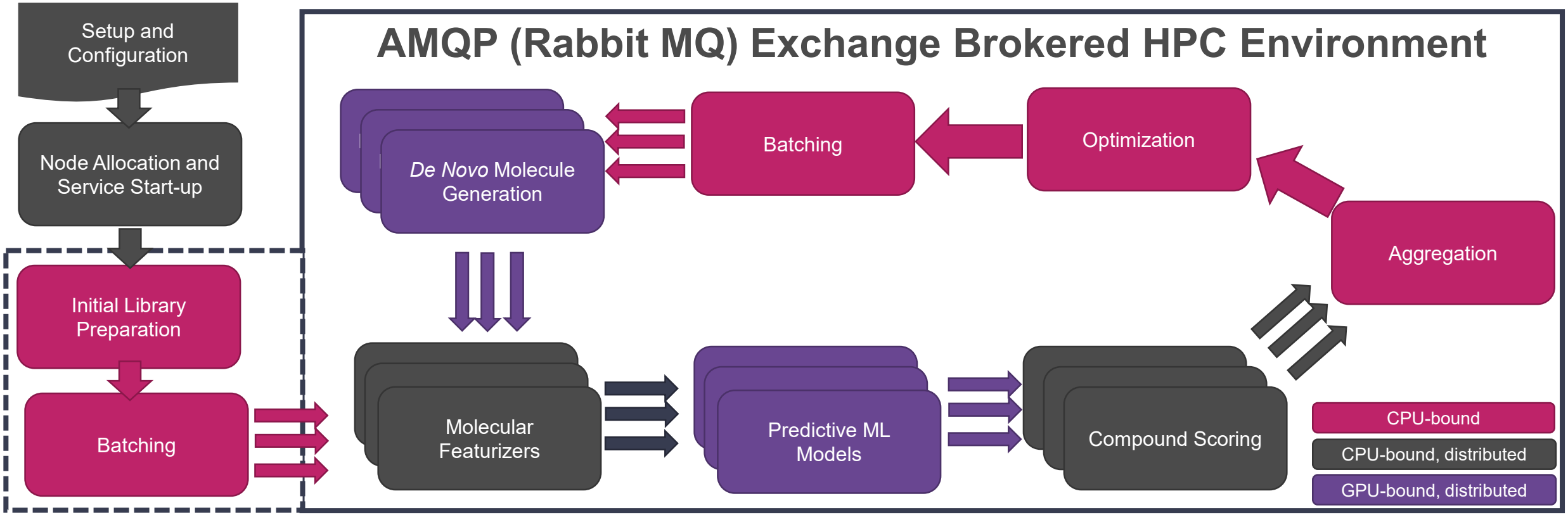
Developability
Solubility (>10uM)
SAS: Synthetic Accessibility Score
QED: Quantitative Estimation of Druglikeness

*In Silico Design Target of <4 adjusted upwards due to current assay LLQs

High Performance Compute Facilitates Large Scale Search

Enables Scalable Management of Heterogeneous Compute Tasks

- Facilitated ideation and evaluation of **>3 million** compounds in **24 hour run time**
- Future **scaling by 10x or more** achievable on currently utilized ~100 node clusters

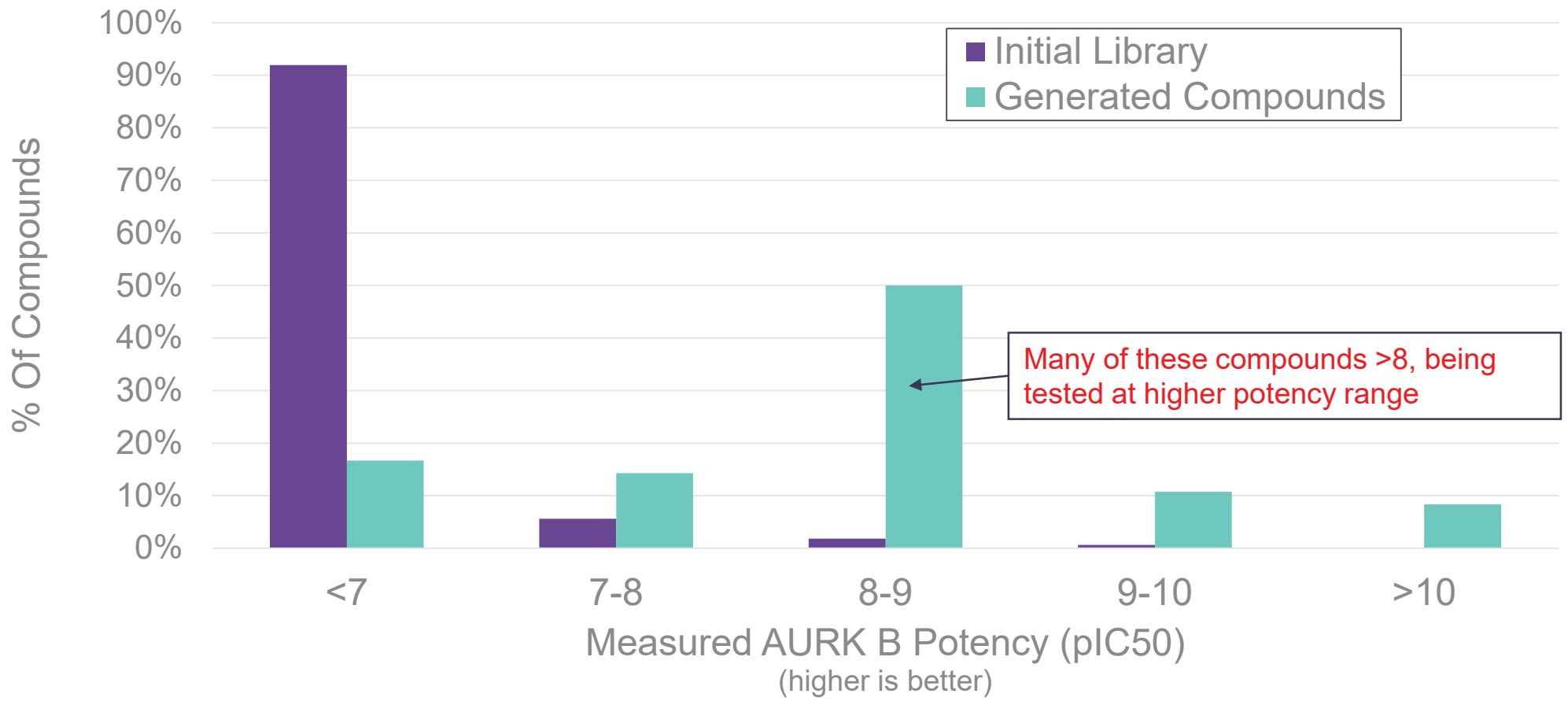


In Silico Optimization Proposed Library of Compound With High Quality *Predicted* Properties

Compound	AURK B pIC50	AURK B/A Selectivity (fold)	hERG pIC_{50}	BSEP pIC_{50}	hLM Clearance (mL/min/g)	Solubility (uM)	SAS
Compound 1	9.2	5287	3.3	4.3	3.6	1096	2.5
Compound 2	9.3	3233	3.2	4.2	2.5	399	2.4
Compound 3	9.6	11512	3.6	4.4	2.2	412	2.6
Compound 4	9.6	2449	3.2	4.3	2.5	60	2.3
Compound 5	9.7	3068	3.3	4.3	2.0	1155	2.5
Compound 6	9.6	5756	3.7	4.5	4.3	232	2.3
Compound 7	9.3	3296	3.3	4.4	2.6	33	2.4
Compound 8	9.1	1197	3.3	4.2	2.4	268	2.5
Compound 9	9.2	7724	3.3	4.3	2.3	733	2.7
Compound 10	10.1	2270	3.2	4.5	2.6	139	2.4

Make Test Results Confirm High On-Target Potency

70% Of Tested Compounds with $pIC_{50} > 8$



De Novo Synthesis & Testing Confirms Enrichment of High Potency Compounds

Where To Next?

- Increase in problem scope
  - Scaling evaluation criteria: full safety and PK panels
  - Scaling of generative framework: Recently increased unique search space by >5x
- Application on realistic lead optimization scenarios
  - Reduction in target specific data
  - Integration of uncertainty quantification
  - Active learning for explore/exploit selection of compounds and model retraining
- Integration of structure based design techniques for virtual hit finding
- Integration of systems modelling for therapeutic window-based optimization

Acknowledgements

ATOM Generative Molecular Design Team

- Jason Deng
- Kevin McLoughlin
- Tom Sweitzer
- Jeff Mast
- Juliet McComas
- Margaret Tse
- Derek Jones

ATOM Joint Research Committee

- Tom Rush
- Jim Brase
- Stacie Calad-Thomson
- Michelle Arkin
- Dwight Nissley

Backup Slides

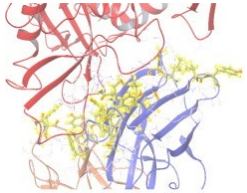
Drug Discovery As a Search Problem

Physical, multi-parametric, multi-fidelity process leads to long, expensive process

Hit Finding

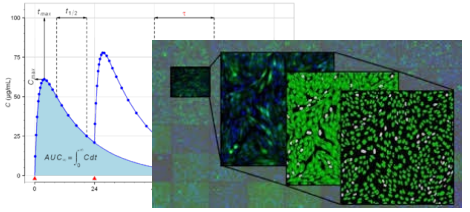
Brute-force, diverse
“pseudo-random” search

Hit to Lead



Greedy selection
Local search

Lead Optimization



Rule based/ heuristics
Multi-parametric local search
(PK, Safety, Efficacy)

Preclinical Profiling

Higher Fidelity Model
Confirmation

Compound Synthesis Requirement
Cost of Incorrect Solution

# of Evaluable Solutions

Issues:

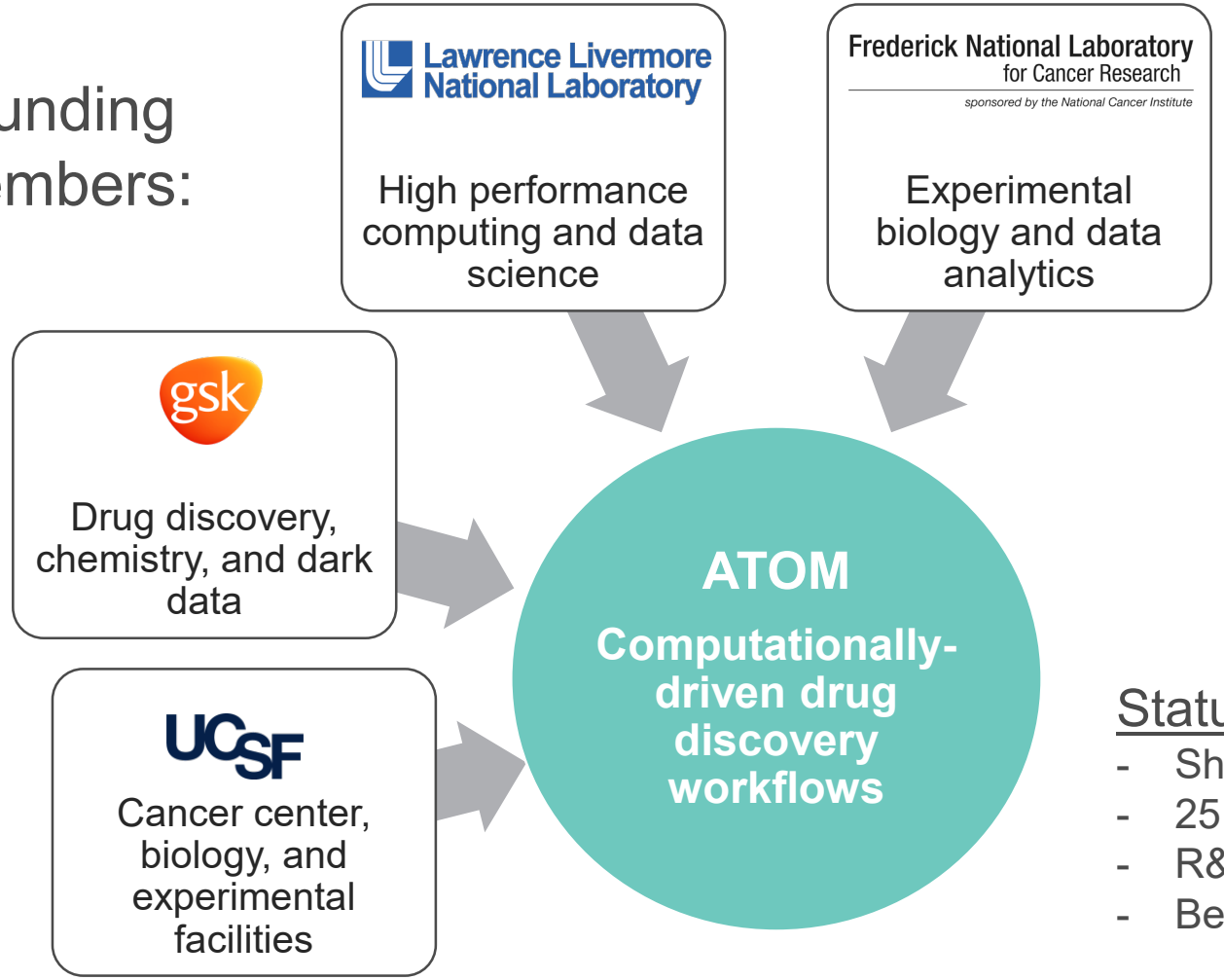
- Large search space(10^{20} – 10^{60} options)
- Long cycle times for *de novo* chemistry
- Difficult multi-factorial decisions

Hypothesis: Significant acceleration through *in silico* generative search with integration of automated QSAR/CADD and systems modelling techniques

Accelerating Therapeutics for Opportunities in Medicine

ATOM Consortium

Founding Members:



Approach: An open public-private partnership

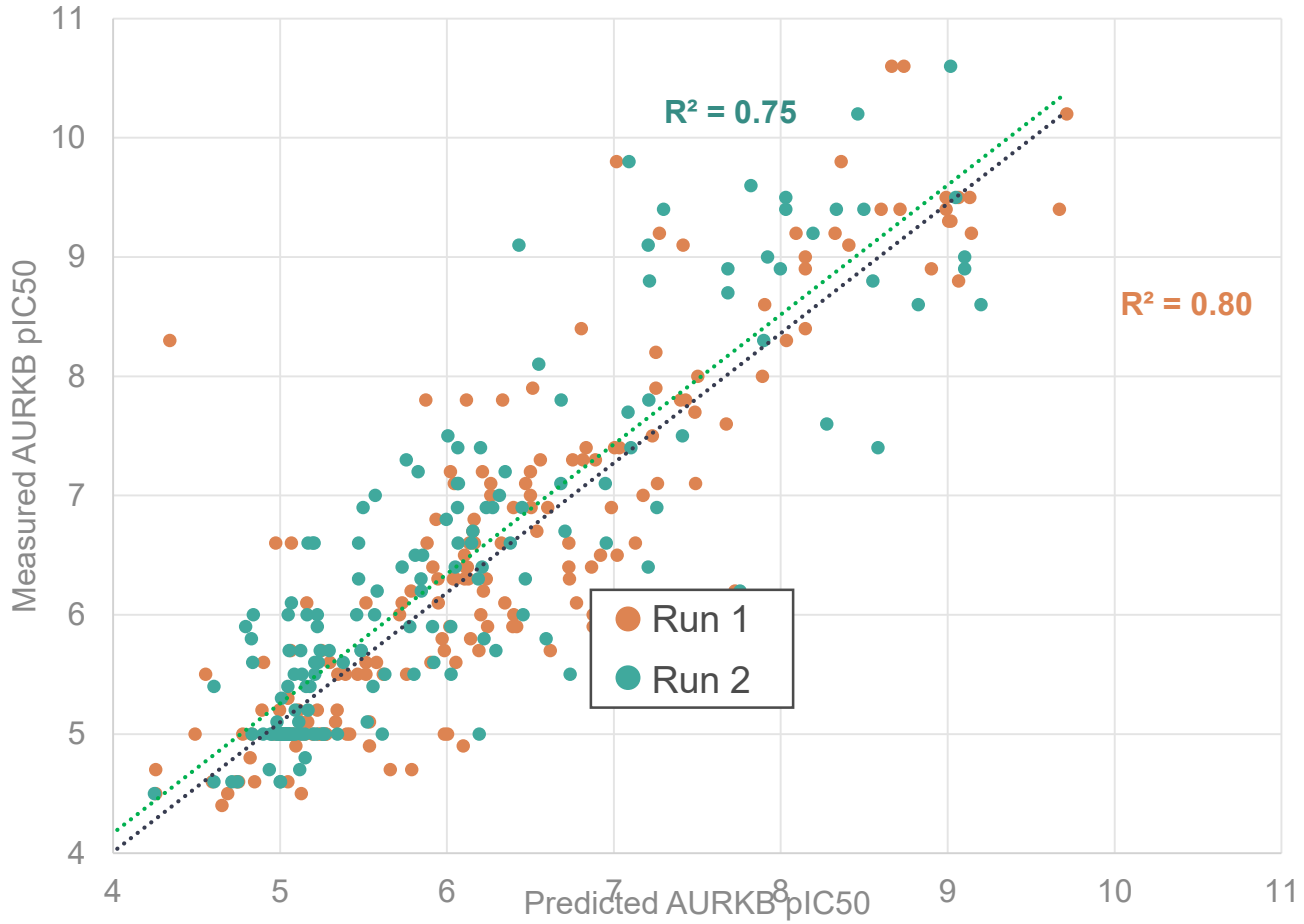
Tactics: Integrated data, computation, experiment, and active learning

Product: An open-source framework of tools and capabilities

Status:

- Shared collaboration space at Mission Bay, SF
- 25 FTEs engaged across the partners
- R&D started February 2018
- Beginning to engage new partners

Additional “Rediscovered” Compounds with Existing Data Validates Predative Accuracy of The Models*



*Rediscovered compounds from non-ATOM library to confirm model accuracy