ATOM: Computationally-Driven Drug Discovery

Leonard P. Freedman, Ph.D.

Chief Science Officer

FNLAC Meeting—June 27, 2019

ATOM: Accelerating Therapeutics for Opportunities in Medicine

Frederick National Laboratory for Cancer Research

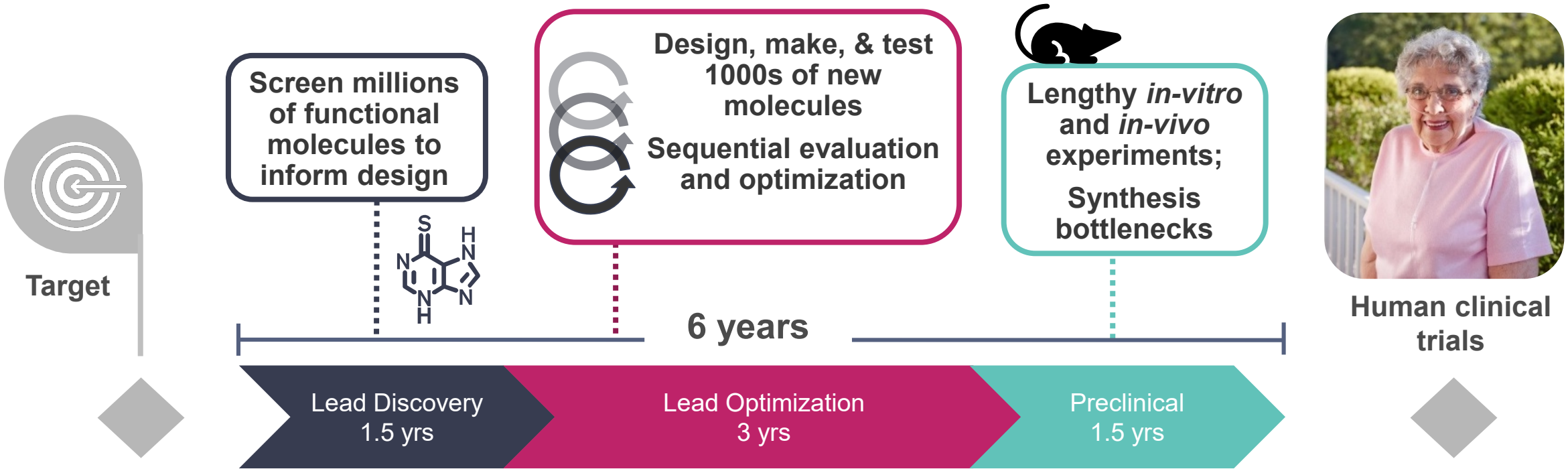
Presentation outline:

- Research approach
- Data and modeling capabilities
- Pilot project—Aurora kinase inhibitor
- ATOM milestones—current and future

Current drug discovery: long, costly, high failure

Is there a better way to get medicines to patients?

Frederick National Laboratory for Cancer Research



- 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

Accelerating Therapeutics for Opportunities in Medicine

ATOM Consortium

Frederick National Laboratory for Cancer Research

Founding Members:

Lawrence Livermore National Laboratory

High performance computing and data science

Frederick National Laboratory for Cancer Research
sponsored by the National Cancer Institute

Cancer biology and data analytics

gsk

Drug discovery, chemistry, and dark data

UCSF

Cancer center, biology, and experimental facilities

ATOM

Computationally-driven drug discovery workflows

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

ATOM – FNLCR Team

Frederick National Laboratory for Cancer Research

Governing Board	Joint Research Committee	Operations & R&D Management	Technical Team

Ethan Dmitrovsky
GB Member

Eric Stahlberg
GB Member

Dwight Nissley
JRC Member

Izumi Hinkson
Scientific PM

Beth Winger
ATOM Fellow

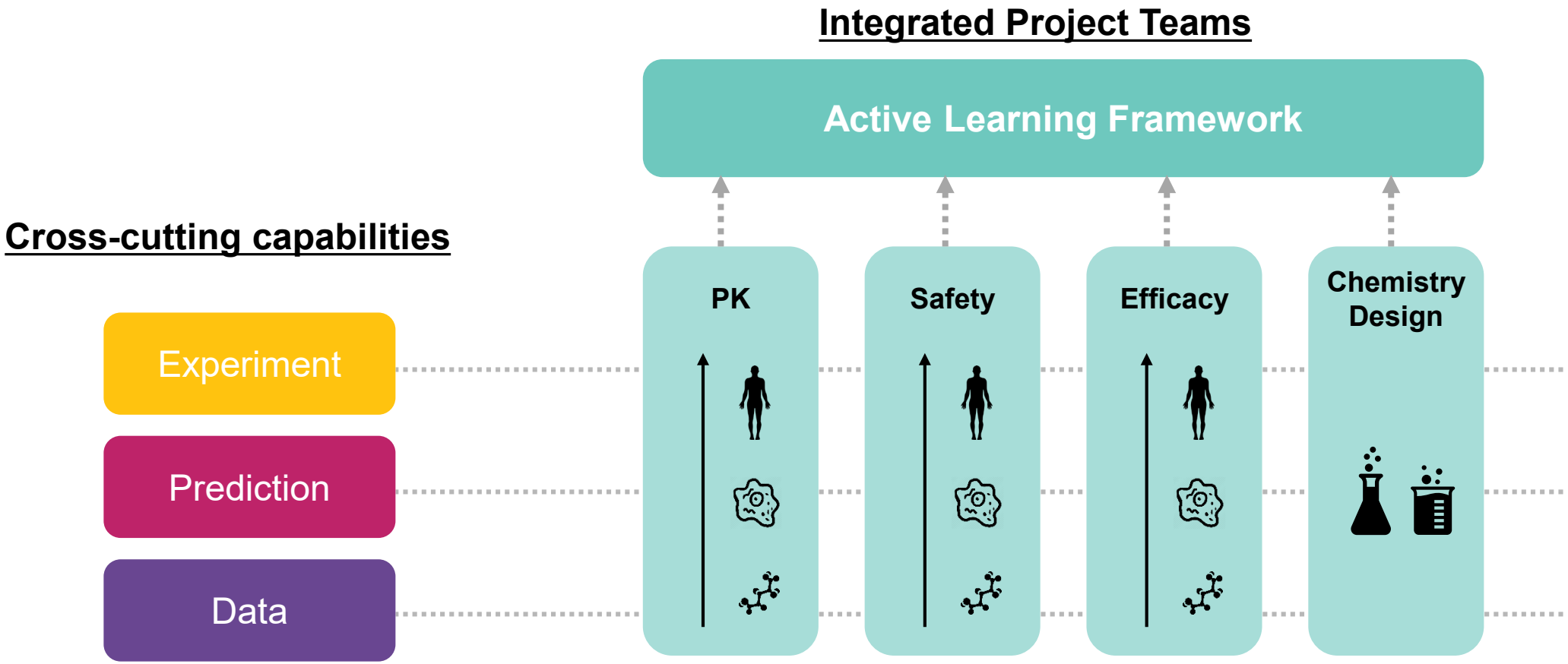
Len Freedman
GB Observer

Ben Madej
Data Scientist

Research and Development Approach

Matrix approach integrating experiment and computation

Frederick National Laboratory for Cancer Research



Accelerated Drug Discovery Concept

The ATOM workflow

Frederick National Laboratory for Cancer Research

Patient-specific
data in – therapeutic out

ATOM Platform

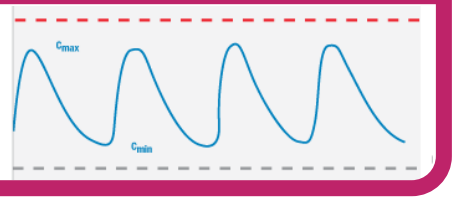
design
make
Empirical
test
active learning
In Silico
simulate

OPEN SOURCE RELEASES

Software, models and generated data

Members use workflow for internal drug discovery

Models of drug behavior in humans



Therapeutic Candidates

Commercialization by members for patient benefit

The ATOM Platform

Active Learning Drug Discovery Framework

Frederick National Laboratory for Cancer Research

Machine Learning Pipeline

Efficacy
Safety
PK
Developability

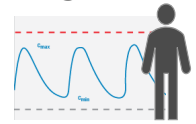
Multi-level models

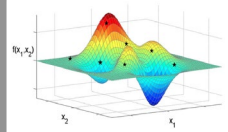
Systems models

Multi-Parameter Optimization Loop

Working Compound Library

Design Criteria





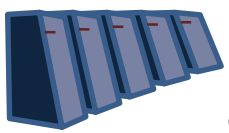
Generative Molecular Design

proposes new molecules with optimized properties

Active learning

decides if/when a simulation or experiment is needed to improve or validate models

Retrain property prediction models



Molecular Feature Simulations

Simulation

Human-relevant assays, complex in vitro models

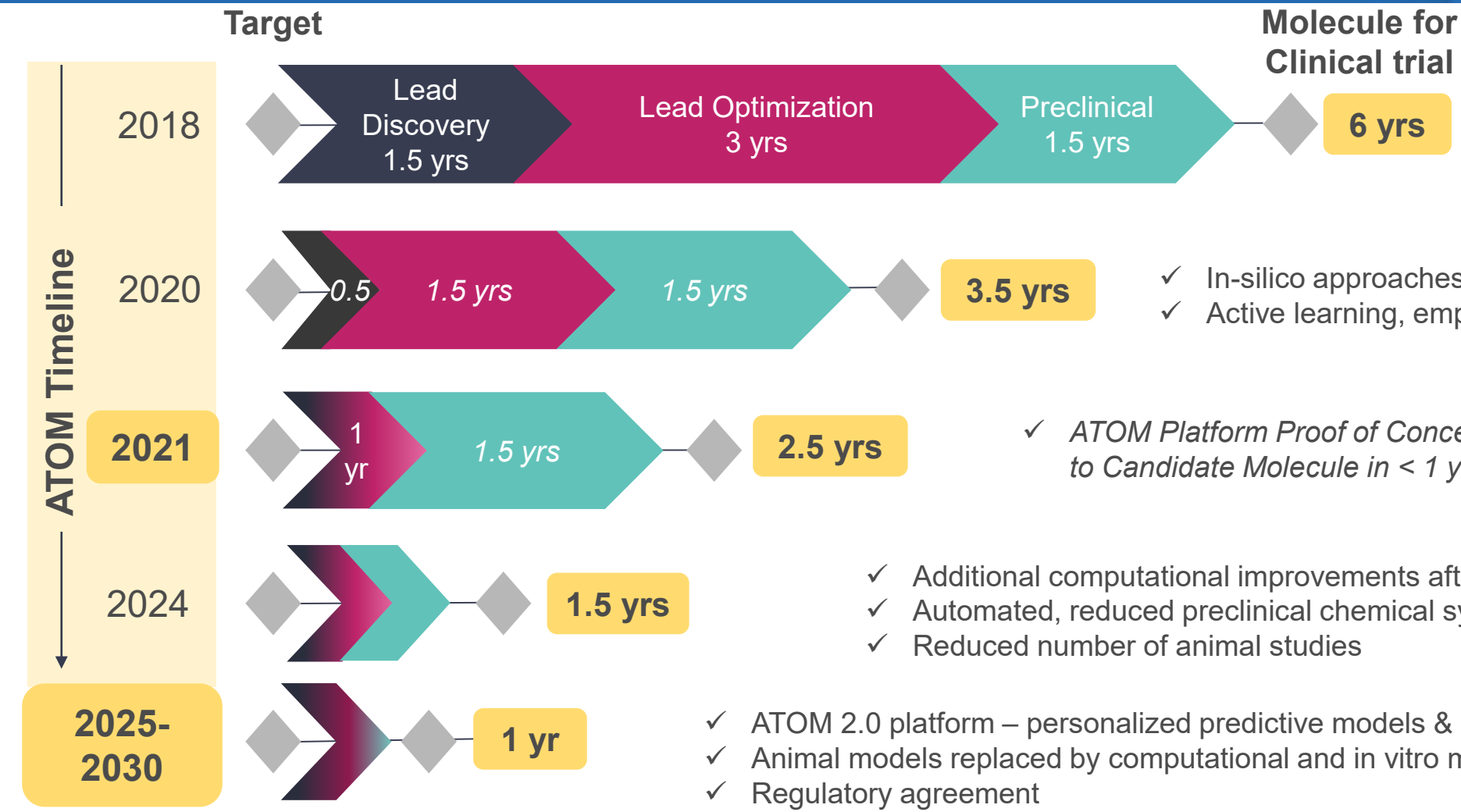
Chemistry Design & Synthesis

Experiment

ATOM target-to-clinical trial roadmap

Active learning approaches to accelerate timeline and reduce experimentation

Frederick National Laboratory for Cancer Research



- ✓ In-silico approaches to improve cycle time
- ✓ Active learning, empirical integration

- ✓ *ATOM Platform Proof of Concept (POC) initiated: Target to Candidate Molecule in < 1 year*

- ✓ Additional computational improvements after POC
- ✓ Automated, reduced preclinical chemical synthesis
- ✓ Reduced number of animal studies

- ✓ ATOM 2.0 platform – personalized predictive models & precision medicine
- ✓ Animal models replaced by computational and in vitro models
- ✓ Regulatory agreement

Data and Modeling Capabilities

Prediction

Data

Example data from 2 million GSK compounds in ATOM

Frederick National Laboratory for Cancer Research

Source	# Compounds	Specific Data Insights
Post-candidate selection programs no longer of interest to GSK	500	• In vitro and in vivo results (see below) • ~100 compounds with anonymized human clinical data
Unique compounds synthesized in lead optimization over last 17 years	515 k	• Structure-activity relationships with learnings on protein target pocket
Retired High Throughput Screening (HTS) compounds	1 M	• in vitro assays against diverse protein targets and physicochemical properties
Commercially available compounds in current HTS collection	420K	• in vitro assay data gained over the past ten years for a diversity of protein targets

Above data sets include, as available:

- in vitro enzyme and cell-based assay screening data against
- physicochemical experimental and calculated properties
- ex vivo ADME data
- in vivo pharmacokinetic, toxicokinetic, and animal safety data
- protein ligand crystal structures

ATOM has curated ~170 model-ready data sets

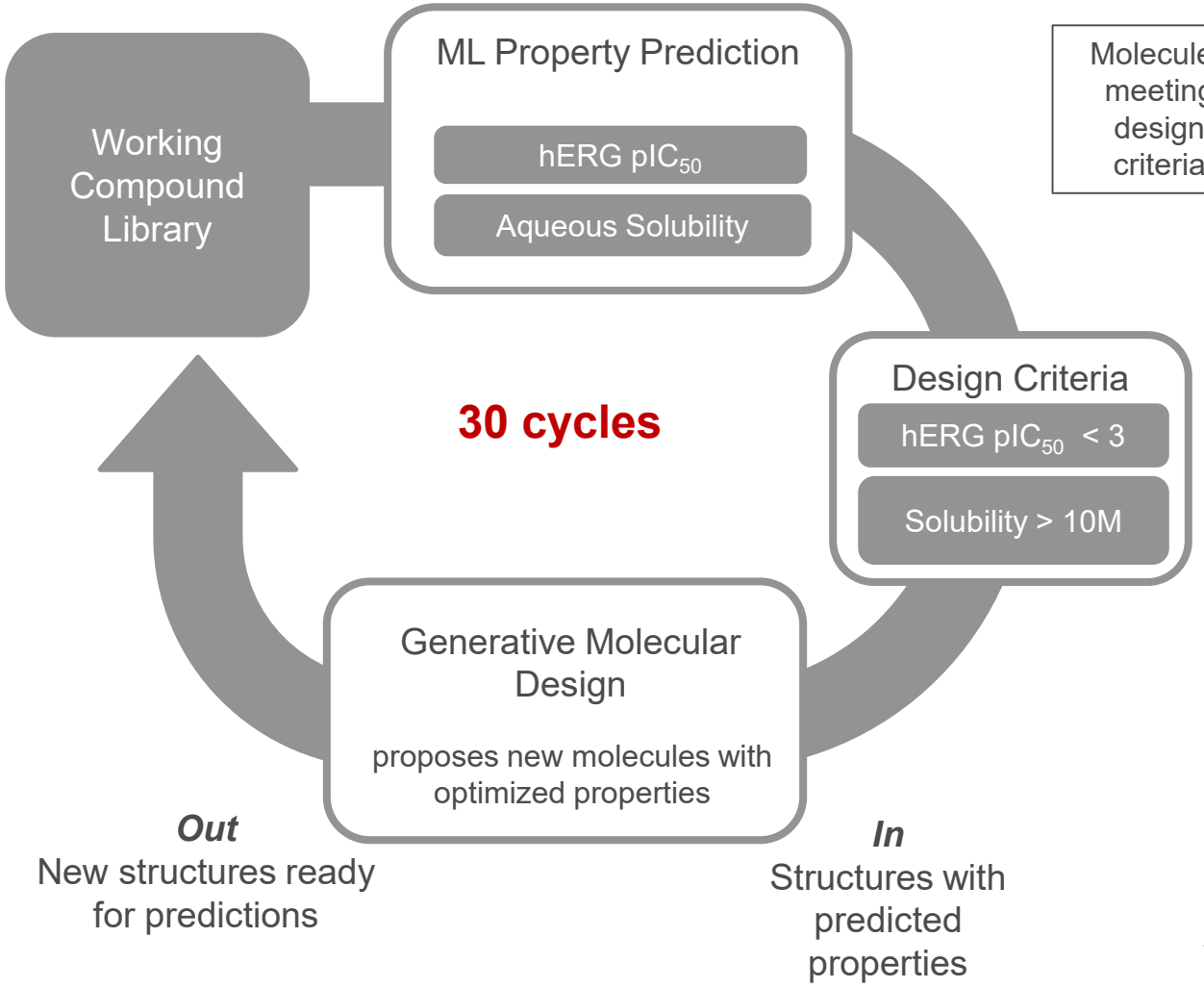
- **Safety Datasets**
  - 100-100,000 compounds per safety data type
  - Safety examples
    - ABCB11, bile salt export pump
    - Cytochrome P450 3A4 (CYP3A4), liver enzyme

- **Pharmacokinetic Datasets**
  - 100-100,000 compounds per PK data type
  - PK examples in rat, dog, and/or human
    - Parallel artificial membrane permeability assay
    - Hepatic clearance
    - Plasma protein binding

Example of *multi parameter optimization* loop

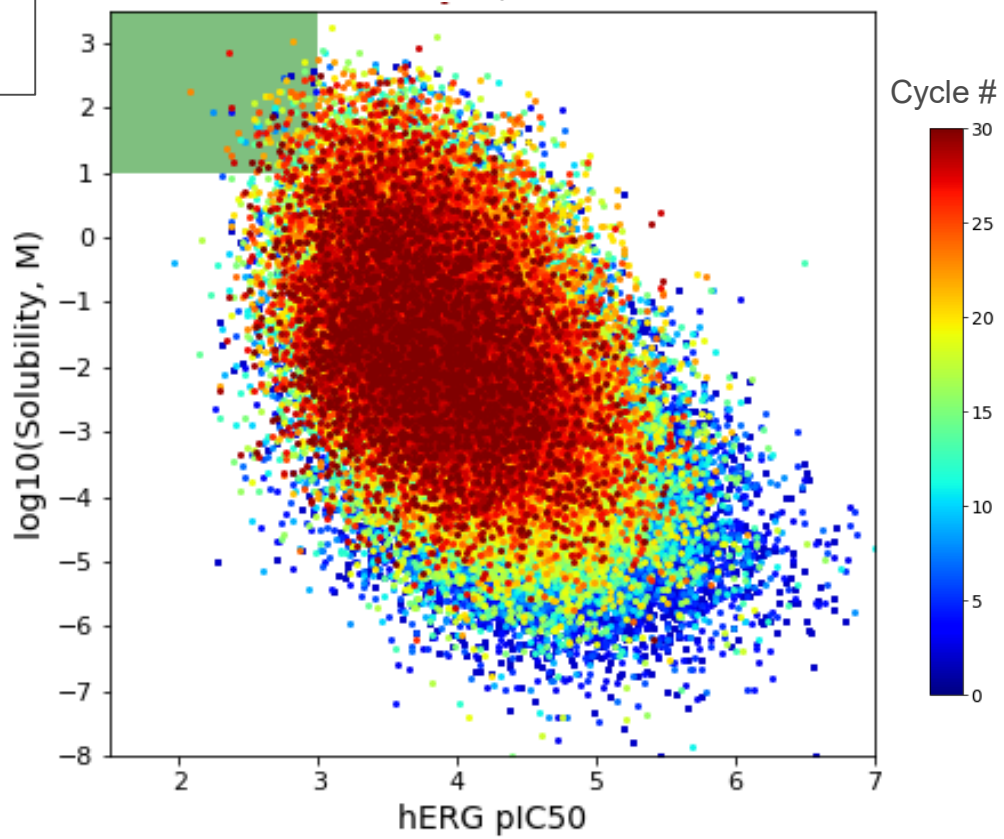
Iteratively run through loop to generate new compounds with better properties

Frederick National Laboratory for Cancer Research



Molecules meeting design criteria

Working compound library improving at each cycle



Active learning drives collection of new, targeted data when *in silico* improvements saturate

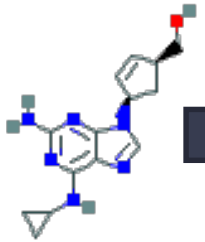
Current PK modeling activities

A limited first case – Single ML model

Frederick National Laboratory for Cancer Research

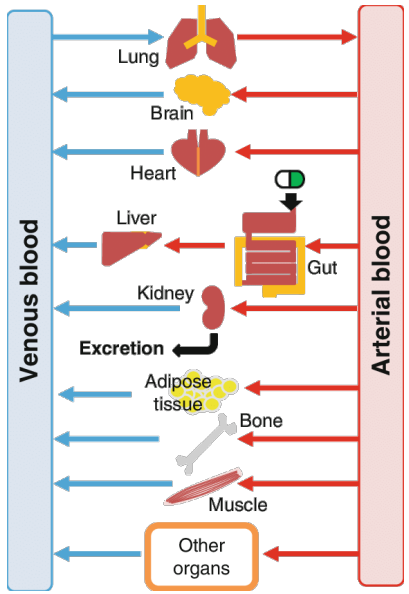
Structure to Input Parameters → **Human PK Simulator (PBPK)** → **Time-Concentration Profiles** → **Summary PK Parameters**

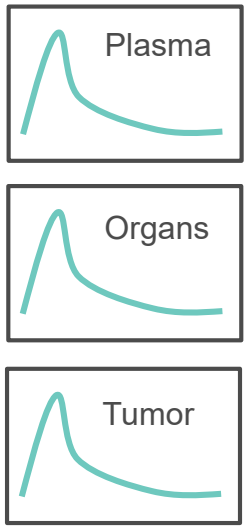
Parent Molecule(s)

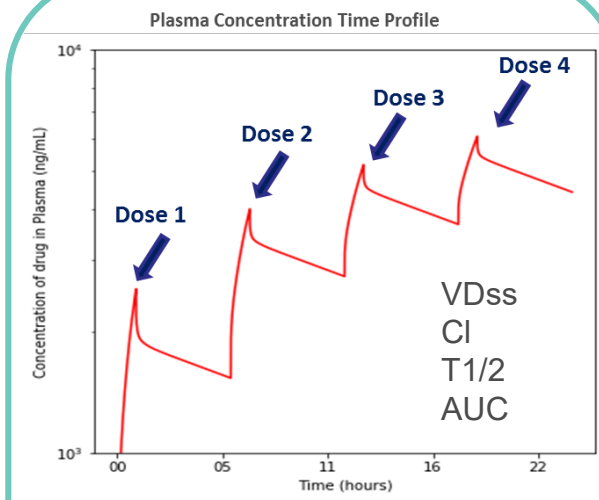


ATOM ML Models

- Parameters: solubility, permeability, pKa, B/P...
- Data curation of *in vitro* parameters from existing datasets
- *in vitro* experiments to fill data gaps

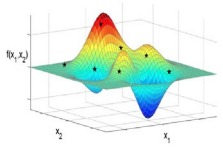






Curated human *in vivo* data (Obach, 2008; Lombardo, 2018)

Molecular Design & Active Learning



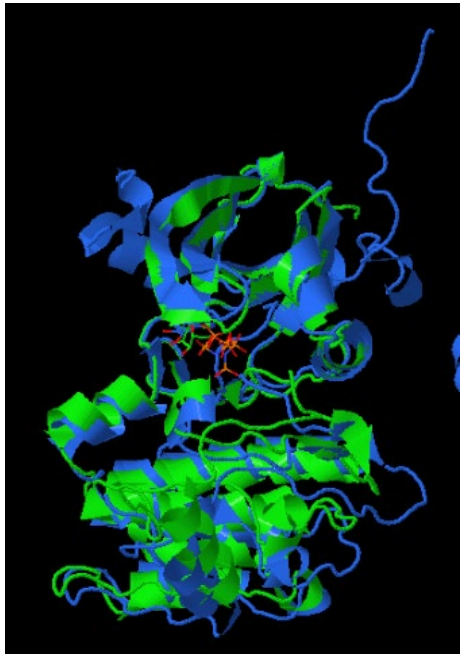
Generative Molecular Design

Pilot project for design loop build and validation

Generative molecular design (GMD) of potent, selective AURK B inhibitors

Frederick National Laboratory for Cancer Research

PILOT 1 | Starting point: Early program data | Lead Optimization | End point: Experimental validation

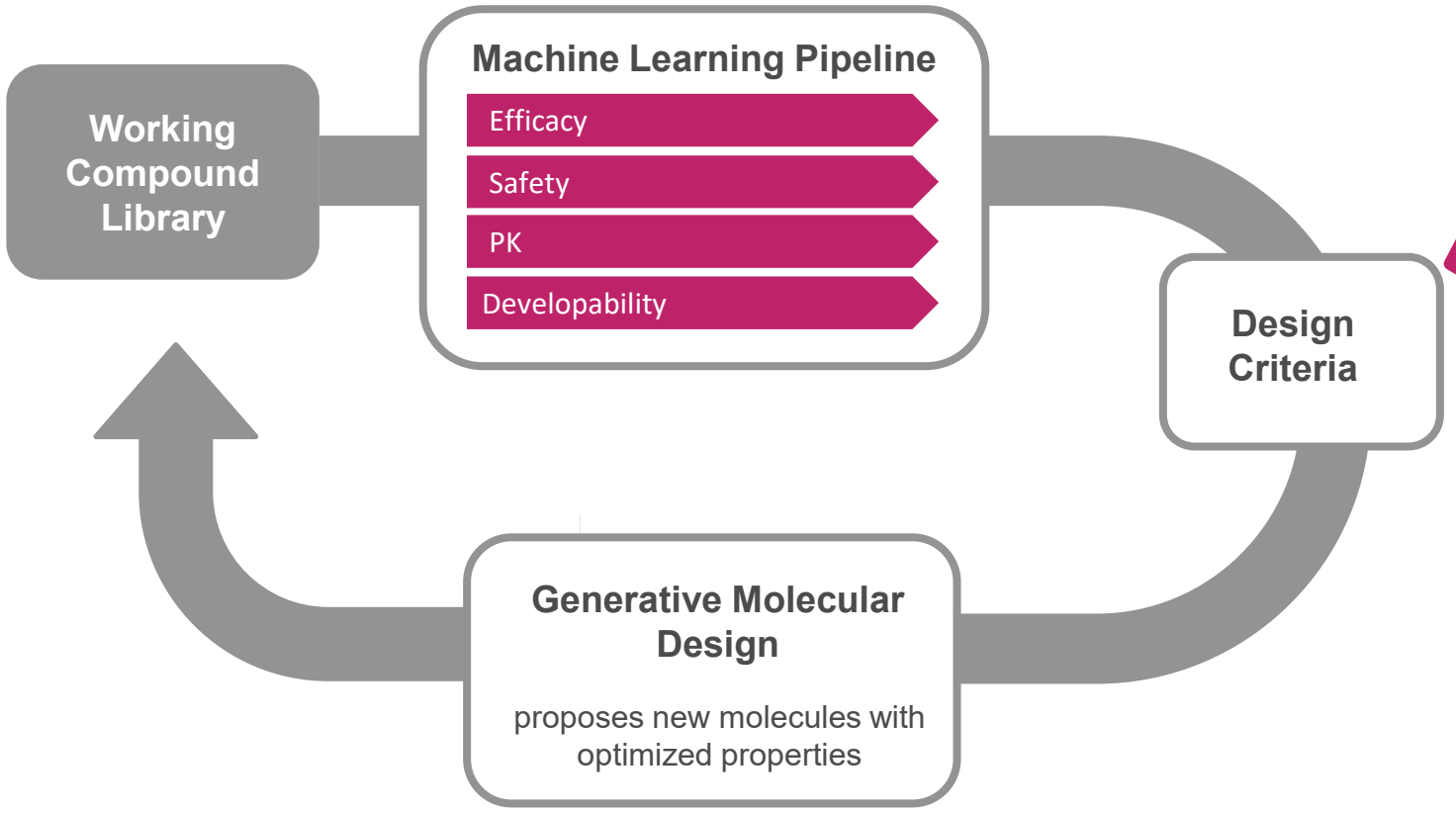


Structure overlay of AURK A and AURK B

Why Aurora Kinase?

- Cancer relevant: >30 clinical trials are ongoing or completed for AURKA selective, AURKB selective, and AURKA/B dual inhibitors
- Data available at ATOM: Potency data on ~24k compound available for AURK B and/or AURK A
- Pharmaceutical discovery relevant problem: Selectivity between kinases is an important and common pharmaceutical discovery problem

AURK Pilot Design Criteria



Candidate Quality Panel

Efficacy
AURK B (pIC_{50} > 9)
AURK B/A Selectivity > 1000

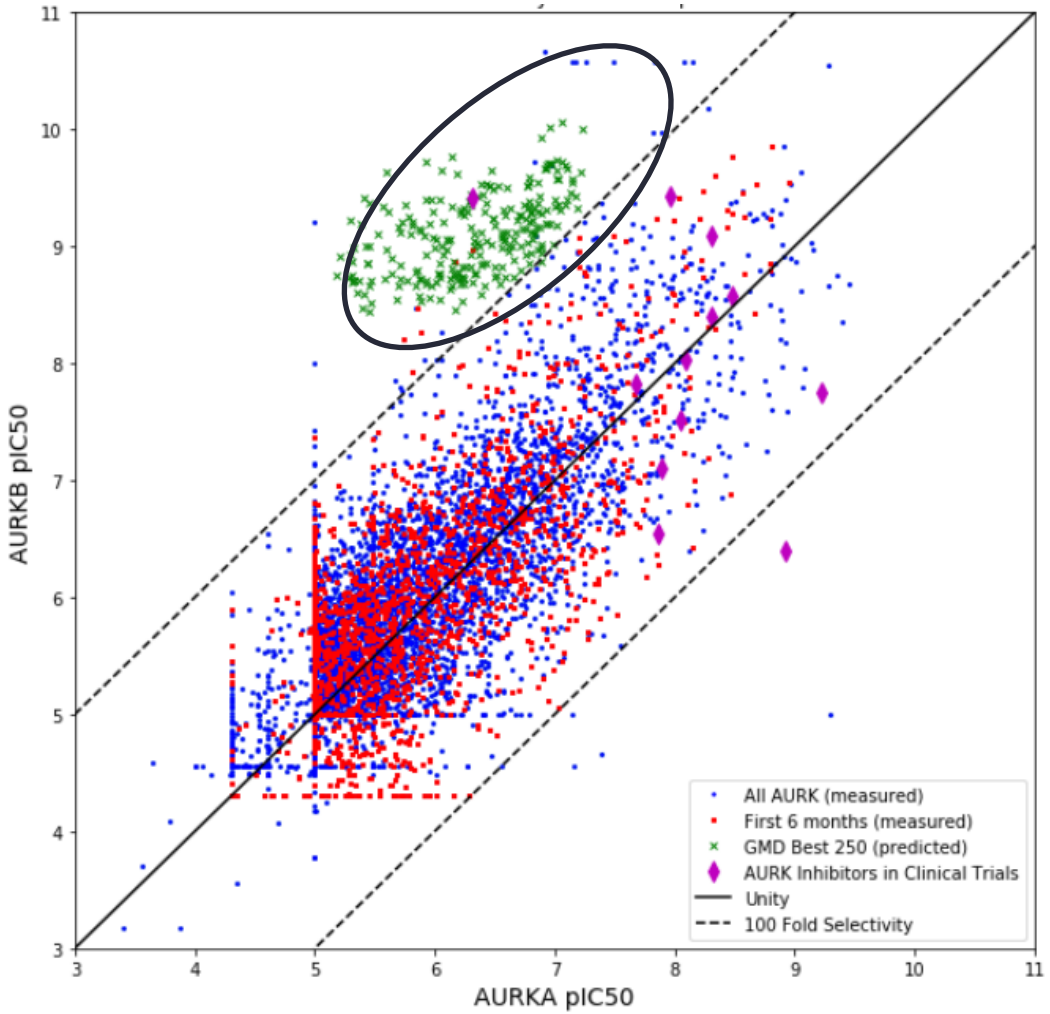
Safety
BSEP (pIC_{50} < 4)
hERG (pIC_{50} < 4)

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

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

Initial pilot results: >200 new potent, selective AURK B compounds with favorable other properties

Frederick National Laboratory for Cancer Research



Multi-Parameter Optimization:

AURKB pIC50	AURKA pIC50	A/B Selectivity (fold)	hERG pIC50	BSEP pIC50	Solubility	hLM CLint	Solubility	SAS
9.627	5.60	10772	3.260	4.010	6.022	1.819	412.492	2.640
9.724	5.92	6381	3.202	4.029	4.241	1.338	69.457	2.632
9.762	6.14	4174	3.197	4.027	4.535	1.322	93.249	2.410
9.298	5.98	2065	3.198	3.969	5.988	1.455	398.809	2.392
9.209	5.73	3024	3.200	4.027	7.000	4.371	1096.282	2.498
9.208	5.81	2477	3.195	4.027	5.413	1.868	224.400	2.397
9.626	6.18	2784	3.868	3.982	5.447	1.434	232.073	2.332
9.407	5.41	9984	3.259	4.018	3.704	1.252	40.620	2.784
9.353	5.75	4028	3.199	4.018	4.470	1.835	87.357	2.339
9.517	6.45	1160	3.223	3.976	4.353	2.024	77.733	2.222
9.252	5.79	2922	3.794	3.977	5.207	1.405	182.459	2.441
9.293	5.61	4851	3.197	3.994	4.006	1.479	54.916	2.627
9.334	5.56	5926	3.198	4.043	6.552	0.986	700.482	2.818
9.393	5.93	2911	3.198	4.026	5.343	1.595	209.163	2.624
9.397	6.05	2247	3.199	4.016	4.017	1.421	55.541	2.640
9.399	5.97	2682	3.211	3.993	3.554	1.632	34.955	2.255
9.193	5.96	1720	3.646	3.970	5.044	1.816	155.047	2.472
9.222	5.30	8342	3.215	4.048	5.936	0.888	378.391	2.628
9.327	6.25	1205	3.198	4.055	6.356	1.498	575.970	2.380
9.440	6.39	1116	3.380	3.968	4.635	1.775	103.039	2.361
9.129	5.88	1775	3.657	4.070	7.134	1.553	1254.501	2.278
9.338	6.14	1579	3.198	3.967	3.507	1.269	33.360	2.369
9.516	6.46	1136	3.202	4.067	6.818	0.858	913.920	2.464
9.278	6.21	1171	3.416	4.069	4.565	1.777	96.042	2.330
9.090	5.91	1509	3.210	4.052	6.827	0.880	922.242	2.433
9.365	6.43	869	3.210	4.020	6.059	0.936	427.917	2.686
9.107	5.53	3788	3.545	3.993	6.339	3.112	565.978	2.501
9.375	5.45	8340	3.199	4.022	2.663	1.340	14.338	2.754
9.650	6.05	3951	3.205	3.990	2.503	1.604	12.224	2.426
8.896	5.64	1821	3.209	4.027	6.561	1.374	707.083	2.181
9.648	6.48	1482	3.244	4.021	4.026	1.318	56.041	2.577
9.389	6.28	1284	3.198	4.055	5.250	1.037	190.604	2.754
9.075	5.98	1235	3.199	4.055	6.027	1.711	414.475	2.527
9.422	6.35	1179	3.202	4.014	3.214	1.569	24.878	2.503
9.298	6.27	1063	4.026	4.058	5.720	1.863	304.910	2.474
9.108	5.80	2028	3.198	4.115	5.448	0.973	232.219	2.608
8.969	5.60	2333	3.925	3.987	5.674	3.901	291.332	2.224
9.162	5.70	2884	3.198	4.069	4.387	0.800	80.426	2.520
9.154	5.87	1907	3.198	4.024	3.459	1.411	31.775	2.319
9.294	6.41	767	3.253	4.000	4.356	0.939	77.924	2.522

Steps for validation

- Comparison to held out, ground truth data (r^2 =0.74)
- Experimentally make and test top compounds

ATOM

ATOM Milestones Towards Proof of Concept

Frederick National Laboratory for Cancer Research

2017

ATOM established as a place

R&D plan developed & initiated

2018

Curated 2M GSK compounds into 169 model-ready sets

Data-Driven Modeling Pipeline built and validated.

Baseline PK & Safety models

Generated new data to expand models

2019

Completion and initial validation of human system-level PBPK models

Application on pilot project(s)

New data and models for DILI

Scale design loop

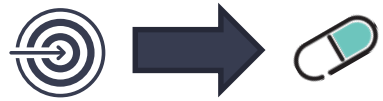
2020

Active Learning platform integrated, validated

Proof of concept project initiated

2021

POC project delivered



<12 months

Target to molecular candidate

ATOM Publications for 2019

Frederick National Laboratory for Cancer Research

(Working) Title	Article Type
Artificial Intelligence & Pharmacometrics: Time to Embrace, Capitalize and Advance?	Commentary doi: 10.1002/psp4.12418
Rethinking Drug Design – The Impact of Artificial Intelligence	Perspective Submitted
Data-Driven Modeling (DDM) for Drug Discovery*	Research, DDM
Evaluation of *in silico* Methods for Predicting Volume of Distribution in Humans for Greater than 300 Structurally Diverse Compound*	Research, PK
ATOM Consortium: A New Paradigm for Drug Discovery	Perspective
Using Manifold Learning to Predict Inhibition of the Bile Salt Export Pump (BSEP)*	Research, BSEP
Generative Molecular Design (GMD) for Lead Optimization: Demonstration by Discovery of a Novel Potent, Selective AURK B Inhibitor*	Research, GMD Pilot
Predictive *in vitro* and *in silico* strategies for drug induced liver injury	Review, DILI

*Associated data & model releases are subject to approval by Governing Board – up to one (1) year after Technology Readiness

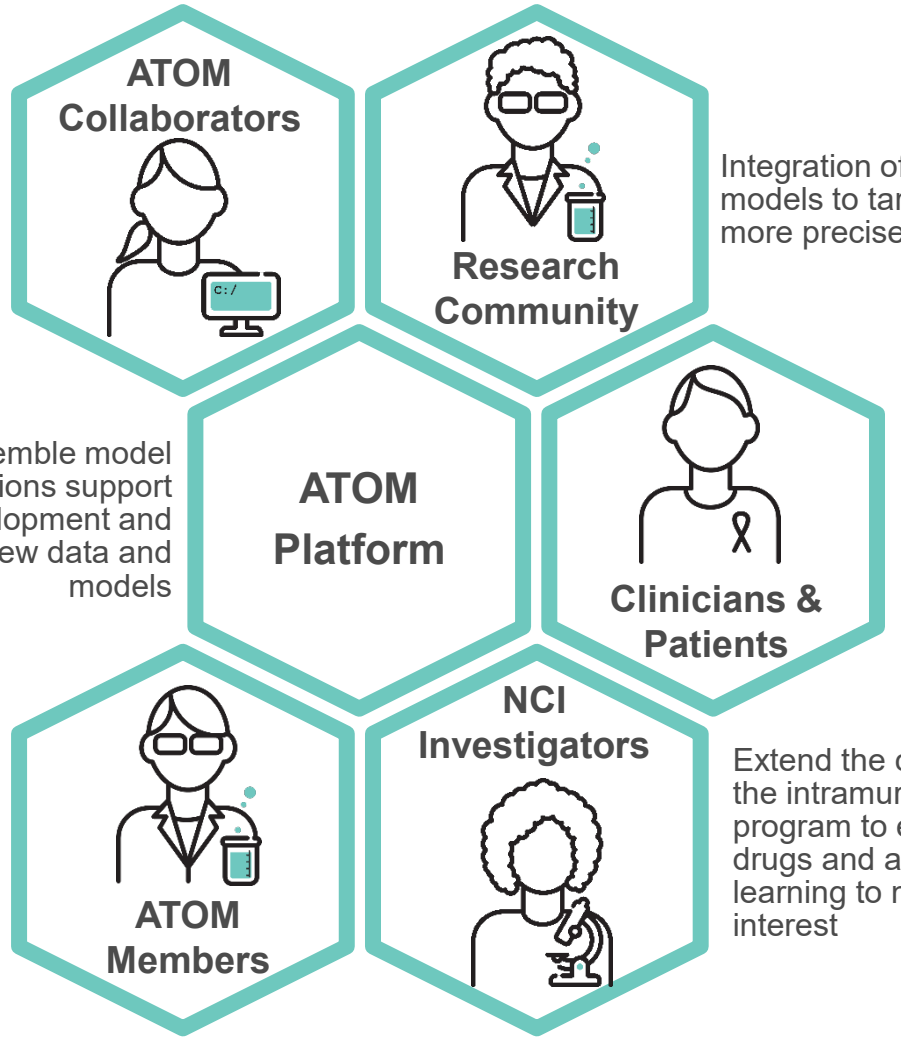
Summary/accomplishments

Frederick National Laboratory for Cancer Research

- Built a data-driven modeling computational pipeline to serve as the foundation of the ATOM platform
- Generated baseline PK and safety parameter prediction models
- Demonstrated Generative Molecular Design capability with the AURK Pilot (follow-up experiments in progress)
- Expanding engagement at the Member and Collaborator levels

ATOM Platform: a resource for research community

Frederick National Laboratory for Cancer Research



Benchmarking and refinement of models and experimental technologies

Integration of multiple models to target cancer more precisely

Ensemble model predictions support development and integration of new data and models

Patient specific data used as input to enable precision medicine

Employing active learning to accelerate drug discovery and generate novel compounds

Extend the capabilities for the intramural research program to explore new drugs and apply active learning to new areas of interest

The ATOM Team

Frederick National Laboratory for Cancer Research

ATOM One Year Anniversary | October 22, 2018
San Francisco, CA | atomscience.org

Supplemental Slides

Organizations under CDA in BD pipeline

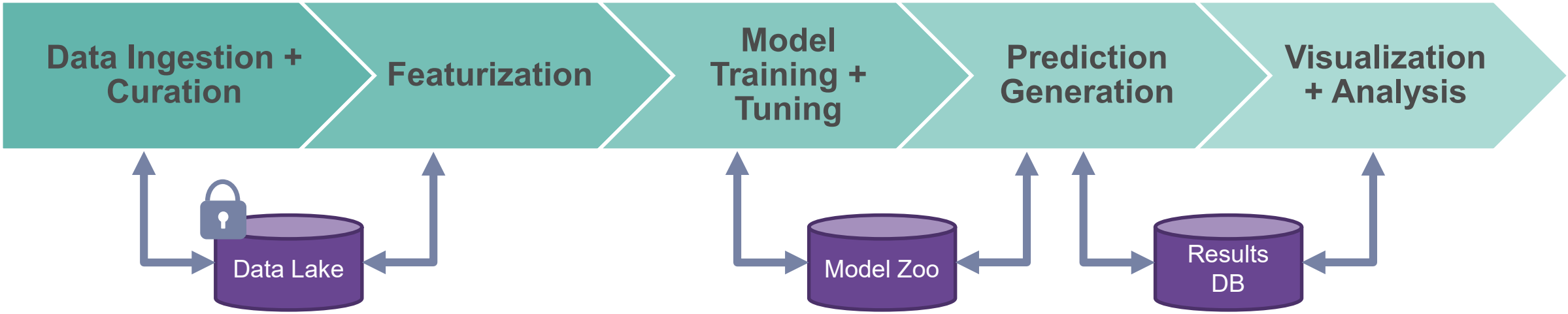
Frederick National Laboratory for Cancer Research

Sector	Vision & Technical Alignment	Consortium Contributions / SOW	Term Sheet	Contract In Preparation	Contract Executed
Pharma	Pharma 1 Pharma 2 Pharma 3				
Gov't Org				DOE Lab 1 DOE Lab 2 DOE Lab 3	
Hardware Tech				NVIDIA	
Academia		University 1 University 2 University 3			
Software	Software 4	Software 2 Software 3	Software 1	AI modeling 1 AI modeling 2 AI modeling 3	DeepChem Advisor
Data and Experimental		Exp. 2	Exp. 1		

End-to-End Data-Driven Modeling (DDM) Pipeline

Common infrastructure in place and ready to receive/transform new data

Frederick National Laboratory for Cancer Research



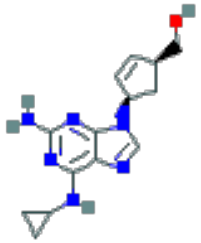
Benefits:

- Easy integration of diverse datasets
- Enables portability of models and reproducibility of results
- Rapid exploration of descriptors and model architectures for each data set

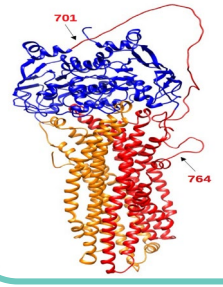
DILI Experimental & Computational Workflow

Frederick National Laboratory for Cancer Research

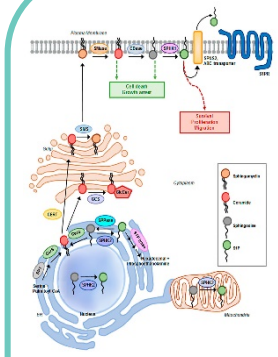
~200 Compounds



In vitro Screening



Protein level transporter inhibition assays



Cellular and pathway level assays:

- Cell painting
- Cell health
- Mitochondrial stress
- hepatocytes & HepG2 cells

In silico Modeling

Protein level modeling of *in vitro* data

Integrate with DILIsym software:

- ML reduction of DILIsym
- Integration of ML models of *in vitro* screening assays into DILIsym

Outcomes

- Human level prediction of hepatotoxicity
- Most predictive in vitro stratefy
- Most predictive in silico modeling platform
- Construct quantitative structure activity relationship DILI model
- Integrate with Generative Molecular Design loop

ATOM Platform supports an integrated workforce

Frederick National Laboratory for Cancer Research

ATOM Platform

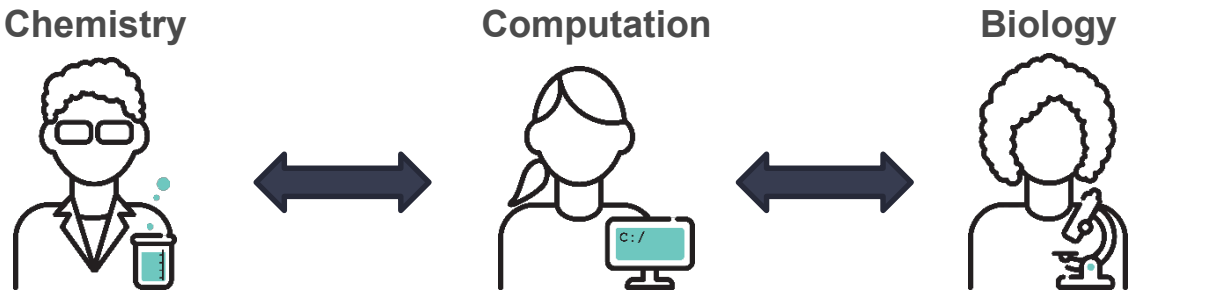
Pharma, public data, new consortium data → Simultaneous predictions and optimized molecules → Active learning and automated experimentation

New data

AI won't replace chemists, but chemists with AI will replace chemists without it.

Jay Bradner, President, Novartis Institutes for BioMedical Research

Integrated Research Workforce



More challenging synthetic problems

Proprietary modelling efforts

Target specific assays