CDD VAULT®
Complexity Simplified

Computation in Drug Discovery — An Insider's View

December 1, 2022

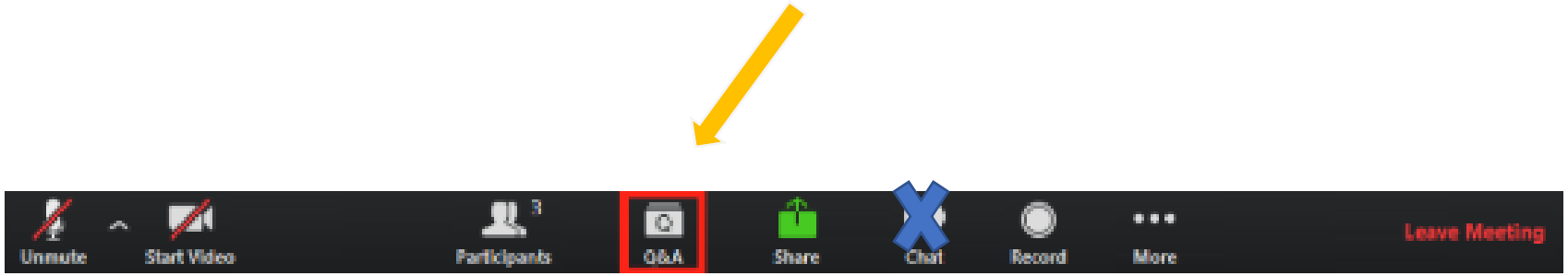
Featuring these leading scientists...

Eric Gifford, PhD
Customer Success Scientist,
Collaborative Drug Discovery

Joshua Horan, PhD
VP - Medicinal Chemistry,
Nuvalent, Inc.

Michael Kappler, PhD
Director, Head of Research
Infomatics,
IDEAYA Biosciences

Have a question to ask our panel?
Open the ZOOM Q&A and type in your question at anytime !

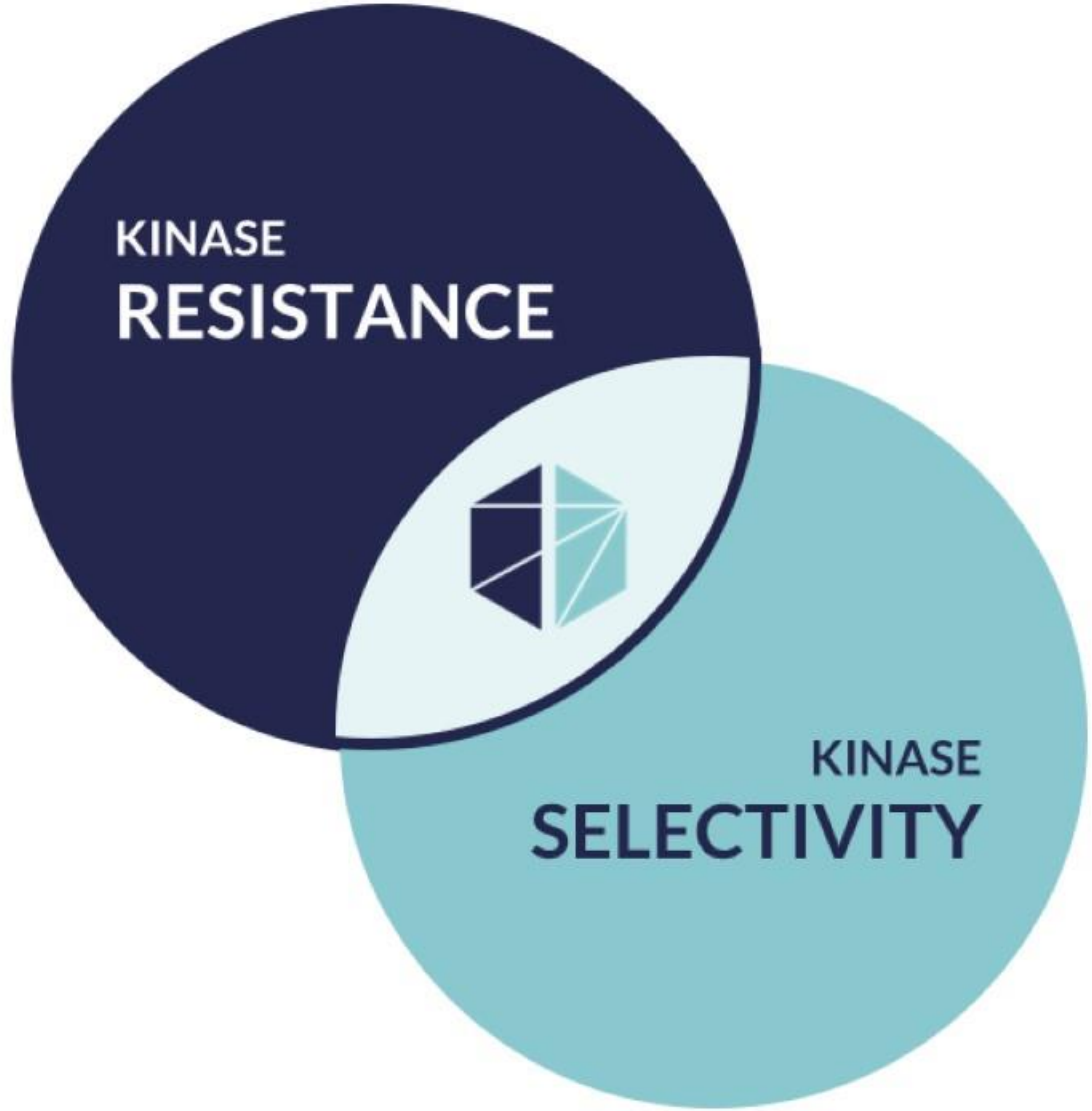


Saving your Questions to the end

PRECISELY Targeted Therapies
for patients with cancer

Nuvalent is focused on creating ***precisely*** **targeted therapies** to overcome key limitations of existing therapies for clinically proven kinases and renew hope for patients in need

- Expertise in structure-based drug design to create innovative small molecules
- "**Threading the needle**": Aim to achieve high affinity for drug-resistant kinases while avoiding off-target kinases in the central nervous system (CNS) and in the periphery
    - Potential to minimize adverse events **AND** drive more durable responses



Nuvalent

Nuvalent Pipeline – Ongoing Programs

Advancing parallel lead programs in ROS1-positive and ALK-positive NSCLC, and multiple early-stage discovery programs

LEAD INDICATION	PRODUCT CANDIDATE	SELECTED MUTATION(S)	DISCOVERY / IND ENABLING / PHASE 1 / PHASE 2 / PHASE 3	ANTICIPATED MILESTONES	WORLDWIDE RIGHTS
ROS1 NSCLC	NVL-520	G2032R, S1986Y/F, L2026M, D2033N	***Open & Enrolling:*** Phase 1 portion of Phase 1/2 trial → ARROS-1	Preliminary dose-escalation data presented at EORTC-NCI-AACR 2022	Nuvalent
ALK NSCLC	NVL-655	G1202R G1202R/L1196M G1202R/G1269A G1202R/L1198F	***Open & Enrolling:*** Phase 1 portion of Phase 1/2 trial → ALKoVE-1		Nuvalent
ALK NSCLC		I1171X / D1203N (X = N, S, or T)	→		Nuvalent
HER2 NSCLC	NVL-330	Exon 20 Insertions	→	Preclinical profile presented at EORTC-NCI-AACR 2022	Nuvalent
Additional Discovery Research Programs Ongoing					

Nuvalent

IDEAYA Synthetic Lethality Drug Discovery Platform

Structure-Based Drug Design & Proprietary Chemical Library Enable "Hard to Drug" Targets

Structural Biology & Structure Based Drug Design

Full suite of capabilities in structural biology, biophysics, & computational chemistry

Ligand bound co-crystal structures resolved to enable Structure Based Drug Design for programs

Multiple potential "first-in-world" co-crystal structures resolved, including for PARG, Pol Theta Helicase and Werner Helicase

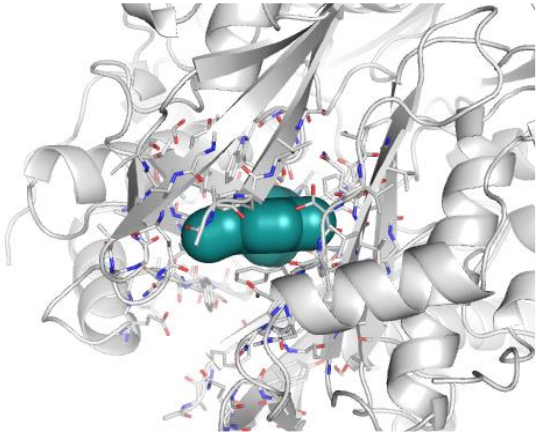
Harmony-ML™ Proprietary Machine-Learning

Our internal ML engine empowers our discovery platform through effective prioritization leading to efficient cycle times

INQUIRE™ Proprietary Chemical Library

Expert-curated HTS library to enhance hit discovery capabilities against novel SL targets classes, such as helicases and endonucleases

Enhances IDEAYA's SL Drug Discovery Platform and competitive differentiation



IDEAYA's Precision Medicine Oncology Pipeline

Building the Industry Leading Synthetic Lethality Focused Biotechnology Company

Precision Medicine Pipeline

	Modality/Indication	Biomarker	Preclinical	IND Enabling	Phase 1	Phase 2	Program Goals	Collaborations	Commercial (IDEAYA)
Darovasertib *PKC*	+cMET[1] Combination MUM, Basket	GNAQ/11					Daro + Crizo Clincal Update in MUM √ Daro + Crizo Reg Trial in MUM Q1 2023	Pfizer (1)	WW Commercial Rights
	Adjuvant UM	GNAQ/11					(Neo)Adjuvant UM – Phase 1 IST √ (Neo)Adjuvant UM – IDEAYA Phase 1 Q4 2022		
	+cMET[1],+KRAS Combo NSCLC, HCC	MET KRAS					+cMET[3] - IND Q1 2023 KRAS – Preclinical Evaluation Q4 2022	Pfizer (1)	
IDE397 *MAT2A*	Monotherapy NSCLC, Esophagogastric	MTAP					Mono Expansion Phase 2 Initiation √		WW Commercial Rights
	Combinations Solid Tumors	MTAP					Combination Cohorts Ph1 Initiation √	AMGEN (2)	
IDE161 *PARG*	Ovarian, Gastric, Breast Cancers	HRD					IND Q4 2022	CANCER RESEARCH UK (3)	WW Commercial Rights
Pol Theta	Small Molecule Helicase Inhibitor	HRD					Development Candidate Q2 2022 √ First-in-Human Studies H1 2023	GSK (4)	Global Royalties
WRN	GI Cancers	High-MSI					Development Candidate 2023	GSK (4)	US 50/50 Profit Share Ex-US Royalties
MTAP-SL	Solid Tumors	MTAP					Lead Series		WW Commercial Rights
SL Platform	Solid Tumors	Defined Biomarker					Lead Series New Target / Biomarker Validation		WW Commercial Rights

(1) Pursuant to Pfizer Clinical Trial Collaboration and Supply Agreements for Darovasertib/Crizotinib Combination in MUM and in cMET-driven Tumors; IDEAYA retains all Darovasertib Commercial Rights
(2) Pursuant to Amgen Clinical Trial Collaboration and Supply Agreement for IDE397 + AMG 193, an investigational MTA-cooperative PRMT5 inhibitor; Amgen will sponsor the study and the parties will jointly share external costs of the study
(3) Pursuant to CRUK Evaluation, Option and License Agreement; IDEAYA controls all PARG Commercial Rights
(4) Pursuant to GSK Collaboration, Option and License Agreement: Polθ: Global Royalties; WRN: 50/50 US Profits + ex-US Royalties

= Target Program Milestones

MAT2A=methionine adenosyltransferase 2a, MTAP=methylthioadenosine phosphorylase, MTA=methylthioadenosine, PRMT5=protein arginine methyltransferase 5 (PRMT5), PARG= poly (ADP-ribose) glycohydrolase, DDT = DNA Damage Target, WRN = Werner Helicase, Polθ = DNA Polymerase Theta, HRD = homologous recombination deficiency, MSI = microsatellite instability, PKC = protein kinase C, MUM = metastatic uveal melanoma, cMET = tyrosine kinase protein MET , Crizo = crizotinib, NSCLC = non-small cell lung cancer, HCC= hepatocellular carcinoma WW = worldwide

Requires Authentication Published by De Gruyter June 16, 2022

BioChemUDM: a unified data model for compounds and assays

Michael A. Kappler, Christopher T. Lowden and J. Chris Culberson

From the journal Pure and Applied Chemistry
https://doi.org/10.1515/pac-2021-1004

Abstract

We present a simple, biochemistry data model (BioChemUDM) to represent compounds and assays for the purpose of capturing, reporting, and sharing data, both biological and chemical. We describe an approach to register a compound based solely on a stereo-enhanced sketch, thereby replacing the need for additional user-specified "flags" at the time of compound registration. We describe a convention for string-based labels that enables inter-organizational compound and assay data sharing. By co-adopting the BioChemUDM, we have successfully enabled same-day exchange and utilization of chemical and biological information with various stakeholders.

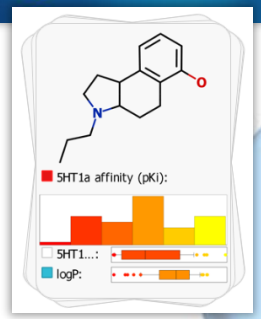
Keywords: Cheminformatics; data sharing; pharmaceutical informatics; unified data model

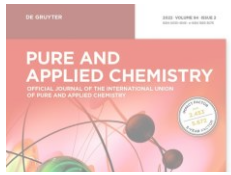
Modern Architecture to Broadly Leverage Information Across Domains

Platform & Visualization Landscape

optibrium

StarDrop™





DE GRUYTER Pure Appl. Chem. 2022; aop

Editorial

Michael A. Kappler*, Christopher T. Lowden and J. Chris Culberson

BioChemUDM: a unified data model for compounds and assays

https://doi.org/10.1515/pac-2021-1004

Abstract: We present a simple, biochemistry data model (BioChemUDM) to represent compounds and assays for the purpose of capturing, reporting, and sharing data, both biological and chemical. We describe an approach to register a compound based solely on a stereo-enhanced sketch, thereby replacing the need for additional user-specified "flags" at the time of compound registration. We describe a convention for string-based labels that enables inter-organizational compound and assay data sharing. By co-adopting the Bio-ChemUDM, we have successfully enabled same-day exchange and utilization of chemical and biological information with various stakeholders.

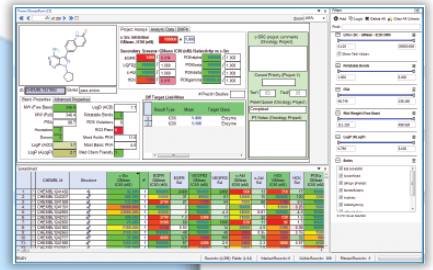
Keywords: Cheminformatics; data sharing; pharmaceutical informatics; unified data model.

SCHRÖDINGER.

LiveDesign®

CERTARA

D360



Data Lake

CDD VAULT®

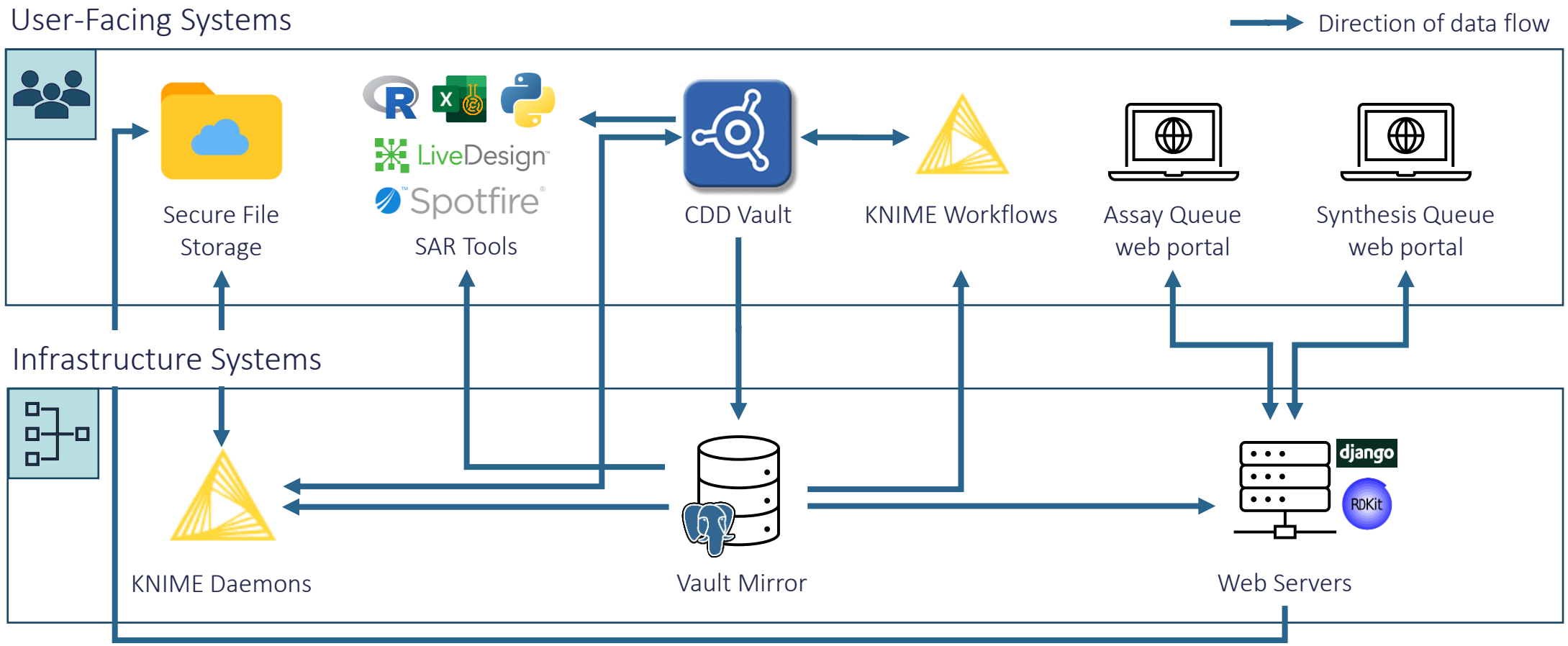
mosaic

titian

IDEAYA BIOSCIENCES

Nuvalent Informatics Overview

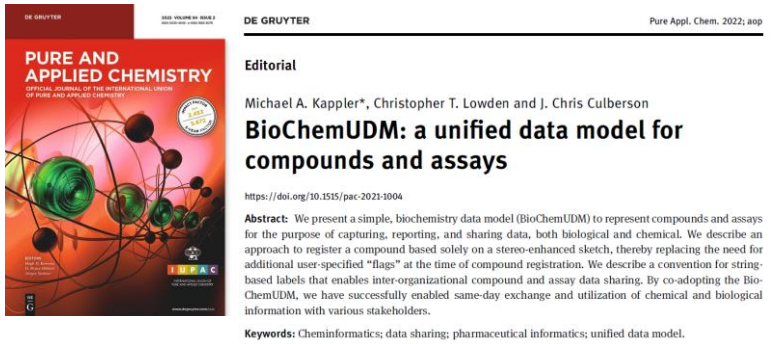
Multiple systems center around a central Vault data repository

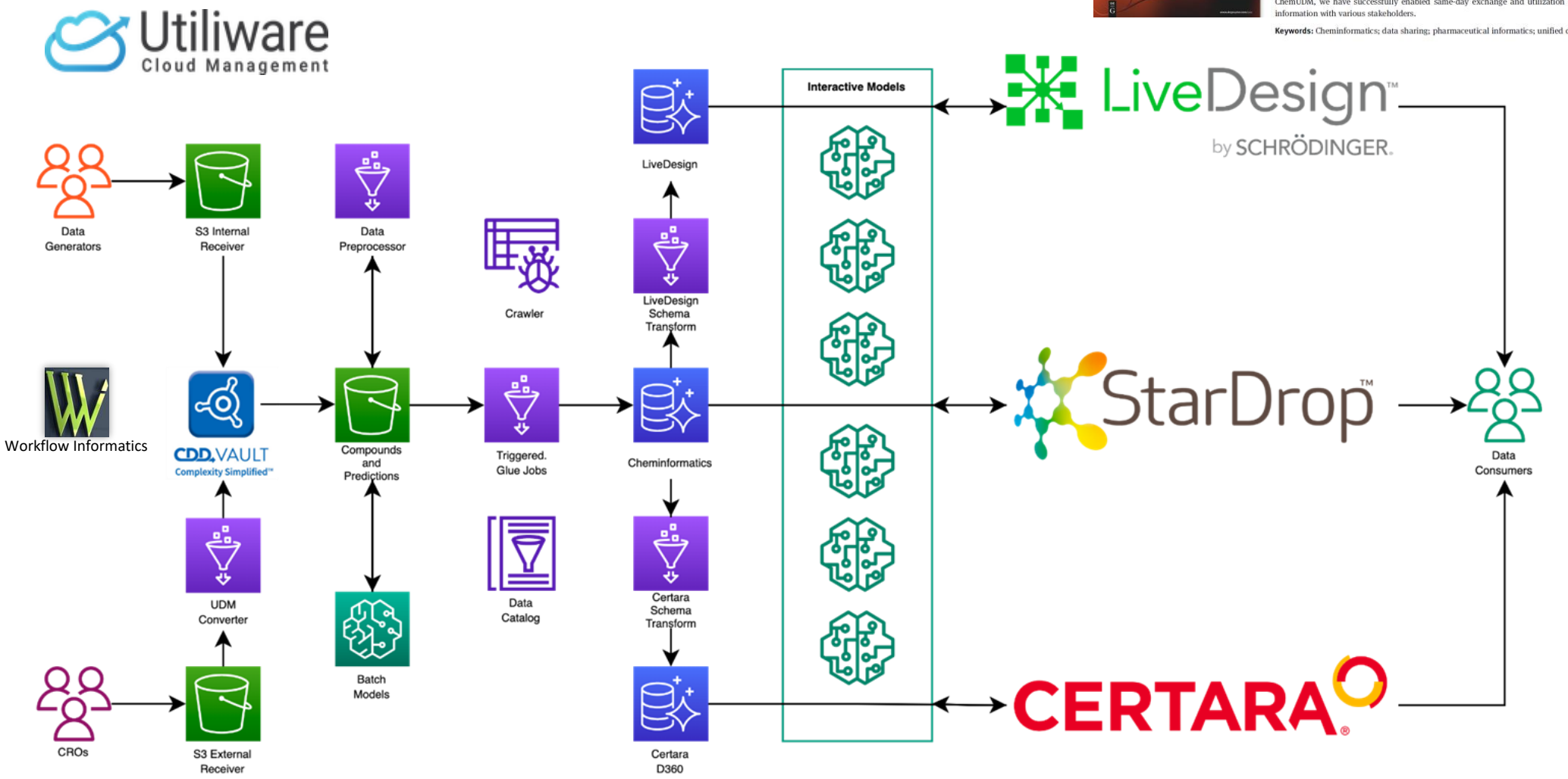


Nuvalent

Machine-Learning Framework

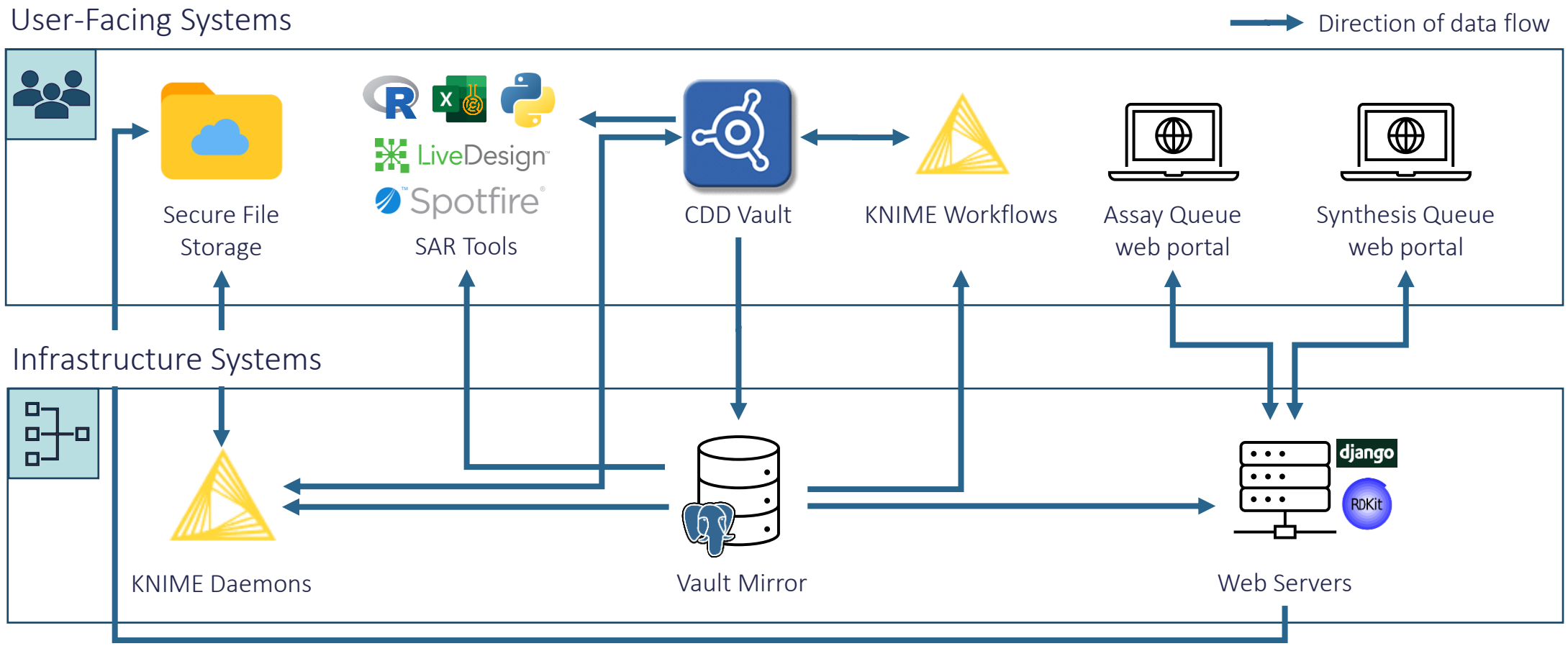
Factory Concept based on AWS SageMaker





Nuvalent Informatics Overview

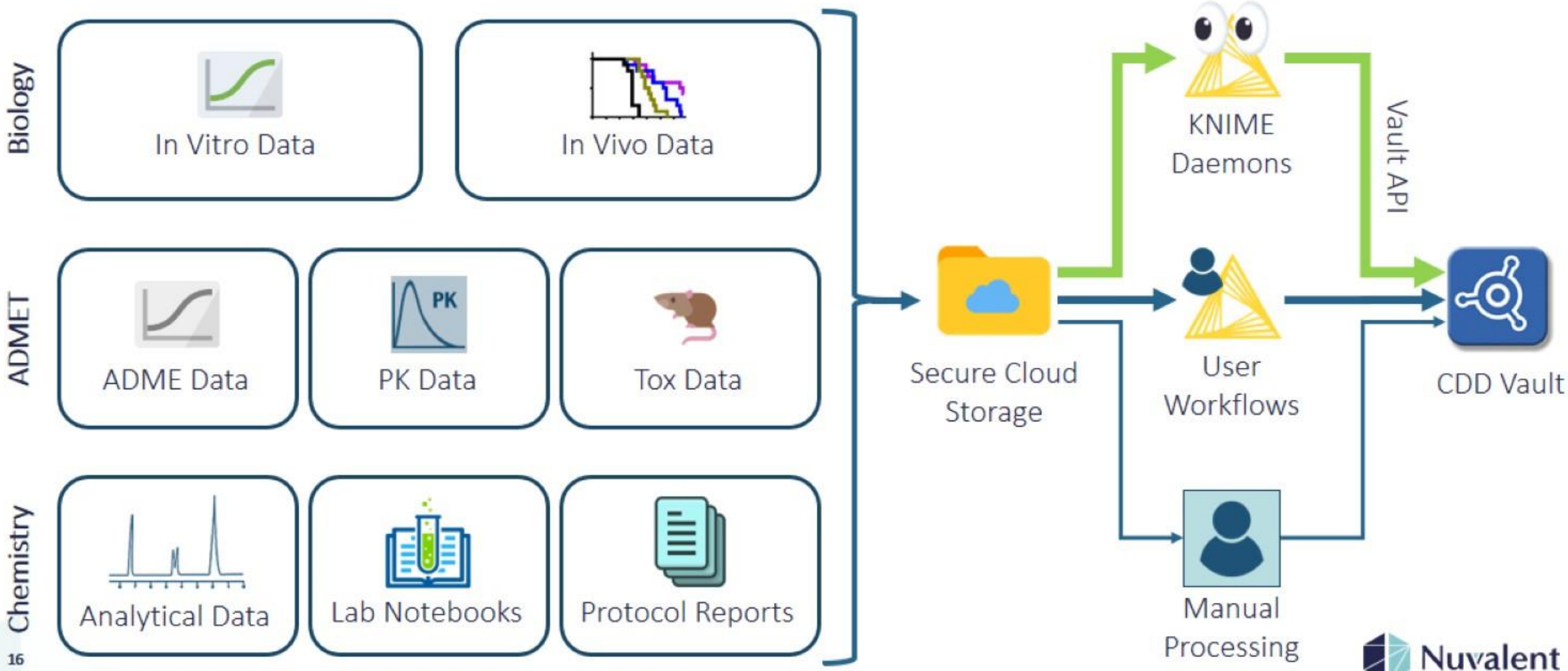
Multiple systems center around a central Vault data repository



Nuvalent

Managing the “Data Firehose” of Drug Discovery

Organizational principle: prioritize datasets and automate as much as possible



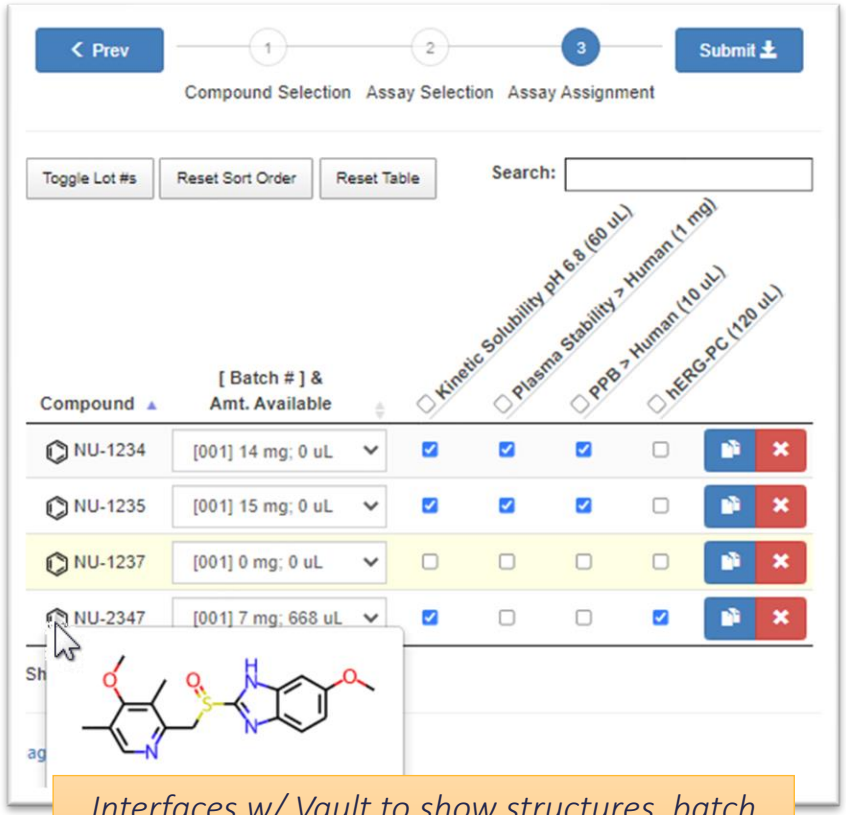
Nuvalent

Enhancing Collaboration: Assay Requests

Assay queue web-portal brings transparency to assay ordering and management

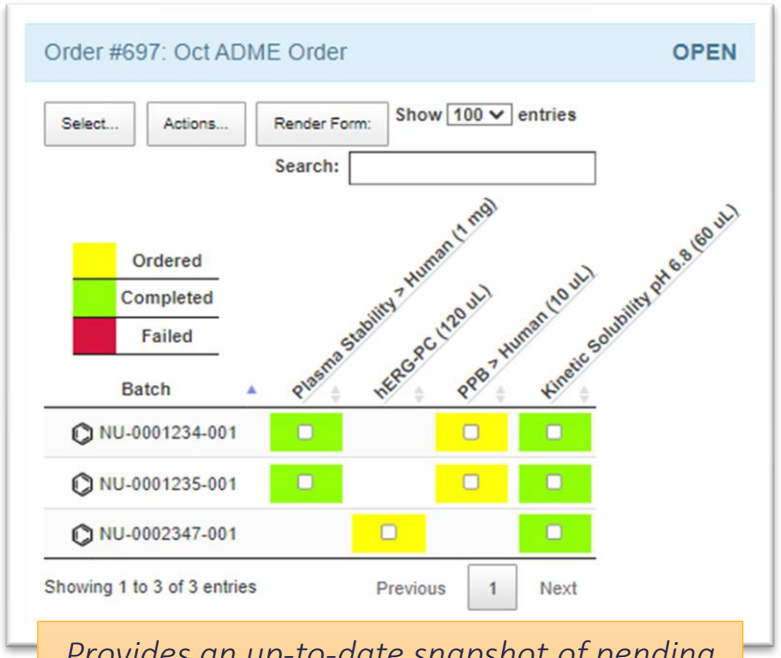
Team Members

Request assays:



Interfaces w/ Vault to show structures, batch info, and available amounts

Monitor request status:



Provides an up-to-date snapshot of pending requests and tracks ADME spending

Secure File Storage

CRO Teams

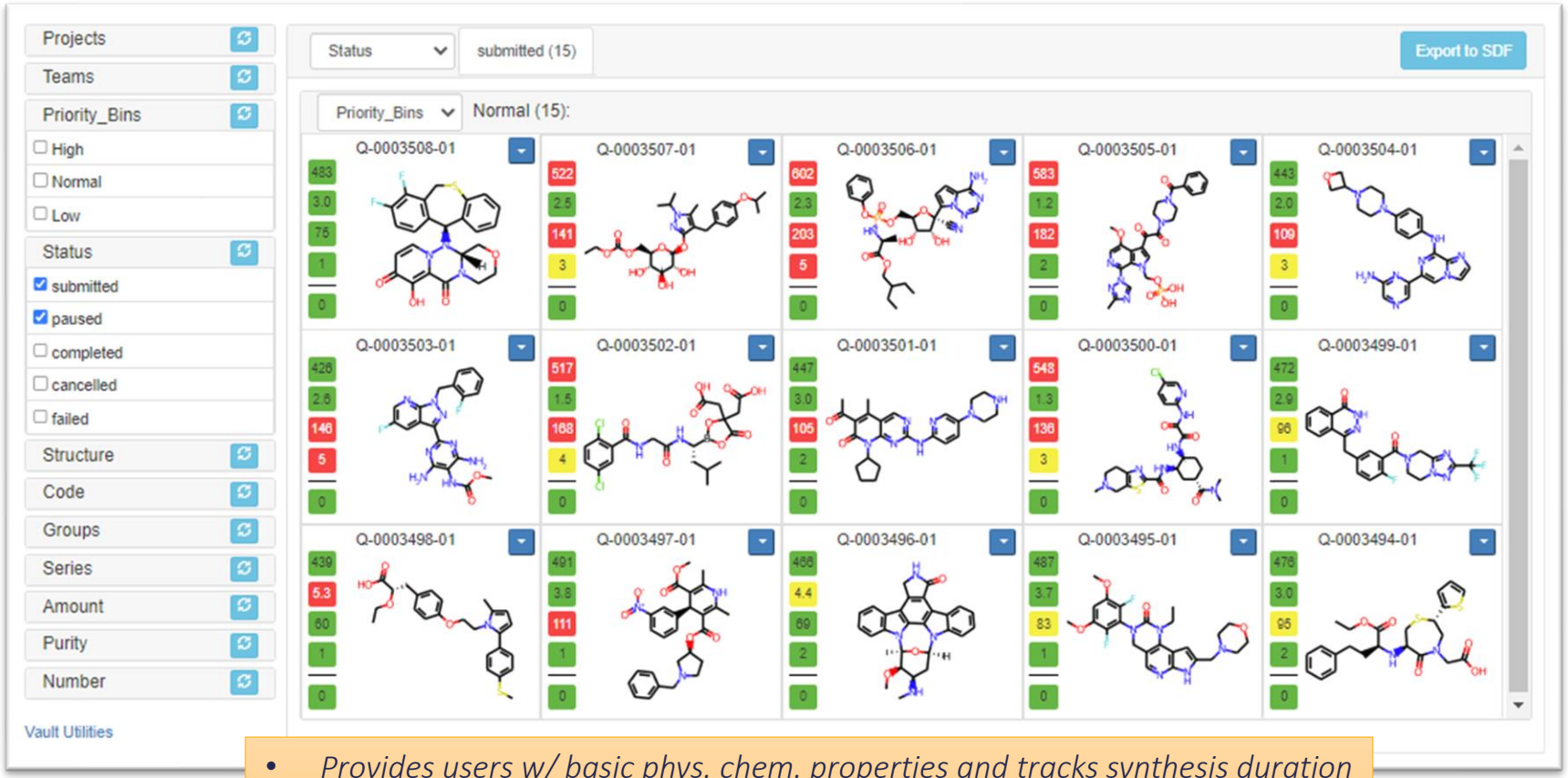
Nuvalent

Enhancing Collaboration: Synthesis Requests

Synthesis queue web-portal brings transparency to synthesis requests and management

Manage separate synthesis queues for each project and team:

Team Members



Secure File Storage

CRO Teams

- *Provides users w/ basic phys. chem. properties and tracks synthesis duration*
- *Syncs w/ Vault to track active and completed targets*

*Chemical structures and codes are shown solely for illustrative purposes

Nuvalent

Real-Time Machine Learning for Prioritization of Compounds in Assays

Progress compounds with higher confidence of success

Depending on the observed data, the false negative rate of a given assay model (Z=0) may be higher than desired.

Evaluate z=(C-P)/U

Increase confidence and decrease false negatives by raising the Activity Bar (Z).

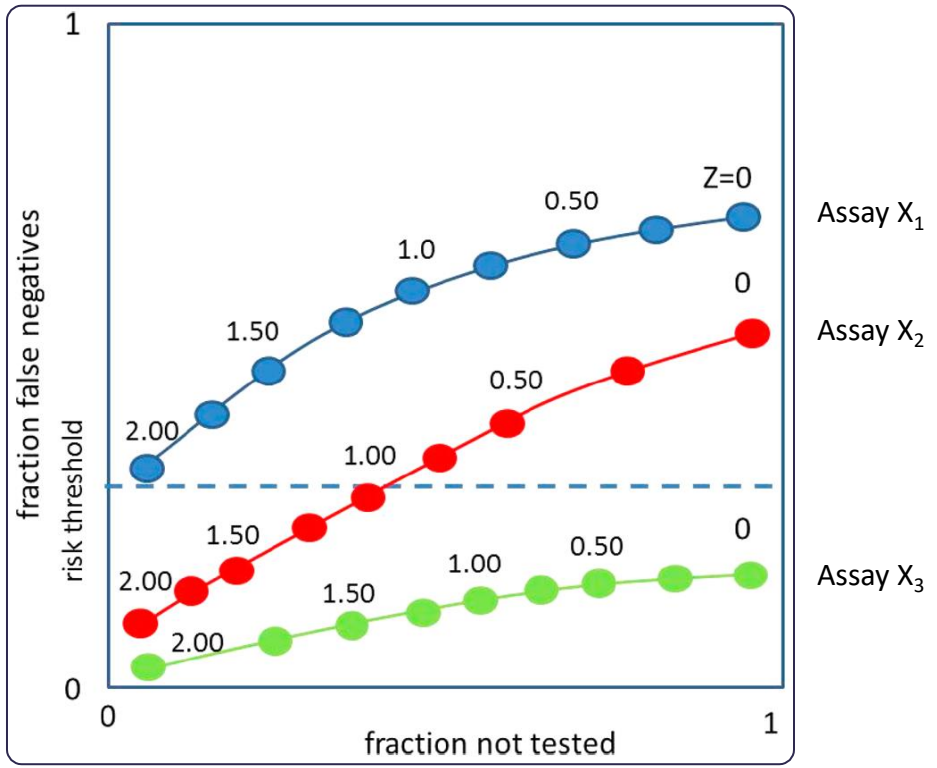
X_1: Categorize Activity=T/F (z>=2)

X_2: Categorize Activity=T/F (z>=1)

X_3: Categorize Activity=T/F (z>=0)

Risk Threshold: FN/(FN+X)

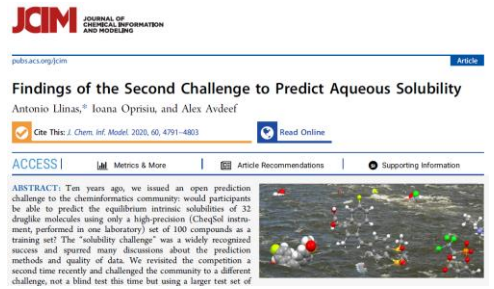
False Negatives (Risk) vs. Cost Savings (Benefit) Varies by Assay and Uncertainty



J. Chem. Inf. Model. 2015, 55, 2, 231–238. https://pubs.acs.org/doi/10.1021/ci500666m

Aqueous Solubility Prediction

Best-in-class prediction is within 1 log unit of experiment


JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Findings of the Second Challenge to Predict Aqueous Solubility

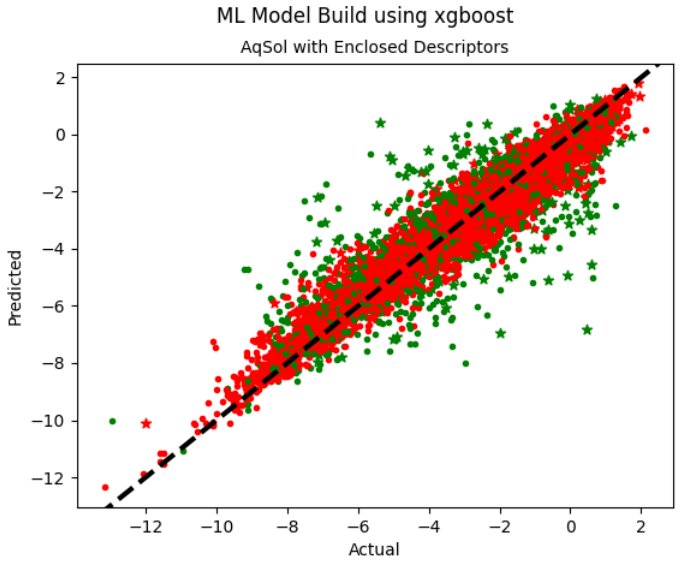
Antonio Llinas,* Ioana Oprisiu, and Alex Avdeef

Cite This: J. Chem. Inf. Model. 2020, 60, 4791−4803 | Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Ten years ago, we issued an open prediction challenge to the cheminformatics community: would participants be able to predict the equilibrium intrinsic solubilities of 32 druglike molecules using only a high-precision (CheqSol instrument, performed in one laboratory) set of 100 compounds as a training set? The "solubility challenge" was a widely recognized success and spurred many discussions about the prediction methods and quality of data. We revisited the competition a second time recently and challenged the community to a different challenge, not a blind test this time but using a larger test set of

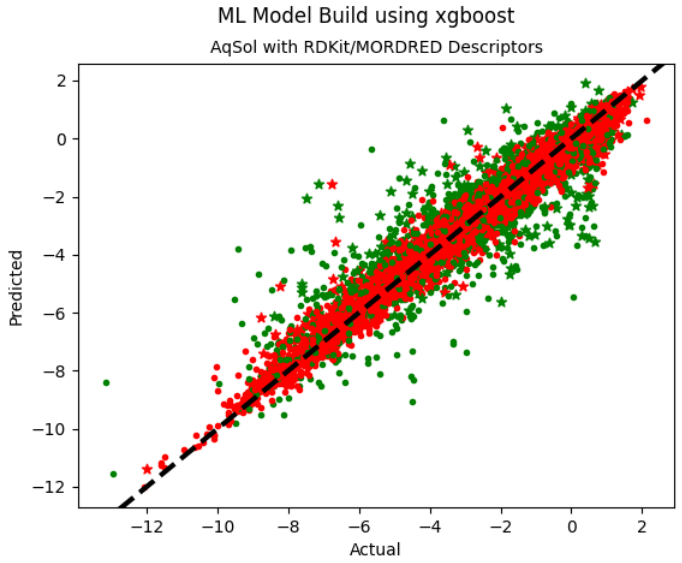
Best Published Model (Public Dataset)



Model	Fitness (R^2)	Error (RMSE)
Training	0.94	0.56
Testing	0.76	1.13

J. Chem. Inf. Model. 2020, 60, 4791-4803. https://doi.org/10.1021/acs.jcim.0c00701

Recapitulated Model with Data in our Hands



Model	Fitness (R^2)	Error (RMSE)
Training	0.97	0.40
Testing	0.82	0.98

IDEAYA BIOSCIENCES

Application of Machine-Learning to Prioritization of Compounds in Assays

Progress compounds with higher confidence of success

Build a machine-learning engine to predict the activity of compounds in assays.

Let M=molecule, X=experiment

Compute P=Prediction(M,X)

Determine U=Uncertainty(M,X)

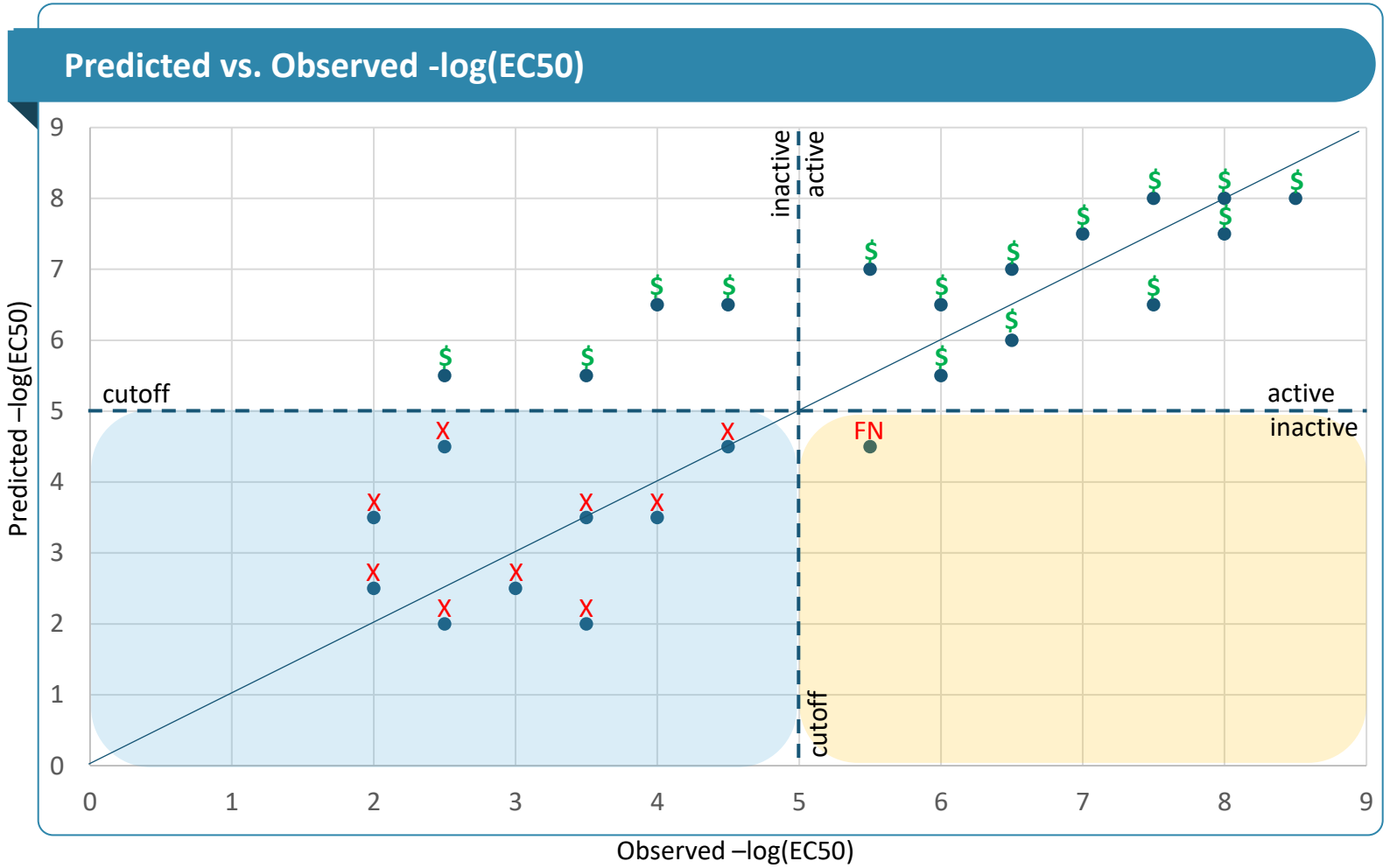
Set C=activity cutoff, e.g., 10 uM

Evaluate z=(C-P)/U

Categorize Activity=T/F (z>=0)

Perform Assay when Activity=T

Risk Threshold: FN/(FN+X)

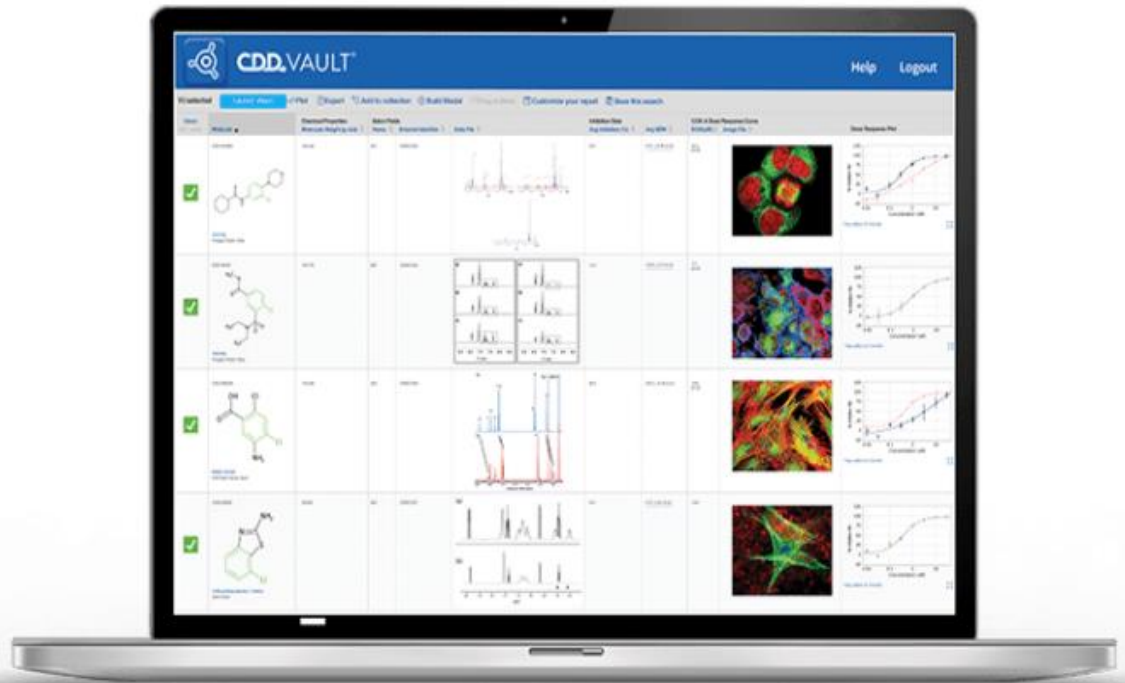


J. Chem. Inf. Model. 2015, 55, 2, 231–238. https://pubs.acs.org/doi/10.1021/ci500666m

Questions?

Activity & Registration
Store and organize your data

Inventory
Keep Track of Compounds



ELN
Document all your research

Visualization
Plot datasets and mine them

Learn More:
https://www.collaborativedrug.com

Personalized Demo & Free Trial:
info@collaborativedrug.com