

Applying Machine Learning and Automation in Novartis Drug Discovery

Cara Brocklehurst

Ischia Advanced School of Organic Chemistry
(IASOC)

September 2024

 **NOVARTIS** | Reimagining Medicine



Novartis is an innovative medicines company

Deliver high-value medicines that alleviate society's greatest disease burdens through technology leadership in R&D and novel access approaches

Our focus

Core therapeutic areas¹

Cardiovascular, renal and metabolic; immunology; oncology; and neuroscience



Technology platforms

Chemistry; biotherapeutics; xRNA; radioligand; gene and cell therapy



4 priority markets

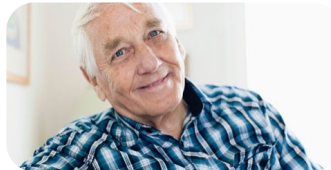
US, Germany, China, Japan



Our priorities

Accelerate growth

Deliver **high-value medicines**



Deliver returns

Embed **operational excellence**



Strengthen foundations

Unleash the power of our **people**

Scale **data science and technology**

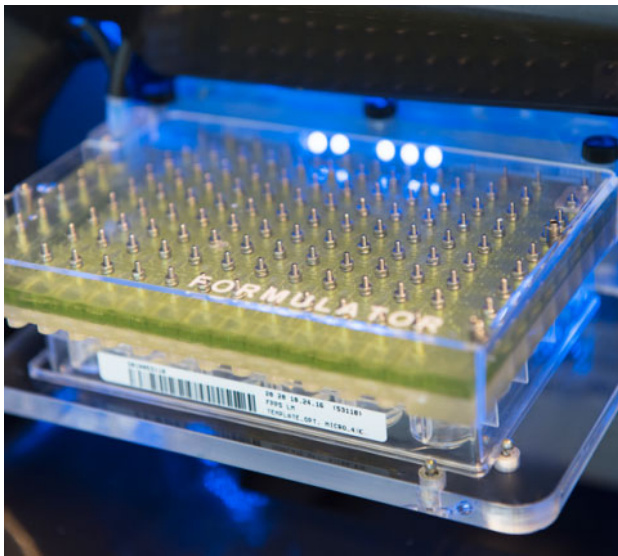
Build trust with **society**



1. Other TAs opportunistically.

Data science driven drug discovery in Biomedical Research

Data Capture @ Novartis



Automated platforms to rapidly deliver high quality data
Med-chem assay & synthesis data

Data Use @ Novartis



ML and AI to guide decisions and inform projects
Designing molecules & synthetic routes

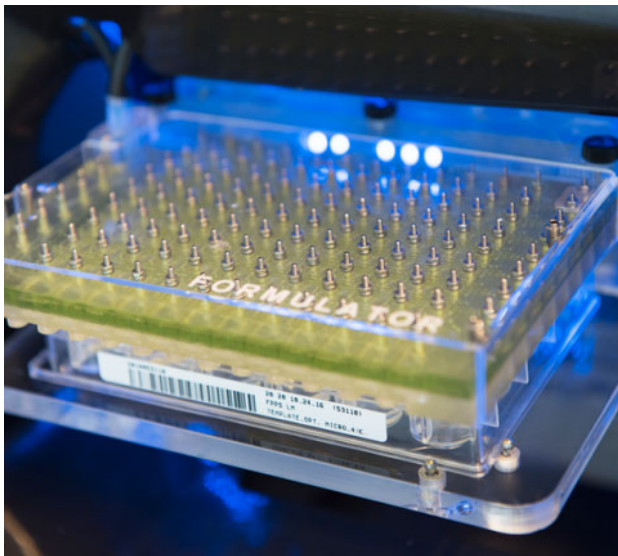
Future Vision



What's next and what are the challenges we face?

Data science driven drug discovery in Biomedical Research

Data Capture @ Novartis



Automated platforms to rapidly deliver high quality data
Med-chem assay & synthesis data

Data Use @ Novartis



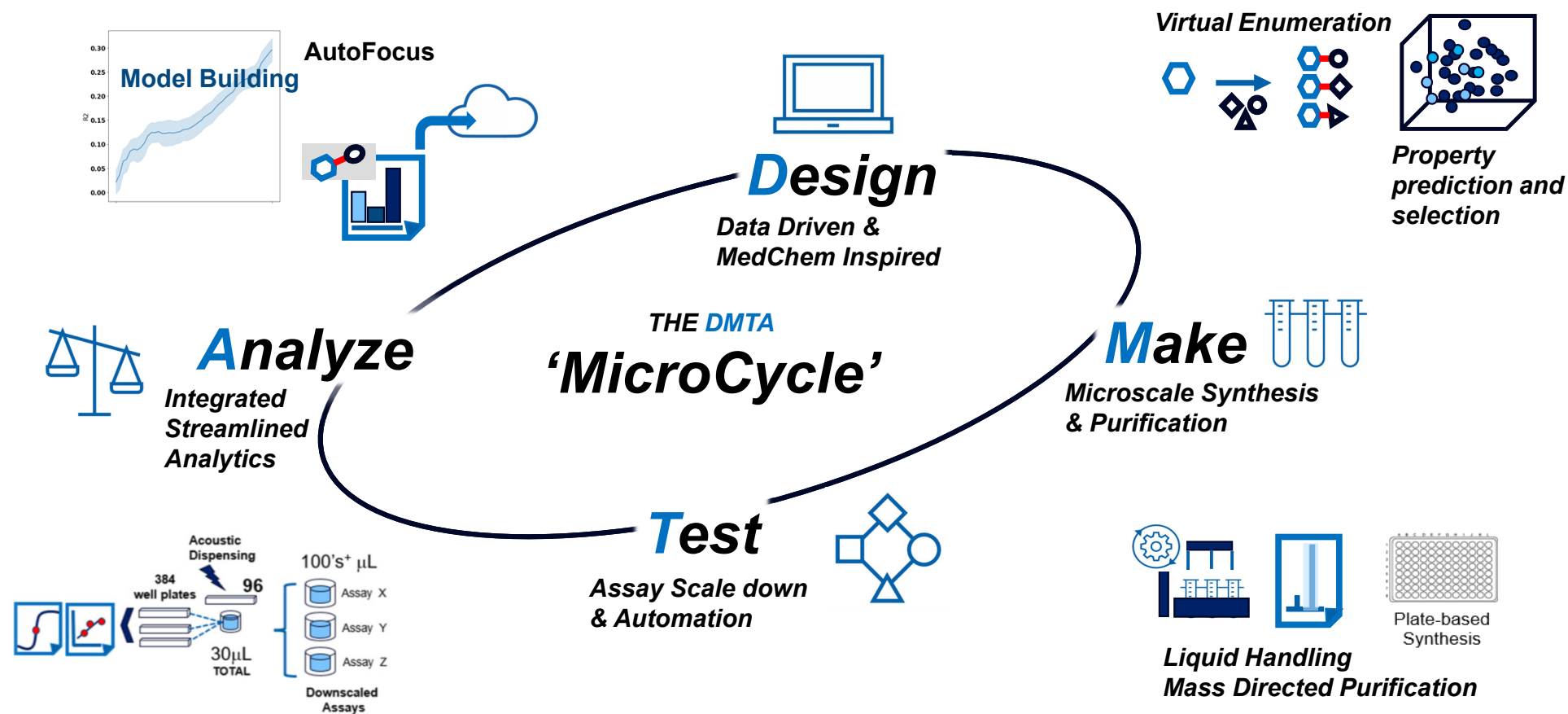
ML and AI to guide decisions and inform projects
Designing molecules & synthetic routes

Future Vision

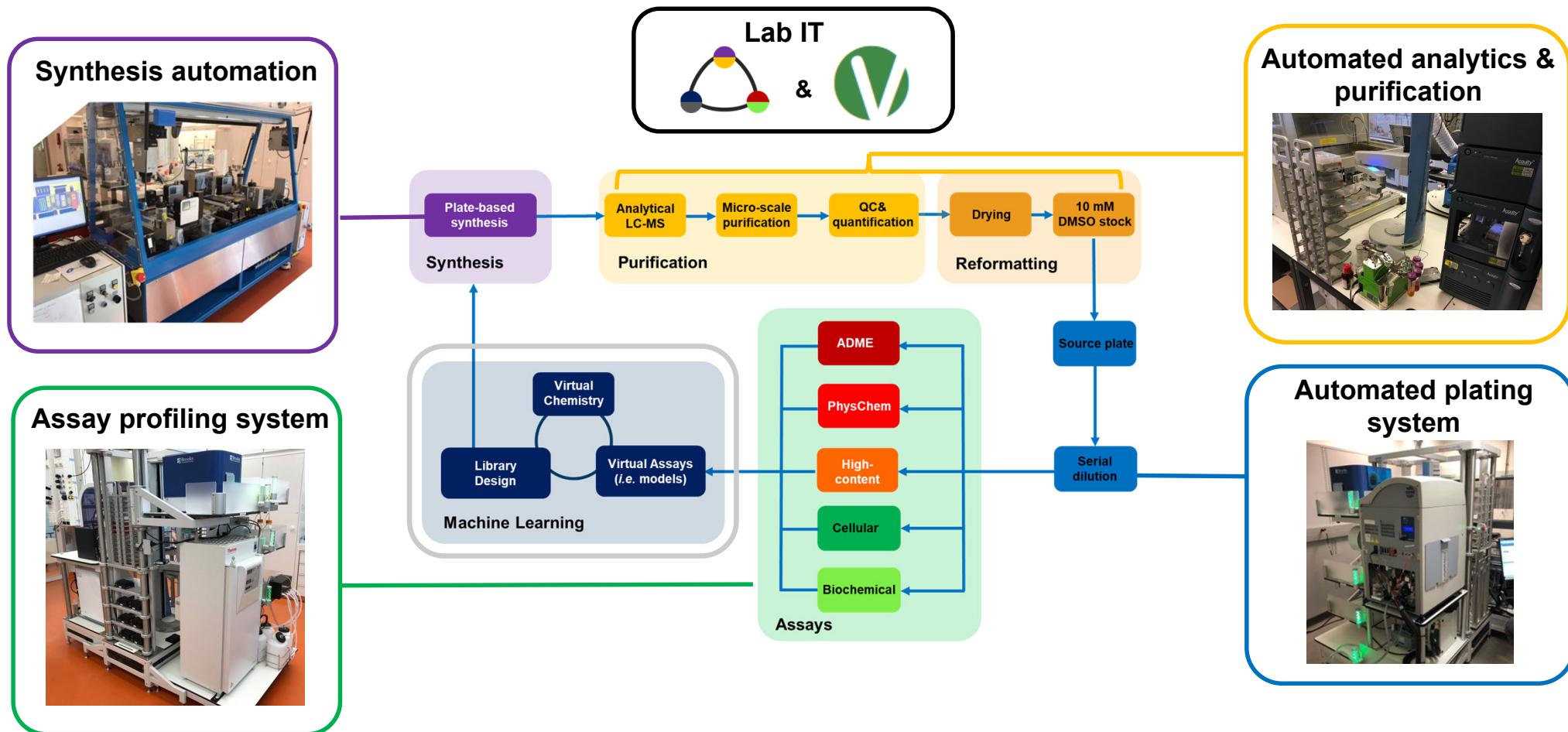


What's next and what are the challenges we face?

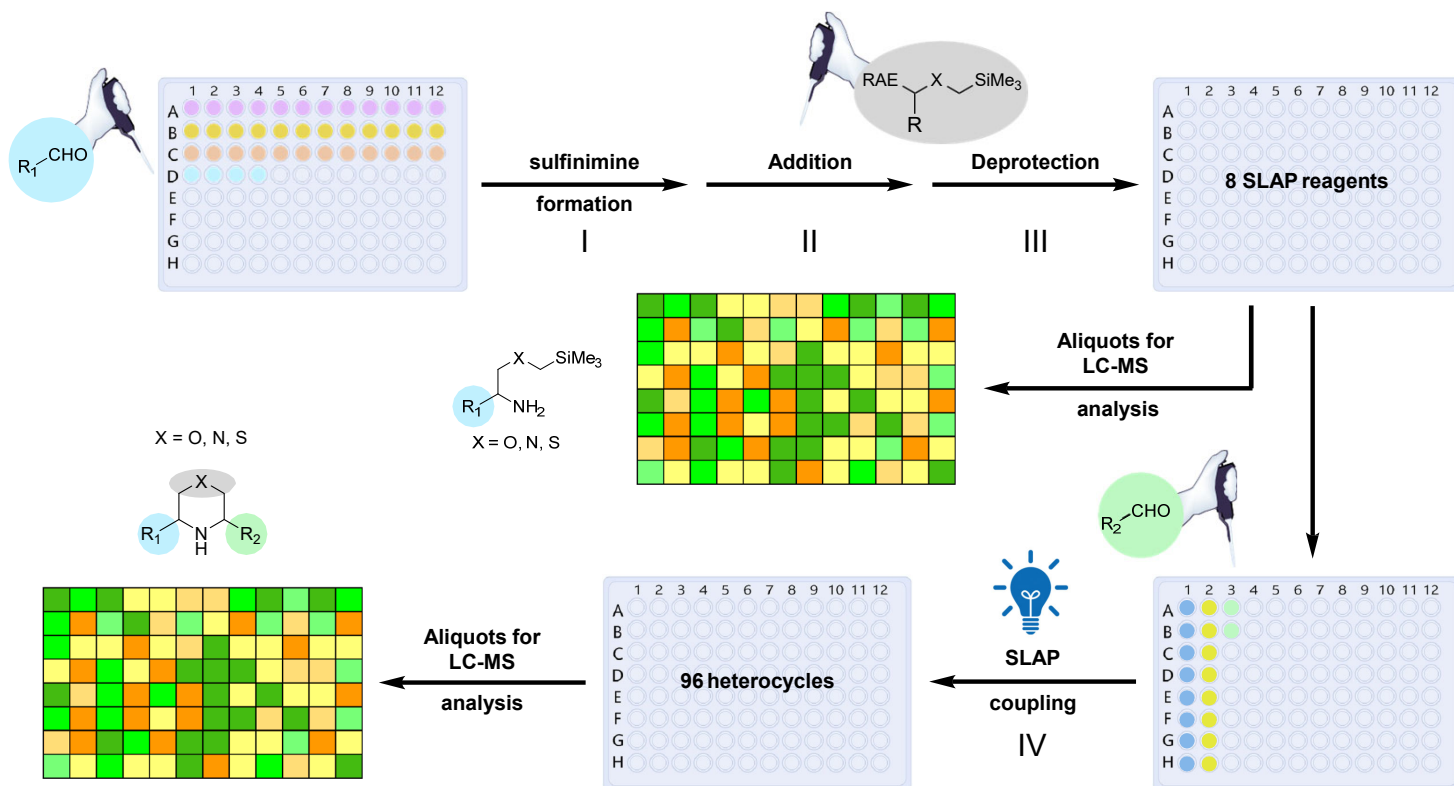
MicroCycle: Coupling data generation with machine learning



Integrated & automated workflow



SLAP plate-based synthesis














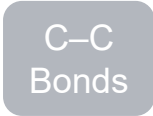




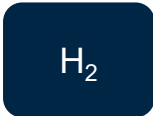





- Morpholines & piperazines as common feature in many marketed drugs
- Access to highly decorated saturated heterocycles rich in sp^3 character
- MicroCycle HT assay capabilities enable us to track properties across a range of heterocycles
- Reaction data used to predict synthesizability

Collaboration with Prof. Jeffrey Bode

'Fostering research synergies between chemists in Swiss academia and at Novartis' *Novartis Chimia (Swiss Chemical Society)* **2021**, 75 (11), 936-942.

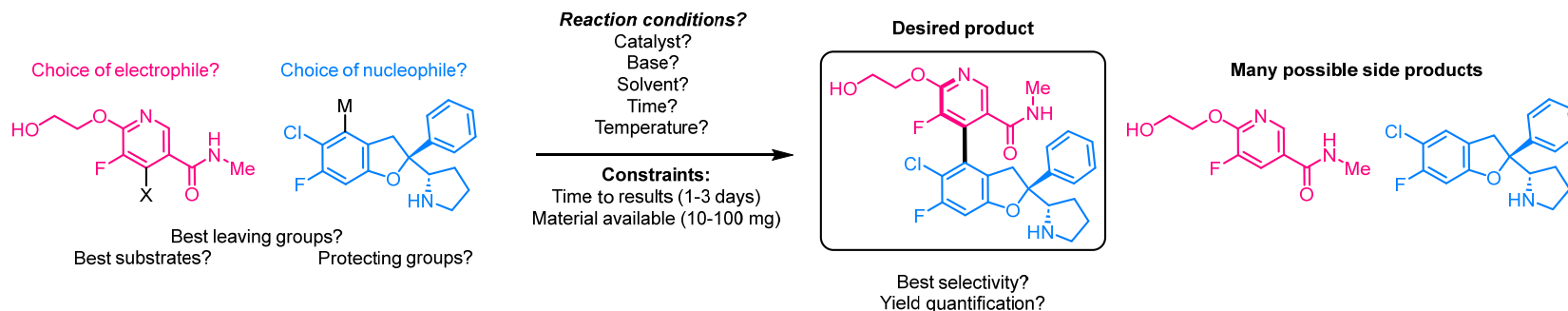
'High-throughput synthesis and data generation for the prediction of molecular properties and synthesizability' *Novartis & J. W. Bode Science Advances* **2023**, 9, eadj2314

SynTech catalysis lab

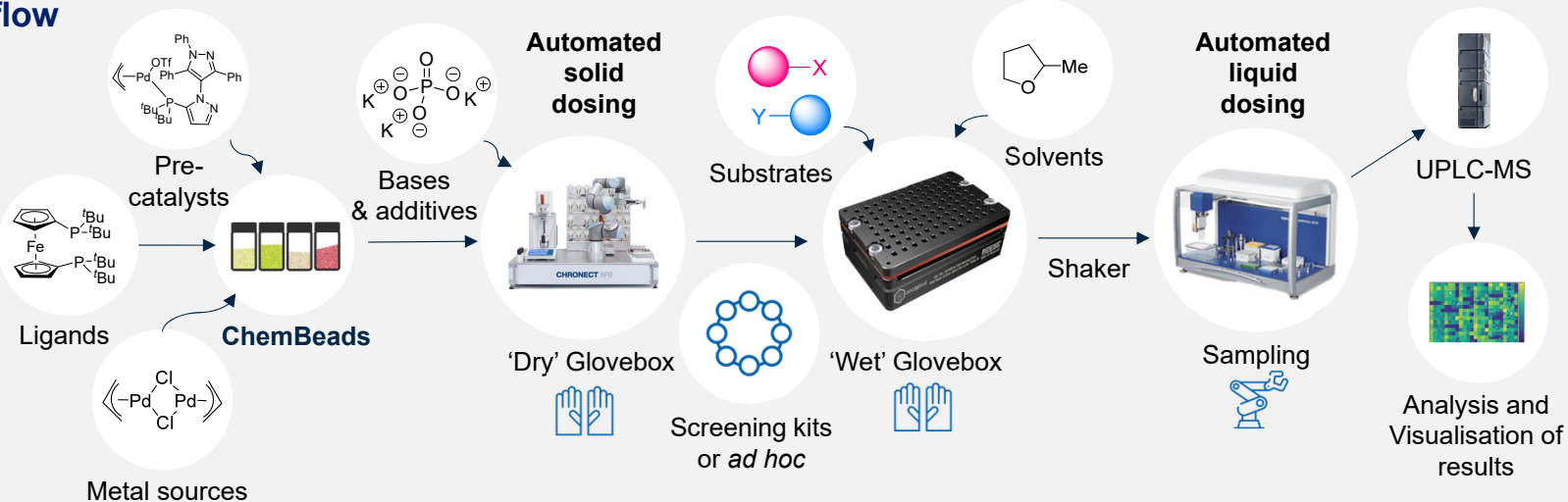
1	Mission	2	Technology Toolbox	3	Reaction Toolbox
	<p> Systematic chemical approach to useful transformations</p> <p> Maximise success of synthesis for desired targets</p> <p> Reduced time to scale-up complex molecules for toxicology studies</p>		<p> Catalyst libraries</p> <p> Glovebox</p> <p> Screening kits</p> <p> Automation</p> <p> Plate-based HTE lab</p> <p> Chembeads</p> <p> Data generation</p> <p> HT analysis</p>		<p>Cross-couplings</p> <p> C-C Bonds</p> <p> C-X Bonds</p> <p> Pd Palladium (46, 106.42) Transition Metals</p> <p> Cu Copper (29, 63.546) Transition Metals</p> <p> Ni Nickel (28, 58.693) Transition Metals</p> <p>Hydrogenations and carbonylations</p> <p> H₂</p> <p> CO</p> <p> Ru Ruthenium (44, 101.07) Transition Metals</p> <p> Rh Rhodium (45, 102.906) Transition Metals</p> <p> Pd Palladium (46, 106.42) Transition Metals</p> <p> Ir Iridium (77, 192.22) Transition Metals</p>

Reaction condition screening in Global Discovery Chemistry

Solving important chemistry questions



Workflow

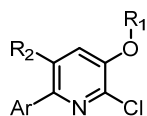


Unlocking difficult targets in lead optimisation

C-X
Bonds

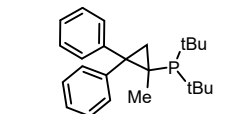
46 106.42
Pd
Palladium
[Kj,4d¹⁰
Transition Metals

**Sterically Hindered
Coupling**

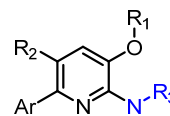


BH Kit
Catalyst Screen

Tested:
Pd-Catalysts (*30)
Solvents (*2)
Bases (*4)
Temp 60-100°C



cBRIDP + [Pd(allyl)Cl]₂
NaOtBu (dry), THF, 60°C

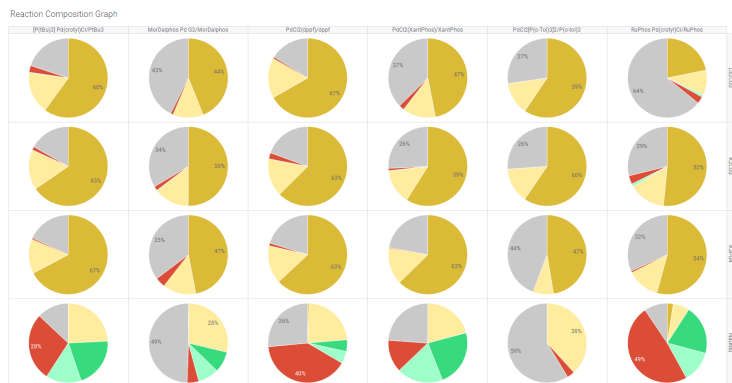


Before: 0% Yield
After: 60% Yield
(60% CRY)

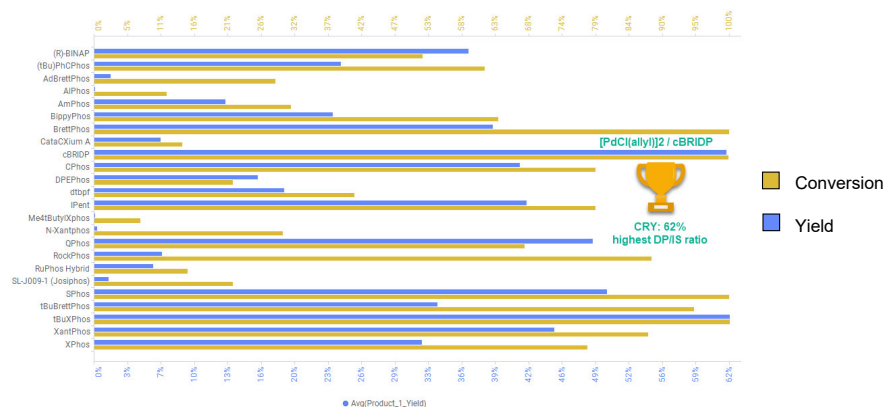
**Potent
inhibitor identification**
(10 times jump in potency in LO)

**HTE chemistry brings high-
quality reaction data**

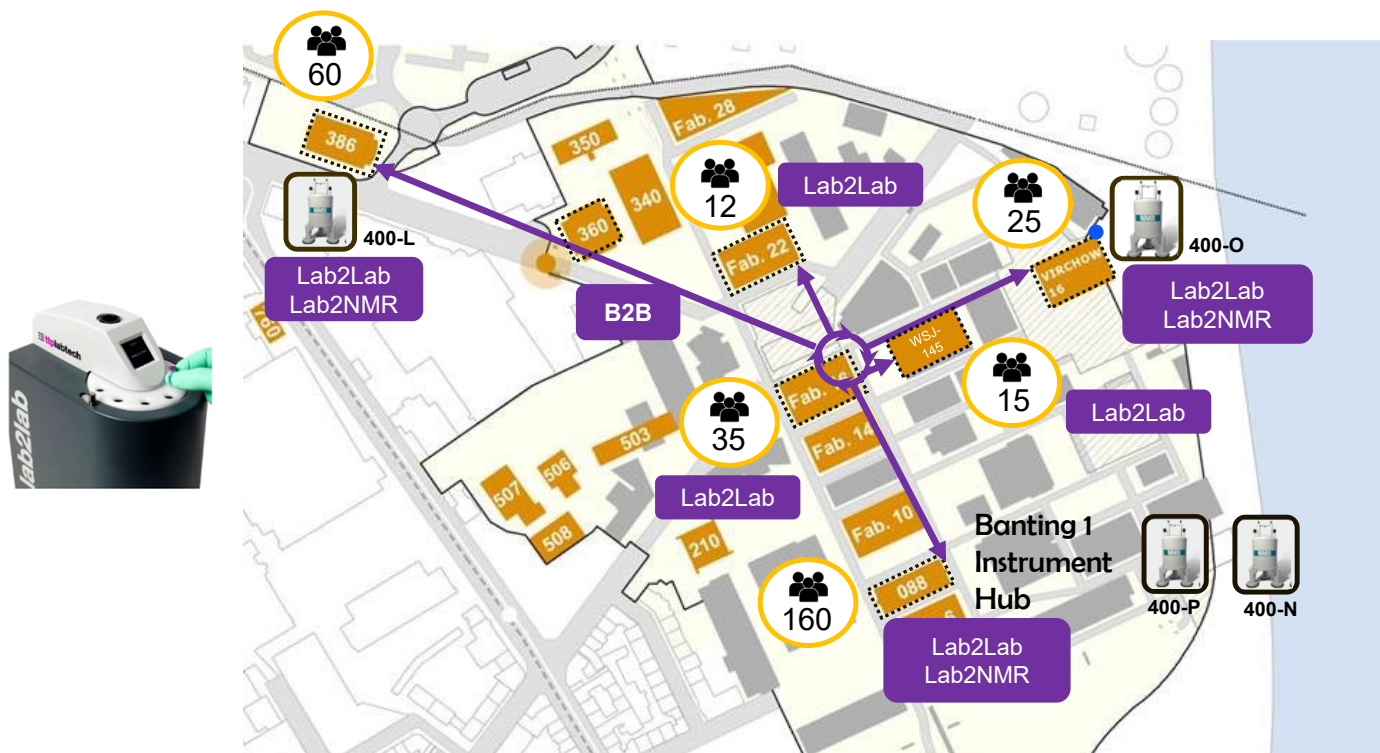
Premade Kit: 6 Catalysts x 4 Bases



24 Catalysts



Basel Lab2Lab: Rapid access to analytical data



Lab2Lab for LCMS and NMR

Drop-off in your lab sender

Sample sent to free machine on campus using scheduling software

Data rapidly available in Signals eLN

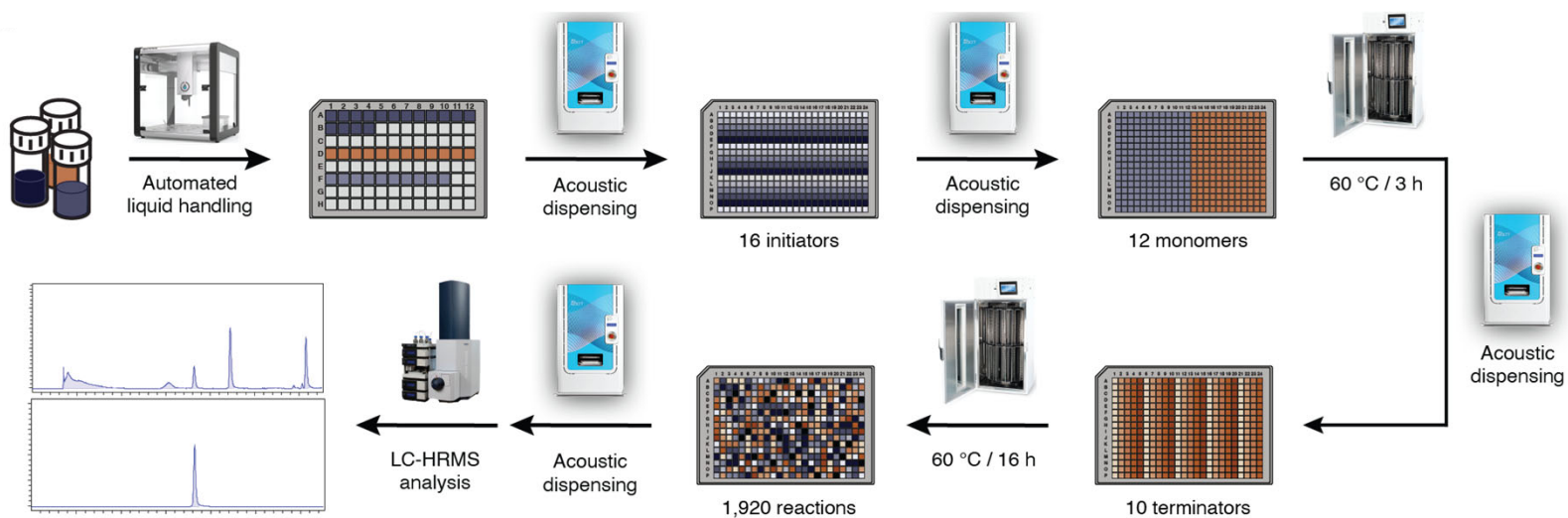
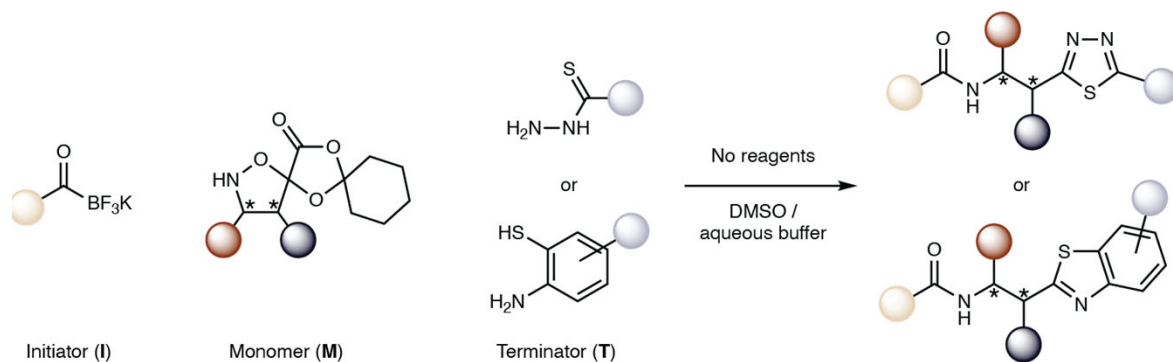
Automated Structure Verification

Confirm structure by NMR

Assign peaks

Send analysis report with calculated measure of confidence

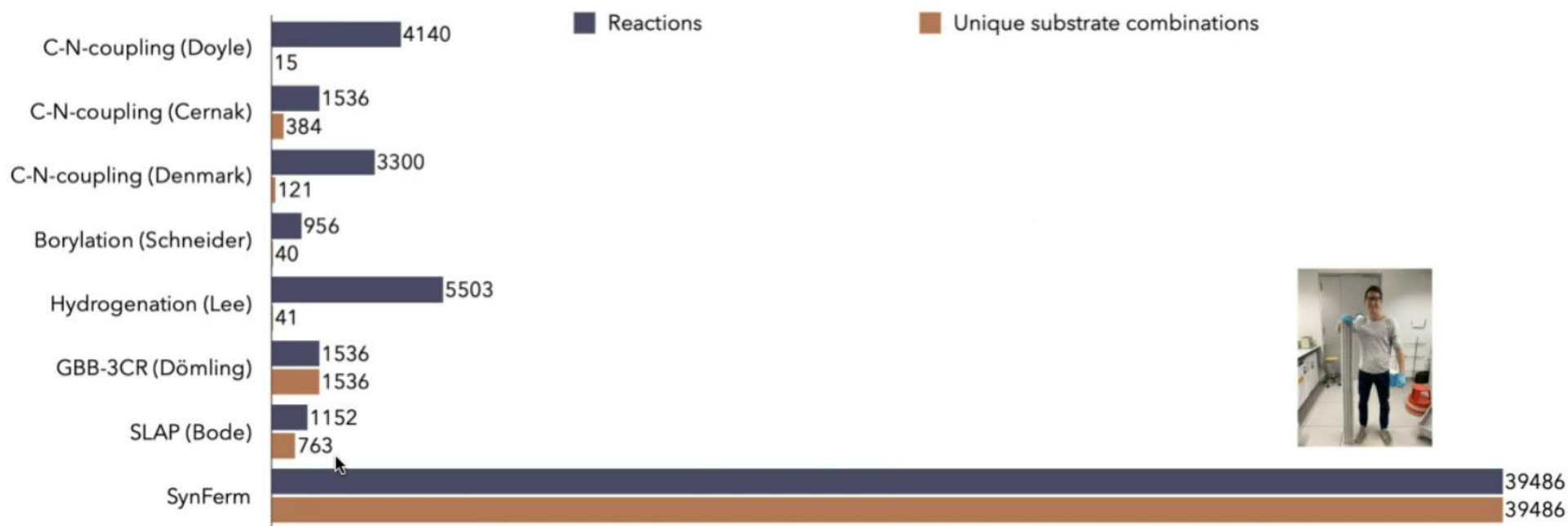
Synthetic fermentation: Bode



Synthetic fermentation: Bode

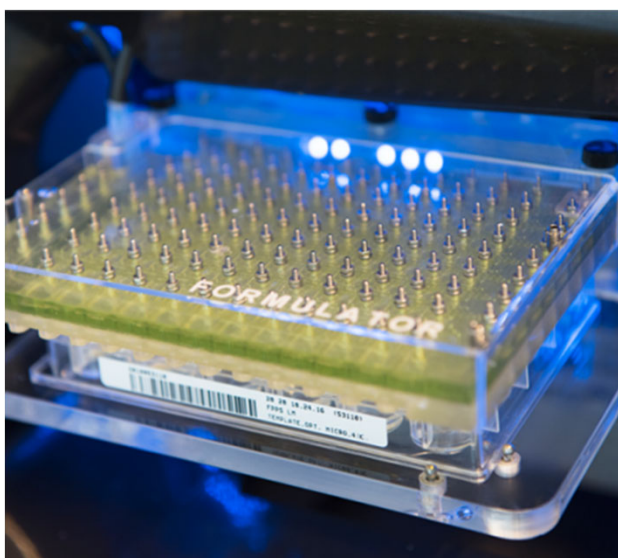
Construction and predictions of reaction outcome data on unprecedented scale

Miniaturisation and automation to conduct and analyse 50,000 reactions performed on a 3 μ L scale with distinct substrates



Data science driven drug discovery in Biomedical Research

Data Capture @ Novartis



Automated platforms to rapidly deliver high quality data
Med-chem assay & synthesis data

Data Use @ Novartis



ML and AI to guide decisions and inform projects
Designing molecules & synthetic routes

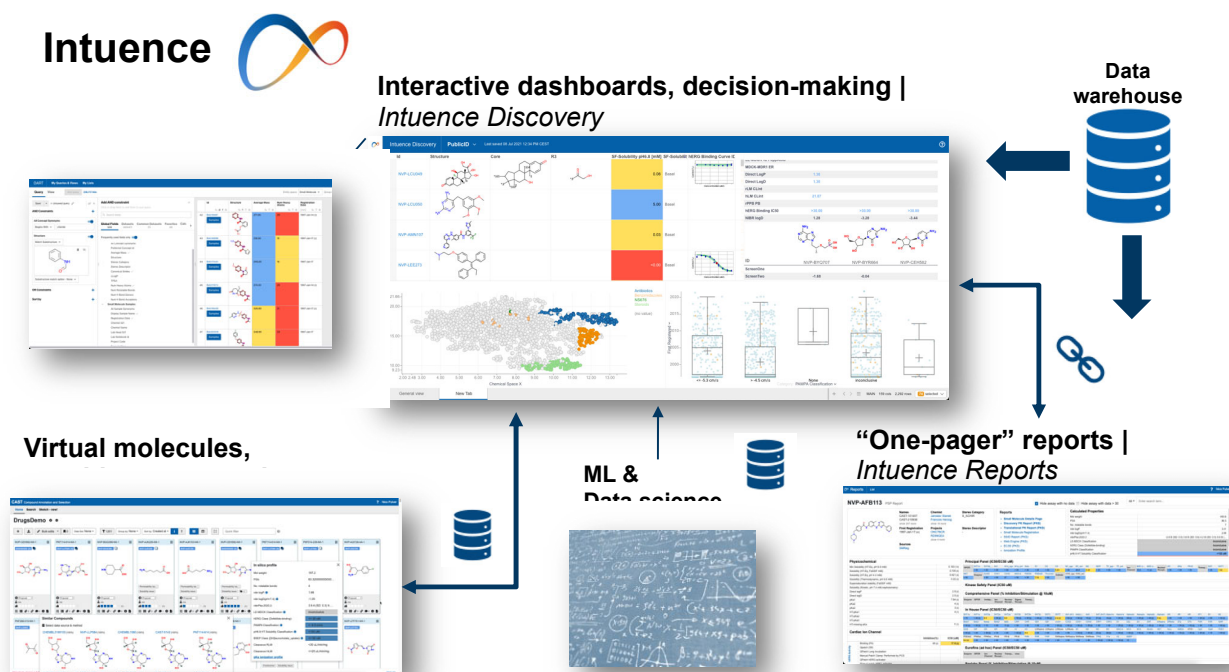
Future Vision



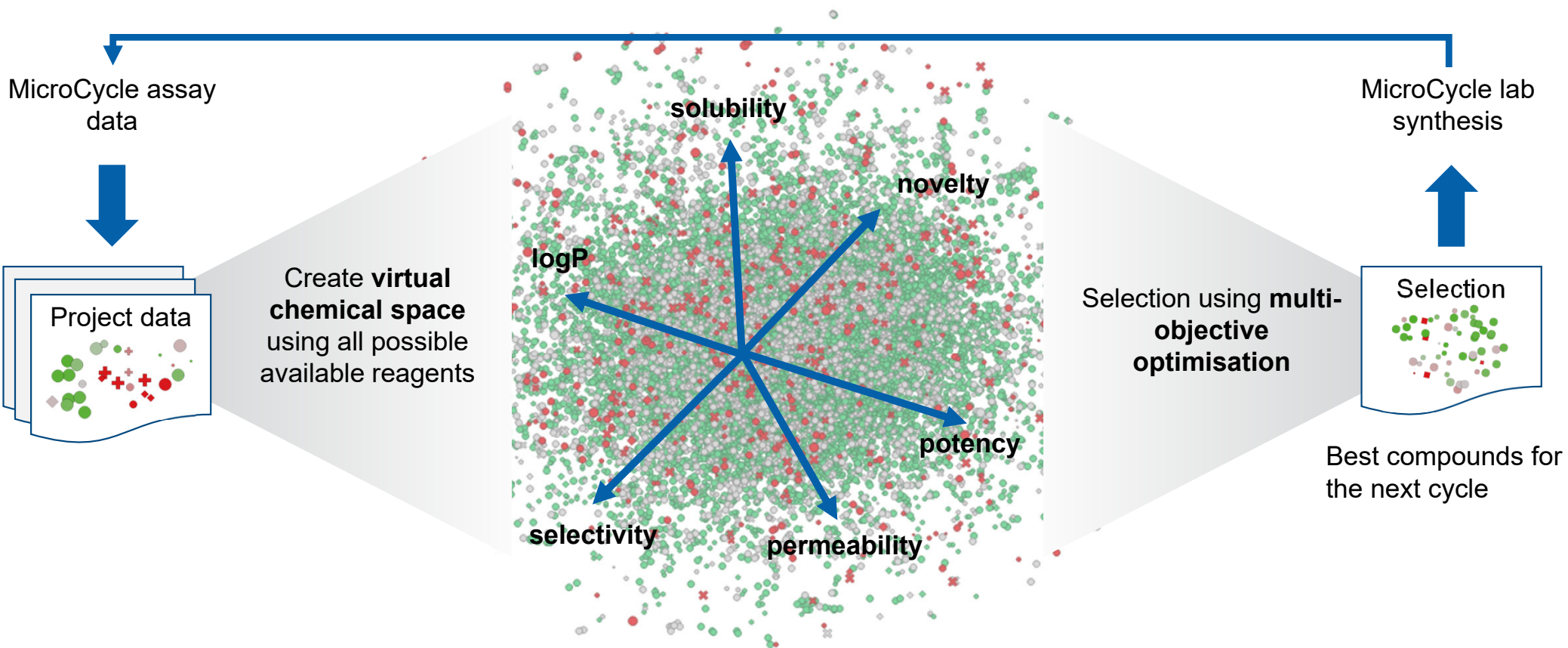
What's next and what are the challenges we face?

Data improvements: Integrative data analysis

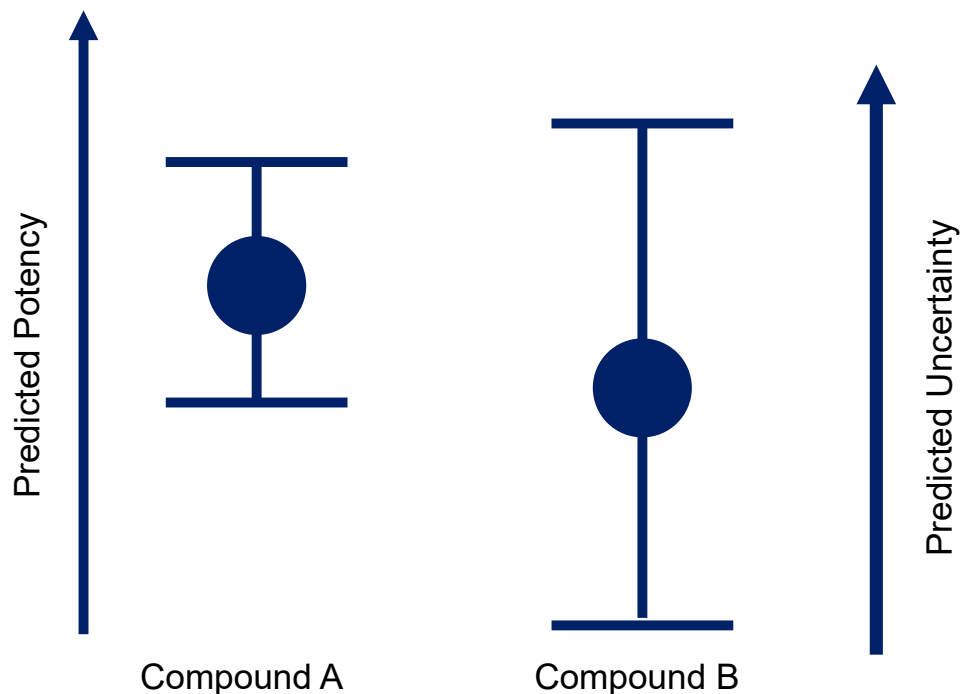
- FAIR principle
- Change the **ownership** paradigm: enable 'Citizen Data Scientist' culture
- **Project-centric** one-stop-shop for all Discovery Data
- **Integrate virtual and real compounds** with measured and *in silico* data



Drug hunting: Data-centric multi-objective library design

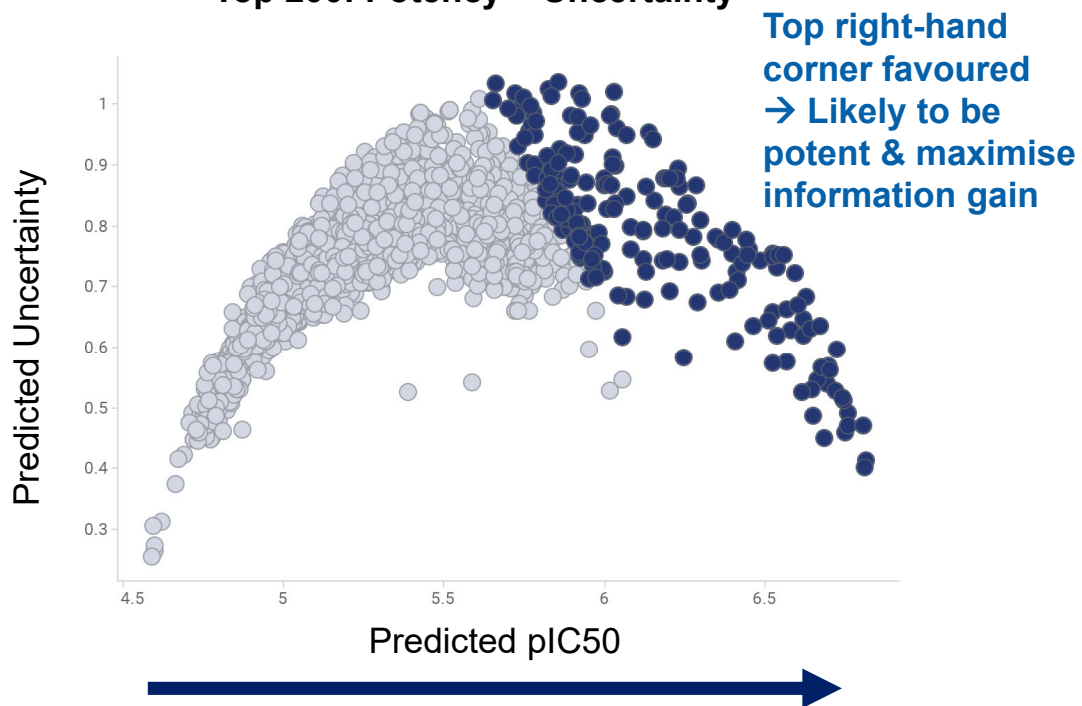


How to best score and rank compounds



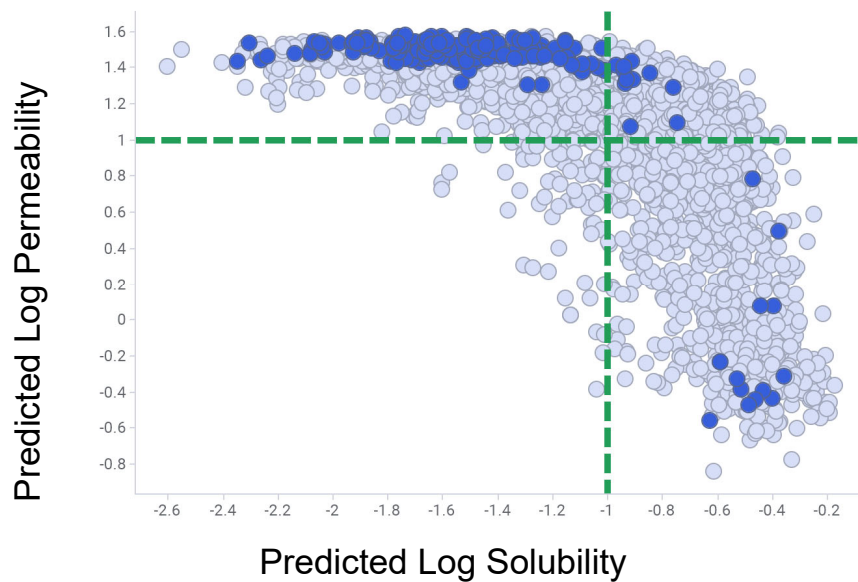
Compound B has a chance of being most potent and improves the model for the next iteration

Virtual library of ~2'800 compounds
Top 200: Potency + Uncertainty

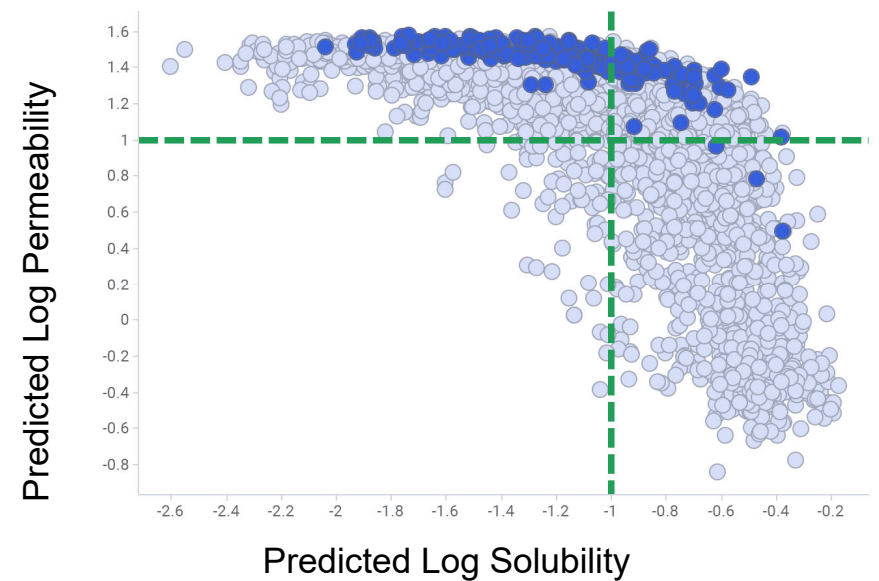


How does multi-parameter optimisation effect our choices?

Top 200: pIC50 + Uncertainty

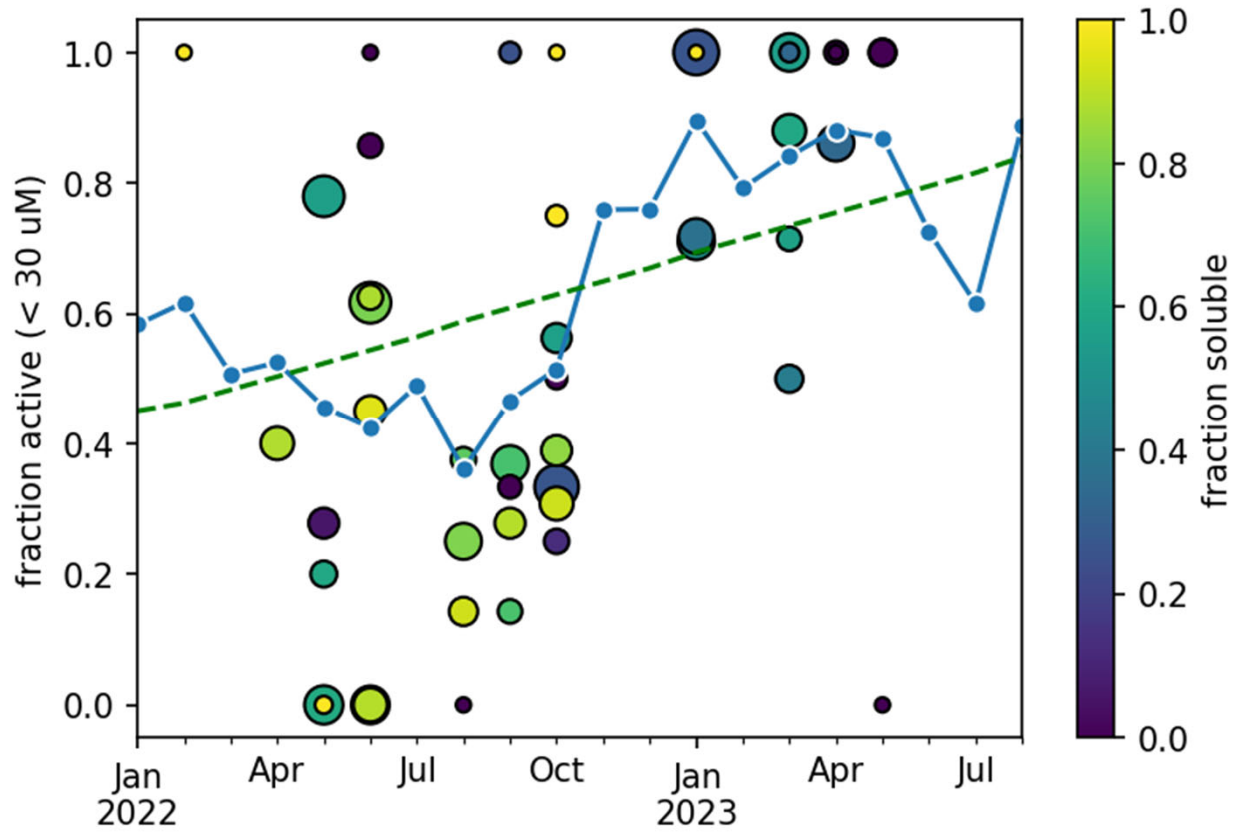


Top 200: MPO
pIC50 + Uncertainty + log Sol. + log MDCK



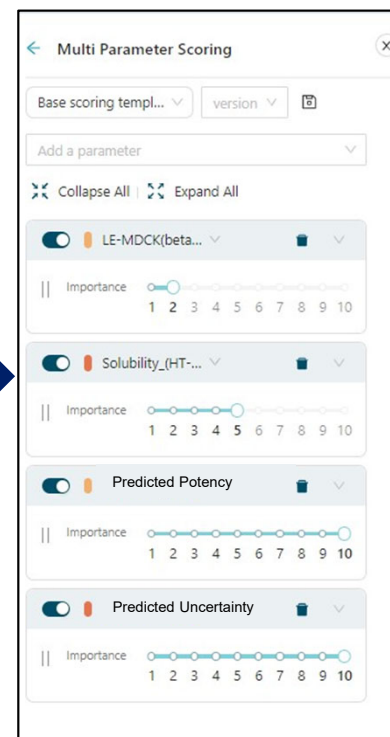
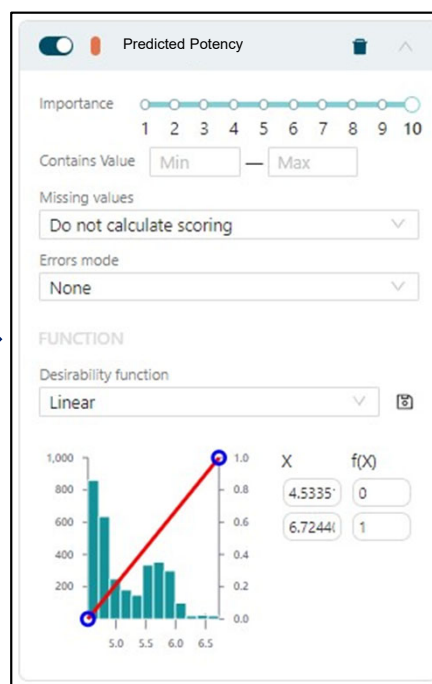
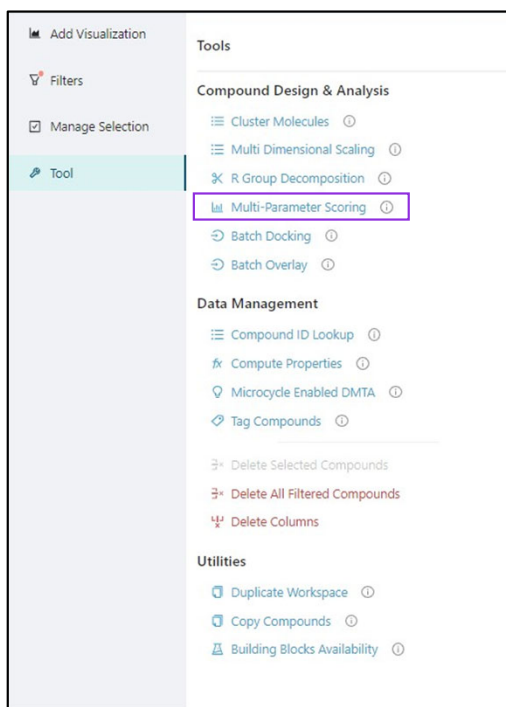
- Final library selection done with **'Human in the loop'**

Were the library designs a success?



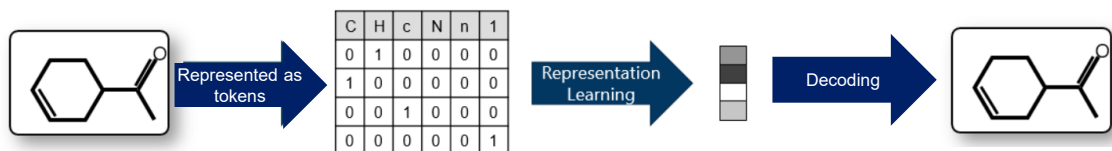
Democratising library design: Autofocus

- Large selection of internal and external building blocks
- Models trained on historic data (potency, selectivity, ADME, etc.)
- Novel selection algorithms used for multi-parameter prioritisation of building blocks



Generative models: How to teach chemistry to computers?

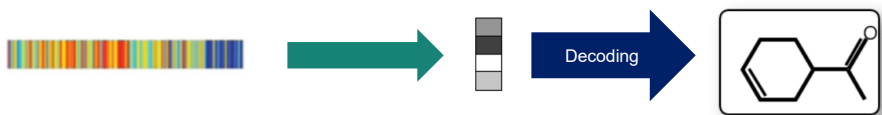
String-based methods (e.g. CDDD¹)



Graph-based methods (e.g. CGVAE², MoLeR³)



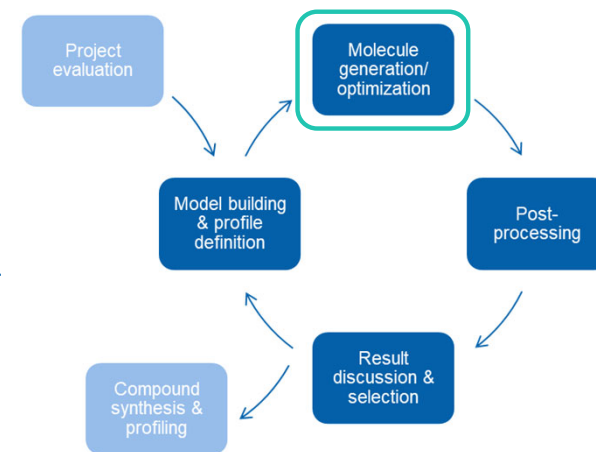
Conditional generation using [signatures, profiles, sequences] (e.g. pqsar2cpd⁴)



Many other approaches exist, major application:

- **Exploration: distribution learning** (reproduce sets of molecules)
- **Exploitation: goal-directed generation** (search latent space without full sampling)

De novo design cycle



[1] Winter, R. *et al. Chem. Sci.* **2019**, *10*, 1692–1701

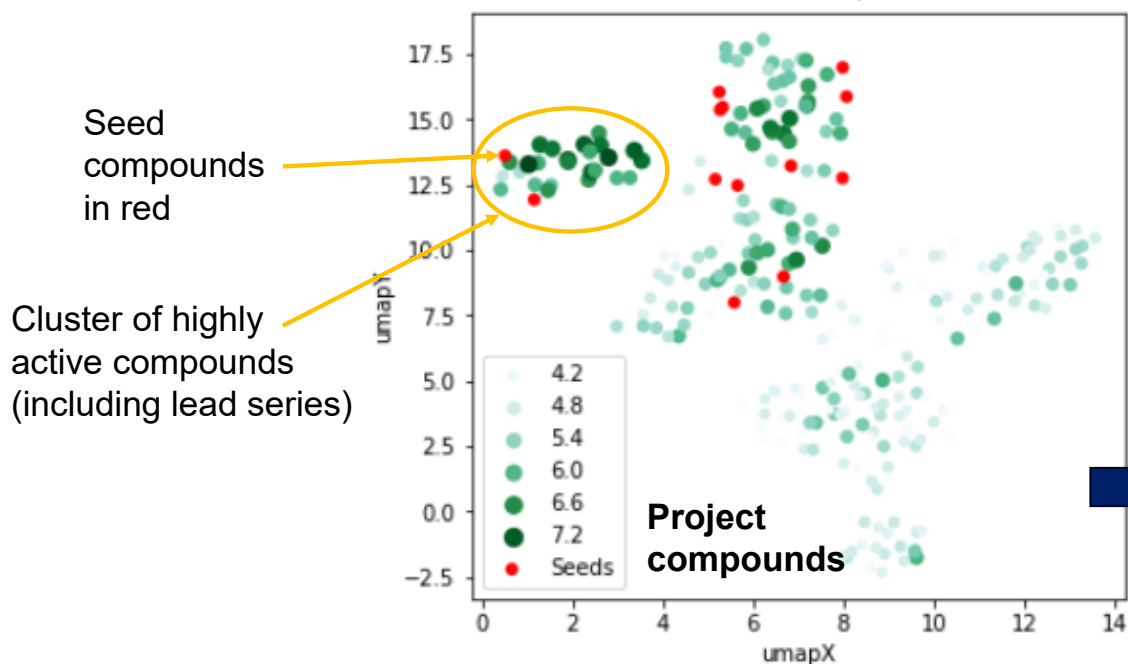
[2] Jin, W. *et al. arXiv* **2019** <https://arxiv.org/pdf/1802.04364.pdf>

[3] Maziarz, K. *et al. arXiv* **2021** <https://arxiv.org/pdf/2103.03864.pdf>

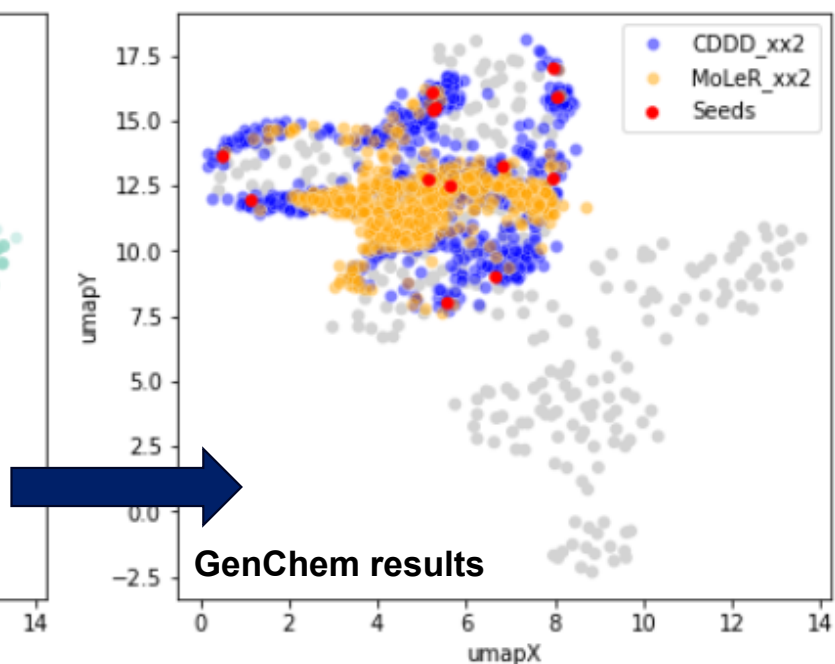
[4] Pikusa M, *et al. bioRxiv* **2022** <https://biorxiv.org/content/10.1101/2021.12.10.472084v1>

GenChem: Chemical space exploration

Pre-defined target property profile guides the search in the latent space



Predictive models used to determine properties of a new point in the latent space



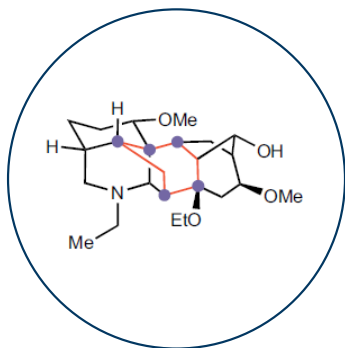
Interesting exploitation of GenChem in areas between current series

Different embeddings and settings provide different exploration profiles

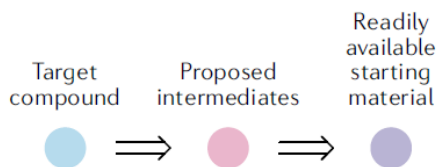
Next steps: Post processing

Next challenge: How do we consider synthesisability?

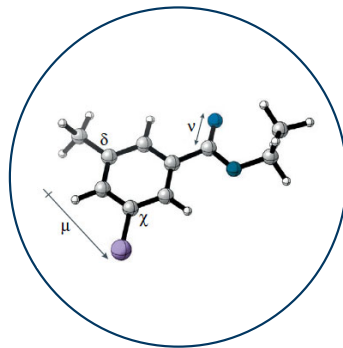
Computer assisted synthesis planning (CASP)



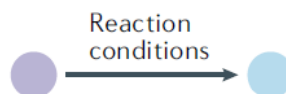
Computer assisted retrosynthesis analysis



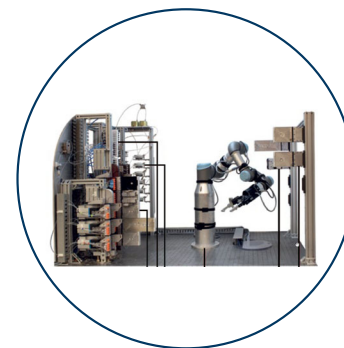
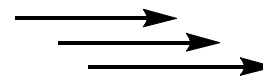
- Most advanced
- Ready for implementation
- e.g. commercially available software



Forward synthesis predictions

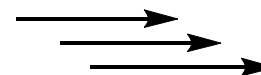


- Early stage
- Data hungry
- e.g. internally built software



Autonomous synthesis

Future goal



Enhanced synthetic chemistry

Synthetic fermentation: Bode

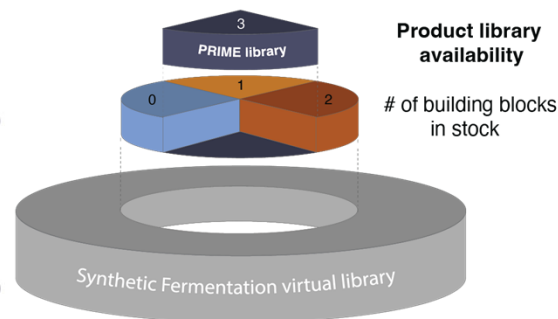
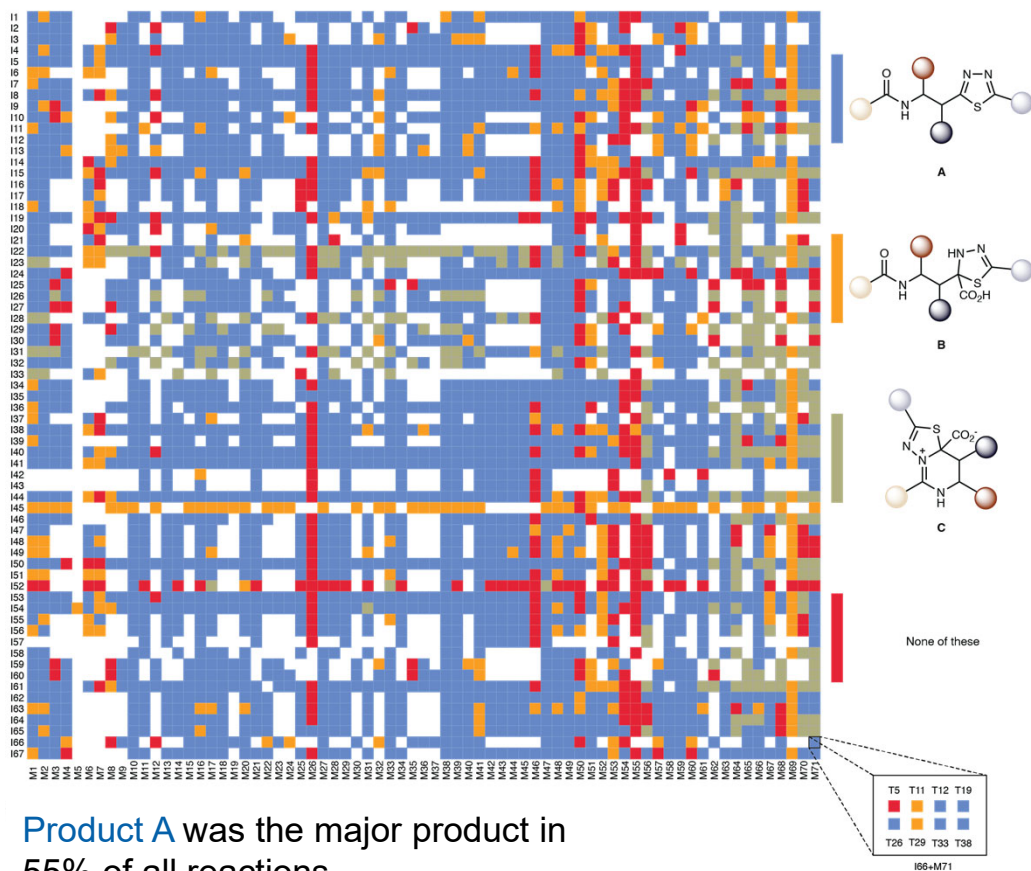
Construction and predictions of reaction outcome data on unprecedented scale

Miniaturisation and automation to conduct and analyse 50,000 reactions performed on a 3 μ L scale with distinct substrates



Synthetic fermentation: Bode

<https://jugoetz.com/synferm-heatmap>

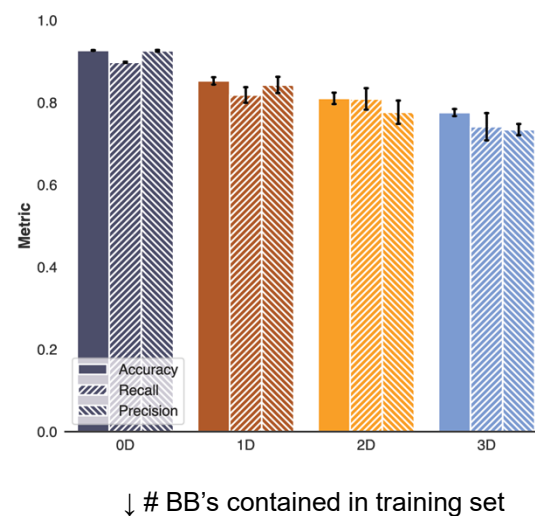


ML used to accurately predict the result of unknown reactions.

With the number of new BB's, the difficulty of predictions increases

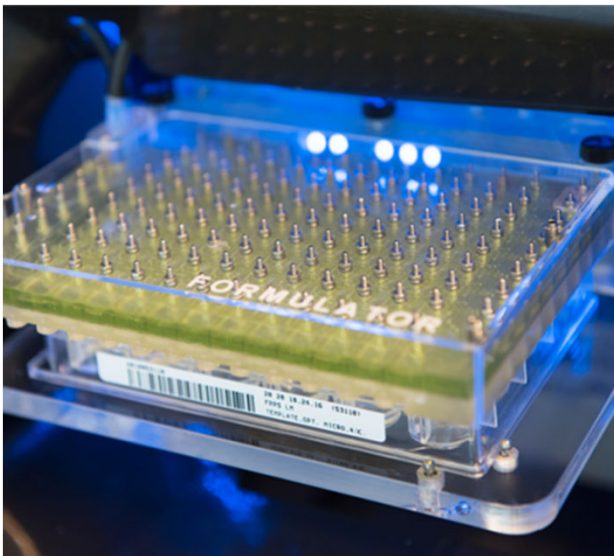
The impact of data set size on model training was analysed

Necessity of tailored data collection for optimising machine learning models in chemical synthesis



Data science driven drug discovery in Biomedical Research

Data Capture @ Novartis



Automated platforms to rapidly deliver high quality data
Med-chem assay & synthesis data

Data Use @ Novartis



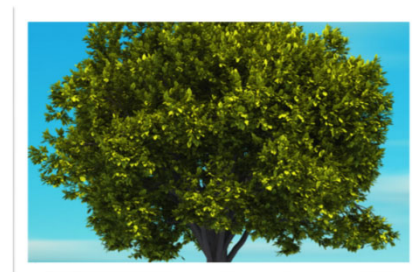
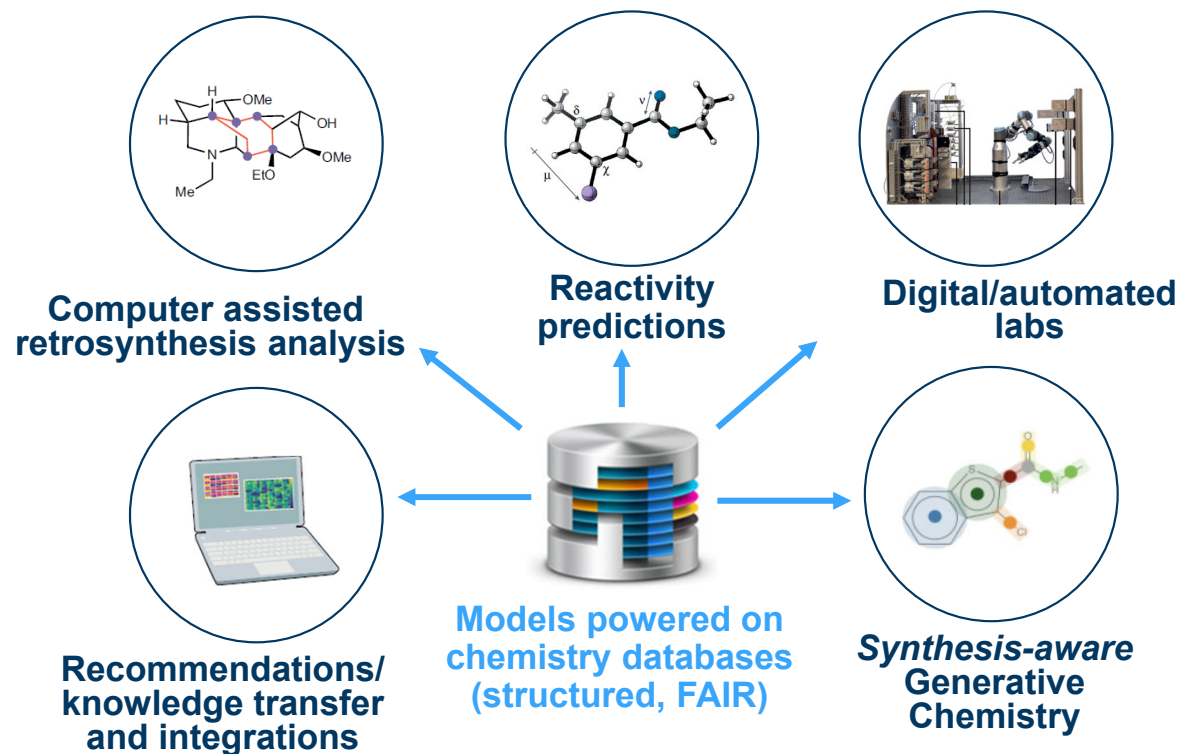
ML and AI to guide decisions and inform projects
Designing molecules & synthetic routes

Future Vision



What's next and what are the challenges we face?

The digital future of synthesis...

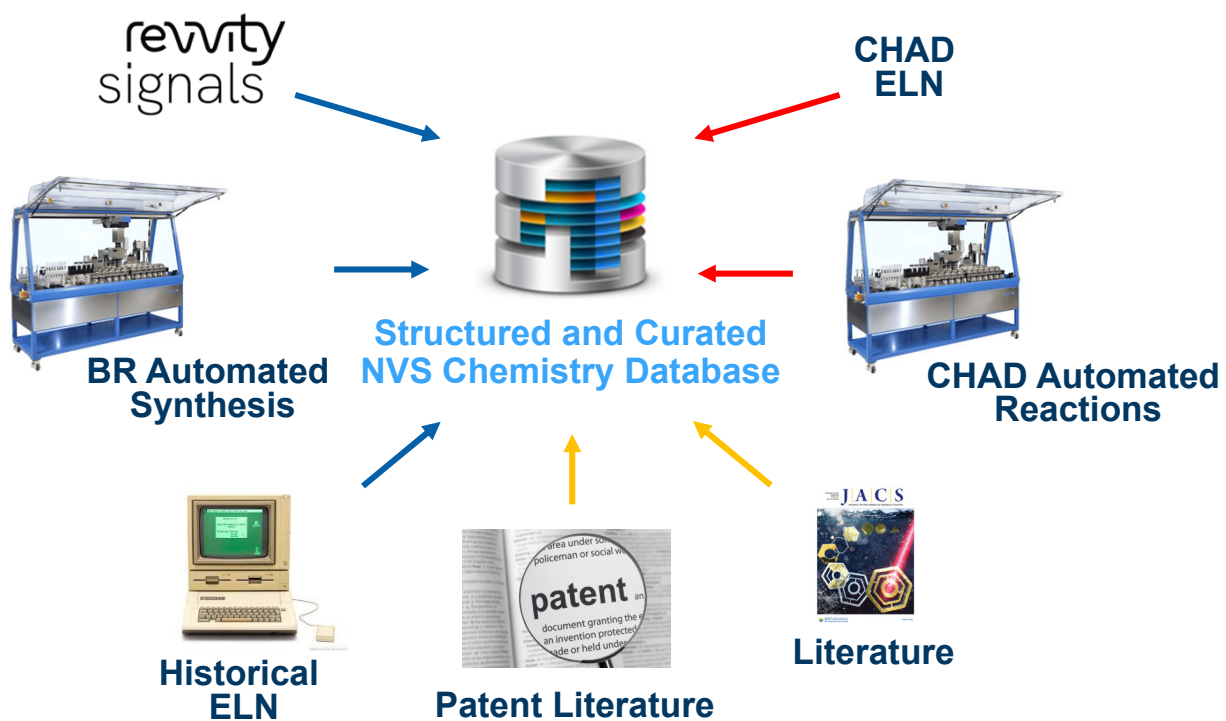


What we want to do...

...needs well organised and curated chemistry data

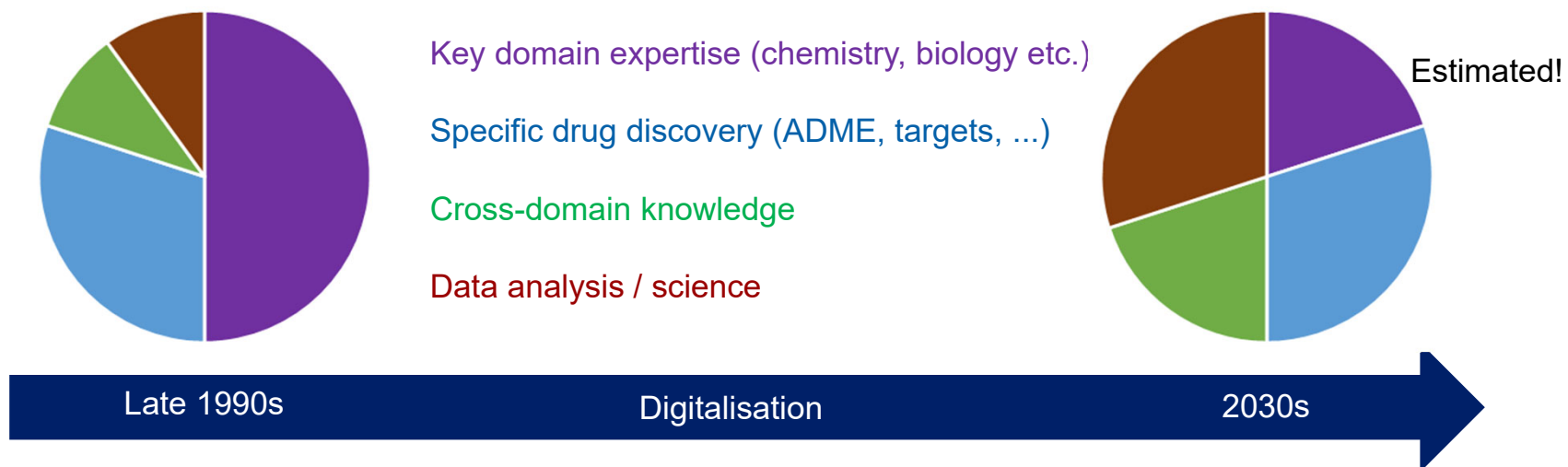


What we **need** to do...



.....and a culture which rewards high quality data input

The changing profile of a drug hunter



“... It is not that machines are going to replace chemists, it’s that the chemists who use machines will replace those that don’t. ...”

Derek Lowe (In the Pipeline)

Acknowledgements

GDC SynTech

Aaron Kendrick
Ali Ross
Angela Vitrey
Ayako Honda
Bill Ulmer
Carina Sanchez
Carol Ginsburg-Moraff
Claudio Bomio
Constanze Hartwieg
Corentin Bon
David Whitehead
Dominik Rufle
Dyuti Majumdar
Fabio Lima
Hansjoerg Lehmann
Heiner Schütz
Holly Davis
Jerome Andre
Jim Brown
Jules Lee

Julia Köhler
Juliette Tinchant
Kian Tan
Markus Furegati
Max Mues
Nate Ware
Netgie Laguerre
Odile Decoret
Olga Mukhina
Pablo Gabriel
Pete Delgado
Reto Hurschler
Richard Robinson
Rohit Duvadie
Sandro Nocito
Scott Plummer
Sophie Racine
Stephanie Rothe-Pöllet
René Wyler
Will Lau

MicroCycle & Automation

Alexandre Luneau
Anna Pelliccioli
Claudio Valente
Daniel Baeschlin
Daniel Gosling
Daniel Meyer
Doug Auld
Ekkehard Goerlach
Gina Geraci
Grant Eastman
Gregori Gerebtzoff
Holger Schlingensiepen
Ian Hunt
Jennifer Poirier
John Reilly
Karl Chin
Laura West
Marcel Reck
Marco Palmieri
Matthias Hübscher

Mike Fortunato

Nadine Schneider
Patricia Gehin
Pei-i Ho
Peter Ertl
Peter Meier
Ritesh Tichkule
Sandra Wildhaber
Stephane Rodde
Thomas Wolf
Tom Caya
Tom Dice
Trixie Wagner
Vincent Unterreiner

GDC & BR Leadership

Academic collaborators

Prof. Jeffrey Bode

Dr. Cara Brocklehurst MChem Ph.D FRSC
cara.brocklehurst@novartis.com

Thank you

