AstraZeneca
IMED Biotech Unit

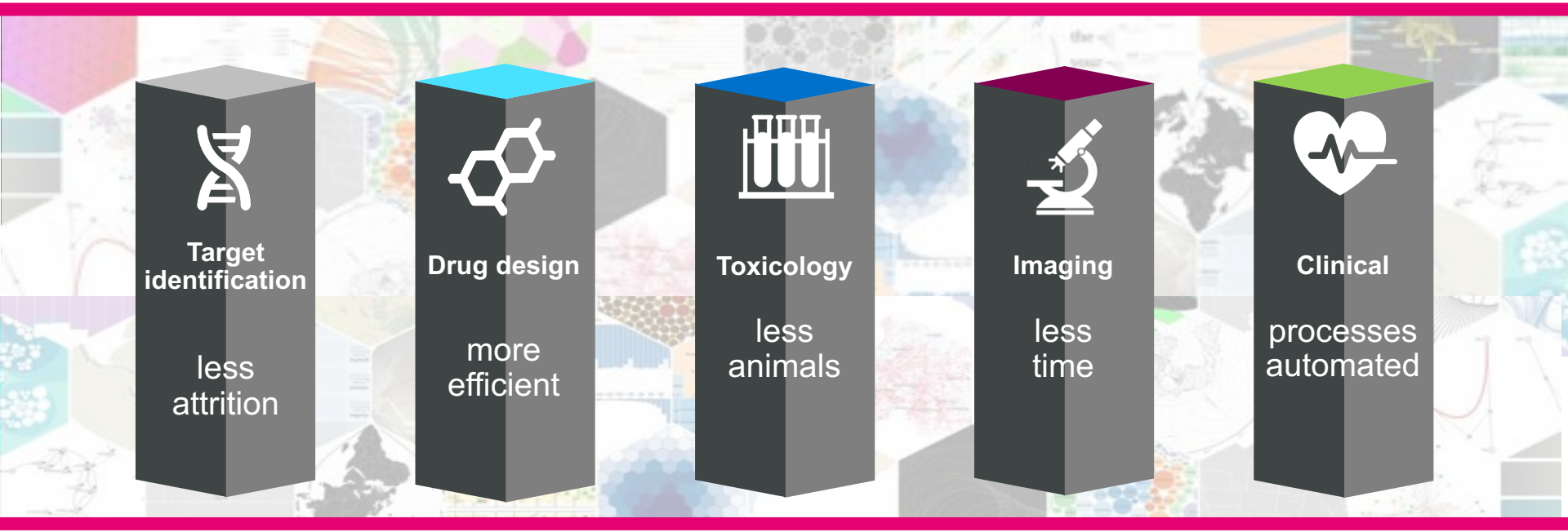
Applying Artificial Intelligence in Drug Design

Ola Engkvist, Hit Discovery, Discovery Sciences, R&D AstraZeneca Gothenburg, Sweden

CompBioMed, London

September 26 2019

Where can AI impact Pharmaceutical R&D



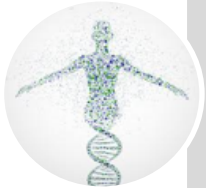
Drug Design

What to make next?

How to make it?

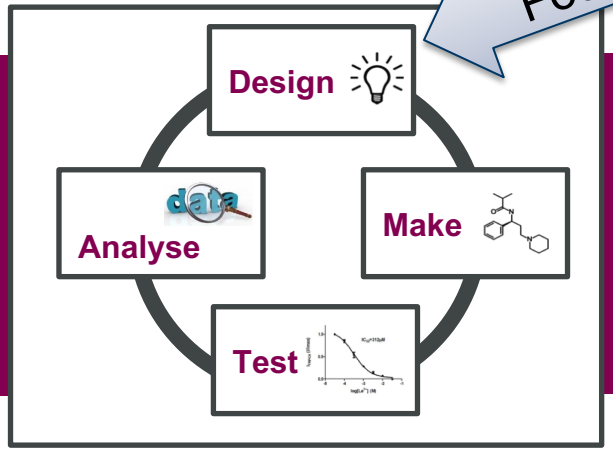
The Design Make Test Analyze cycle in Drug Design

Drug target



Chemical starting point (“Hit”) found through HTS, DEL, fragment screening or knowledge

- Weakly active
- Target unselective
- Toxicity risk
- Low metabolic stability



Candidate drug

- Highly potent
- Effective in *in vivo* models
- Metabolically stable
- No toxicity issues

~3 years

Multiple of DMTA cycles

4-6 weeks per cycle

Hand-overs between multiple labs

The challenge: Find ways to speed up and improve the process using AI

Key priority areas in ML/AI

- Deep learning based de novo molecular design
- Synthesis Prediction
- More accurate property predictions
- Decision making under uncertainty

+

Science @AZ

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2018, 4, 120−131

Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks

Marwin H. S. Segler,*[,†] Thierry Kogej,[‡] Christian Tyrchan,[§] and Mark P. Waller*[,∥]

Commentary

For reprint orders, please contact: reprints@future-science.com

Future Medicinal Chemistry

The convergence of artificial intelligence and chemistry for improved drug discovery

Clive P Green*[,1], Ola Engkvist[2] & Garry Pairaudeau[3]

RESEARCH

Molecular De-Novo Design through Deep Reinforcement Learning

Marcus Olivecrona*, Thomas Blaschke†, Ola Engkvist† and Hongming Chen†

Application of Generative Autoencoder in *De Novo* Molecular Design

Thomas Blaschke,*[a, b] Marcus Olivecrona,[a] Ola Engkvist,[a] Jürgen Bajorath,[b] and Hongming Chen*[a]

The rise of deep learning in drug discovery

Hongming Chen[1], Ola Engkvist[1], Yinhai Wang[2], Marcus Olivecrona[1] and Thomas Blaschke[1]

[1]Hit Discovery, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, AstraZeneca R&D Gothenburg, Mölndal 43183, Sweden
[2]Quantitative Biology, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Unit 310, Cambridge Science Park, Milton Road, Cambridge CB4 0WG, UK

Computational prediction of chemical reactions: current status and outlook

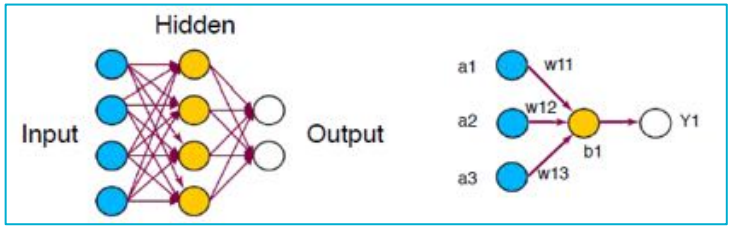
Ola Engkvist[1], Per-Ola Norrby[2], Nidhal Selmi[1], Yu-hong Lam[3], Zhengwei Peng[3], Edward C. Sherer[3], Willi Amberg[4], Thomas Erhard[4] and Lynette A. Smyth[4]

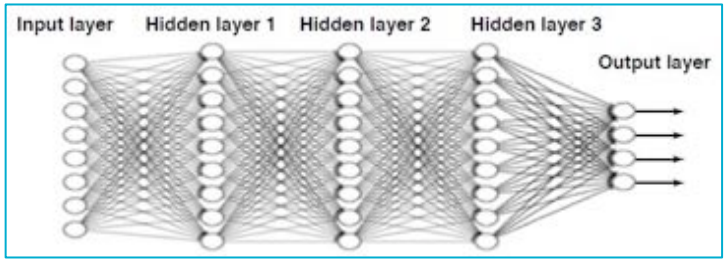
Ola Engkvist was awarded his PhD in computational chemistry by the University of Lund in 1997, and continued with postdoctoral research at

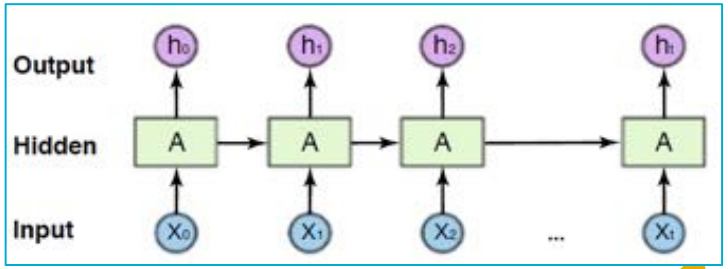
Open Source:
https://github.com/MarcusOlivecrona/REINVENT

Neural Networks & Deep Learning

- **Neural Networks known for decades**
  - Inputs, Hidden Layers, Outputs
  - Single layer NNs have been used in QSAR modelling for years
- **Recent Applications use more complex networks such as**
  - Multi-layer Feed-Forward NNs
  - Convolutional NNs
    - biological image processing
  - Auto-encoder NNs
  - Recurrent NNs
    - Trained using Maximum Likelihood Estimation to maximize the likelihood of next character







Why? Generation of Novel Compounds in the 10^{60} Chemical Space!

10^{10}-10^{12}

10^{60}

Journalist units:

Known space: 0,00017 ng of Hydrogen atoms

Possible space: The Hydrogen atoms in 90 Suns

Where´s the impact?

- Use for de novo Molecular Design
  - Scaffold Hopping
  - Novelty
  - Virtual Screening
  - Library Design

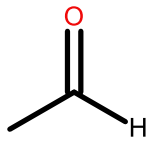
Natural language generation and molecular structure generation

- Can we borrow concepts from natural language processing and apply to SMILES description of molecular structures to generate molecules?

The → grass → is → ?

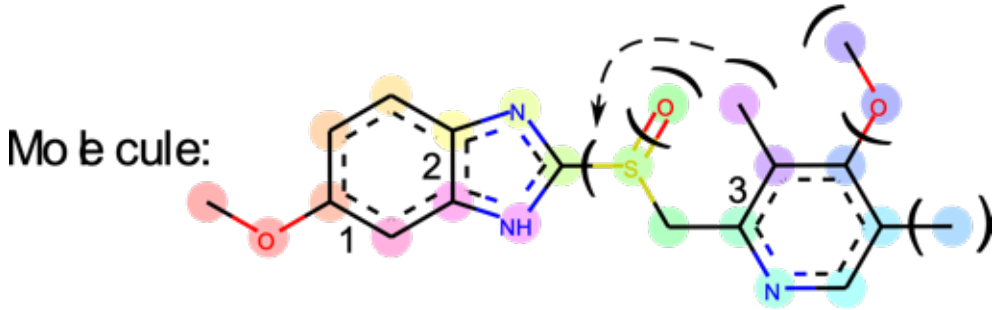
- Conditional probability distributions given context
- $P(green \mid is, grass, The)$

C → C → = → ?



- $P(O \mid =, C, C)$

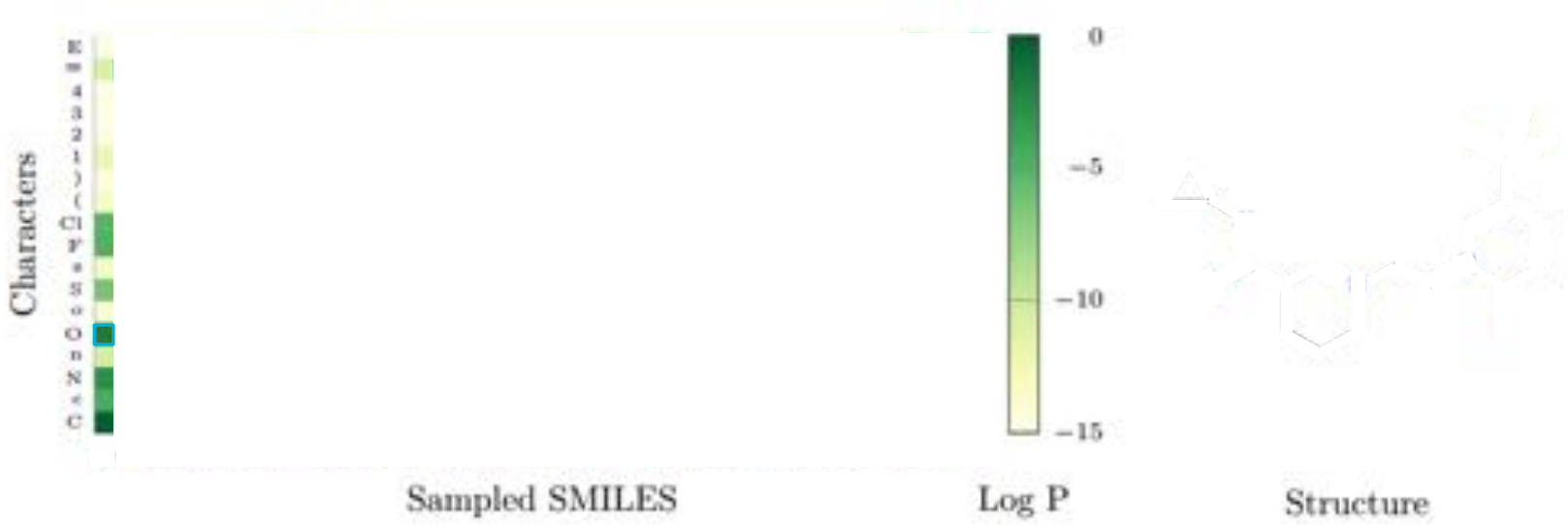
Simplified Molecular Input Line Entry Specification (SMILES)



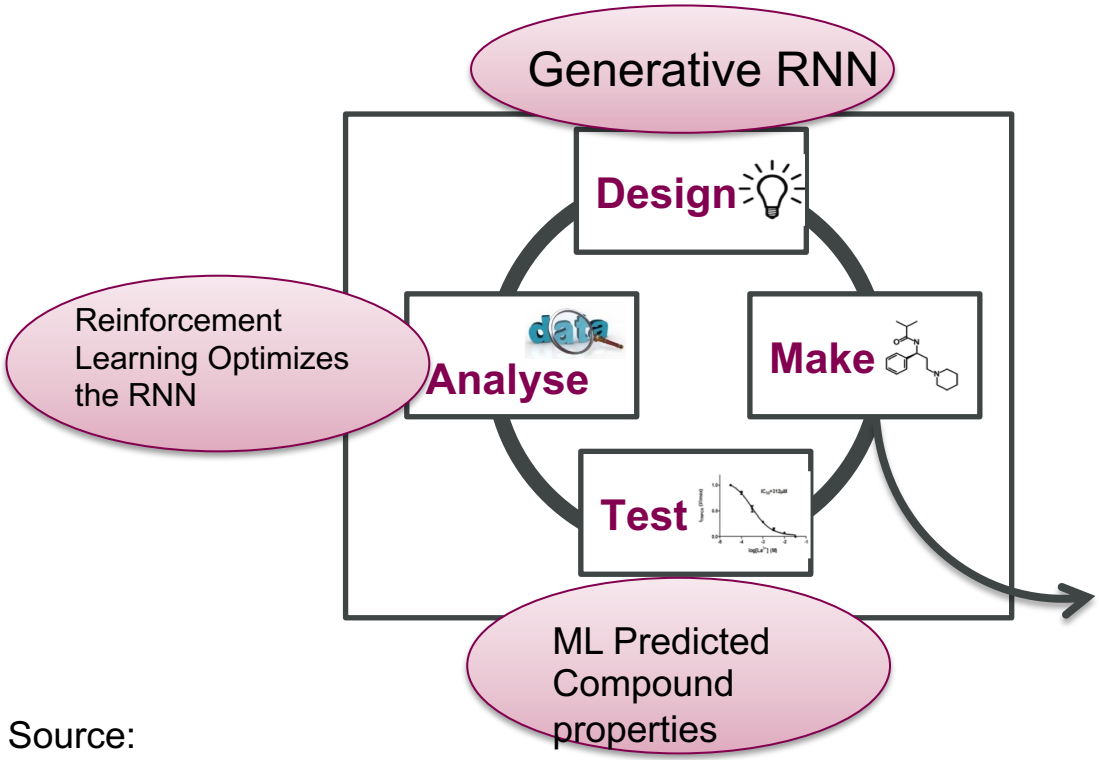
SMILES: COc1ccc2nc(S(=O)Cc3ncc(C)c(OC)c3C)[nH]c2c

- A sequence format for molecules
- Allows us to use the progresses made with natural language processing in the recent years ☺

The generative process



Reinforcement Learning: An *In Silico* mini-DMTA cycle



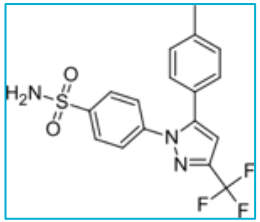
The Value:
Molecules for DMTA cycle

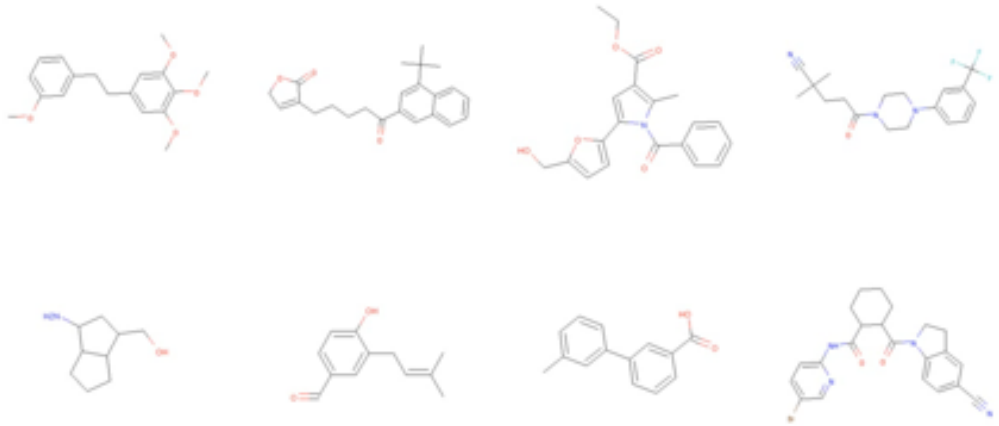
Produces novel scaffolds and improved compound suggestions for drug discovery projects

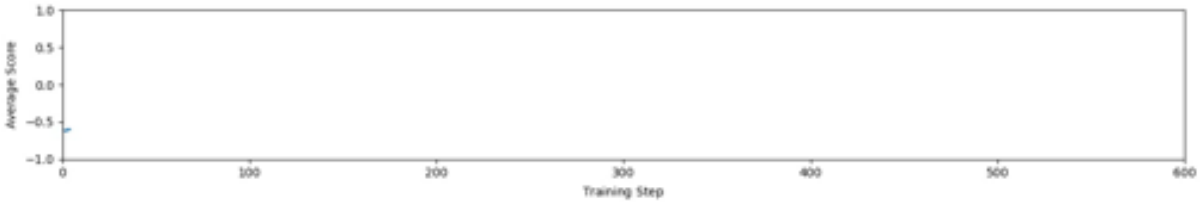
Less real world DMTA cycles => Saved time

Open Source:
https://github.com/MarcusOlivecrona/REINVENT

AI live: Create Structures Similar to Celecoxib



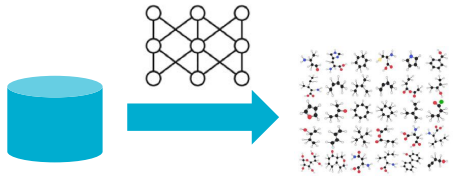




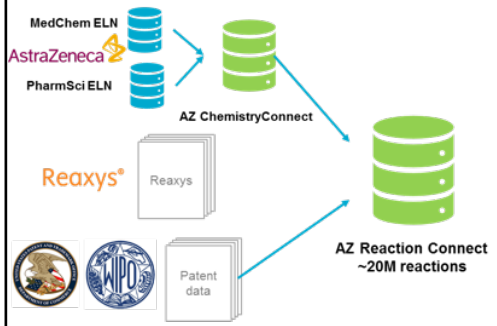
- **Key Message**
  - RNN generates structures similar to Celecoxib
  - Rapid sampling!
  - Average score describes how many learning steps are required to reach similar compounds

Artificial Intelligence Guided Drug Design Platform

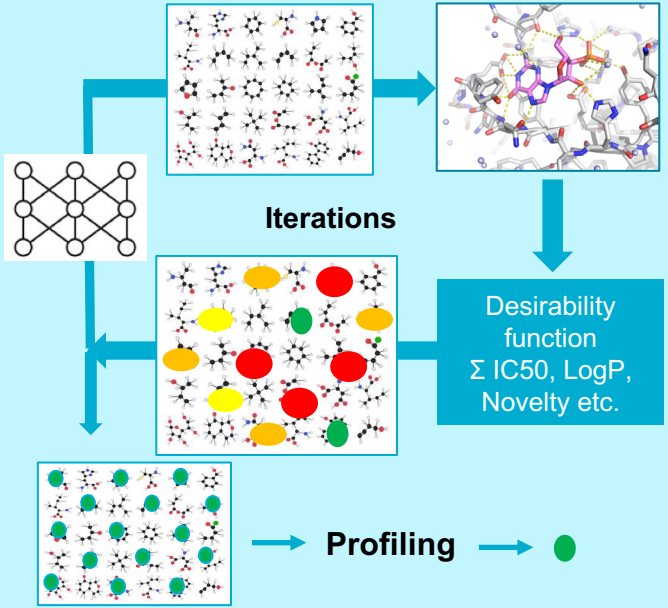
Generation of Novel Chemical Space



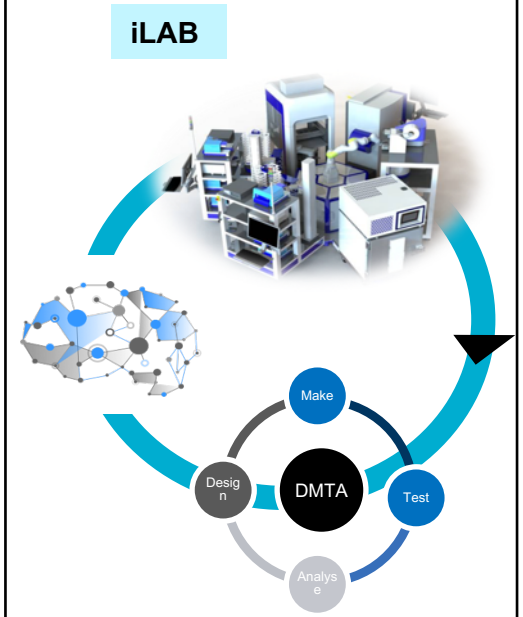
Reaction & Synthesis Prediction



AI Design Platform

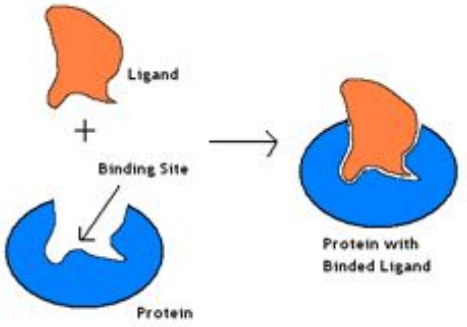


Fully Automated DMTA Cycle



How can we improve affinity prediction?

- Synergize with automation
- Better Machine Learning Models
  - Access to more data (MELLODDY)
  - Experimental descriptors
  - Graph convolution, include protein based information
  - Multi-task modelling
  - Matrix factorization with side information
- Free energy calculations
  - Progress in speed
  - Combine with machine learning
- Confidence estimation
  - Conformal prediction
  - Bayesian methods
- Benchmarking
  - Public Chemogenomics set available (Excape-DB, Pidgin)
  - Blind competitions (SAMPL, D3R)



MACHINE LEARNING LEDGER ORCHESTRATION FOR DRUG DISCOVERY

JUNE 2019 – MAY 2022

MELLODDY

powered by aws

PHARMA PARTNERS

AMGEN · astellas · AstraZeneca · BAYER · Boehringer Ingelheim · gsk · janssen · MERCK · NOVARTIS · SERVIER

PUBLIC PARTNERS

MŰEGYETEM 1782 · IKTOS · KU LEUVEN · loodse · NVIDIA · OWKIN · SUBSTRA FOUNDATION

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N° 831472. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA

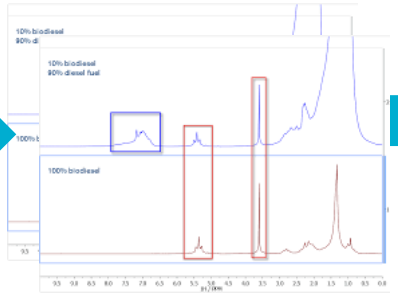
imi innovative medicines initiative · efpia

How can we create a successful AI friendly environment?

Input data
Size, Quality
Historical, Bespoke

Decision making AI
Creativity, Accuracy,
Confidence, Transparency

Experimental Validation
Speed, Quality



Acknowledgements

Discovery Sciences Molecular AI Team

Thierry Kogej
Hongming Chen
Isabella Feierberg
Atanas Patronov
Esben Jannik Bjerrum
Preeti Iyer
Christian Margreiter
Lewis Mervin
Kostas Papadoupoulos
Jiangming Sun (Postdoc 2015-2017)
Noe Sturm (Postdoc 2017-2018)
Philipp Buerger (Postdoc 2017-2019)
Jiazhen He (Postdoc 2019-2022)
Rocio Mercado (Postdoc 2018-2021)
Thomas Blaschke (PhD student 2017-2018)
Josep Arus Pous (PhD student 2018-2019)
Michael Withnall (PhD student 2018-2019)
Oliver Laufkötter (PhD student 2018-2019)
Laurent David (PhD student 2018-2019)
Ave Kuusk (PhD student 2016-2019)
Marcus Olivecrona (AZ GradProgram 2017)
Alexander Aivazidis (AZ GradProgram 2018)
Dhanushka Weerakoon (AZ GradProgram 2018-2019)
Panagiotis-Christos Kotsias (AZ AI GradProgram 2018-2020)
Dean Sumner (AZ AI GradProgram 2019-2020)
Edvard Lindelöf (Master Thesis Student 2018-2019)
Simon Johansson (PhD Student 2019-2024)
Oleksii Prykhodko (Master Thesis Student 2019)
Viktor Norrsjö (Master Thesis Student 2019-2020)

Discovery Sciences

Garry Pairaudeau
Clive Green
Lars Carlsson
Nidhal Selmi
Michael Kossenjans
Anna Tomberg

DSM AI Team

Ernst Ahlberg
Suzanne Winiwarter
Ioana Oprisiu
Ruben Buendia (Postdoc 2018)

PharmSci

Per-Ola Norrby

2018 PoP Pilot Study

Werngard Czechtizky
Ina Terstiege
Christian Tyrchan
Anders Johansson
Jonas Boström
Kun Song
Alex Hird
Neil Grimster
Richard Ward
Jeff Johannes

Academic Collaborators

Marwin Segler (Munster)
Juergen Bajorath (Bonn)
Jean-Louis Reymond (Bern)
Andreas Bender (Cambridge)
Sepp Hochreiter (Linz)
Gunther Klambauer (Linz)
Sami Kaski (Helsinki)

Confidentiality Notice