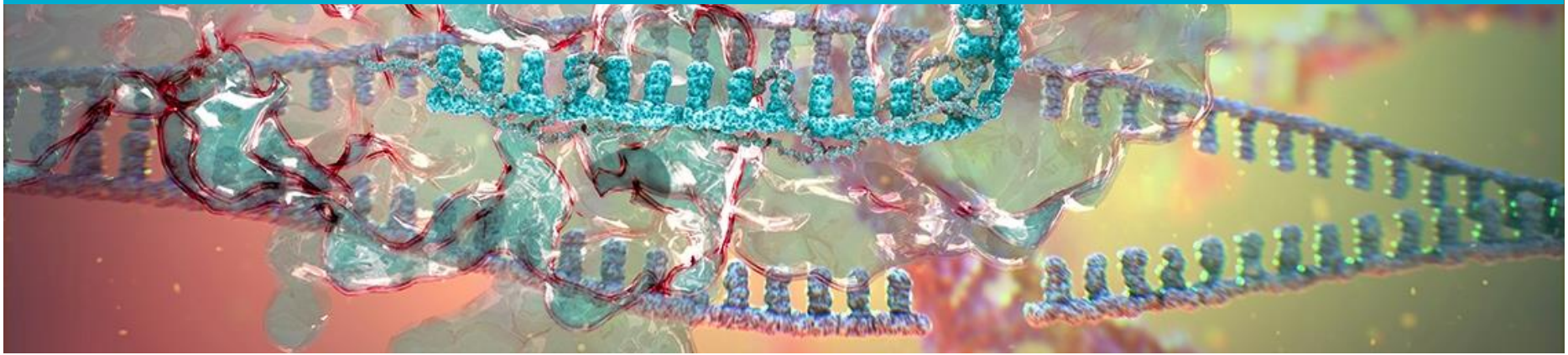
AstraZeneca
IMED Biotech Unit

Artificial Intelligence in Drug Design Progress and Bottlenecks

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

CompBioMed Meeting

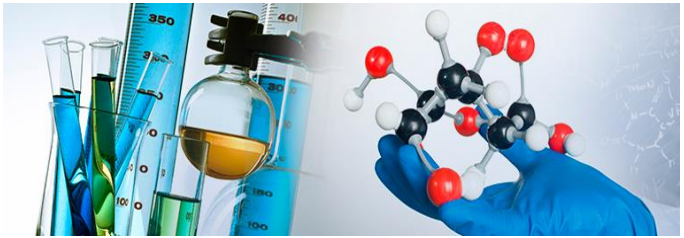
March 16 2020



Drug Design

What to make next?

How to make it?

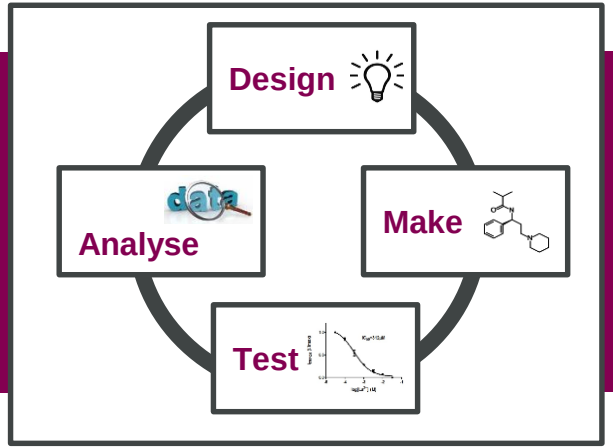


The Design Make Test Analyze cycle in Drug Design

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

10s-100s of DMTA cycles

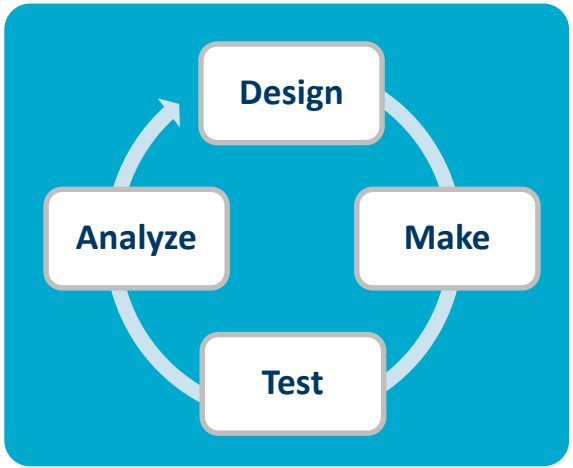
3-6 weeks per cycle

Hand-overs between multiple labs

Augmented Drug Discovery

How can we reduce the time to deliver a clinical candidate?

Select the most efficient synthetic route



Make more compounds in each cycle

Increase speed

Maximize learning

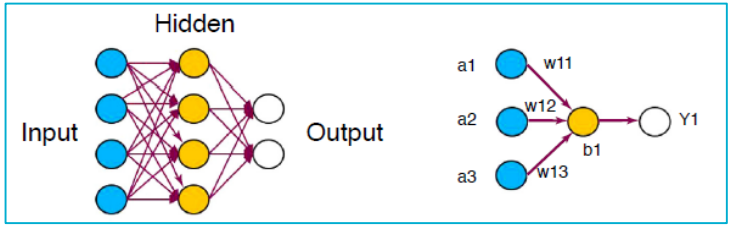
Key priority areas in ML/AI

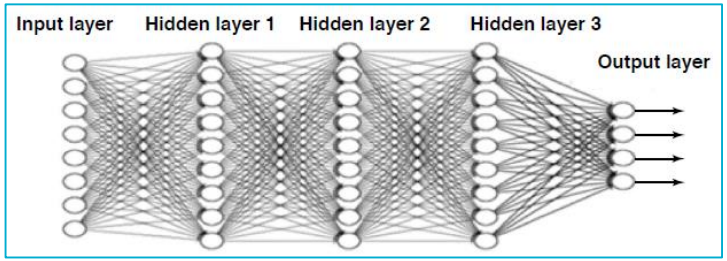
- Deep learning based de novo molecular design
- Synthesis Prediction
- More accurate property predictions
- Decision making under uncertainty

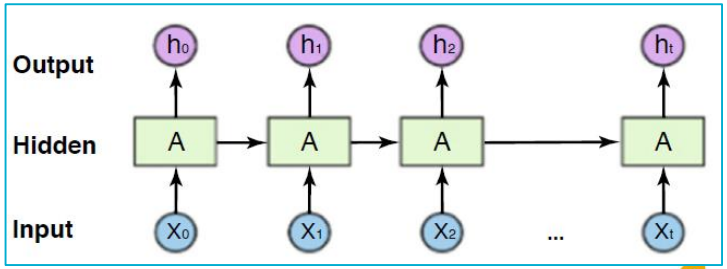
+

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

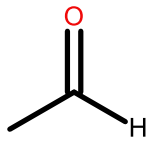
Natural language generation and molecular structure generation

- 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
- $P(green \mid is, grass, The)$

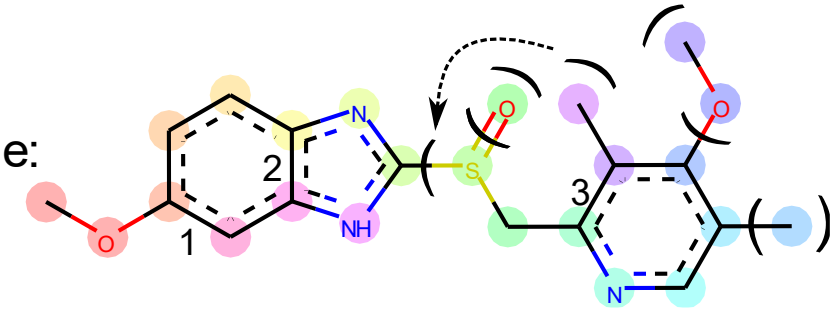
C → C → = → ?



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

Simplified Molecular Input Line Entry Specification (SMILES)

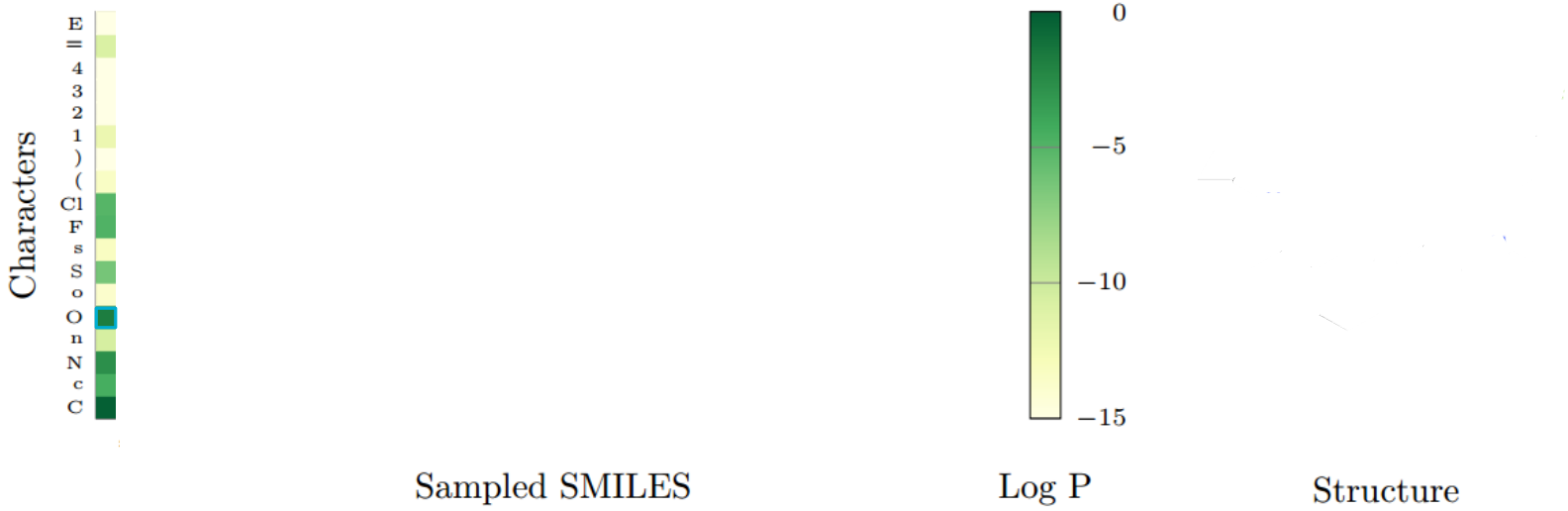
Molecule:



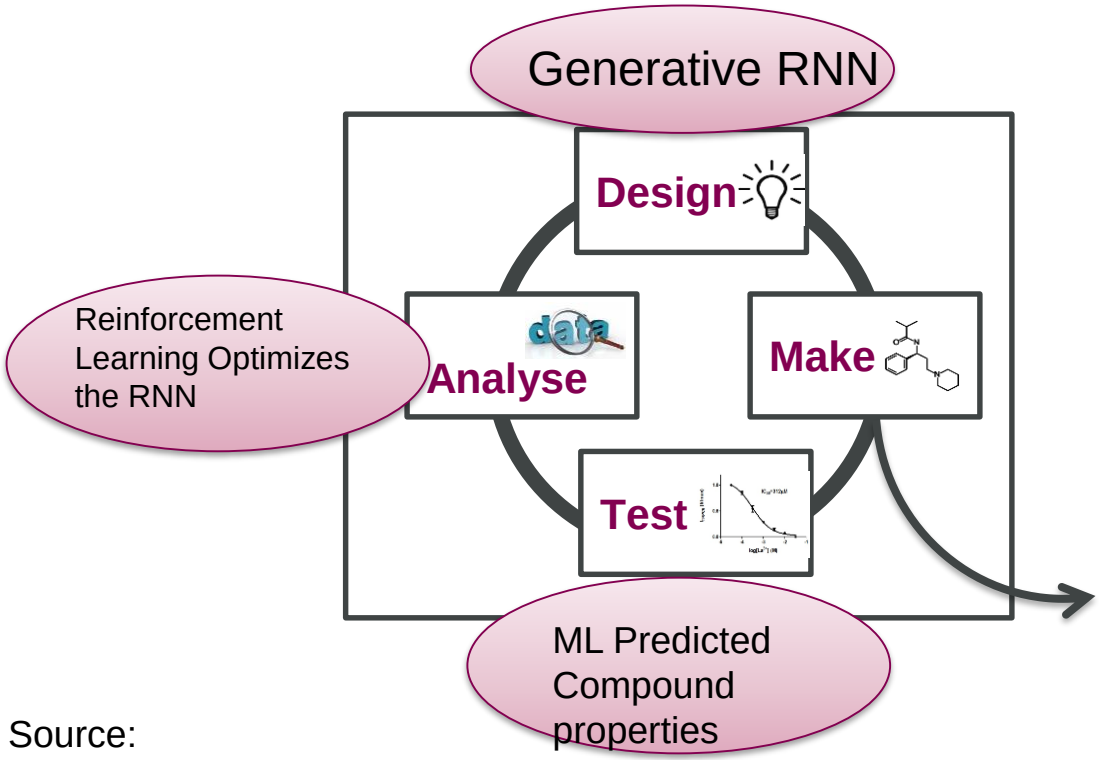
SMILES: COc1ccc2nc(S(=O)Cc3ncc(C)c(OC)c3C)[nH]c2c1

- 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



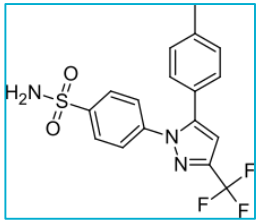
The Value:
Molecules for DMTA cycle

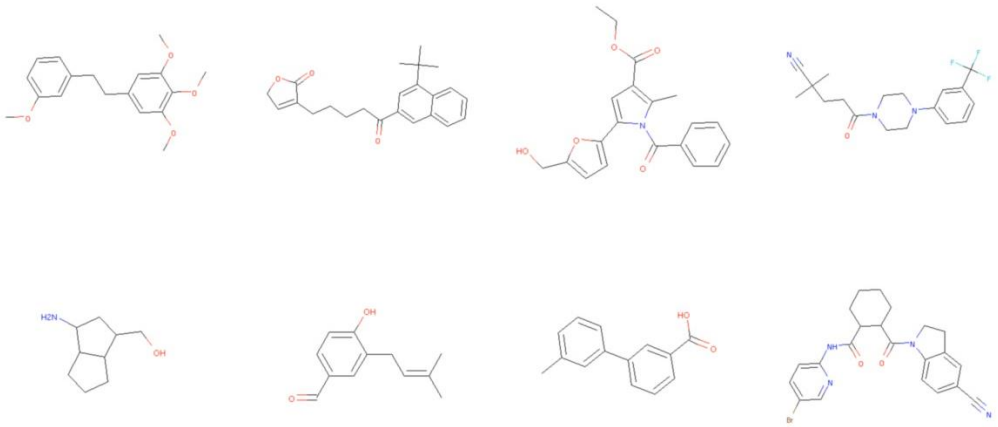
Produces novel scaffolds and improved compound suggestions for drug discovery projects

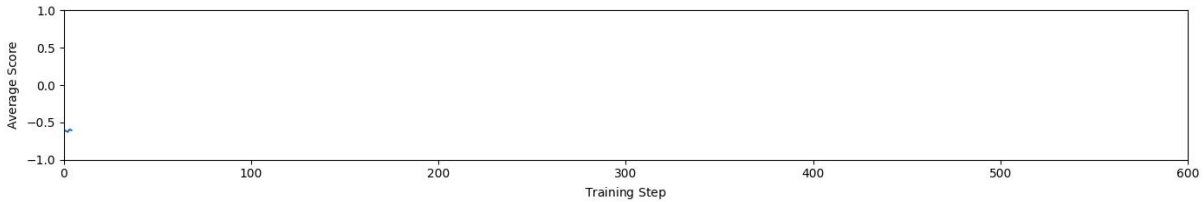
Less real world DMTA cycles => Saved time

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

AI live: Create Structures Similar to Celecoxib



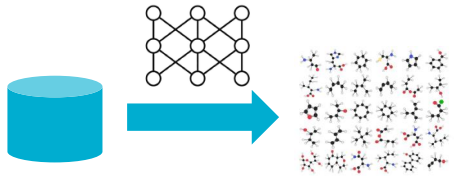




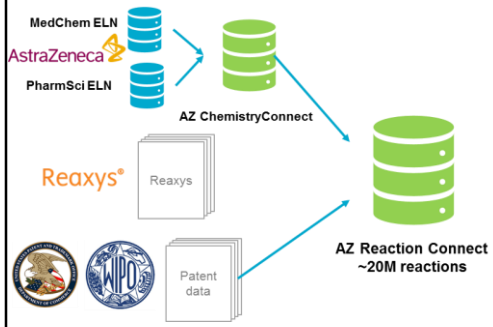
- **Key Message**
  - RNN generates structures similar to Celecoxib
  - Rapid sampling!
  - Average score describes how many learning steps are required to reach similar compounds

Artificial Intelligence Guided Drug Design Platform

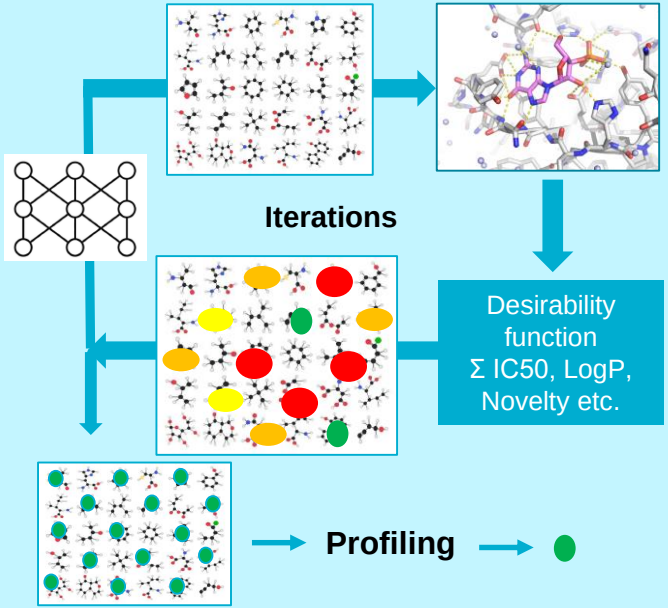
Generation of Novel Chemical Space



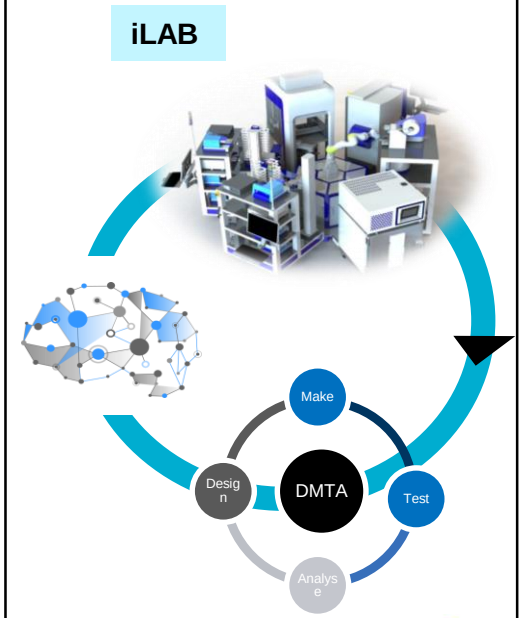
Reaction & Synthesis Prediction



AI Design Platform



Fully Automated DMTA Cycle



Lessons learned from project

- Novel scaffolds were identified in a crowded chemical space
- Compound series could be efficiently optimised
- ADME and especially binding affinity predictions are limiting factors
- Too many ideas might make prioritization for synthesis challenging
- Chemistry resources might be frontloaded to assess the generated ideas
- Currently used in 12-15 LG/LO projects at all sites
- Continuously build of REINVENT platform

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

What are the challenges for AI driven drug design?

- Scaling AI and chemistry automation for drug design to a whole drug discovery project portfolio including projects with low data volume
- Binding affinity and solubility predictions are major bottlenecks
- “Cambrian revolution” of new AI methods makes it difficult to assess progress
- Educational, cultural & logistical challenges besides scientific
- The bar is set high to transform drug design

Acknowledgements

Molecular AI Team

Thierry Kogej
Hongming Chen (2001-2019)
Isabella Feierberg
Atanas Patronov
Esben Jannik Bjerrum
Preeti Iyer
Christian Margreitter
Lewis Mervin
Kostas Papadopoulos
Samuel Genheden
Christos Kannas
Alexey Voronov
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)
Tomas Bastys (postdoc 2019-2022)
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)
Amol Thakkar (PhD student 2019-2020)
Ave Kuusk (PhD student 2016-2019)
Marcus Olivecrona (AZ Graduate Program 2017)
Alexander Aivazidis (AZ Graduate Program 2018)
Dhanushka Weerakoon (AZ Graduate Program 2018-2019)
Panagiotis-Christos Kotsias (AZ AI Graduate Program 2018-2020)
Dean Sumner (AZ AI Graduate Program 2019-2020)
Edvard Lindelöf (Master Thesis Student 2018-2019)
Simon Johansson (PhD Student 2019-2024)
Oleksii Prykhodko (Master Thesis Student 2019)
Viktor Norrsiö (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
Graham Smith
Ruben Buendia (Postdoc)

PharmSci

Per-Ola Norrby
Kjell Johner
David Buttar

AI Projects

Werngard Czechtizky
Ina Terstiege
Christian Tyrchan
Anders Johansson
Jonas Boström
Kun Song
Alex Hird
Neil Grimster
Richard Ward
Jeff Johannes
Graeme Robb
Eva Nittinger
Anna Tomberg
Kathryn Giblin

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

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

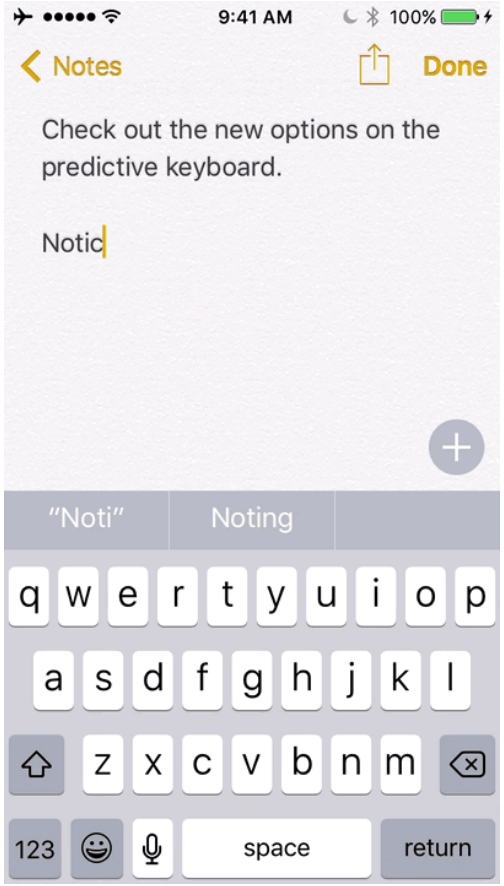
Ola Engkvist[1], Per-Ola Norrby[2], Nidhal Selmi[1], Yu-hong Lam[3], Zhengwei Peng[3], Edward C. Sherer[3], Willi Amberg[4], Thomas Erhard[4] and Lynette A. Smyth[4]

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

BigChem

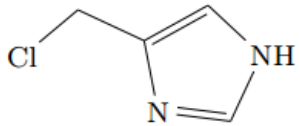
Recurrent Neural Network & Natural language generation



Tokenization of SMILES

- Tokenize combinations of characters like “Cl” or “[nH]”
- Represent the characters as one-hot vectors

Graph:



SMILES: ClCc1c[nH]cn1

One-hot encoding:

	Cl	C	c	1	c	nH	c	n	1
C	0	1	0	0	0	0	0	0	0
c	0	0	1	0	1	0	1	0	0
n	0	0	0	0	0	0	0	1	0
1	0	0	0	1	0	0	0	0	1
nH	0	0	0	0	0	1	0	0	0
Cl	1	0	0	0	0	0	0	0	0