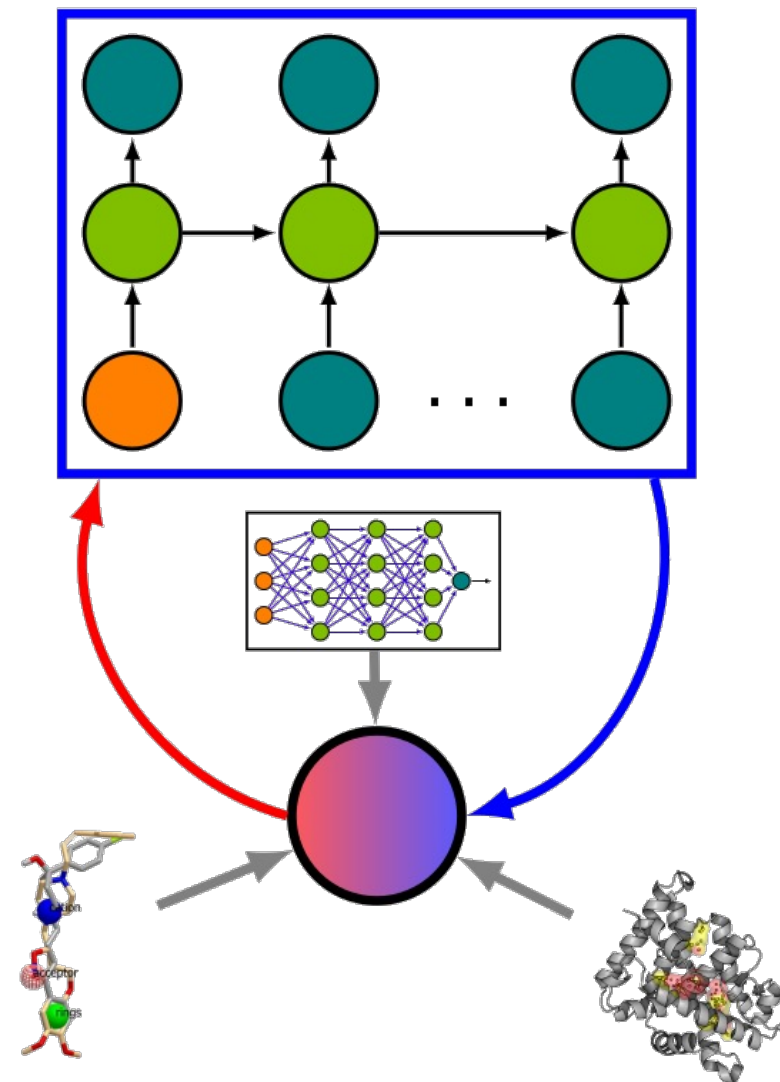


Accelerating drug design with AI & simulation

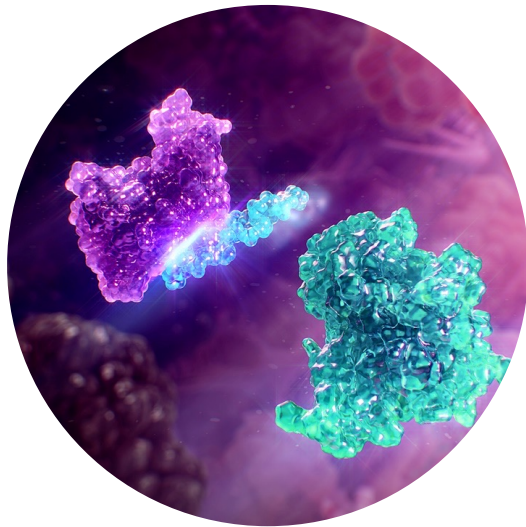
Jon Paul Janet

Molecular AI, Discovery Sciences, BioPharmaceuticals
R&D, AstraZeneca, Gothenburg, Sweden

October 2022



AstraZeneca 2021 global dimensions



\$37.4bn

Total Revenue
(incl. COVID-19 vaccine)

+38%

Total Revenue growth
(23% excl. COVID-19
vaccine)

\$9.7bn

invested in our science

13

medicines with annual
sales of more than
\$1 billion

\$13bn

Oncology Product Sales

\$8bn

Cardiovascular, Renal &
Metabolism Product Sales

\$6bn

Respiratory &
Immunology Product Sales

\$3bn

Rare Disease Product Sales
(from 21/7/2021)

22

Regulatory approvals
and authorisations in
major markets

2.5bn

COVID-19 vaccine doses
supplied to more than
180 countries together
with our partners

110

successful markets
launches

161

projects in clinical phase
of development

83,100

employees
(Dec. 2021)

87%

of employees believe
strongly in our future
direction and key
priorities (Nov. 2021)

59%

Reduction in Scope 1 and
2 greenhouse gas
emissions since 2015

+31m

people reached through
our Access to Healthcare
programmes



Global, science-led, patient-focused biopharmaceutical company



Science and innovation-led



Therapy area focus

- Oncology
- Cardiovascular, Renal & Metabolism
- Respiratory & Immunology
- Vaccines & Immune Therapies
- Rare Disease



Diversified portfolio with broad coverage across primary care, specialty care and rare diseases



Commitment to people, society and the planet



Global strength, with balanced presence across regions

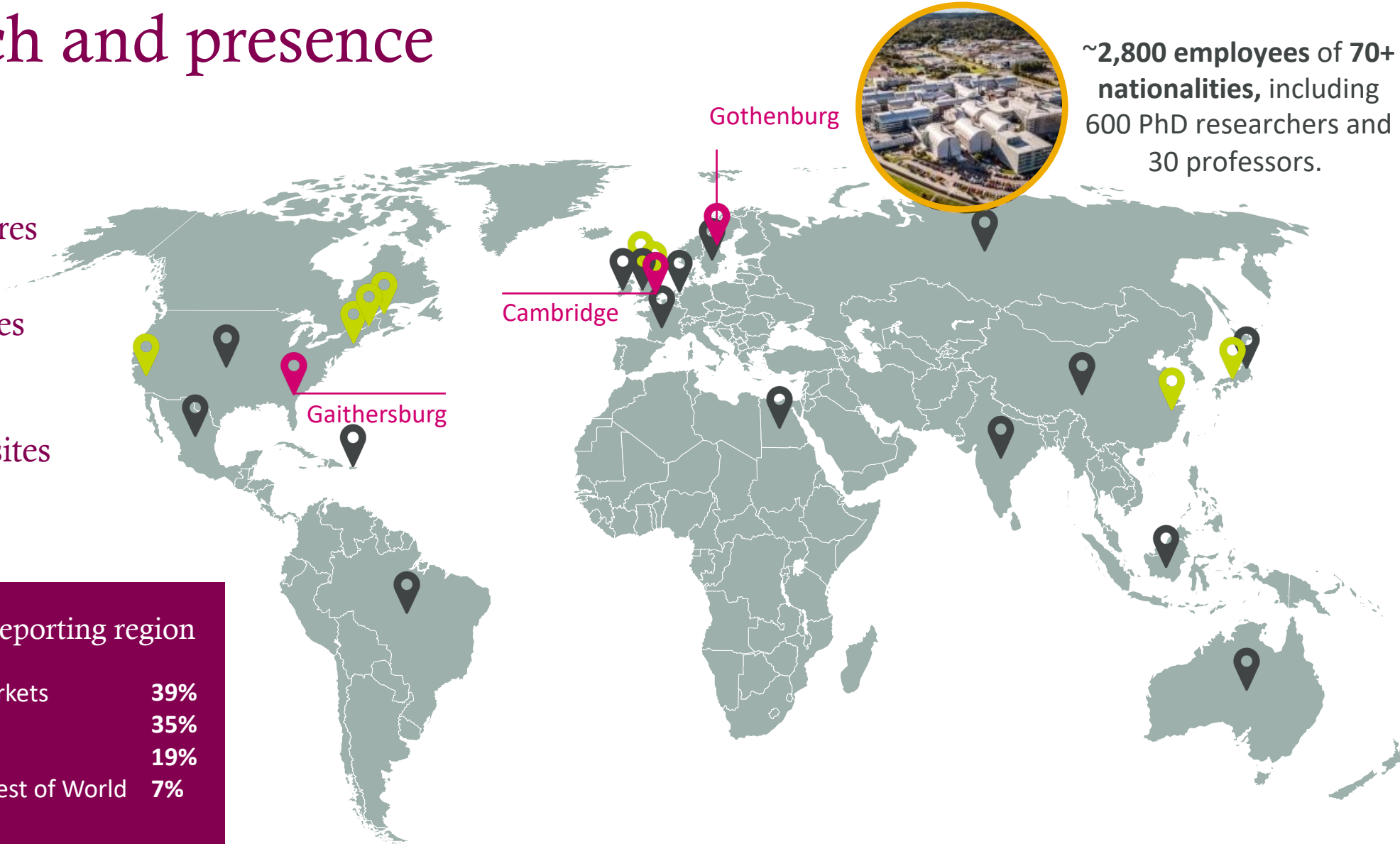


Global reach and presence

 3 global R&D centres

 8 other R&D centres and offices

 28 manufacturing sites in 16 countries

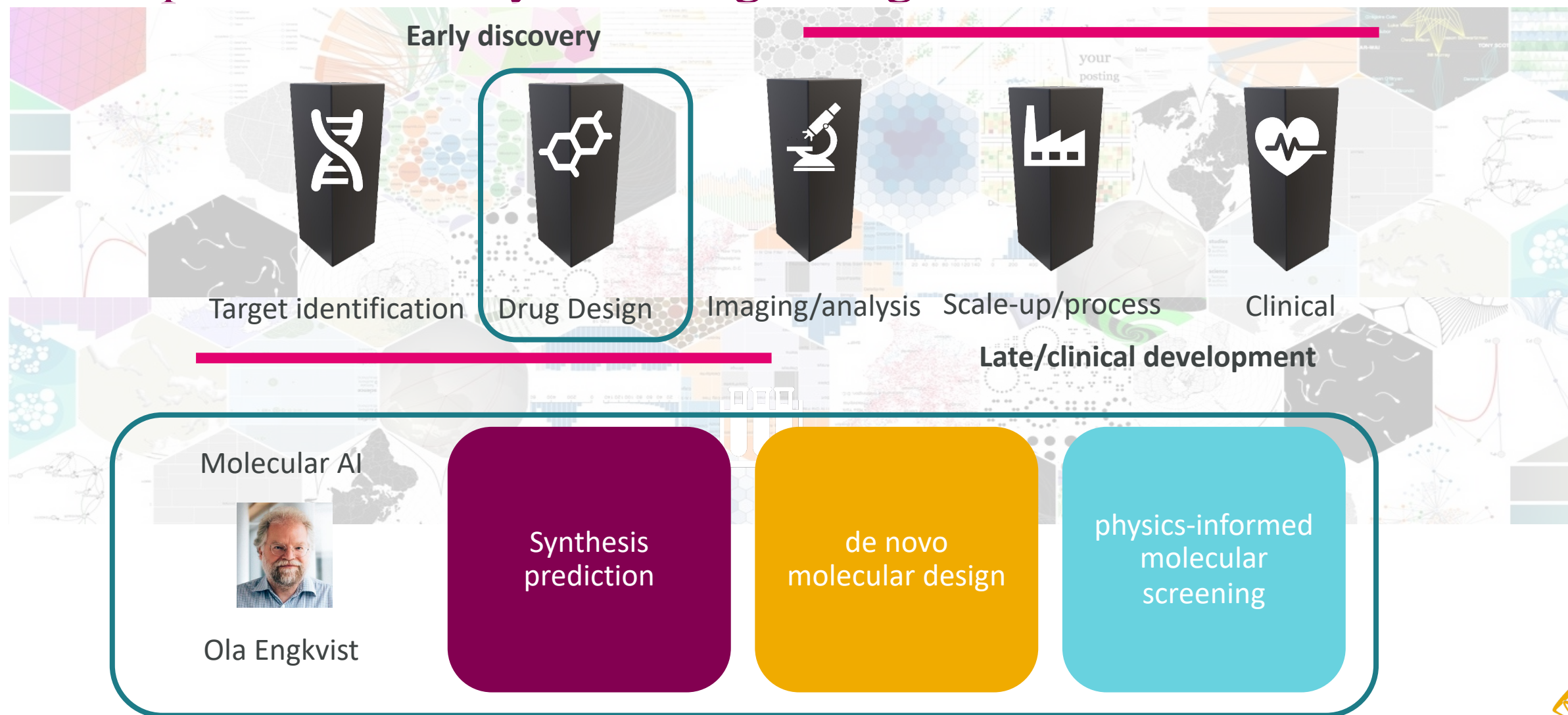


AstraZeneca's significance in numbers

- **17.4** billion SEK AstraZeneca AB's investment in R&D
- **91** billion SEK in Swedish exports
- **1** billion SEK sales in Sweden
- **11.4%** of all shareholders are Swedish
- **7,600** employees across AstraZeneca Sweden,
- **2,800** in Gothenburg
- **4,800** Södertälje



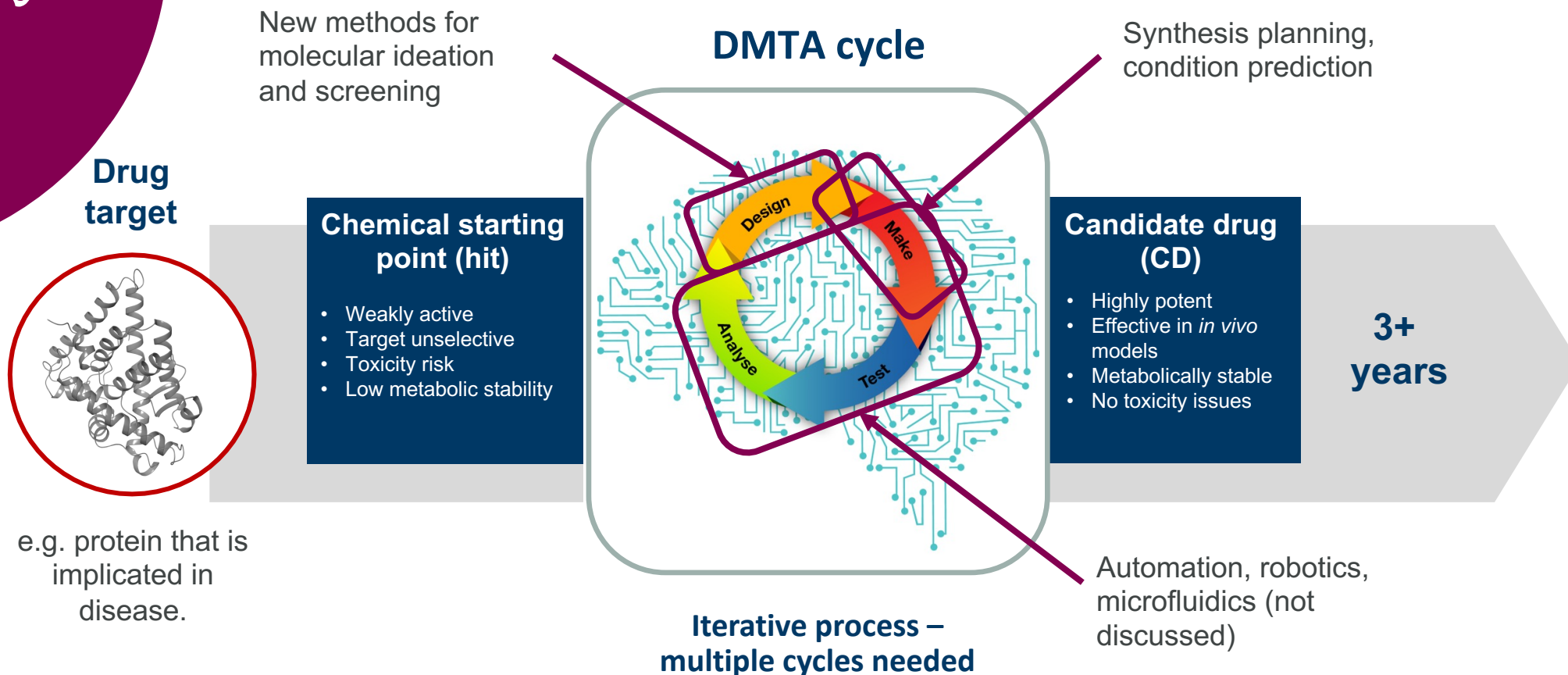
Machine learning techniques are poised to impact pharmaceutical development industry from beginning to end



The Drug Discovery Process

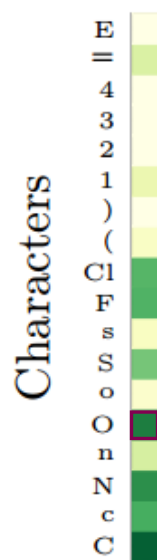
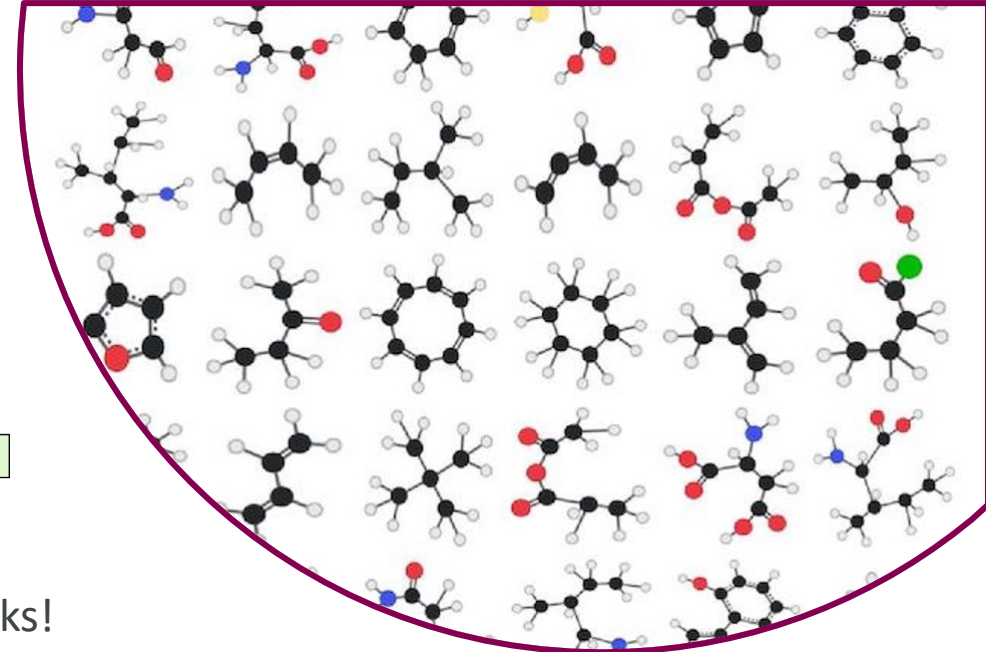
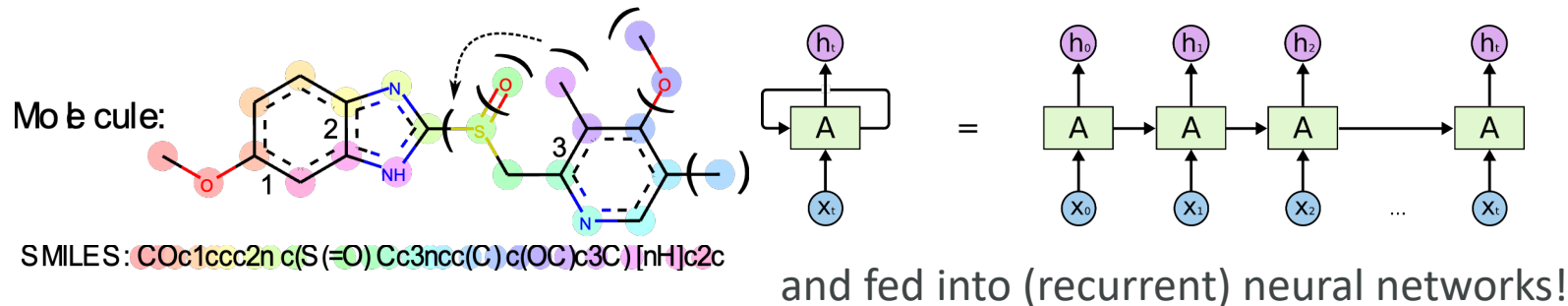
At the heart of the drug design process is the Design, Make, Test and Analyze (DMTA) cycle, which is a core concept for iterative, hypothesis driven design.

How can we accelerate this process?



Chemical language models are central to much of our work

Molecules can be described in the language of SMILES...



Sampled SMILES



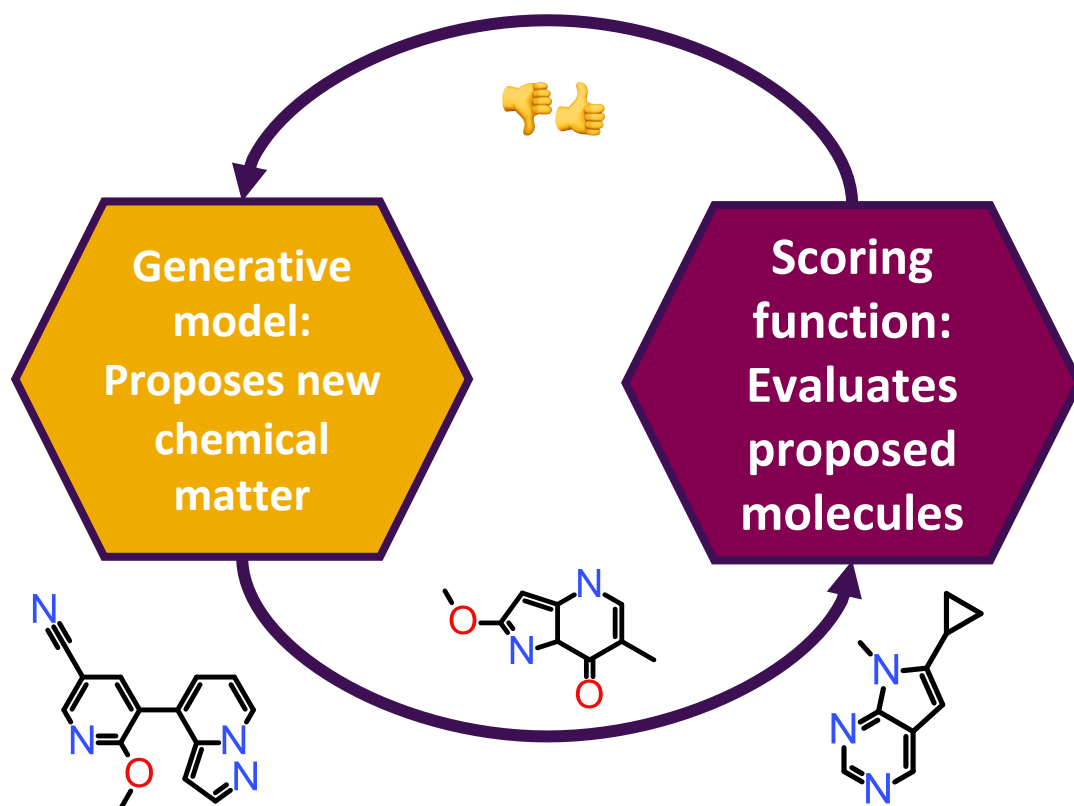
Log P

Structure



REINVENT – designing new molecules with AI

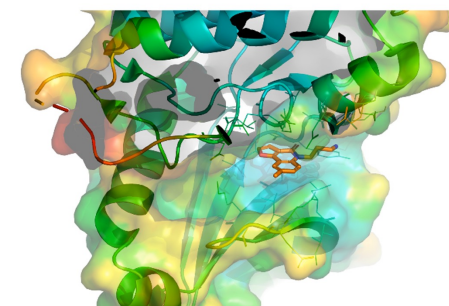
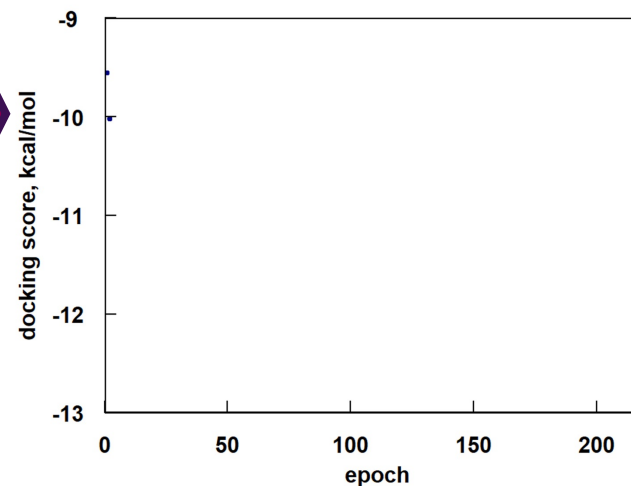
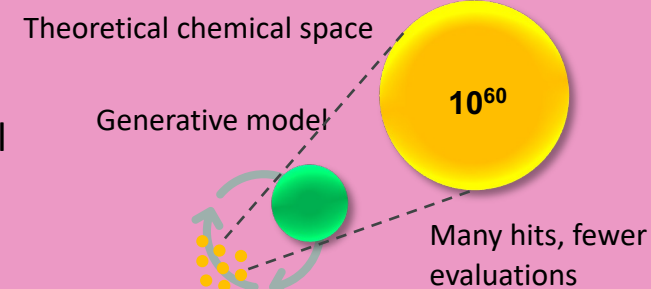
REINVENT is the in-house developed de novo molecular design tool, using generative reinforcement learning to solve *in silico* molecular design tasks



Traditional approaches rely on searching a large database for a small number of suitable hits

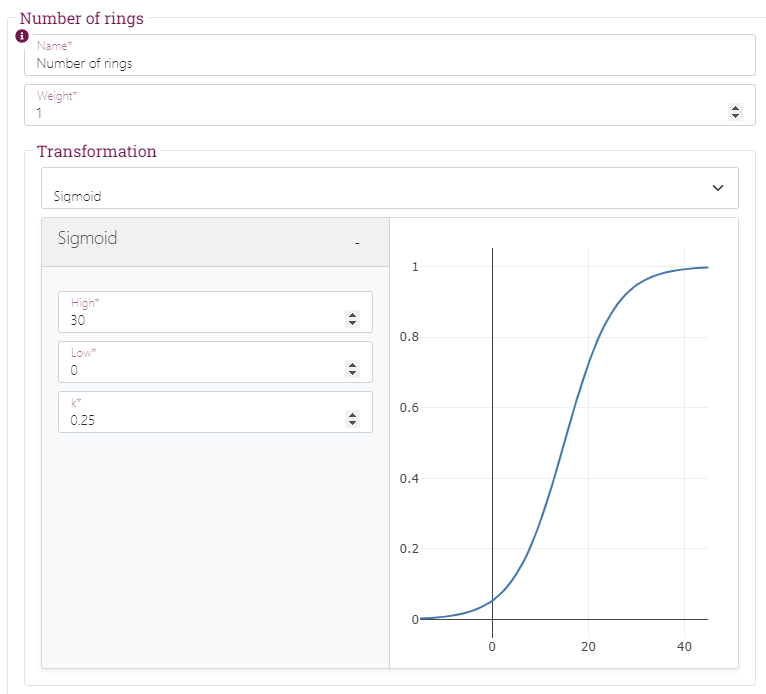


Generative models encode practically unlimited chemical space probabilistically

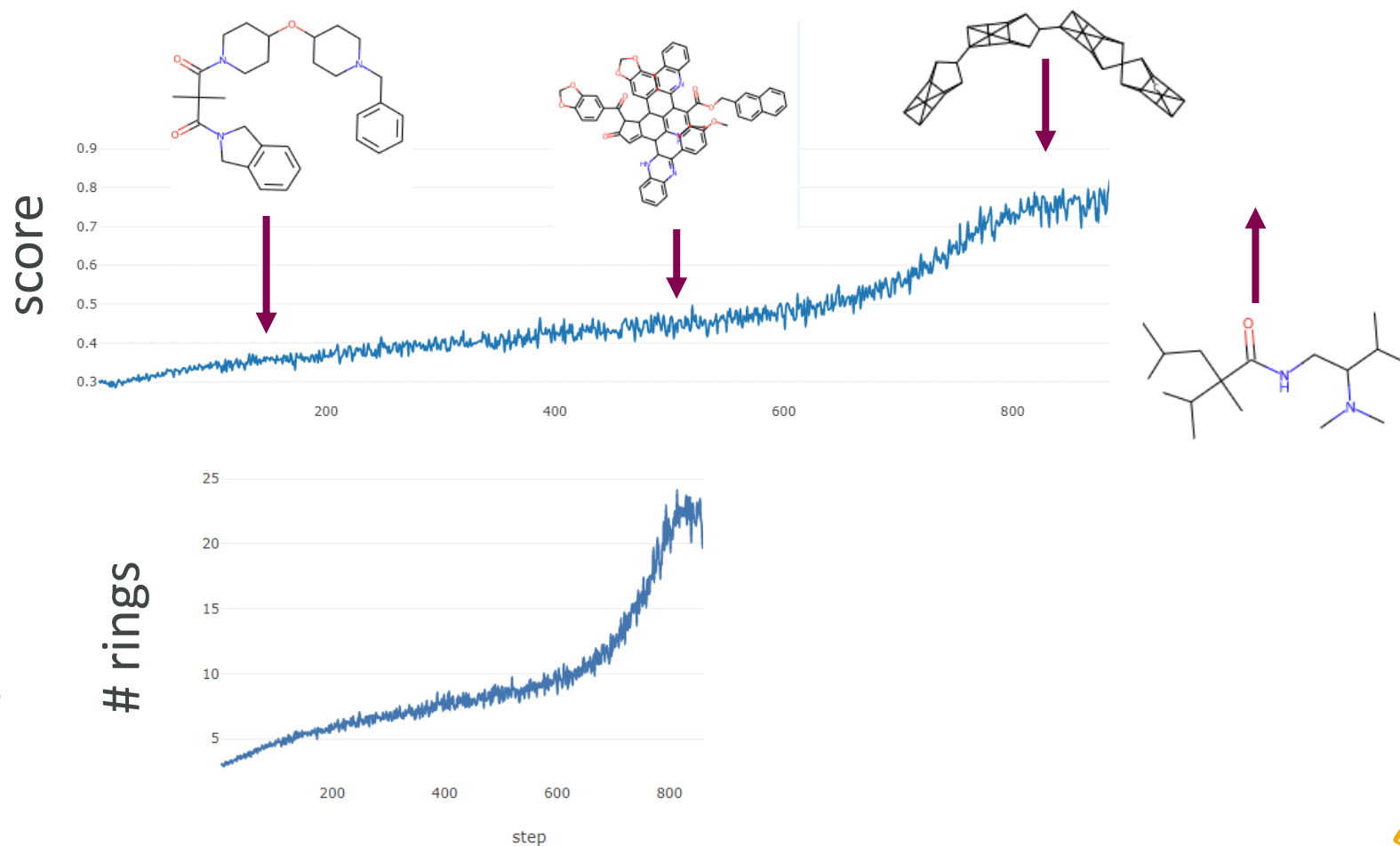


Superpowered molecular optimization engines

Reinvent agents exhibit remarkable plasticity wrt prior and retain adaptability after 100s of epochs. E.g. spend ~800 epochs learning to make as many rings as possible...



Then reverse score transform!



A growing ecosystem of (publis

Open Source:



Bioorg. Med. Chem. 44 (2021) 116308

Contents lists available at ScienceDirect
Journal of Cheminformatics

EL

RESEARCH ARTICLE Open Access

DockStream: a docking wrapper to enhance de novo molecular design

nature machine intelligence

ARTICLES
https://doi.org/10.1038/s42256-022-00494-4

Check for updates

Improving de novo molecular design with curriculum learning

Jeff Guo^{1,3}, Vendy Fialková^{1,3}, Juan Diego Arango¹, Christian Margreitter¹, Jon Paul Janet¹, Kostas Papadopoulos¹, Ola Engkvist^{1,2} and Atanas Patronov¹

He et al. J Cheminform (2021) 13:26
https://doi.org/10.1186/s13321-021-00497-0

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

Molecular optimization by capturing chemist's intuition using deep neural networks

Jiazheng He Werngard
He et al. Journal of Cheminformatics (2022) 14:18
https://doi.org/10.1186/s13321-022-00599-3

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

Transformer-based molecular optimization beyond matched molecular pairs

Jiazheng He^{1*}, Eva Nittinger², Christian Tyrchan², Werngard Czechitzky², Atanas Patronov¹, Esben Jannik Bjerrum¹ and Ola Engkvist^{1,3}

Olivecrona et al. J Cheminform (2017) 9:48
DOI 10.1186/s13321-017-0235-x

Journal of Cheminformatics

RESEARCH ARTICLE

Open Access

Molecular de-novo design through deep reinforcement learning



RESEARCH ARTICLE

Open Access

Exploring the GDB-13 chemical space using deep generative models

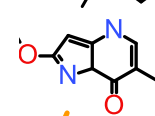
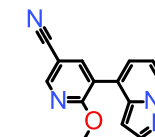


Josep Arús-P and Ola Engkvist

ACS central science

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

Research Article



Traditional REINVENT

REINVENT ecosystem

LibInvent

Link-Invent

Link-INVENT: Generative Linker Design with Reinforcement Learning

Working Paper

Jeff Guo¹ AstraZeneca (United States),
Franziska Knuth¹ Norwegian University of Science and Technology,
Christian Margreitter¹ AstraZeneca (Sweden), Jon Paul Janet¹ AstraZeneca (Sweden),
Kostas Papadopoulos¹ AstraZeneca (Sweden),
Ola Engkvist¹ AstraZeneca (Sweden) & Chalmers University of Technology,
Atanas Patronov¹ AstraZeneca (Sweden)

Gene Recur
Marwin
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Application

REINVENT 2.0: An AI Tool for De Novo Drug Design

Thomas Blaschke, Josep Arús-Pous, Hongming Chen, Christian Margreitter, Christian Tyrchan, Ola Engkvist

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Article

LibINVENT: Reaction-based Generative Scaffold Decoration for *in Silico* Library Design

Vendy Fialková, Jiaxi Zhao, Kostas Papadopoulos, Ola Engkvist, Esben Jannik Bjerrum, Thierry Kogej, and Atanas Patronov*

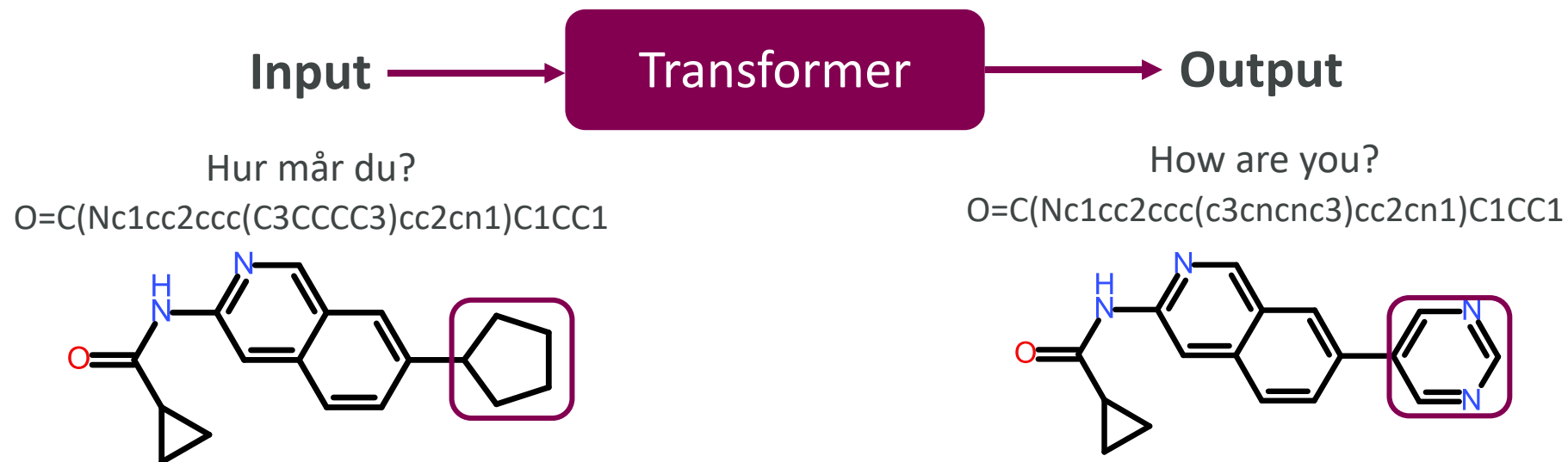


ent

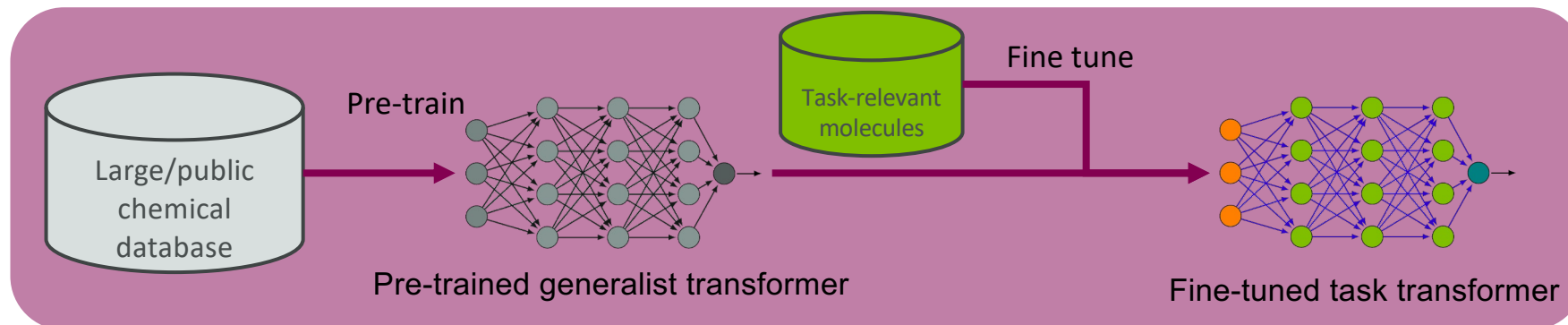


Molecular transformers

Generating similar molecules given a starting molecule can be cast as machine translation in natural language processing

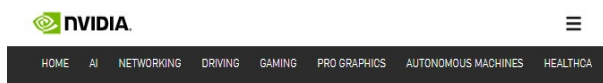


Borrowing from NLP, we can pretrain on large databases and fine-tune as required



Nvidia collaboration: bigger is better

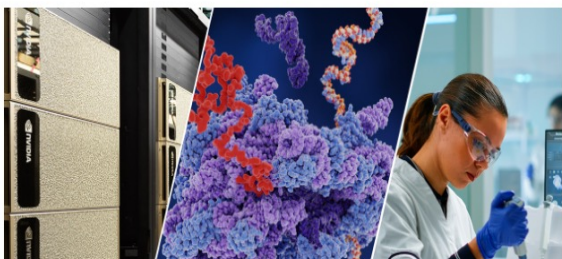
Investigate the potential benefit of pre-training **larger models** on **larger datasets**



Drug Discovery Gets Jolt of AI via NVIDIA Collaborations with AstraZeneca, U of Florida Health

NVIDIA Clara Discovery aims to give researchers tools needed to discover promising pharmaceuticals faster.

April 12, 2021 by KIMBERLY POWELL



“The MegaMolBART drug discovery model being developed by NVIDIA and AstraZeneca ... based on AstraZeneca’s MolBART transformer model ... to enable massively scaled-out training on supercomputing infrastructure.”

AZ transformer model
pre-trained transformer-based
model for chemical language

Nvidia

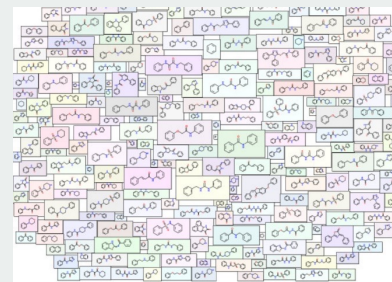
Cambridge-1 supercomputer



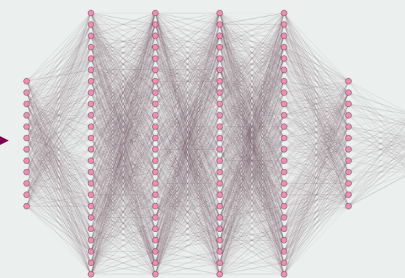
Code for training at large scale

Cutting-edge Science

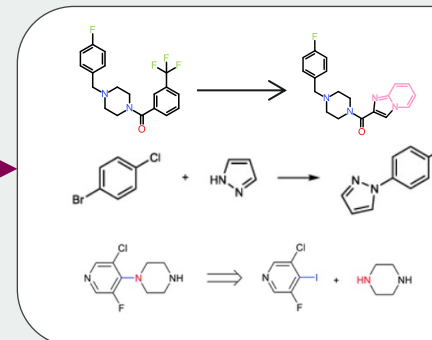
Larger data



Larger pre-trained model



fine tune



Molecular optimization

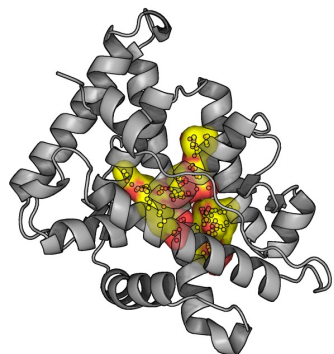
Forward synthesis prediction

Retrosynthesis prediction

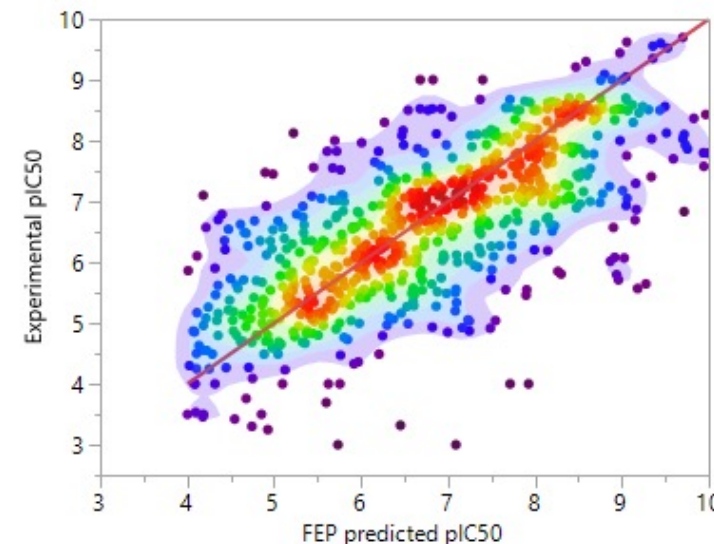


Combining generative models & state-of-the-art simulation

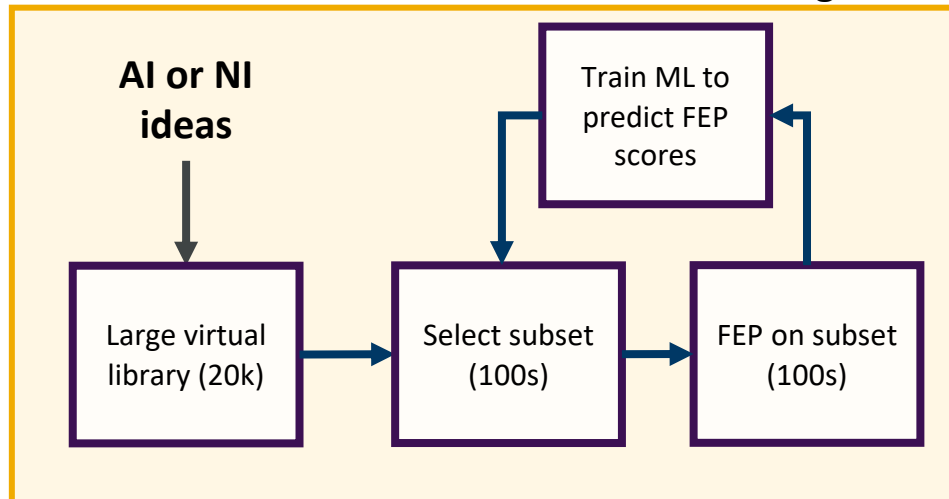
Proteins are dynamic.
This is important for
ligand binding.



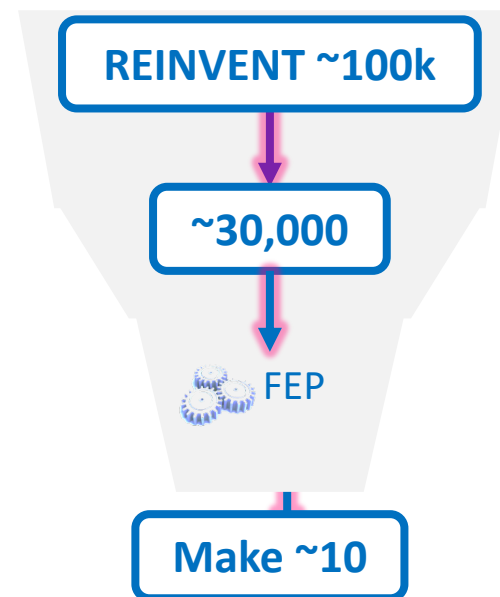
Free energy perturbation (FEP) is an advanced, computationally expensive but accurate way to predict potency of new compounds using molecular dynamics. Validated over 16 targets, 15k compounds at AZ over 3+ years.



Active learning FEP



We combine with
FEP with active
learning to
accelerate to
larger scales

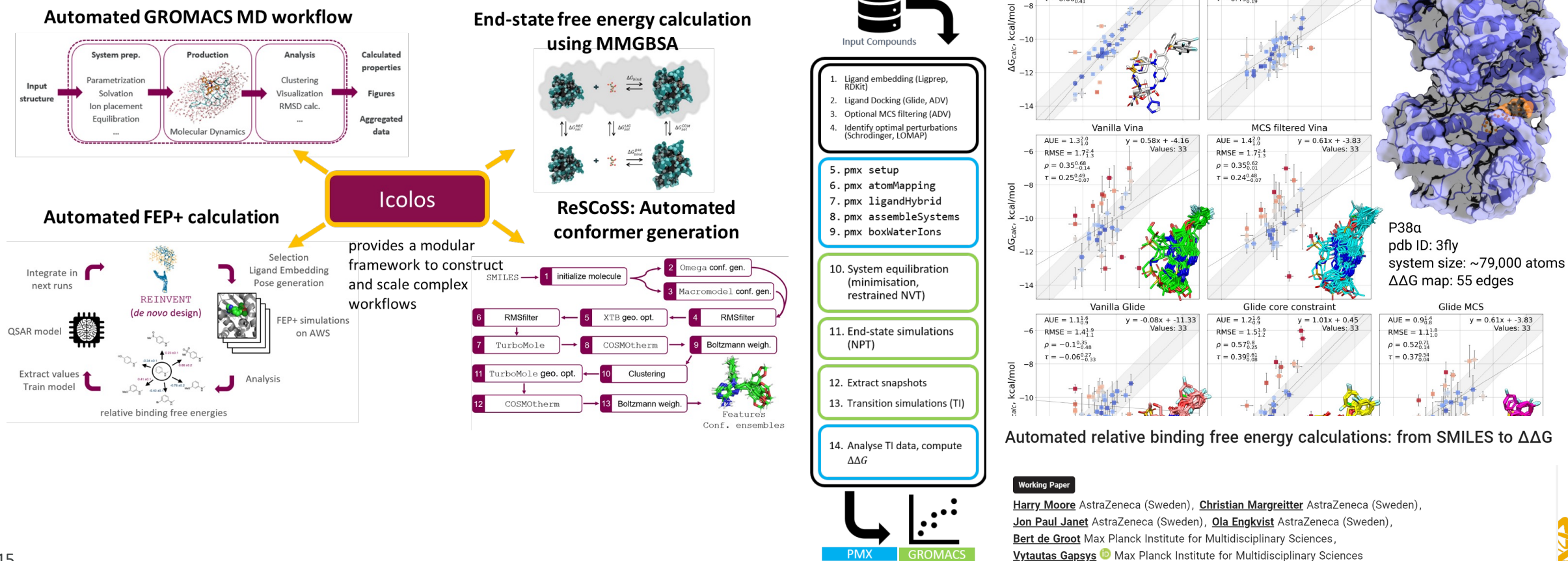


100s of GPU-
hours per
compound
made!



Icolos and automatic molecular simulations

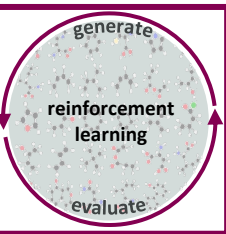
These type of compute-heavy simulations are playing an ever-increasing role in drug design. We have implemented an open-source workflow manager, **Icolos**, to manage complex multi-step simulations and connect them to generative models.



Computer-aided synthesis planning

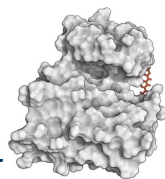
REINVENT:

de novo molecule design
generate right ideas



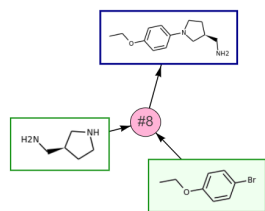
Active-learning FEP+:

dynamic affinity prediction
prioritize with physics & ML



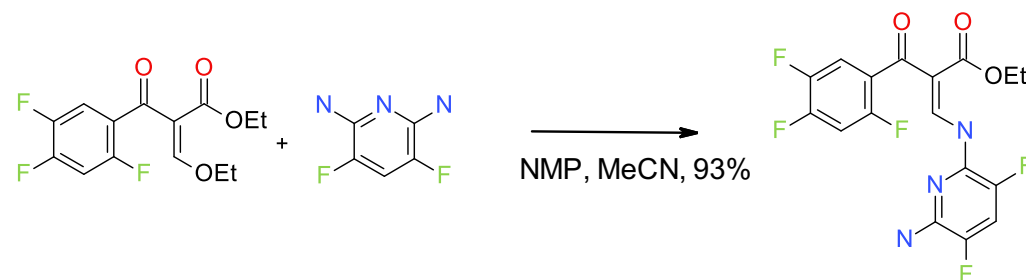
CAZP/AlZynth:

AI synthesis planning
realize ideas quickly

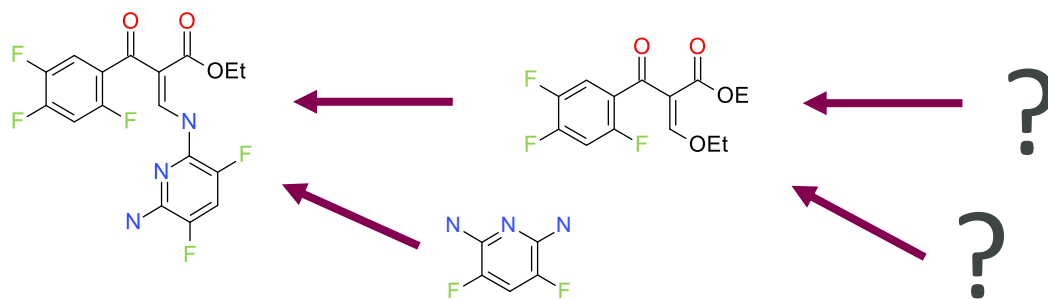


Can we teach computers to produce recipes for molecules of interest? (retrosynthesis)

Lots of data available like this:



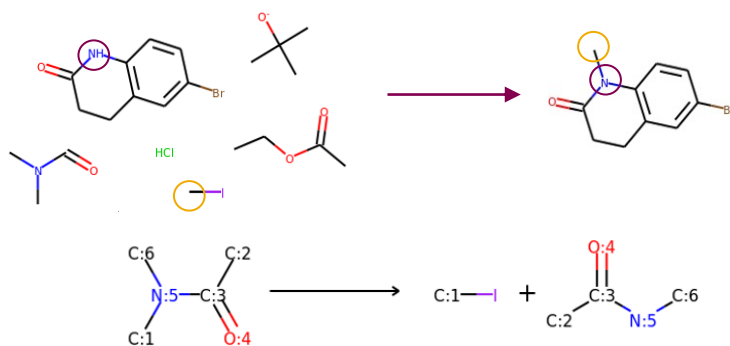
We want to find feasible reactions to make new ideas (amongst other things):



Computer-aided synthesis planning

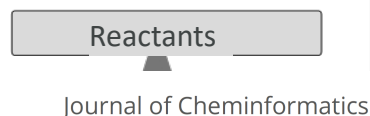
Template-based model

Extract reaction rules from known reactions



Template-free model

Train a transformer model to learn rules automatically



Genheden et al. *J Cheminform* (2020) 12:70
<https://doi.org/10.1186/s13321-020-00472-1>

SOFTWARE

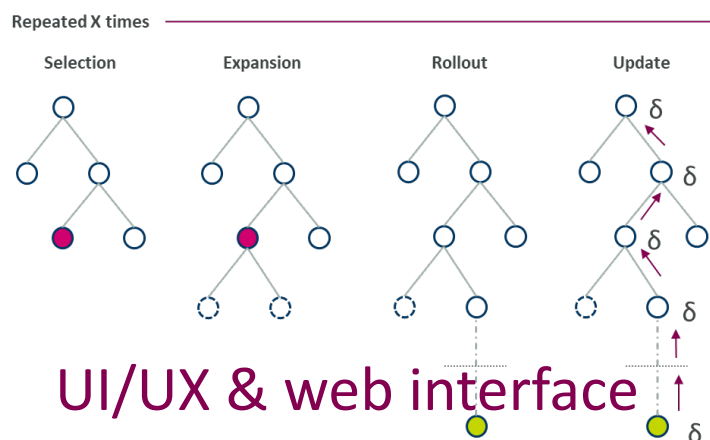
Open Access

AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning

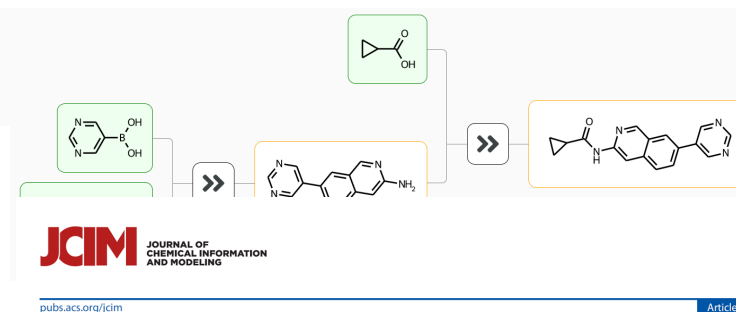
Samuel Genheden^{1*}, Amol Thakkar^{1,2}, Veronika Chadimová¹, Jean-Louis Reymond², Ola Engkvist¹ and Esben Bjerrum^{1*}

Multi-step retrosynthesis

Need to deploy a tree search of one-step reactions



UI/UX & web interface



JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Clustering of Synthetic Routes Using Tree Edit Distance

Samuel Genheden^{*}, Ola Engkvist, and Esben Bjerrum

Cite This: *J. Chem. Inf. Model.* 2021, 61, 3899–3907

[Read Online](#)

Open Source: 

<https://github.com/MolecularAI>

Preloaded with public data



Drug Discovery Today: Technologies

Editors-in-Chief

Kelvin Lam – Simplex Pharma Advisors, Inc., Boston, MA, USA
Henk Timmerman – Vrije Universiteit, The Netherlands

AI-assisted synthesis prediction

MACHINE LEARNING
Science and Technology

PAPER

Fast prediction of distances between synthetic routes with deep learning

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

[View Article Online](#)
[View Journal](#) | [View Issue](#)

[Check for updates](#)

Cite this: *Chem. Sci.*, 2021, 12, 3339

All publication charges for this article have been paid for by the Royal Society of Chemistry

Retrosynthetic accessibility score (RAScore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning†

Amol Thakkar^{*,†}, Veronika Chadimová^{*,†}, Esben Jannik Bjerrum^{*,†}
† This is an open access article published under an ACS authorchoice license, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Journal of Medicinal Chemistry

pubs.acs.org/jmc

"Ring Breaker": Neural Network Driven Synthesis Prediction of the Ring System Chemical Space

Amol Thakkar^{*}, Nidhal Selmi, Jean-Louis Reymond, Ola Engkvist, and Esben Jannik Bjerrum^{*}

Cite This: *J. Med. Chem.* 2020, 63, 8791–8808

[Read Online](#)

Conclusions and outlook

Where we are

AI & computational methods, along with powerful new hardware (GPUS), have emerged as practical new tools to accelerate early-stage discovery.

Where we going

Future drug candidates will require even more computation, both data-driven and physics based, before ever being tested.

Limitations

Many limitations remain in what can predicted & the wet lab will not be replaced. AI design is complementary, and humans remain in the loop





Thank you, any
questions?

Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK, T: +44(0)203 749 5000, www.astrazeneca.com

