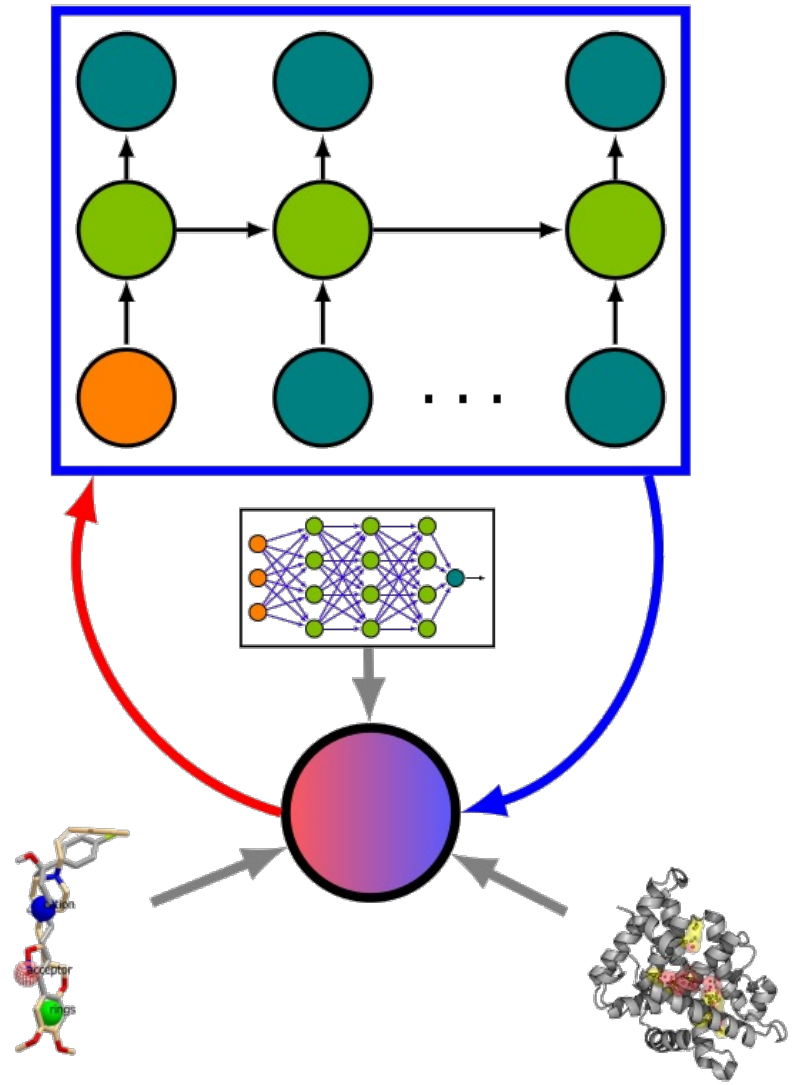
AstraZeneca

Accelerating drug design with AI & simulation

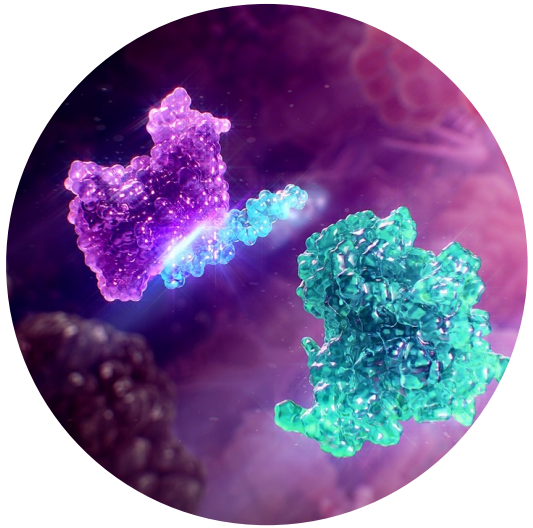
Jon Paul Janet

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

NOTE: All growth rates at Constant Exchange Rates
Source: 2021 Annual Report

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

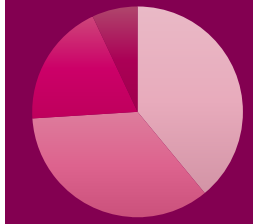
Global strength, with balanced presence across regions

Global reach and presence

~2,800 employees of **70+ nationalities,** including 600 PhD researchers and 30 professors.

- 3 global R&D centres
- 8 other R&D centres and offices
- 28 manufacturing sites in 16 countries

Employees by reporting region



Emerging Markets	**39%**
Europe	**35%**
US	**19%**
Established Rest of World	**7%**

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

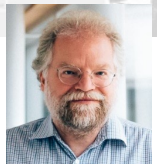
Machine learning techniques are poised to impact pharmaceutical development industry from beginning to end

Early discovery

Target identification | Drug Design | Imaging/analysis | Scale-up/process | Clinical

Late/clinical development

Molecular AI



Ola Engkvist

- Synthesis prediction
- de novo molecular design
- physics-informed molecular screening

The Drug Discovery Process

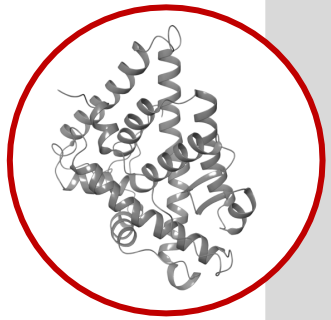
At the heart of the drug design process is the Design, Make, Test and Analyze (DMTA) cycle, which is a core concept for iterative, hypothesis driven design.

How can we accelerate this process?

New methods for molecular ideation and screening

Synthesis planning, condition prediction

Drug target

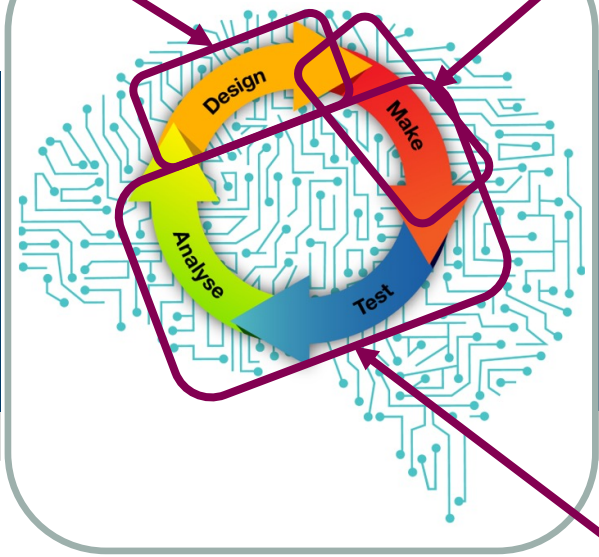


e.g. protein that is implicated in disease.

Chemical starting point (hit)

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

DMTA cycle



Candidate drug (CD)

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

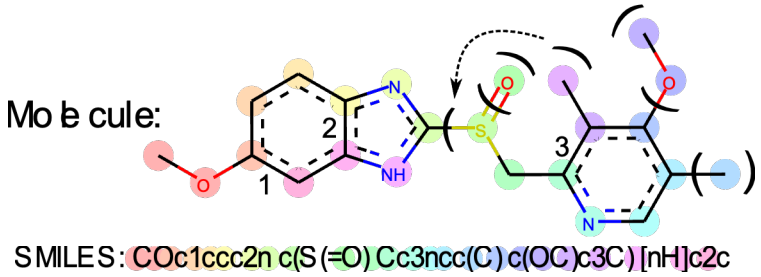
3+ years

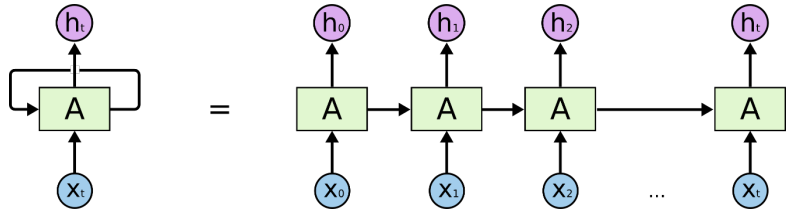
Iterative process – multiple cycles needed

Automation, robotics, microfluidics (not discussed)

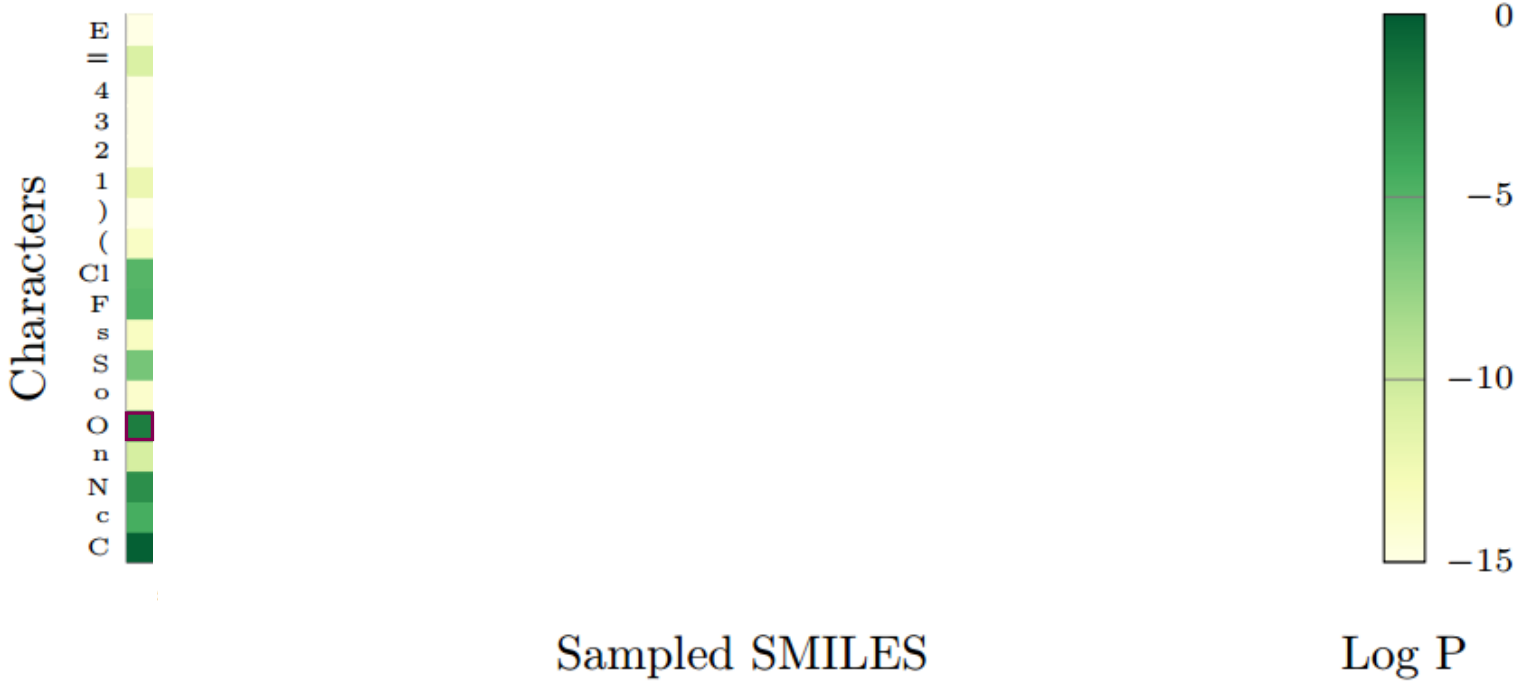
Chemical language models are central to much of our work

Molecules can be described in the language of SMILES…



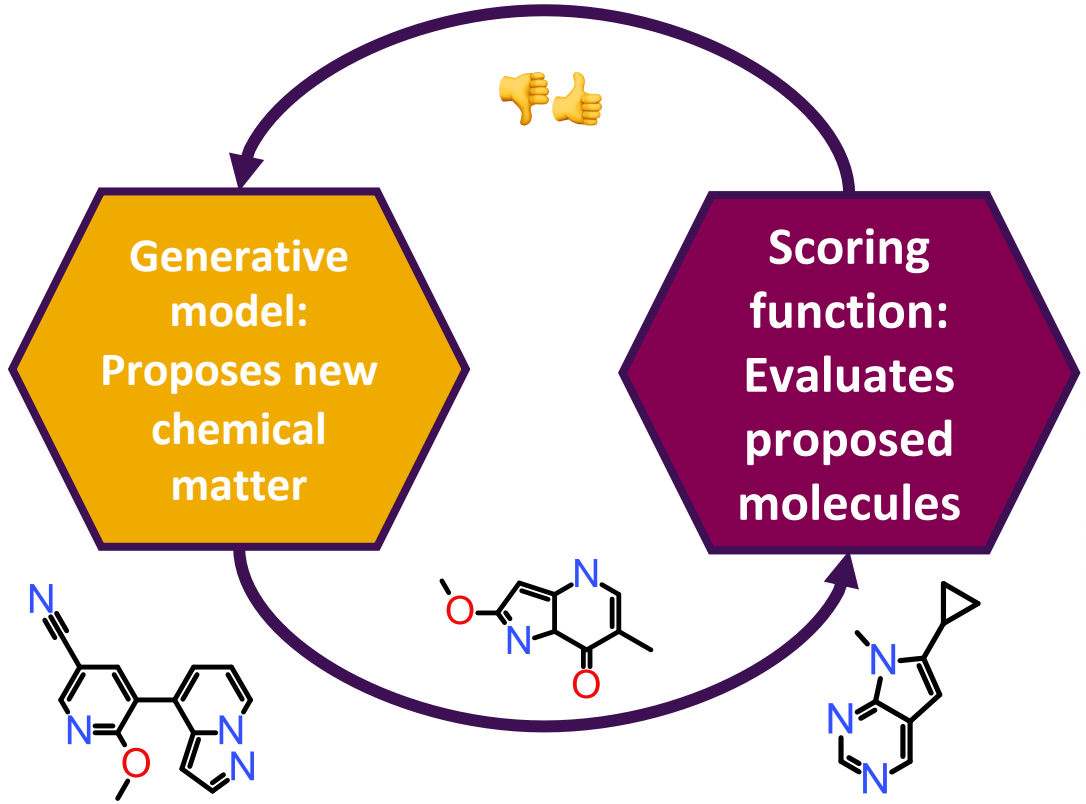


and fed into (recurrent) neural networks!

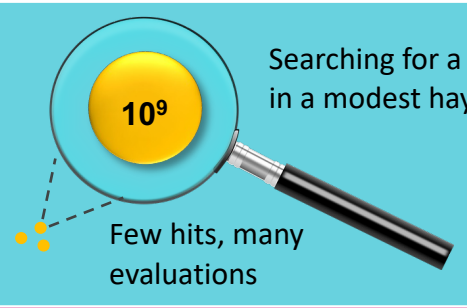


REINVENT – designing new molecules with AI

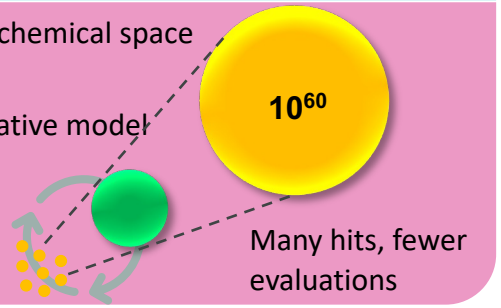
REINVENT is the in-house developed de novo molecular design tool, using generative reinforcement learning to solve *in silico* molecular design tasks

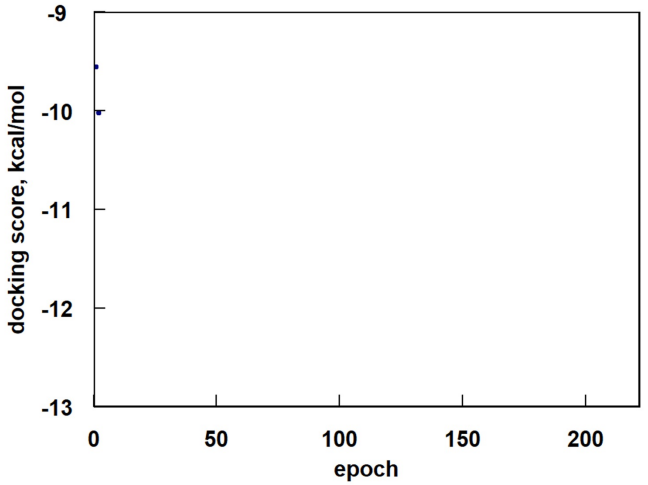


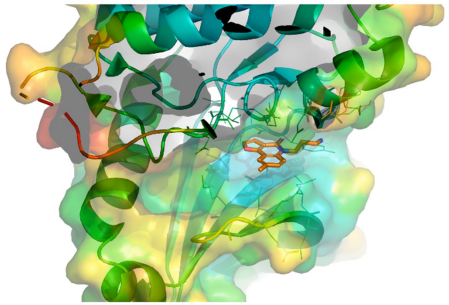
Traditional approaches rely on searching a large database for a small number of suitable hits



Generative models encode practically unlimited chemical space probabilistically

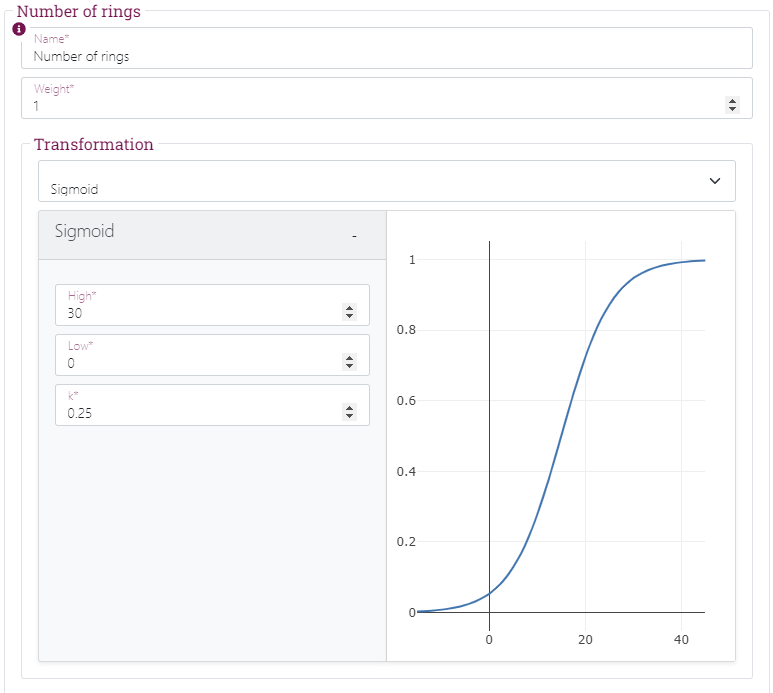




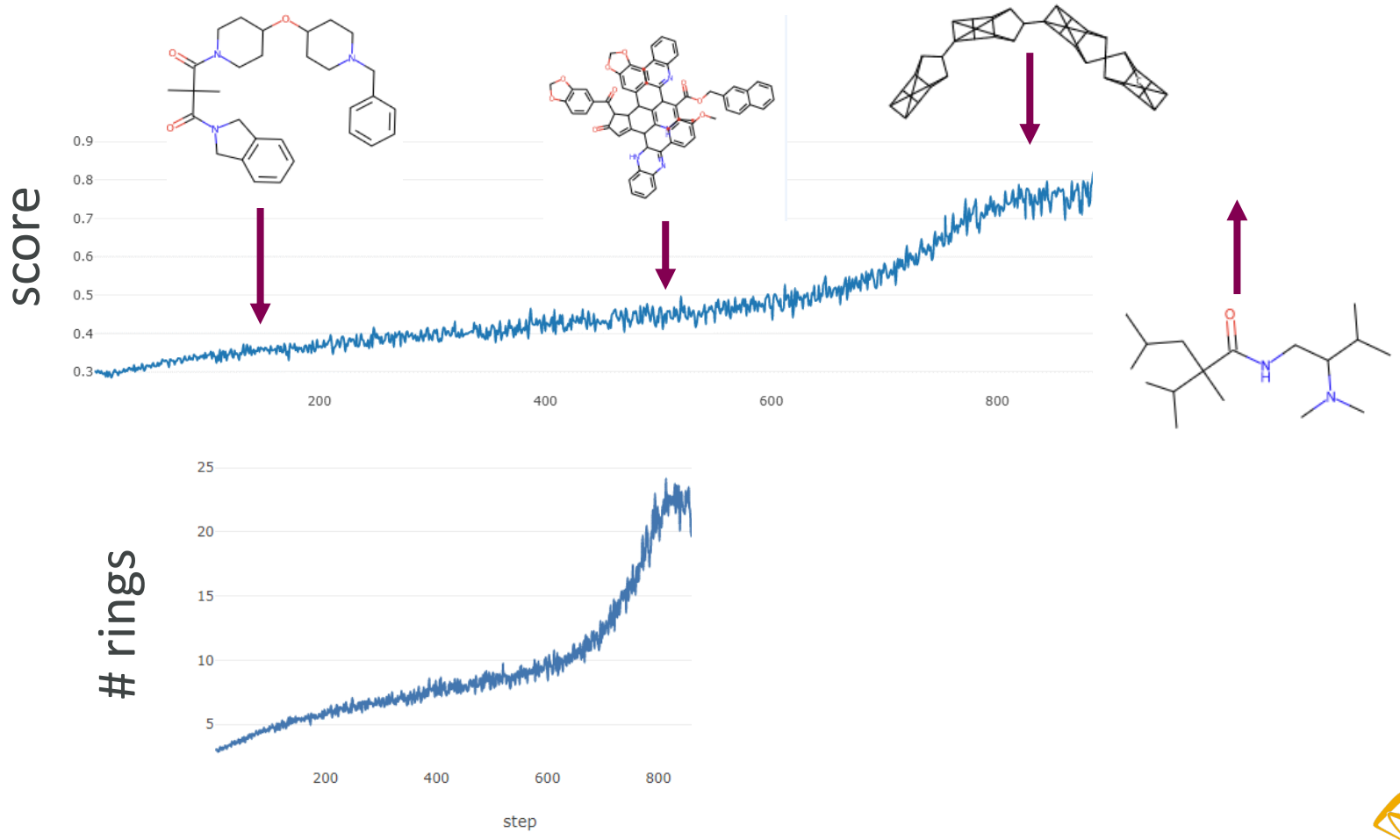


Superpowered molecular optimization engines

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…

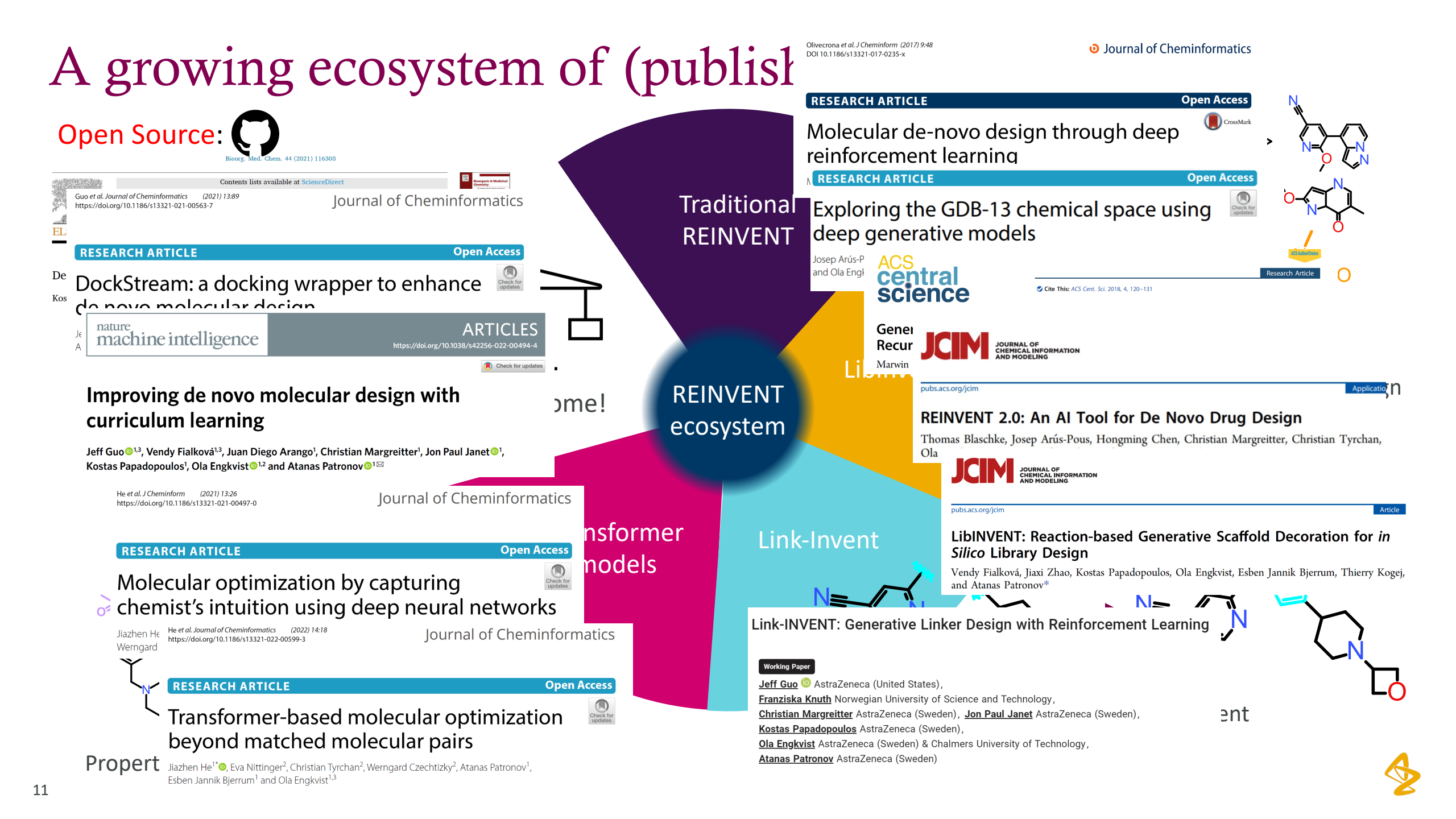


Then reverse score transform!



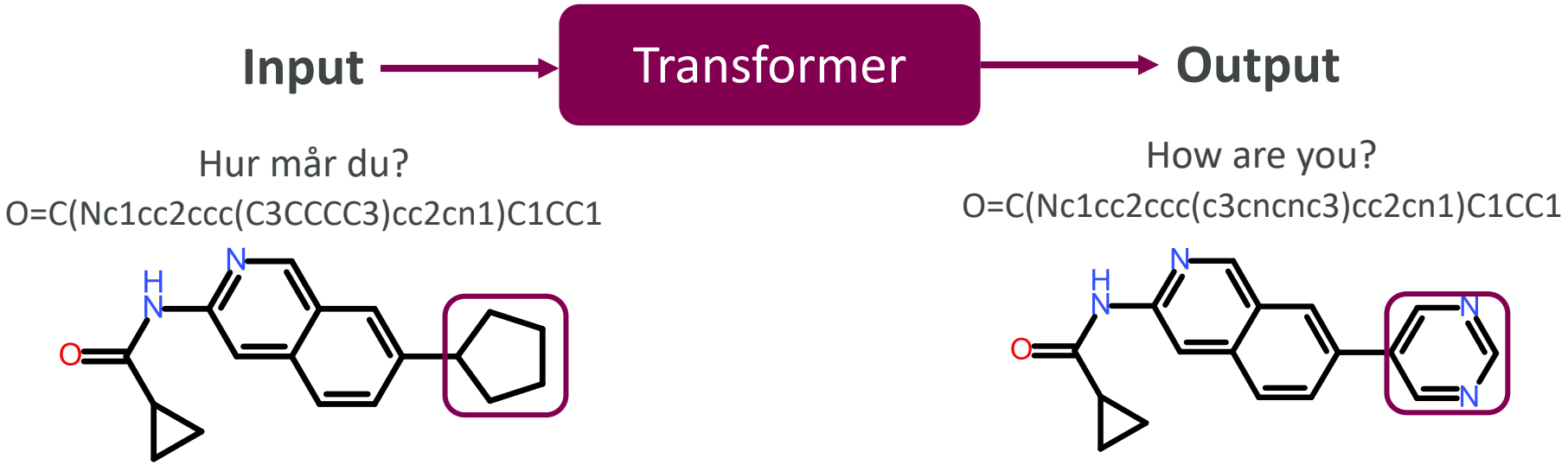
A growing ecosystem of (publish

Open Source:

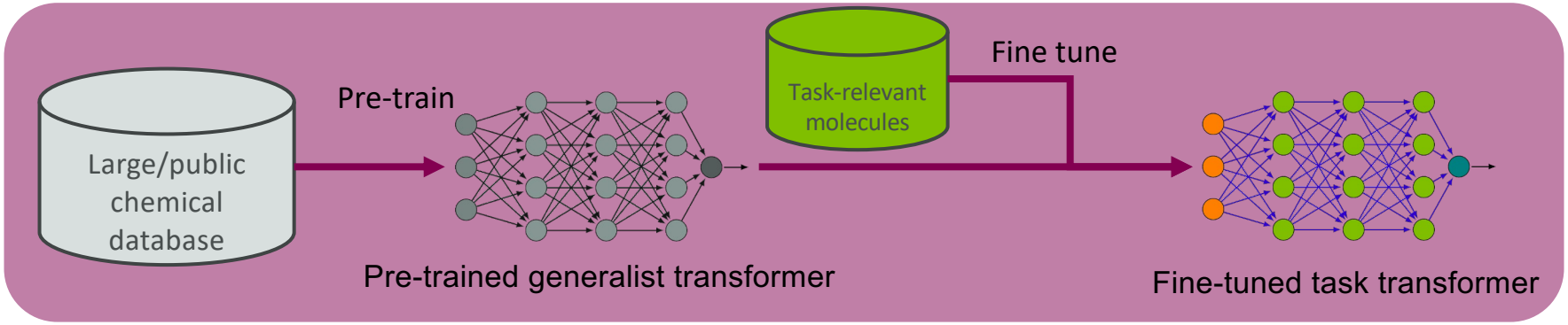


Molecular transformers

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**

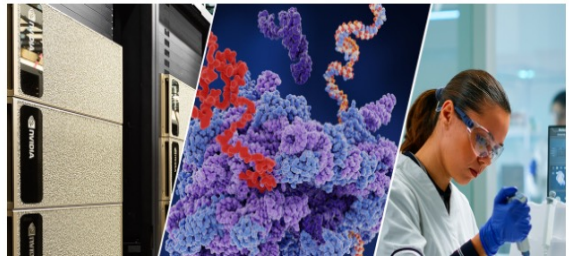
NVIDIA

HOME AI NETWORKING DRIVING GAMING PRO GRAPHICS AUTONOMOUS MACHINES HEALTHCA

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

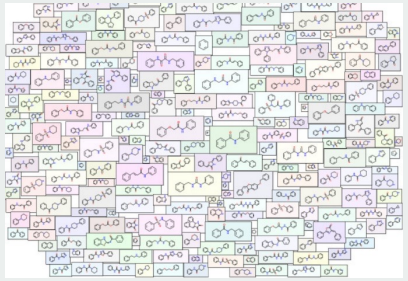
Nvidia

Cambridge-1 supercomputer

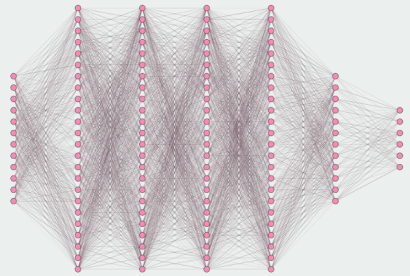
Code for training at large scale

Cutting-edge Science

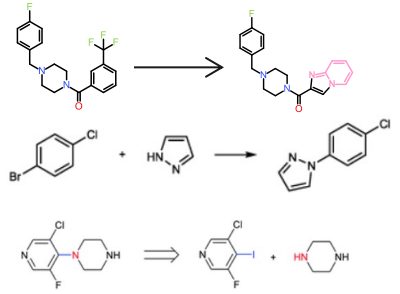
Larger data



Larger pre-trained model



fine tune



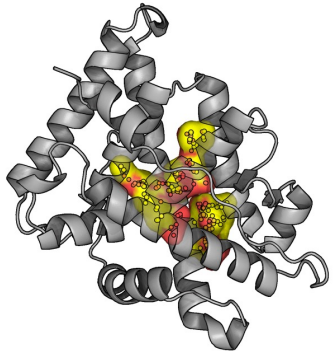
Molecular optimization

Forward synthesis prediction

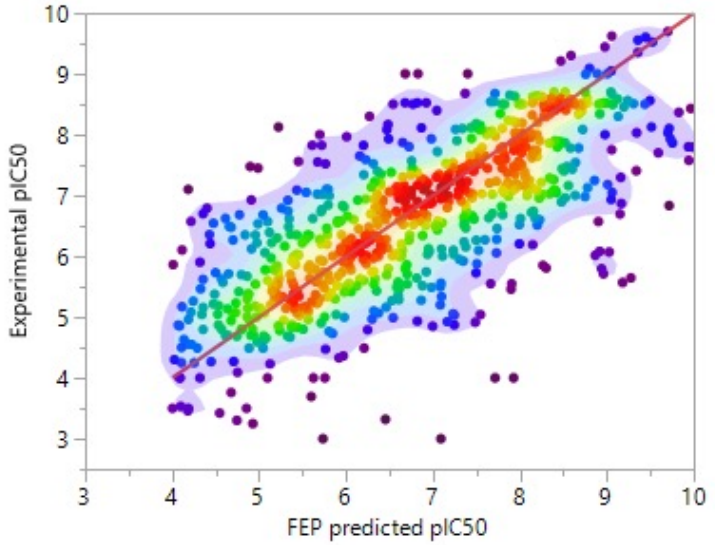
Retrosynthesis prediction

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

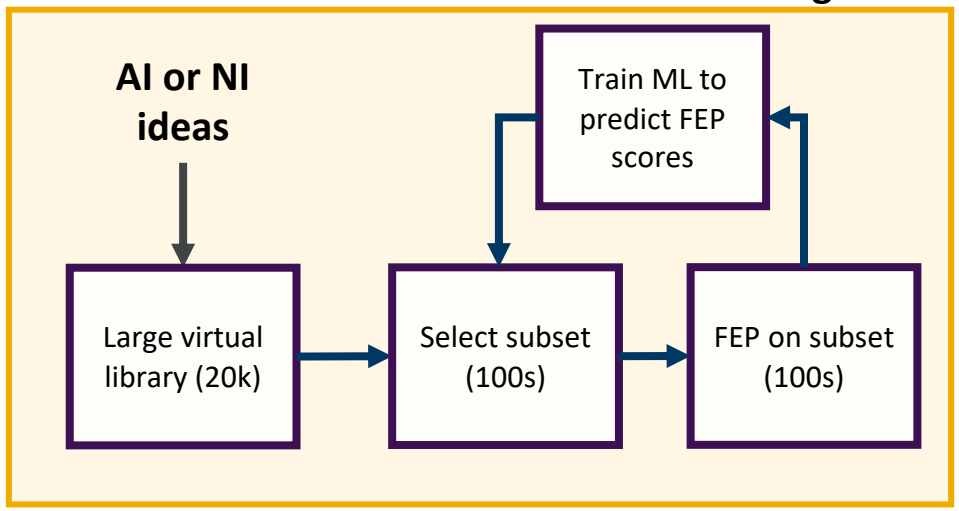
Proteins are dynamic. This is important for ligand binding.



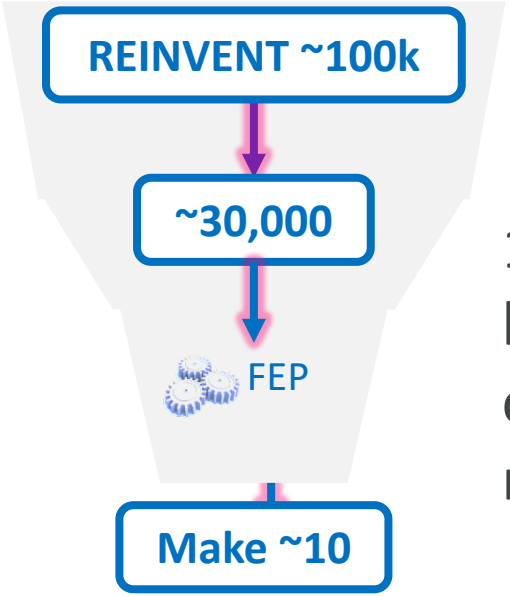
Free energy perturbation (FEP) is an advanced, computationally expensive but accurate way to predict potency of new compounds using molecular dynamics . Validated over 16 targets, 15k compounds at AZ over 3+ years.



Active learning FEP



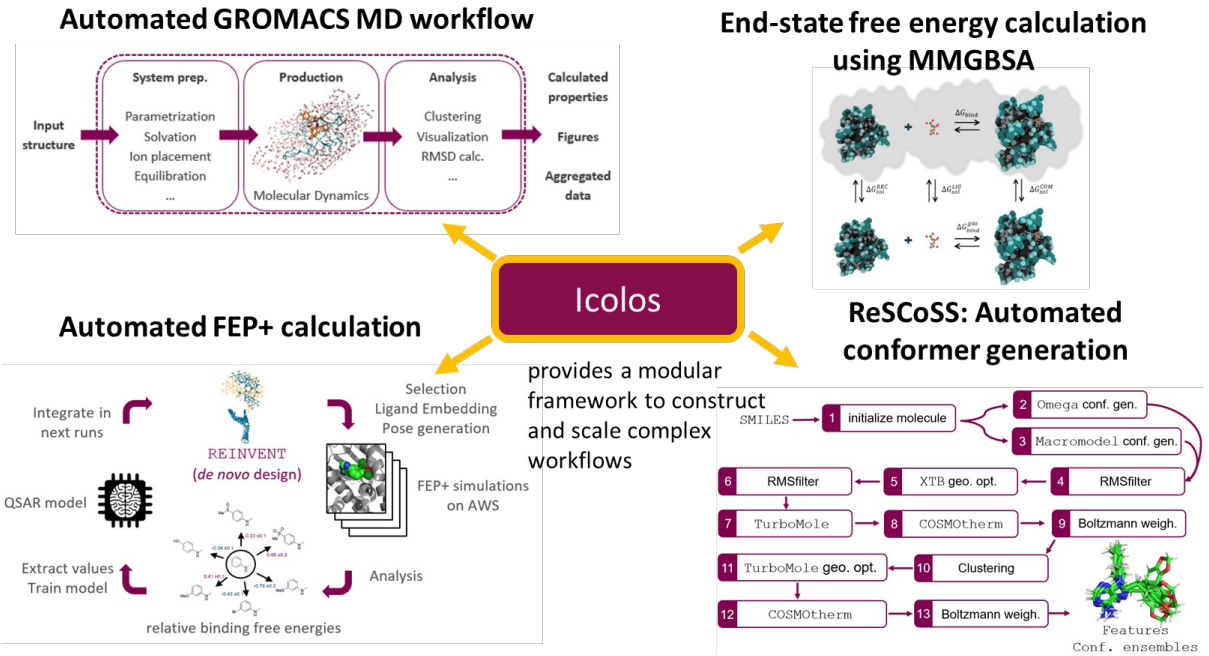
We combine with FEP with active learning to accelerate to larger scales

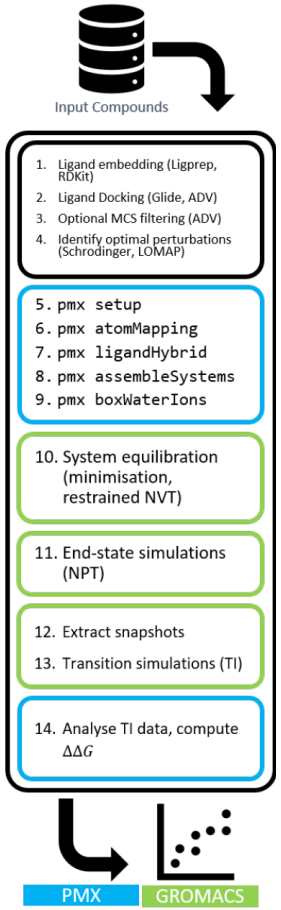


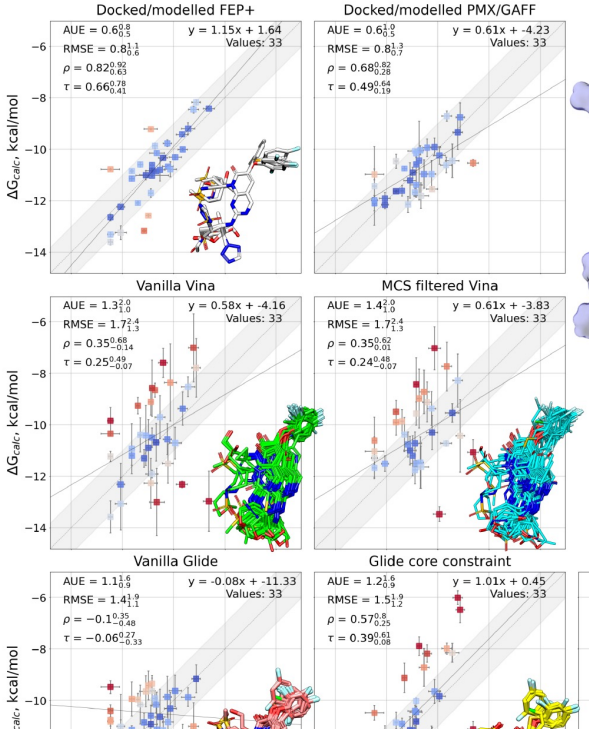
100s of GPU-hours per compound made!

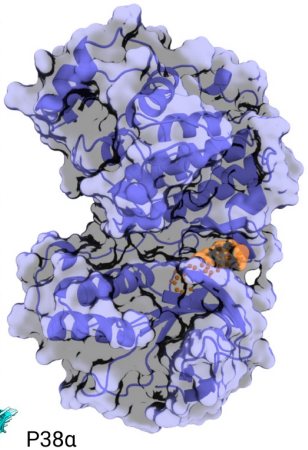
Icolos and automatic molecular simulations

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 mange complex multi-step simulations and connect them to generative models.









P38α
pdb ID: 3fly
system size: ~79,000 atoms
ΔΔG map: 55 edges

Automated relative binding free energy calculations: from SMILES to ΔΔG

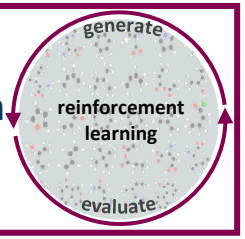
Working Paper

Harry Moore AstraZeneca (Sweden), Christian Margreitter AstraZeneca (Sweden), Jon Paul Janet AstraZeneca (Sweden), Ola Engkvist AstraZeneca (Sweden), Bert de Groot Max Planck Institute for Multidisciplinary Sciences, Vytautas Gapsys Max Planck Institute for Multidisciplinary Sciences

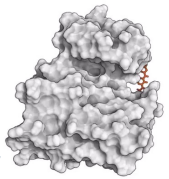
Computer-aided synthesis planning

AI-assisted drug discovery pipeline

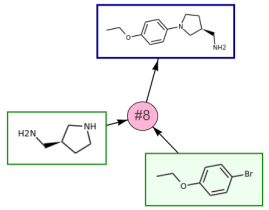
REINVENT:
de novo molecule design
generate right ideas



Active-learning FEP+:
dynamic affinity prediction
prioritize with physics & ML

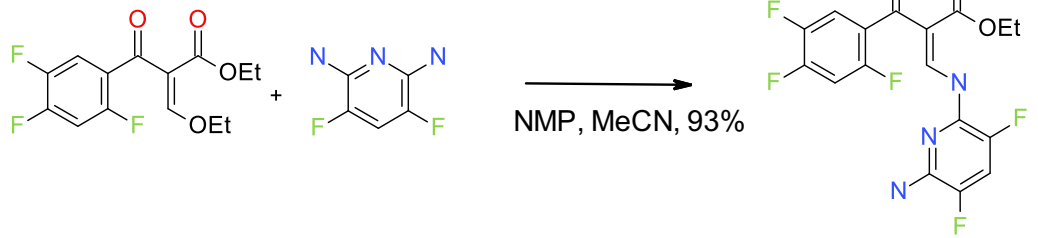


CAZP/AIZynth:
AI synthesis planning
realize ideas quickly

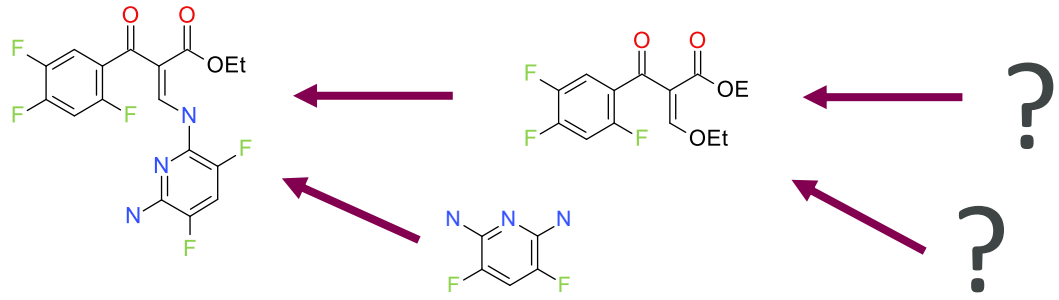


Can we teach computers to produce recipes for molecules of interest? (retrosynthesis)

Lots of data available like this:



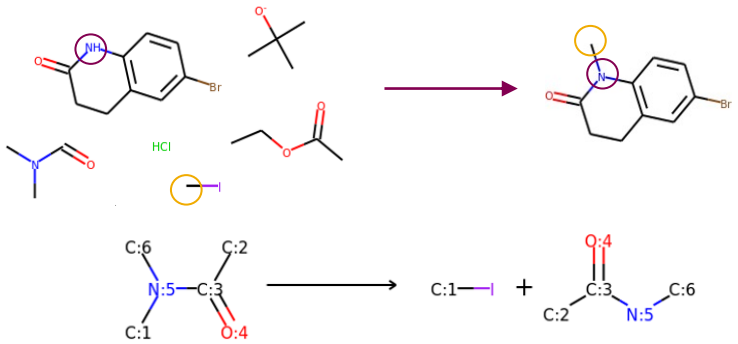
We want to find feasible reactions to make new ideas (amongst other things):



Computer-aided synthesis planning

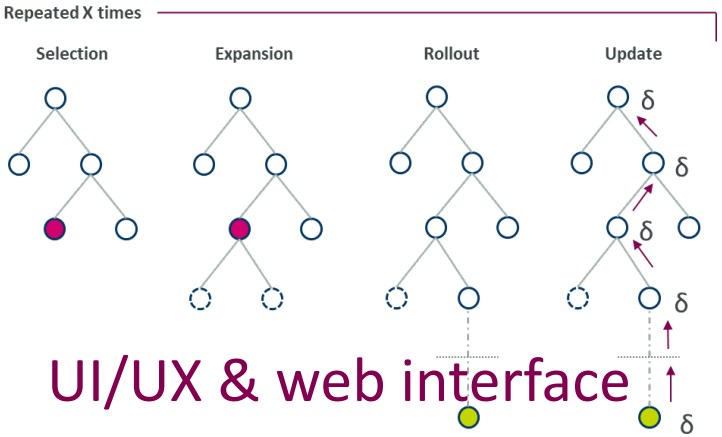
Template-based model

Extract reaction rules from known reactions



Multi-step retrosynthesis

Need to deploy a tree search of one-step reactions



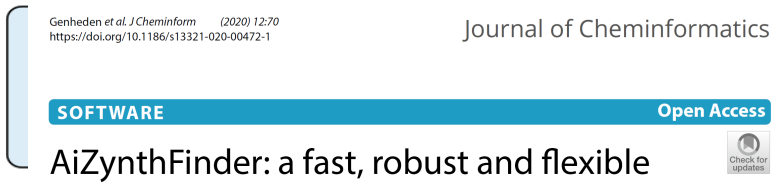
Open Source:

https://github.com/MolecularAI

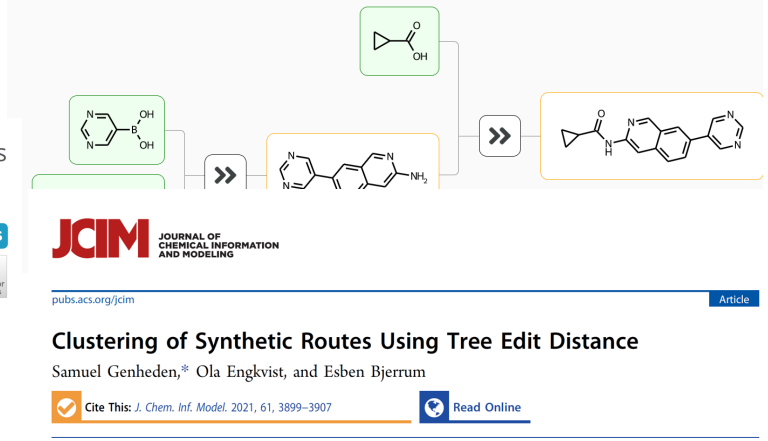
Preloaded with public data

Template-free model

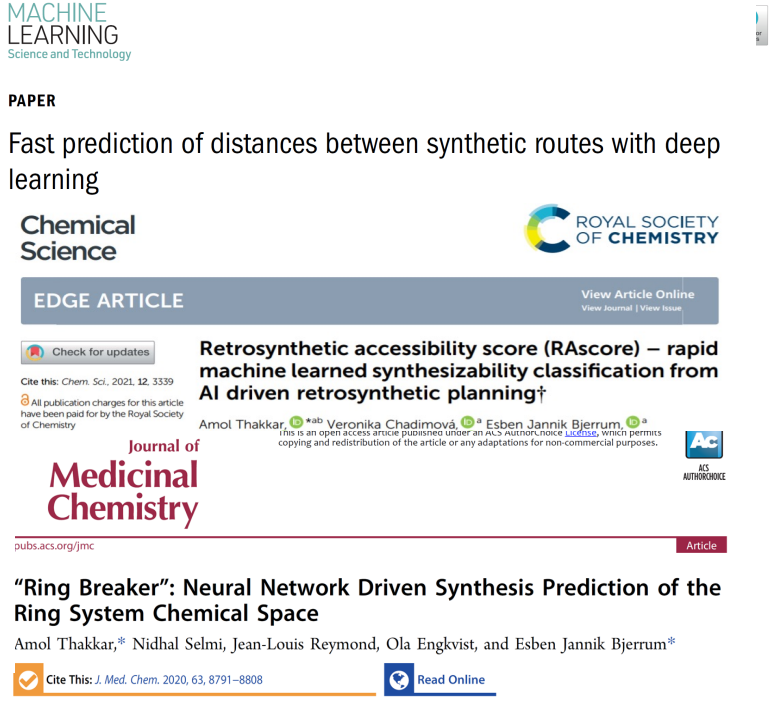
Train a transformer model to learn rules automatically



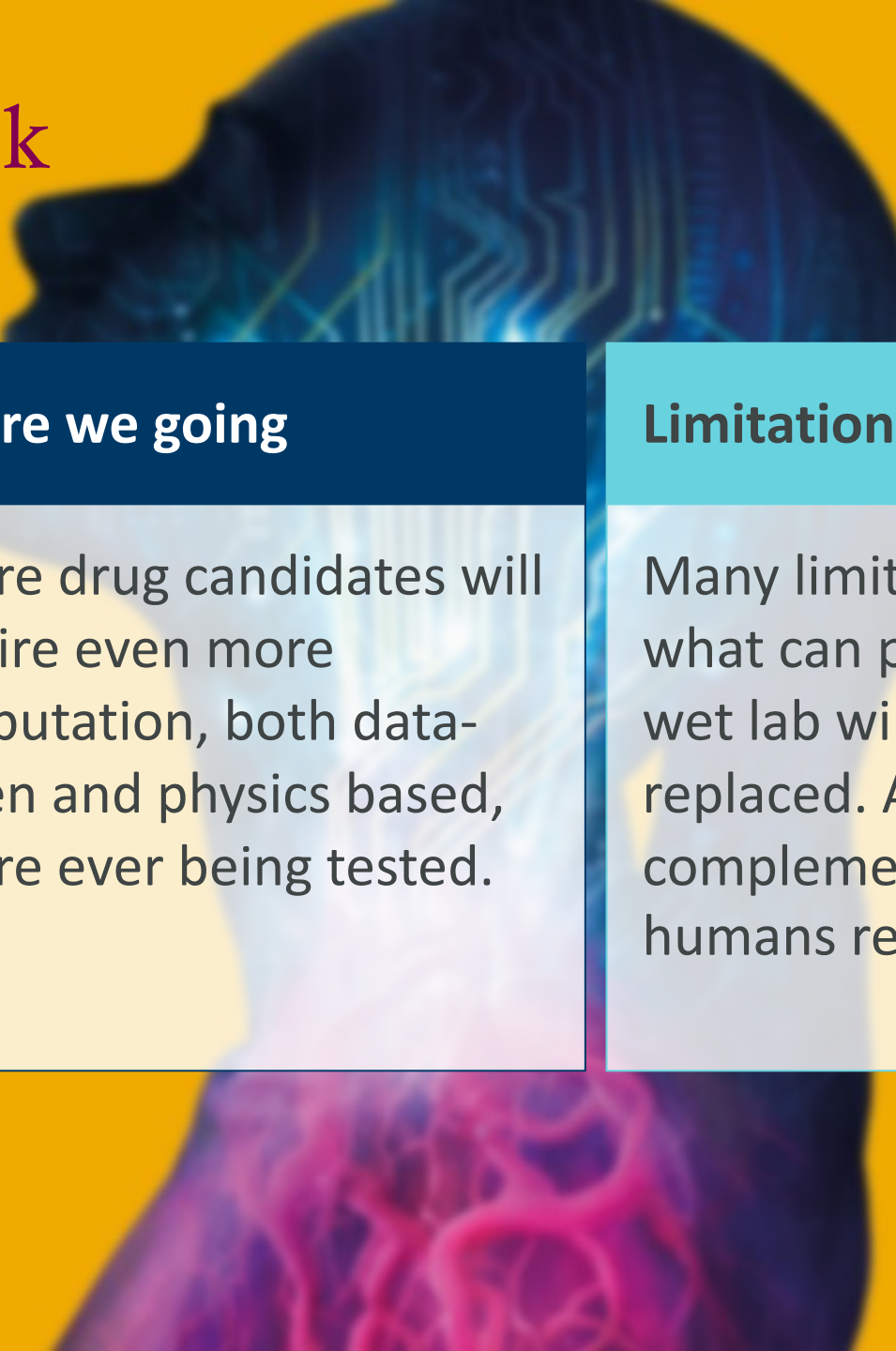
UI/UX & web interface



AI-assisted synthesis prediction



Conclusions and outlook



Where we are	Where we going	Limitations
AI & computational methods, along with powerful new hardware (GPUS), have emerged as practical new tools to accelerate early-stage discovery.	Future drug candidates will require even more computation, both data-driven and physics based, before ever being tested.	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?