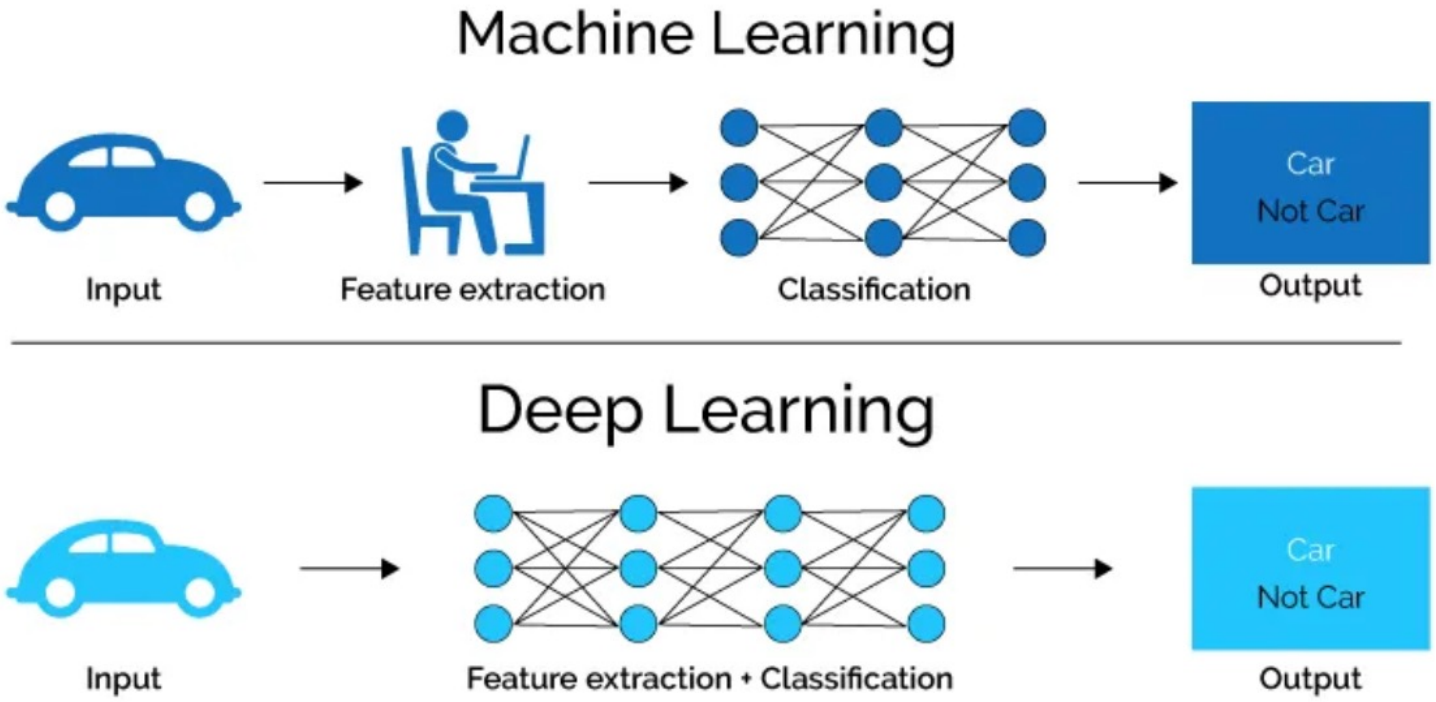
AI in drug discovery

DRIES VAN ROMPAEY

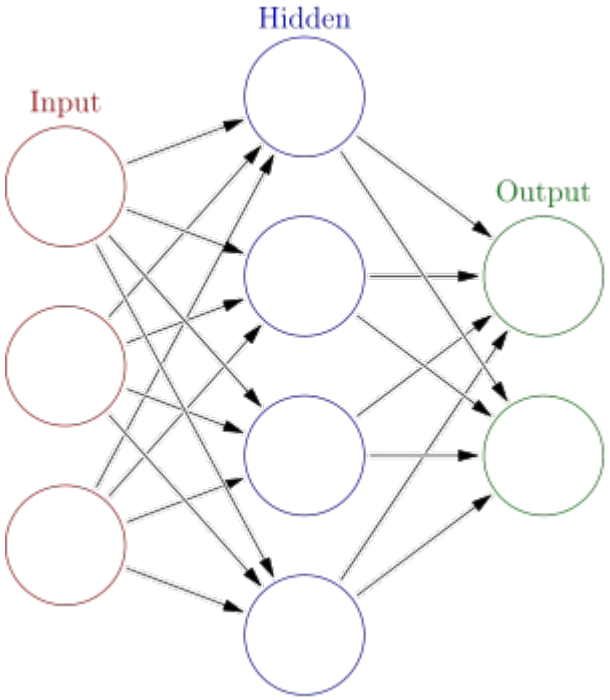
Wie ben ik?

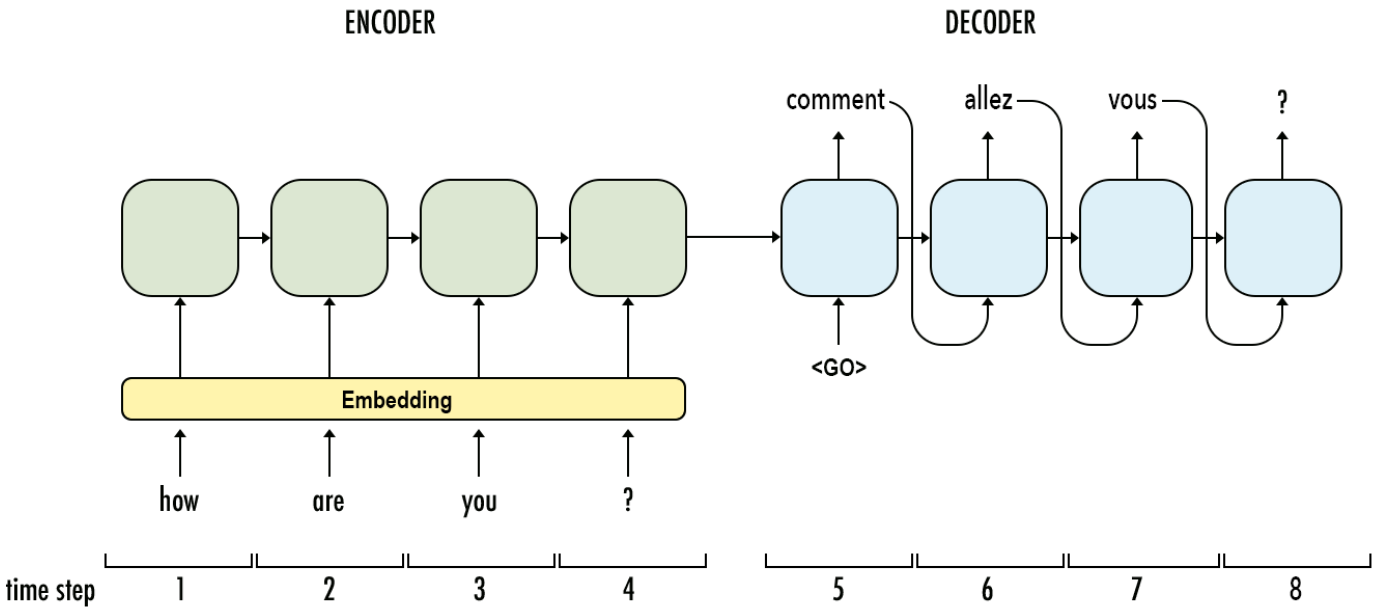
- Master Farmaceutische Wetenschappen
- Doctoraat in medicinale/computationele chemie onder prof. De Winter
- Bij Janssen sinds 2018
  - Ondersteuning projectteams
  - Ontwerp van small molecules
  - Bouwen van machine learning modellen

Machine learning vs deep learning?



Neurale netwerken





ML voor small molecule design

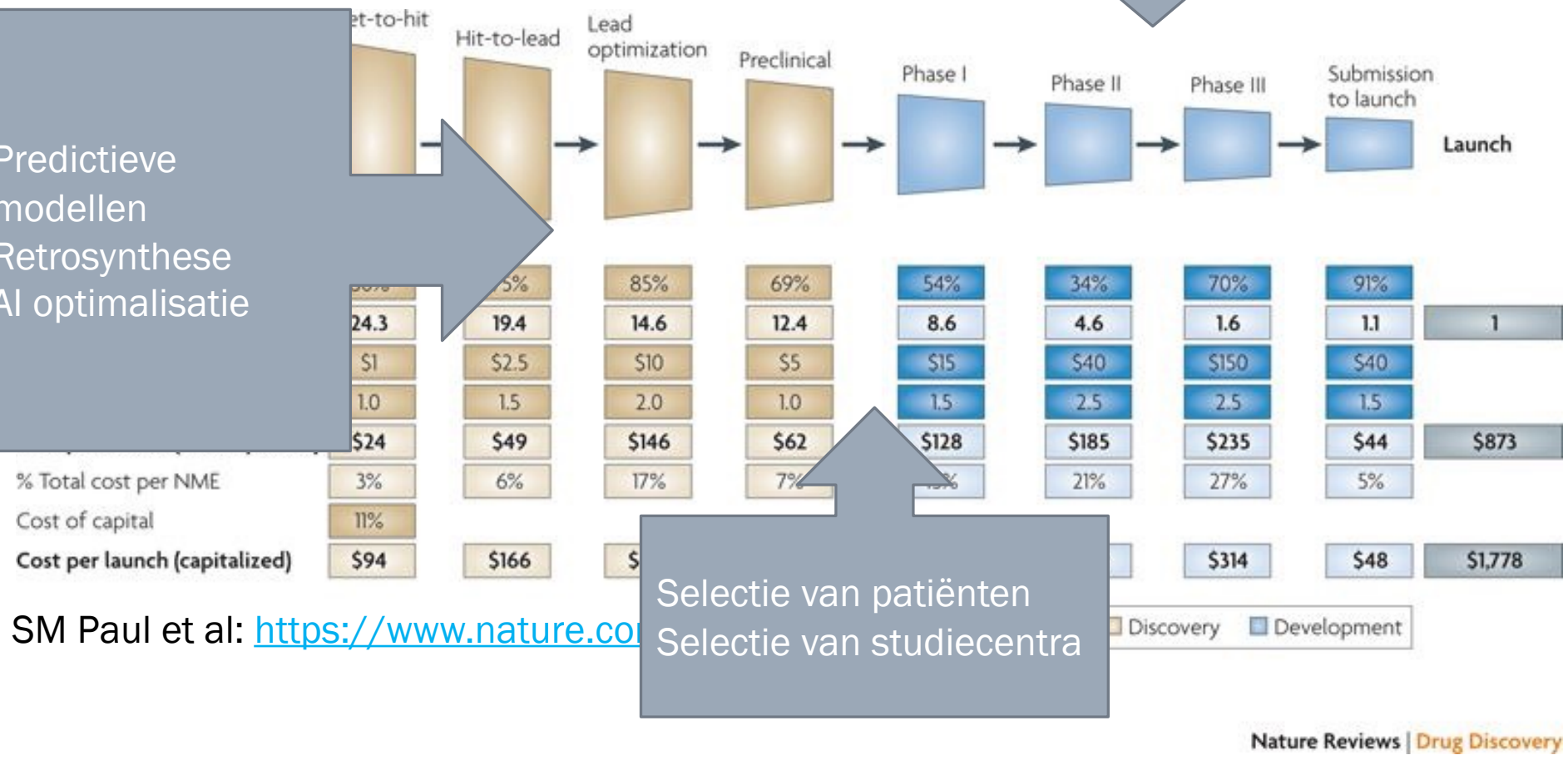
KAN AI MOLECULEN (HELPEN) ONTWERPEN?

ML-based screening
Structure-based generative

AI-gebaseerde endpoints
Digital twin

Drug discovery pipeline

Predictieve modellen
Retrosynthese
AI optimalisatie



Selectie van patiënten
Selectie van studiecentra

SM Paul et al: https://www.nature.co

Nature Reviews | Drug Discovery

Eigenschappen van een drug

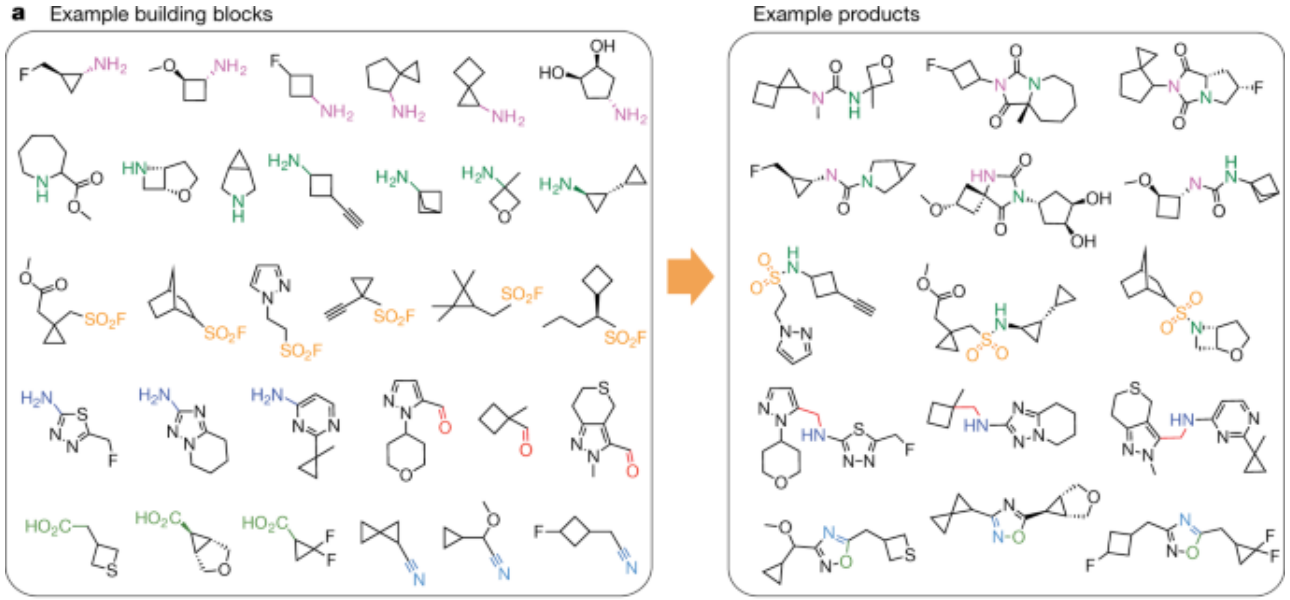
- Potente *on-target* binding
- Weinig *off-target* binding
- ADMET:
  - Permeabiliteit
  - Metabole stabiliteit
  - Half-leven compatibel met toediening
  - Geen toxiciteit: hERG, andere ionkanalen, mitochondriale toxiciteit, ….

Predictieve modellen voor ontwerp

Onmiddelijke feedback voor design ideeën

Selectie van compounds in virtuele libraries

Detectie van risico's (glu/gal, hERG, ...)



Waarom falen drugs in de kliniek?

Efficaciteit – therapeutische hypothese?

Veiligheid – onverwachte toxiciteitsbevindingen (IADR!)

Commerciele redenen – concurrentie is sneller/beter

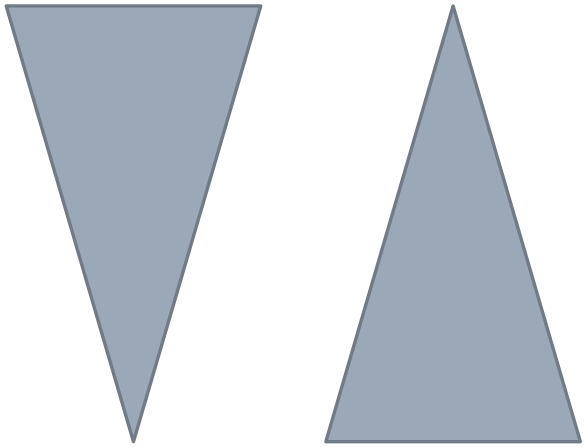
ADME – tegenwoordig zelden

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092479/

Kan AI daarmee helpen?

Vaak hebben we in discovery enkel model systemen:

- geisoleerde eiwitten
- cel-gebaseerde assays
- diermodellen
- humane data



Hoeveelheid data Relevantie

Kan AI daarmee helpen?

Efficaciteit – beperkte impact; kunnen sneller optimaliseren maar translatie van assay > human blijft een groot risico

Veiligheid – ja voor gekende mechanismen (hERG, mitotox), maar moeilijk voor onverwachte toxiciteitsbevindingen (idiosyncratic tox!)

Commerciele redenen – beperkt; acceleratie van programma kan eventueel helpen

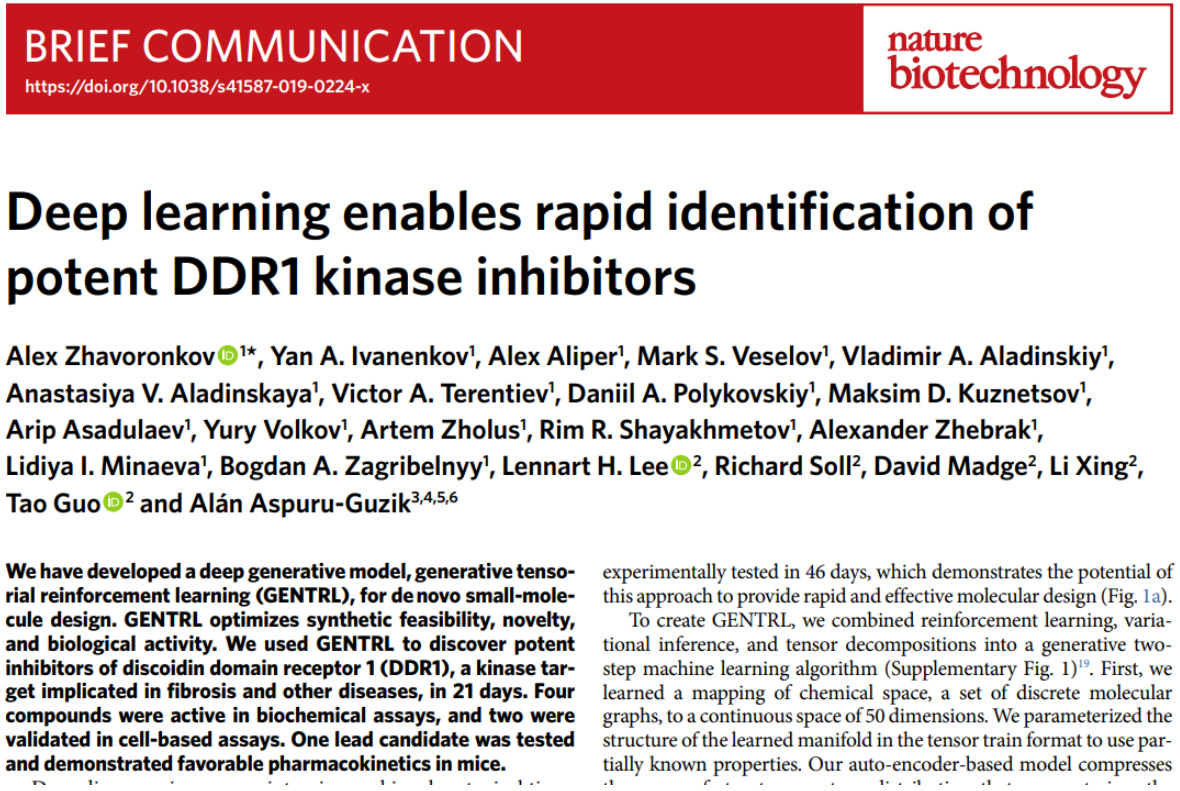
ADME – ja, predictie van ADME parameters werkt goed en translatie is vaak redelijk tot goed

https://www.sciencedirect.com/science/article/pii/S1359644621000428?via%3Dihub

Case study

ZHAVORONKOV ET AL.

Is generatieve ML dé oplossing voor drug design?



BRIEF COMMUNICATION
https://doi.org/10.1038/s41587-019-0224-x

nature biotechnology

Deep learning enables rapid identification of potent DDR1 kinase inhibitors

Alex Zhavoronkov[1*], Yan A. Ivanenkov[1], Alex Aliper[1], Mark S. Veselov[1], Vladimir A. Aladinskiy[1], Anastasiya V. Aladinskaya[1], Victor A. Terentiev[1], Daniil A. Polykovskiy[1], Maksim D. Kuznetsov[1], Arip Asadulaev[1], Yury Volkov[1], Artem Zholus[1], Rim R. Shayakhmetov[1], Alexander Zhebrak[1], Lidiya I. Minaeva[1], Bogdan A. Zagribelnyy[1], Lennart H. Lee[2], Richard Soll[2], David Madge[2], Li Xing[2], Tao Guo[2] and Alán Aspuru-Guzik[3,4,5,6]

We have developed a deep generative model, generative tensorial reinforcement learning (GENTRL), for de novo small-molecule design. GENTRL optimizes synthetic feasibility, novelty, and biological activity. We used GENTRL to discover potent inhibitors of discoidin domain receptor 1 (DDR1), a kinase target implicated in fibrosis and other diseases, in 21 days. Four compounds were active in biochemical assays, and two were validated in cell-based assays. One lead candidate was tested and demonstrated favorable pharmacokinetics in mice.

experimentally tested in 46 days, which demonstrates the potential of this approach to provide rapid and effective molecular design (Fig. 1a).

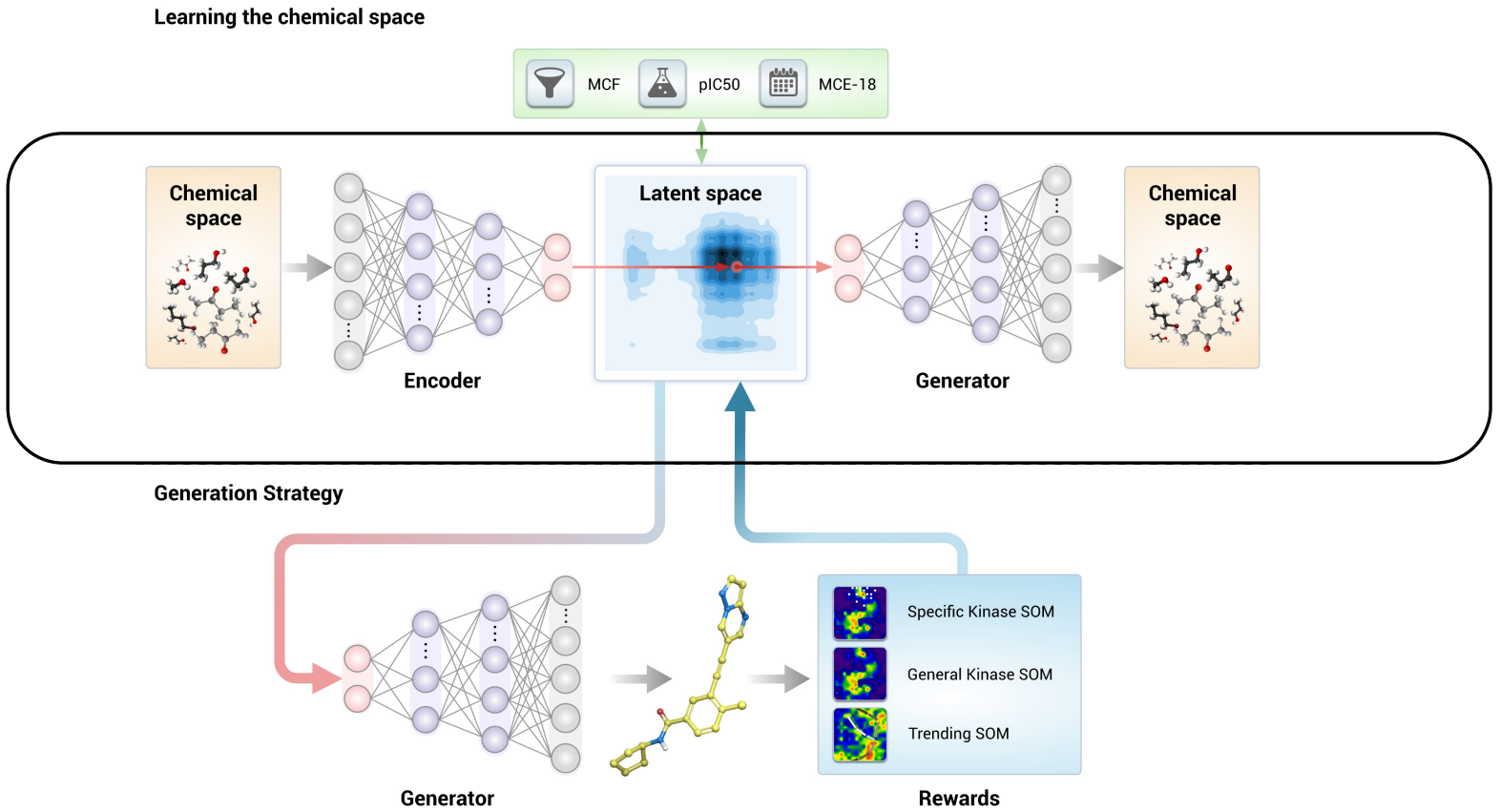
To create GENTRL, we combined reinforcement learning, variational inference, and tensor decompositions into a generative two-step machine learning algorithm (Supplementary Fig. 1)[19]. First, we learned a mapping of chemical space, a set of discrete molecular graphs, to a continuous space of 50 dimensions. We parameterized the structure of the learned manifold in the tensor train format to use partially known properties. Our auto-encoder-based model compresses

- Insilico claims major victory as AI outperforms big pharma in early drug R&D
- AI Goes Where AI Has Never Gone Before: Artificial Intelligence (AI)—in a first-ever—has designed and validated a new drug candidate in days
- Pharma's AlphaGo Moment: For the First Time AI Has Designed and Validated a New Drug Candidate in Days

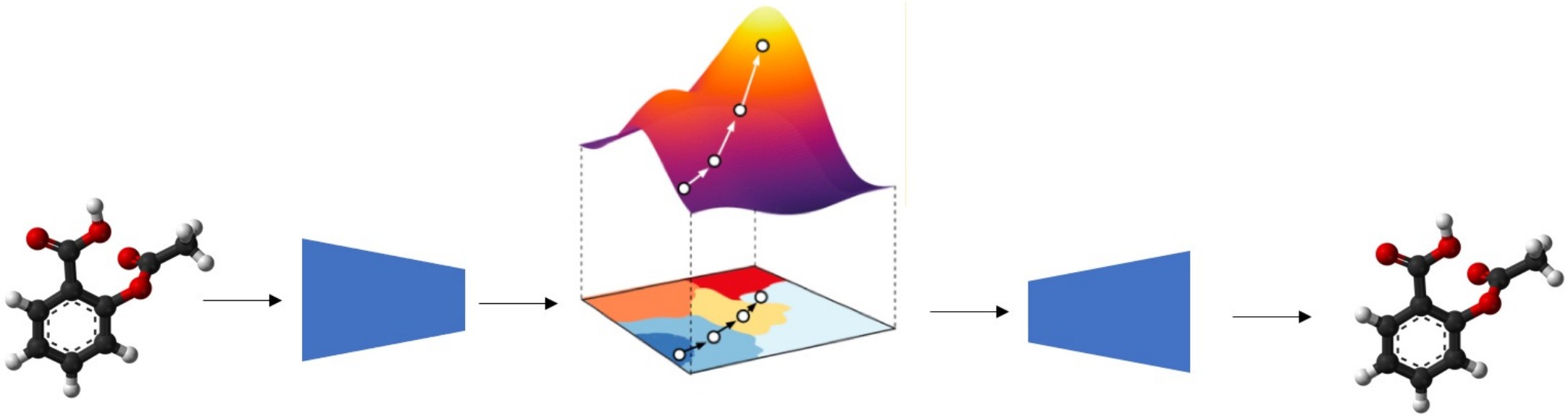
https://pharmaphorum.com/news/insilico-claims-major-victory-as-ai-drastically-cuts-drug-development-time/
https://asianroboticsreview.com/home301-html
https://www.linkedin.com/pulse/pharmas-alphago-moment-first-time-ai-has-designed-new-colangelo

Generative ML – *New Hope* of *New Hype*?

Een blik op de onderliggende technologie

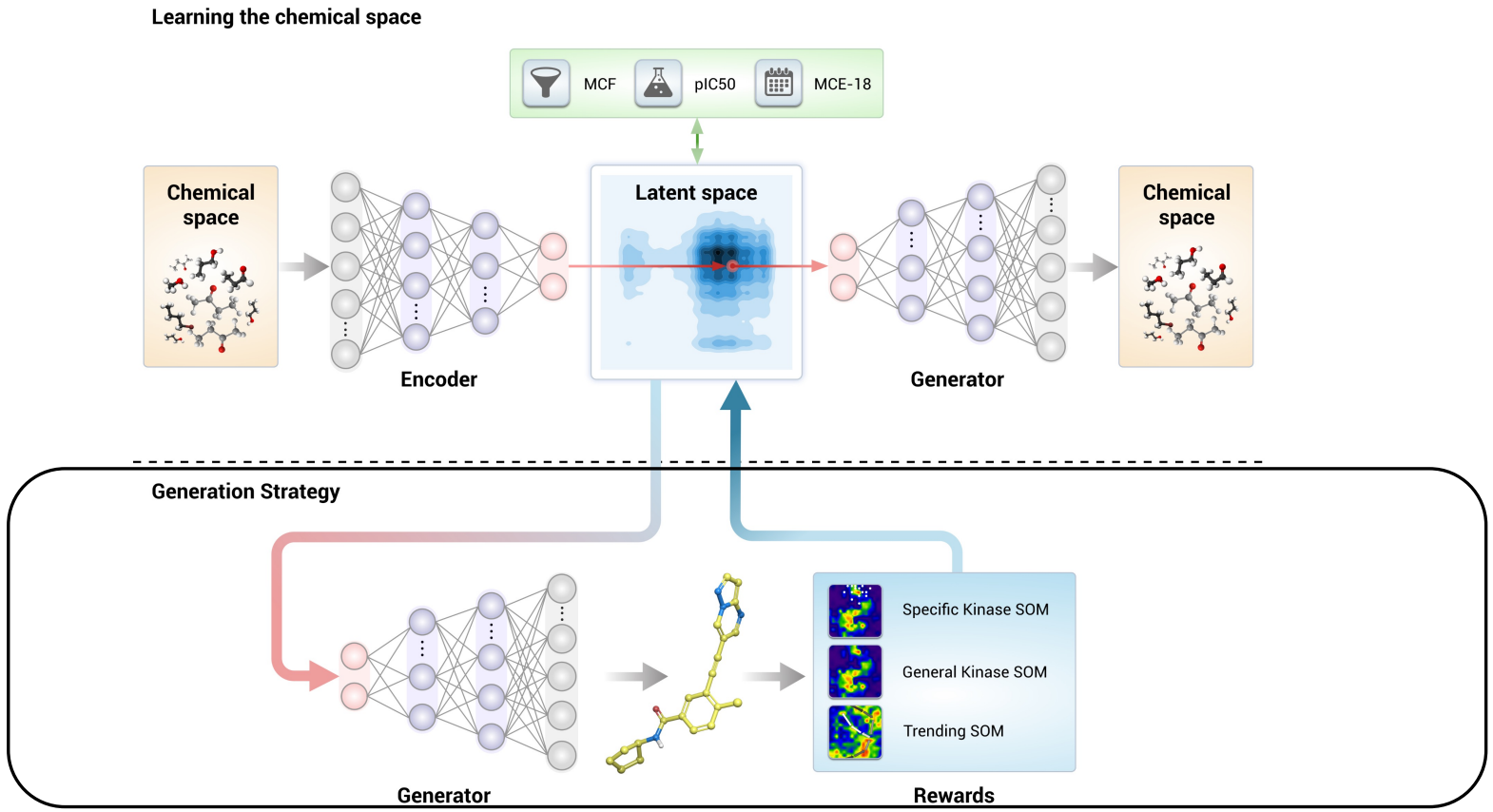


Smiles-based autoencoders mappen chemie naar een continue representatie

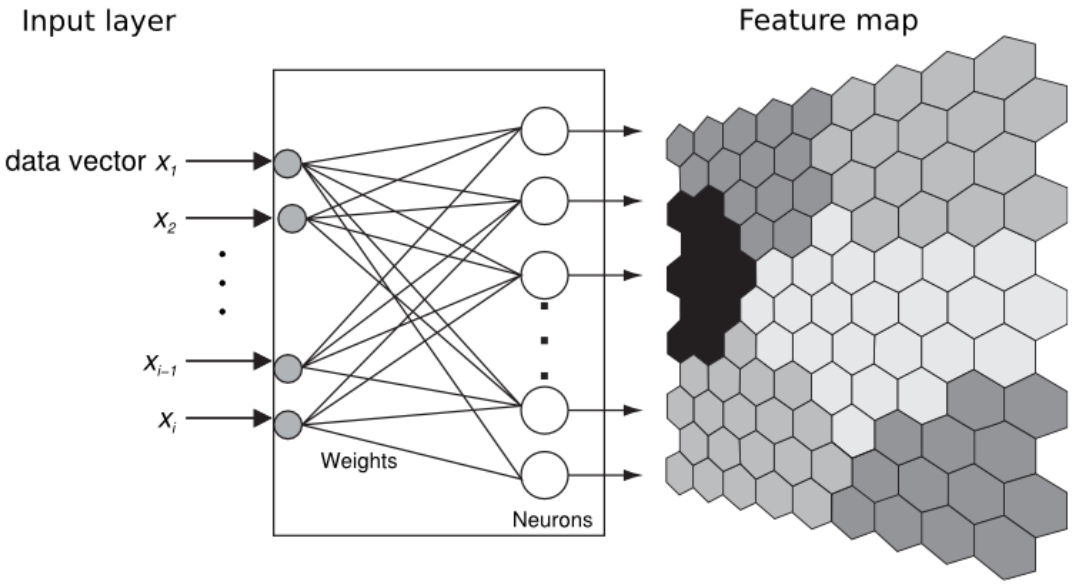


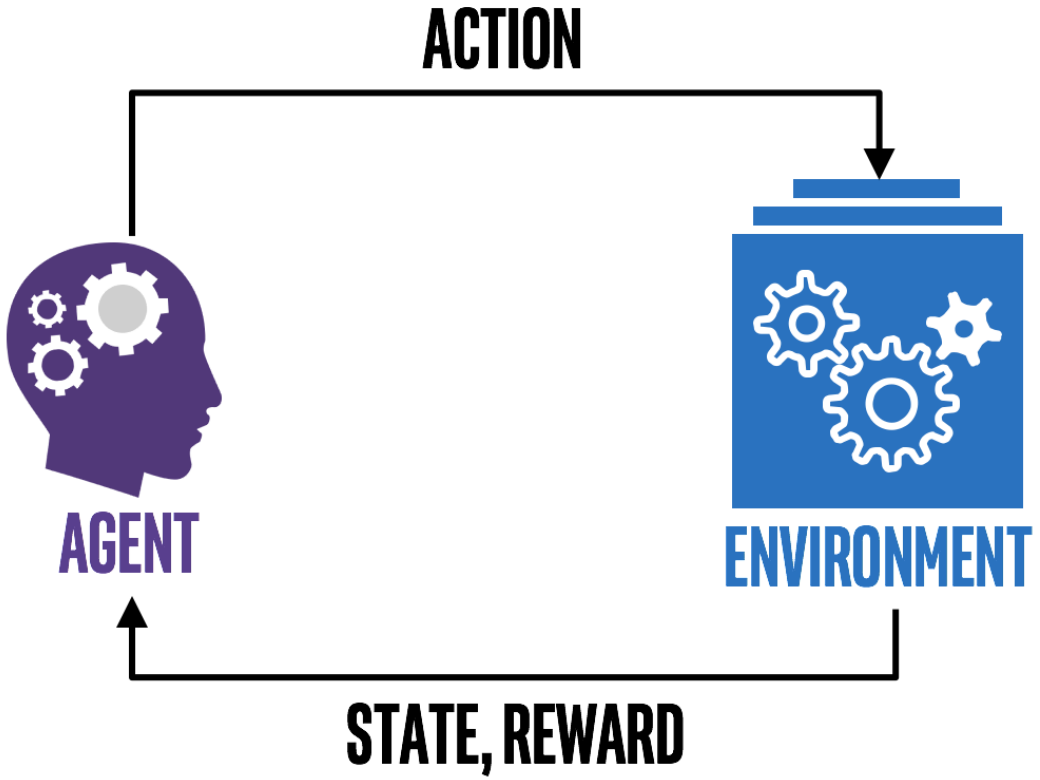
Gomez-Bombarelli et al: https://arxiv.org/abs/1610.02415

Een blik op de onderliggende technologie



Self-organizing maps & reinforcement learning





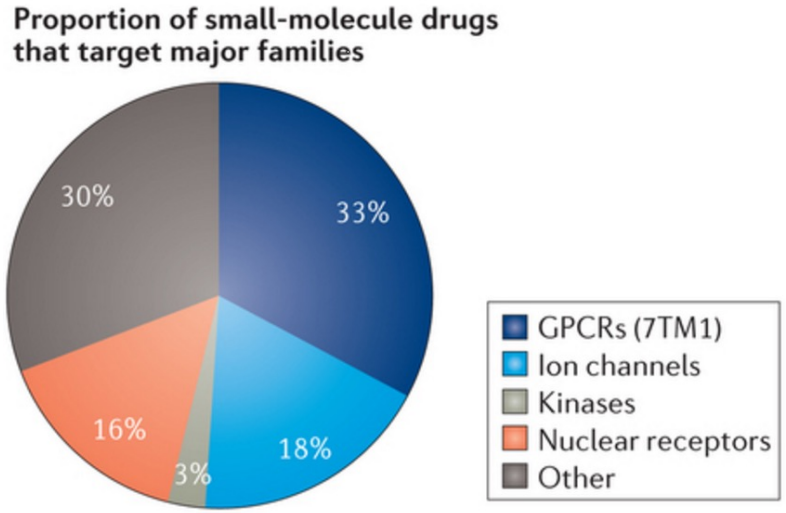
Neural network voor dimensionaliteitsreductie

https://nervanasystems.github.io/coach

Kinases als drug target

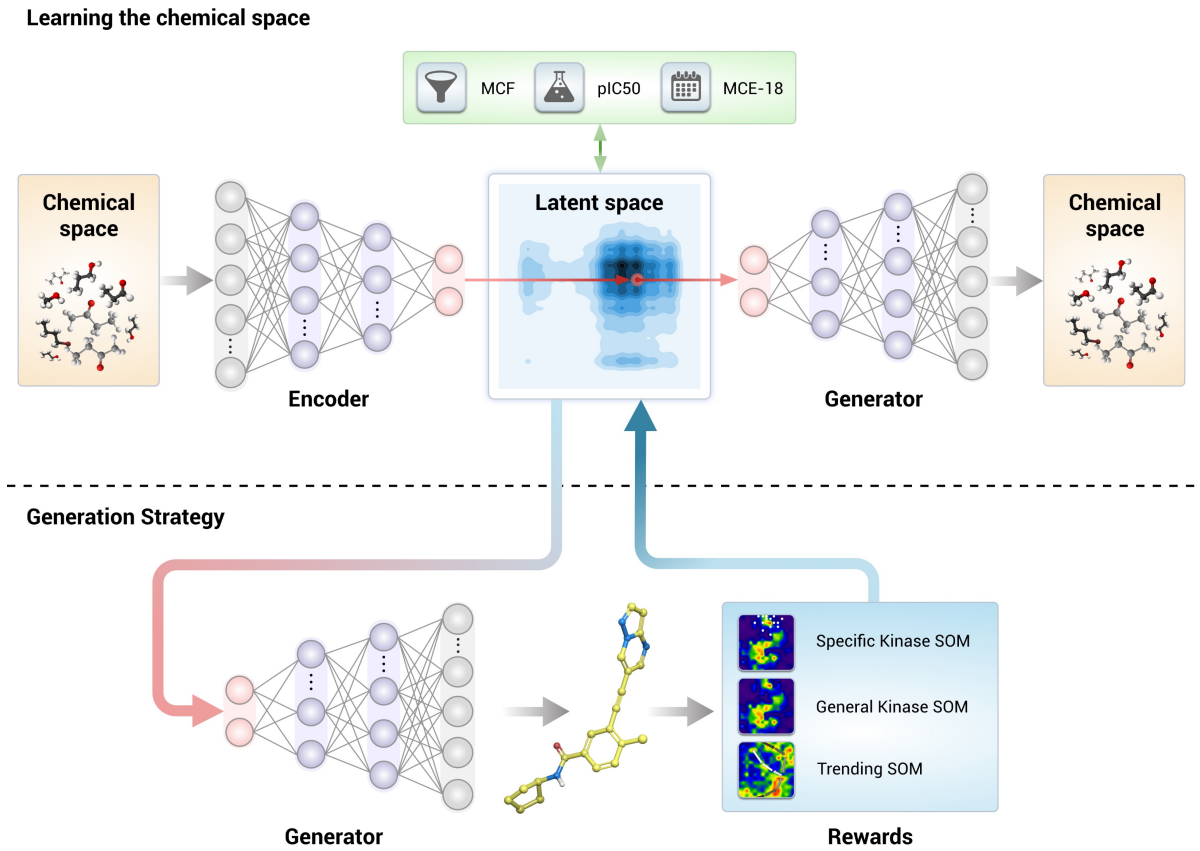
- Kinase inhibitoren worden gebruikt voor diverse aandoeningen:
  - Oncologie
  - Autoimmuunaandoeningen
  - Degeneratieve aandoeningen

- Eenvoudig om hits te vinden tegen kinases – selectiviteit is moeilijk. Meer dan 500 humane kinasen gekend.



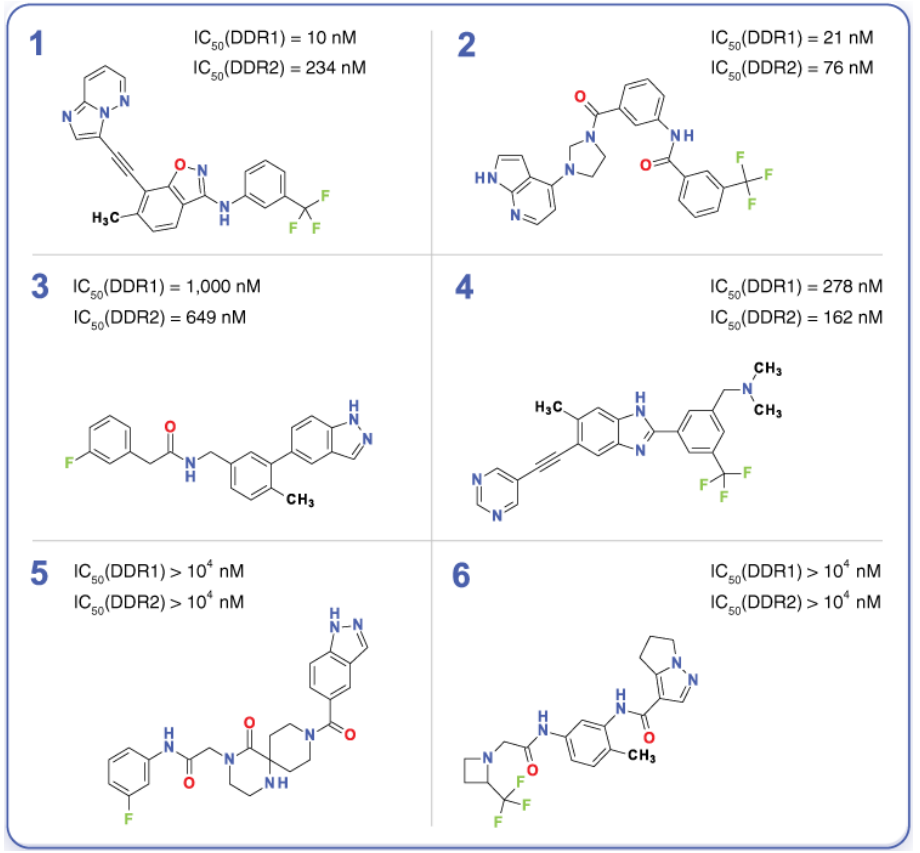
https://www.nature.com/articles/nrd.2018.21/figures/1

Een blik op de onderliggende technologie

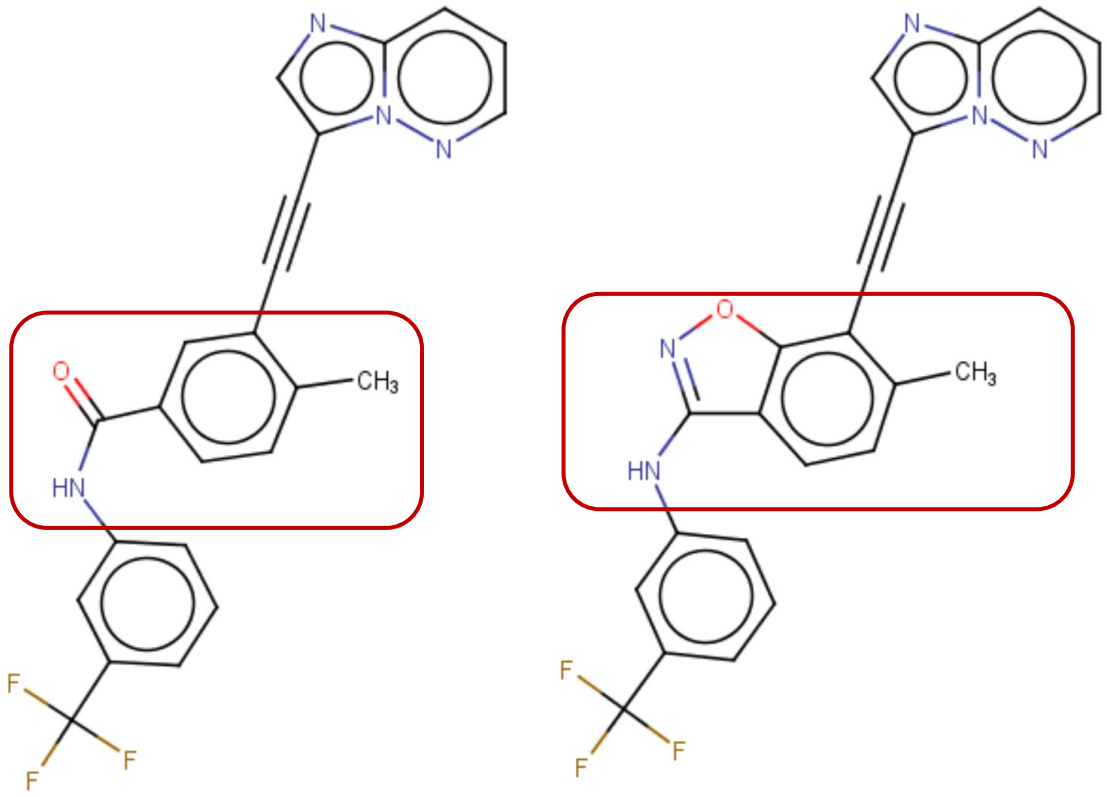


Verdere selectie en synthese

- Filteren van moleculen door middel van:
  - Molecular descriptoren
  - Verwijderen reactive and undesirable groups
  - Clusteren
  - Kinase SOM
  - Farmacofoor screening
- 6 moleculen getest
- Compound 1 werd getest in een muismodel, met een halfleven van 3.5h



Compound 1 lijkt op bekende DDR1 inhibitor!



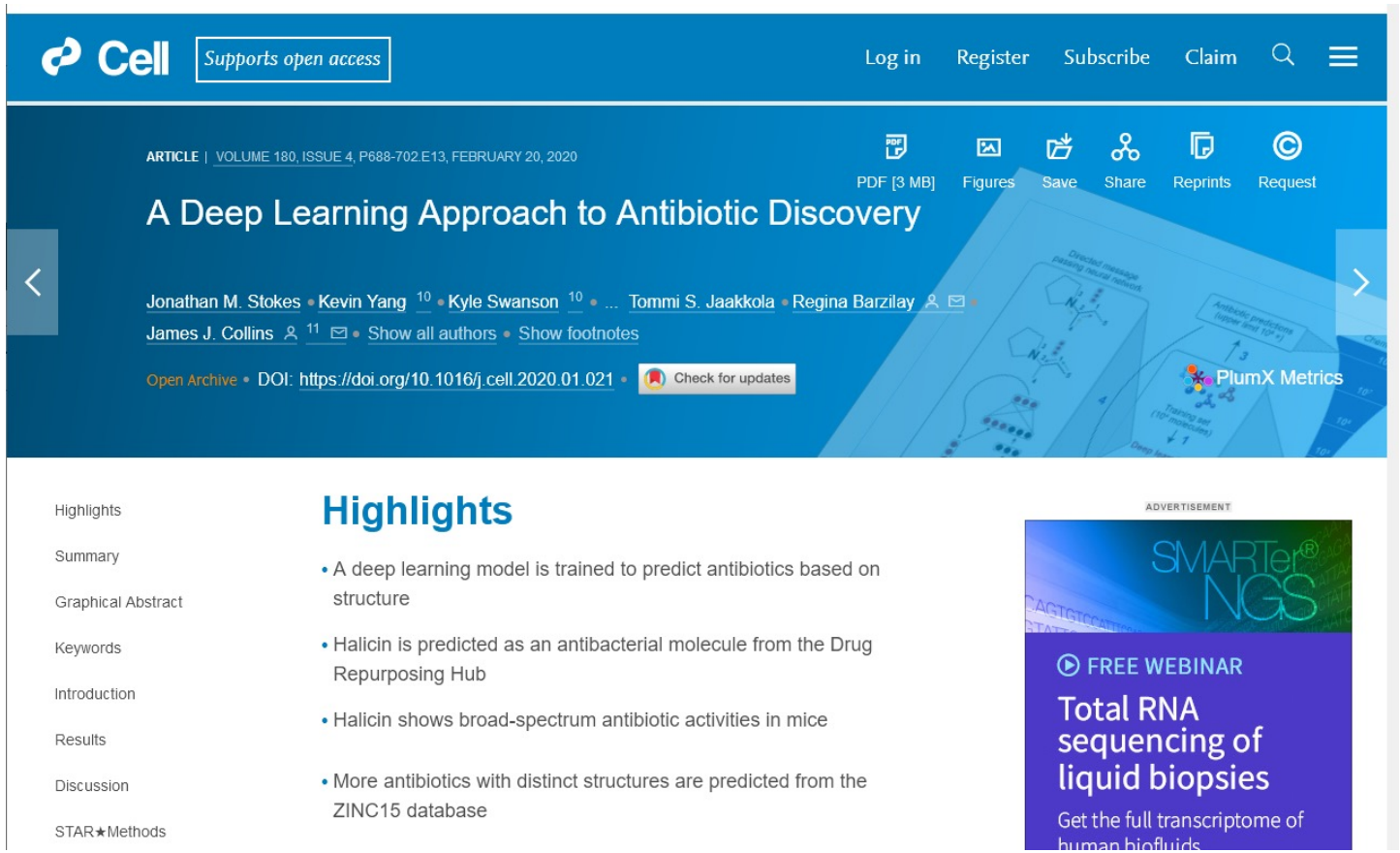
Conclusie

- Veel papers in generatieve veld optimaliseren simpele descriptoren zoals logP – biologische validatie ligt dichter bij de echte toepassing
- DDR1 is een “eenvoudig” target: veel chemische materie gekend + kristalstructuur beschikbaar
- Generatieve machine learning is een interessante tool, maar zal alleen drug discovery niet oplossen!
- Modellen moeten meer synthesis-aware worden om breder inzetbaar te zijn.

Case study

STOKES ET AL.

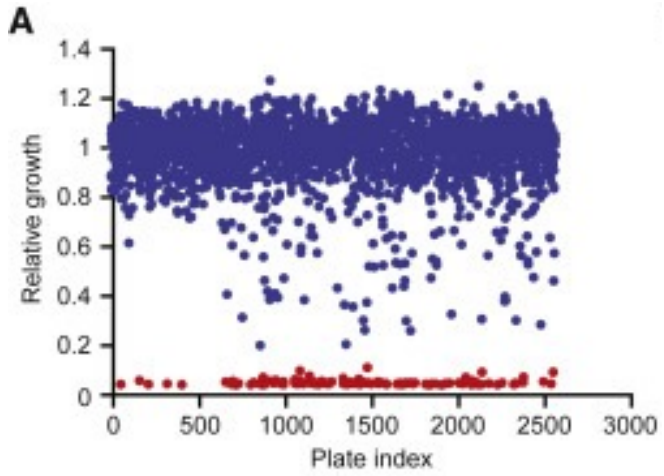
Deep learning for antibiotic discovery



Deep learning for antibiotic discovery

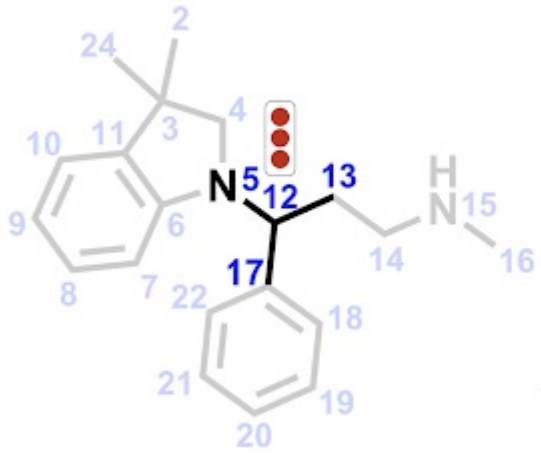
Doel: drug repurposing via AI om antibioticaresistente bacteriën te bestrijden

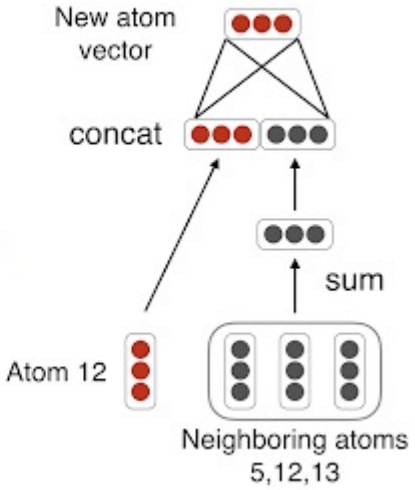
Trainingsdata: 2335 moleculen uit natuurlijke producten + approved drugs

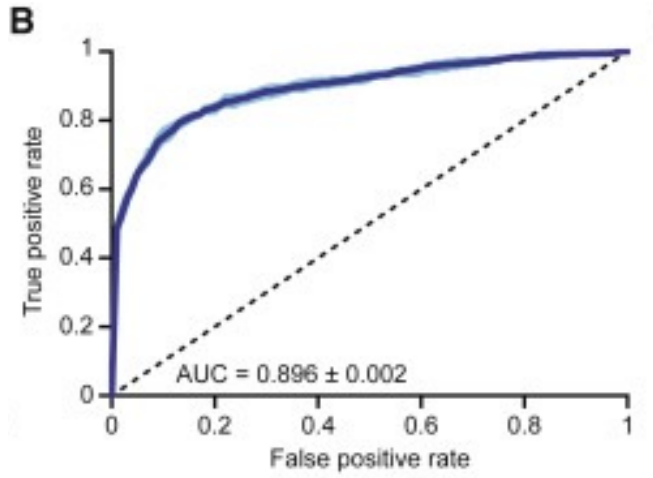


ChemProp model

Message passing neural network

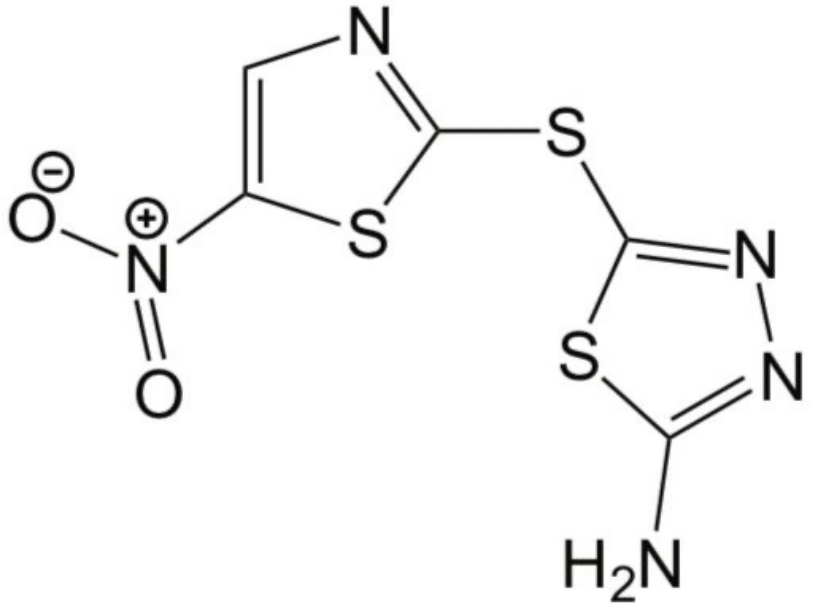






Screenen naar nieuwe compounds

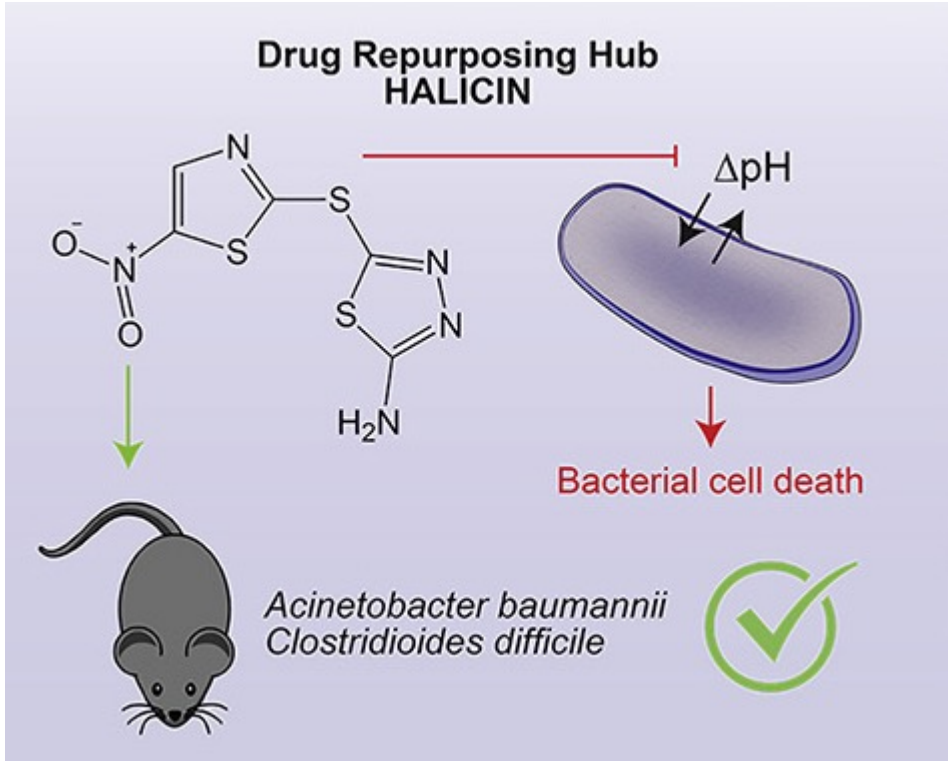
- 6111 moleculen uit drug repurposing hub
- Top 99 geselecteerd en getest
- 51 moleculen vertonen groei-inhibitie ten opzichte van E.Coli!
- Meest interessente compound: Halicin
  - Oorspronkelijk ontwikkeld voor diabetes, preklinisch molecule
  - Lage voorspelde toxiciteit



Profiling van Halicin

Werkzaam in muismodel van infectie

Nieuw mechanisme: intracellulaire ijzersequestratie wat pH balans verstoort



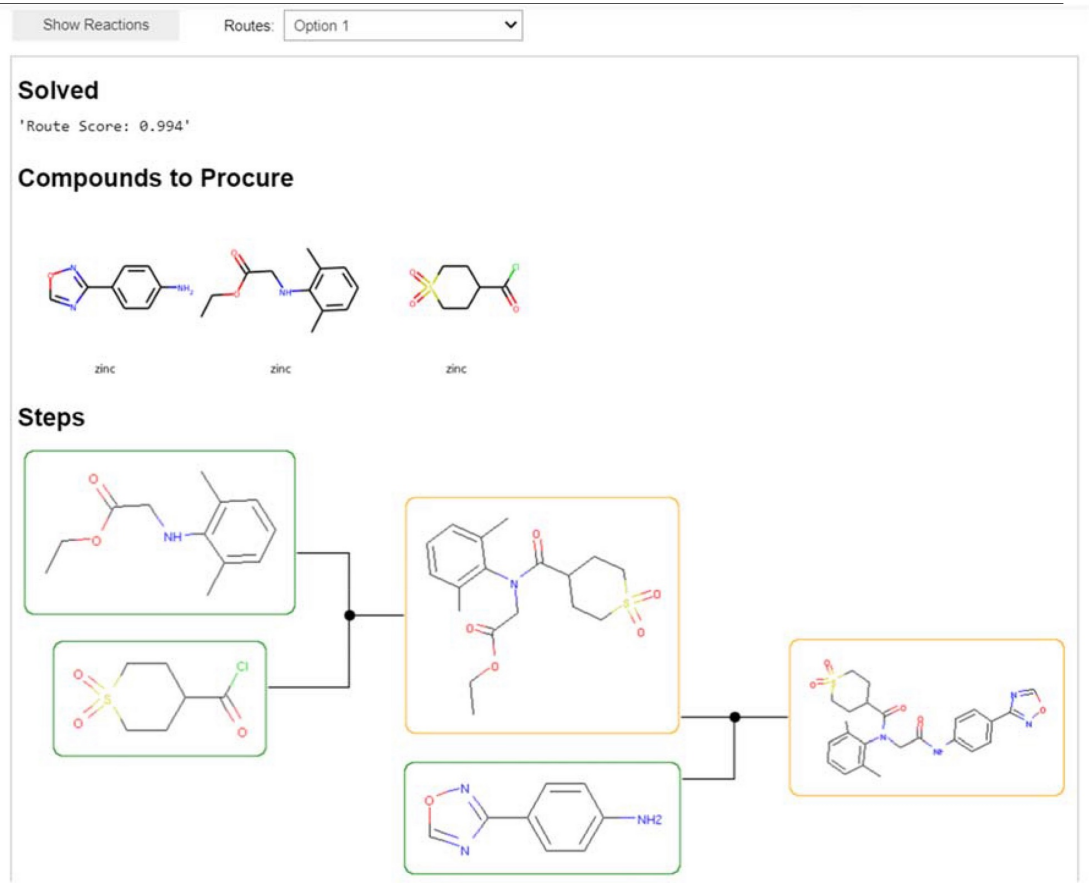
Andere toepassingen

AI IN DRUG DISCOVERY

AI voor retrosynthese

AiZynthfinder gebruikt reactie templates en een databank aan reagentia om syntheseroutes voor te stellen

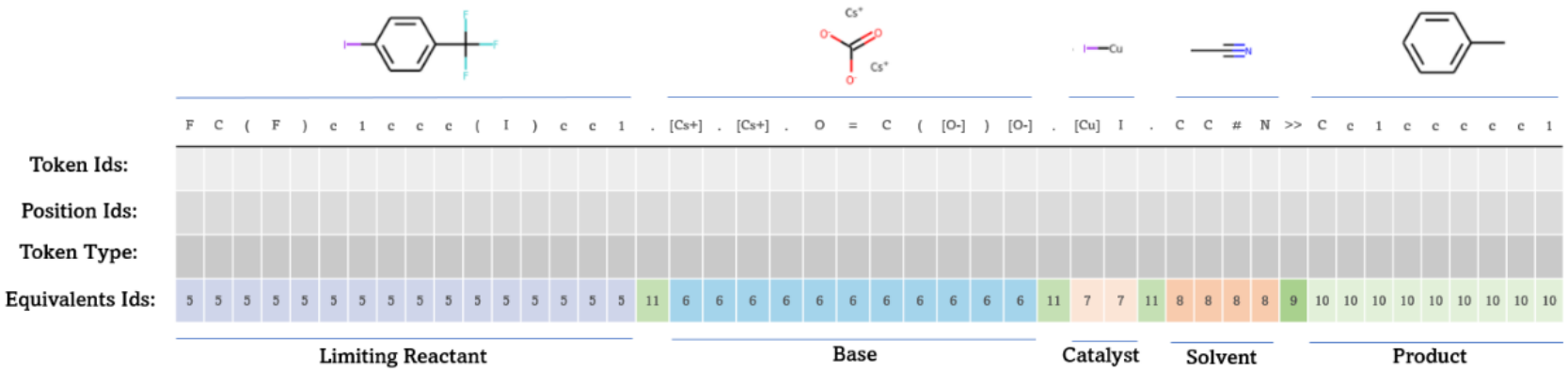
Monte carlo tree search als algoritme



Voorspellen van reactie succesrate

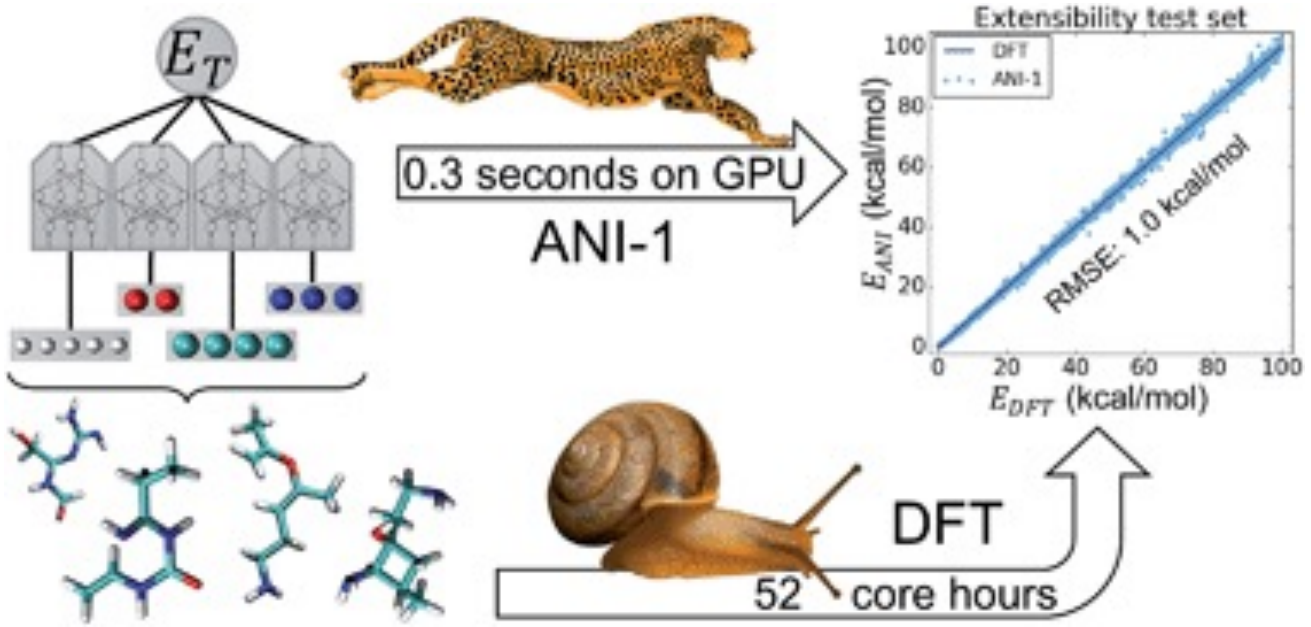
Model voorspelt of reactie werkt of niet (>5% yield)

Kan gebruikt worden als reagent recommender



Global Reactivity Models are Impactful in Industrial Synthesis Applications, arxiv

Acceleratie van QM



QM-berekeningen zijn nuttig voor vele farmaceutische eigenschappen zoals chemische stabiliteit, conformationele energie, covalente binding, …

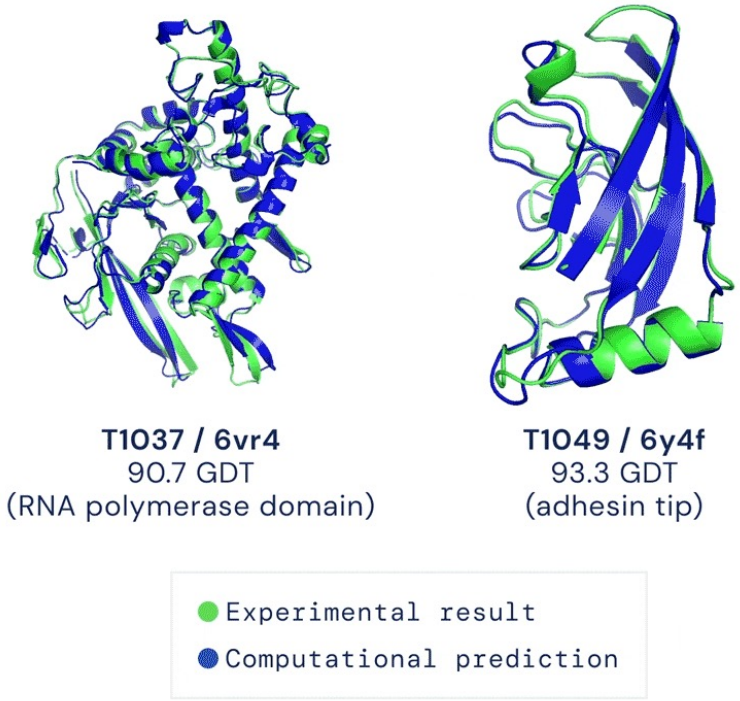
Computationeel kostelijk en moeilijk *at scale* te gebruiken!

QM-ML modellen bieden DFT-like performance at forcefield cost

Alphafold

Bijzonder performante structuurvoorspelling met behulp van AI

Globale structuur komt goed overeen, maar kleinere side chains zijn vaak nog niet voldoende accuraat (bv voor docking).



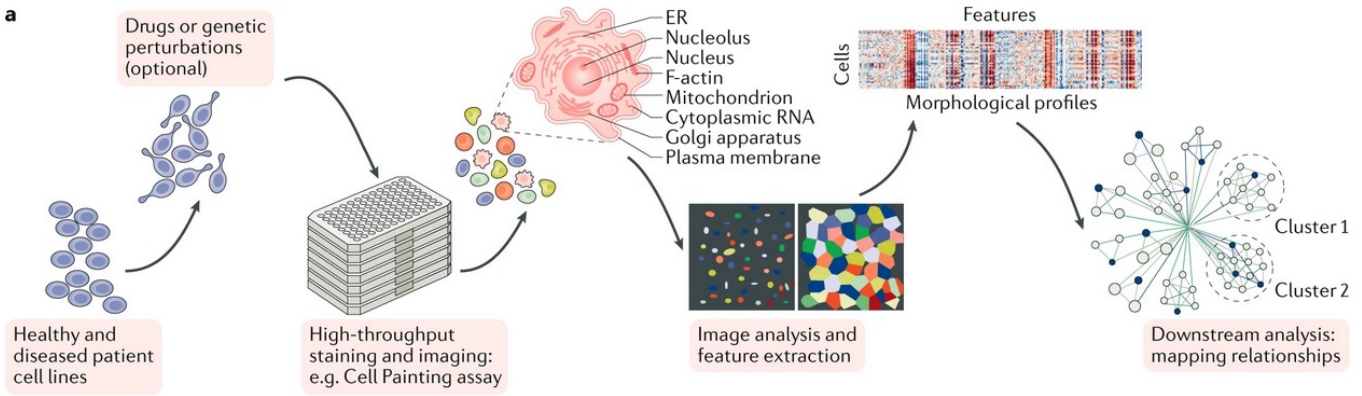
https://www.nature.com/articles/s41586-021-03819-2
https://chemrxiv.org/engage/chemrxiv/article-details/6316a14149042ab863cd0481

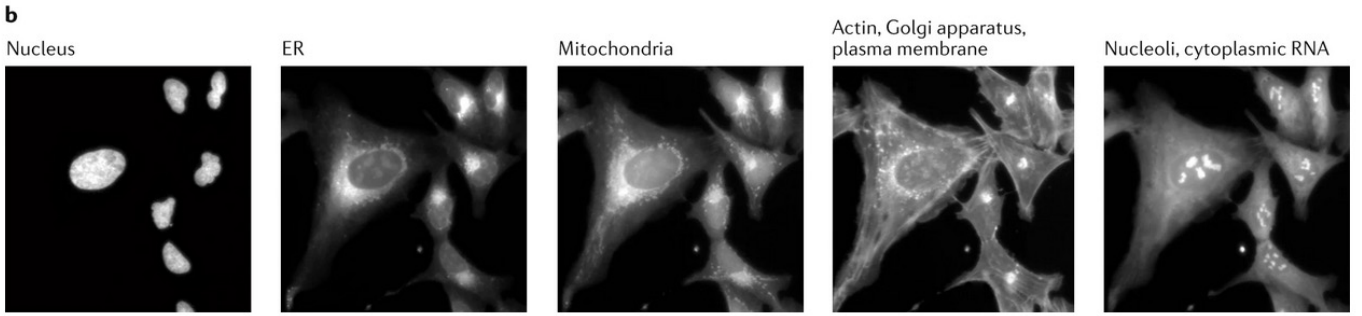
Biologie begrijpen via AI

Cell painting data is multidimensioneel – moeilijk te interpreteren

AI kan helpen: clustering of supervised machine learning

Cell painting + perturbaties om therapeutische interventies te begrijpen





a | Overview of the typical steps in the workflow for generating image-based profiles from biological samples. **b** | Example images from the Cell Painting assay often used for image-based profiling. It includes six stains labelling eight cellular components, which are imaged in five channels[20]. ER, endoplasmic reticulum.

Disclaimer

De standpunten op deze slides zijn de persoonlijke mening van de auteur en staan los eventuele meningen/standpunten van Janssen Pharmaceutica NV.