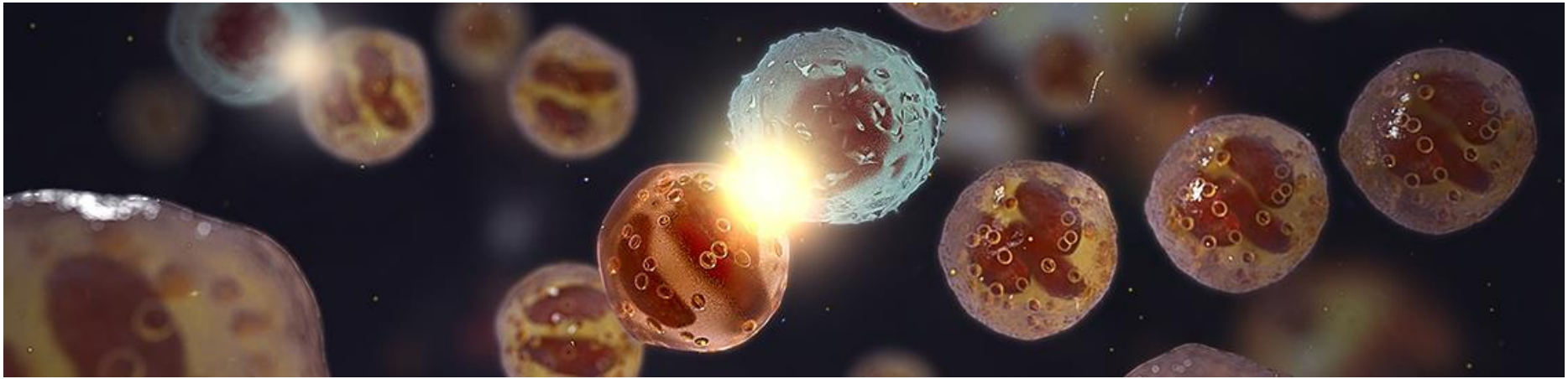
Cheminformatics in Drug Discovery, an Industrial Perspective

Hongming Chen, Thierry Kogej, Ola Engkvist
Hit Discovery, Discovery Sciences, Astrazeneca R&D Gothenburg



Drug Discovery Today

An industry perspective of Today's Challenge

We need to become **better**, **faster** & **cheaper**

Source: Paul et al., Nat. Rev. Drug Disc. (2010) 9:March

Source: PhRMA profile 2016

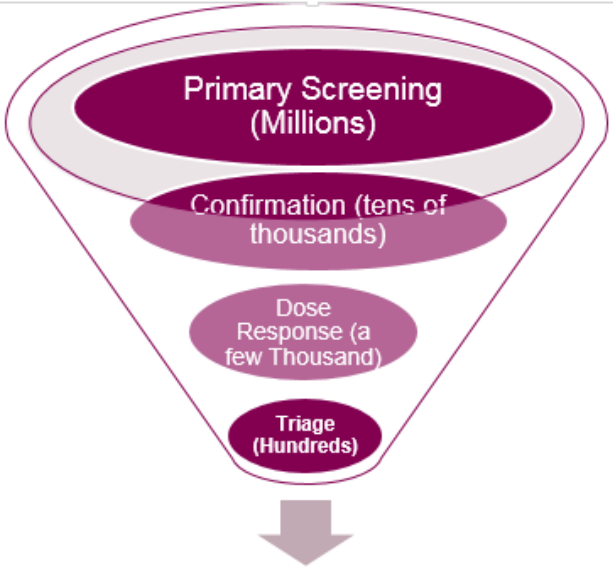
Cheminformatics @ AstraZeneca

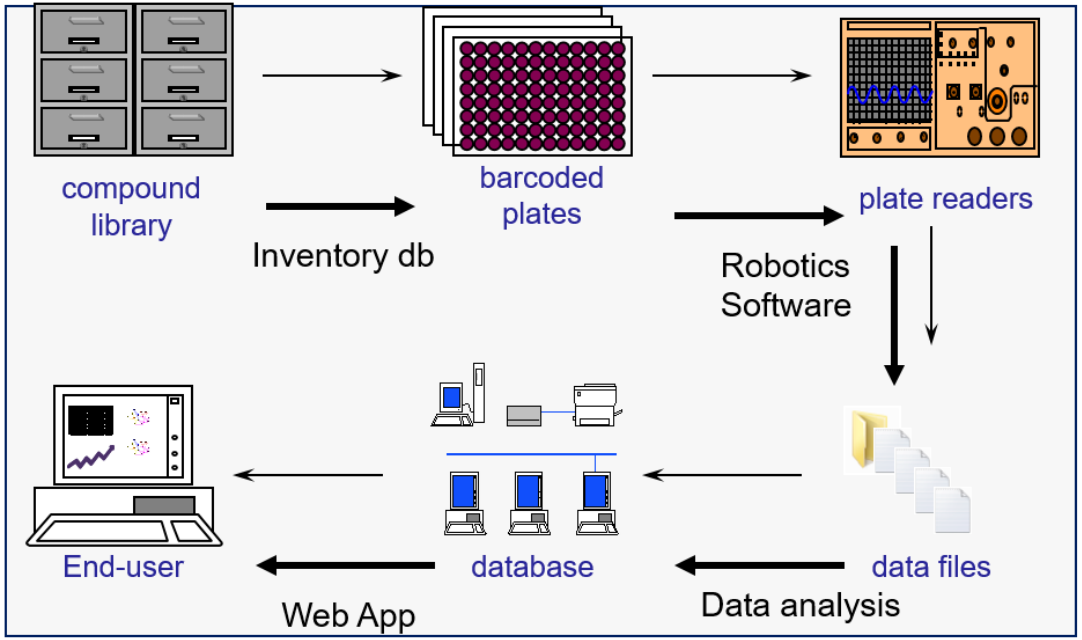
- HTS work-up
- Library design
- Virtual screening
- Machine learning & AI

High Throughput Screening

From Millions to just a few

Low cost/compound





High cost/compound

Slide modified from Mark Wigglesworth, AZ, with permission

HTS Analysis: Clustering analysis

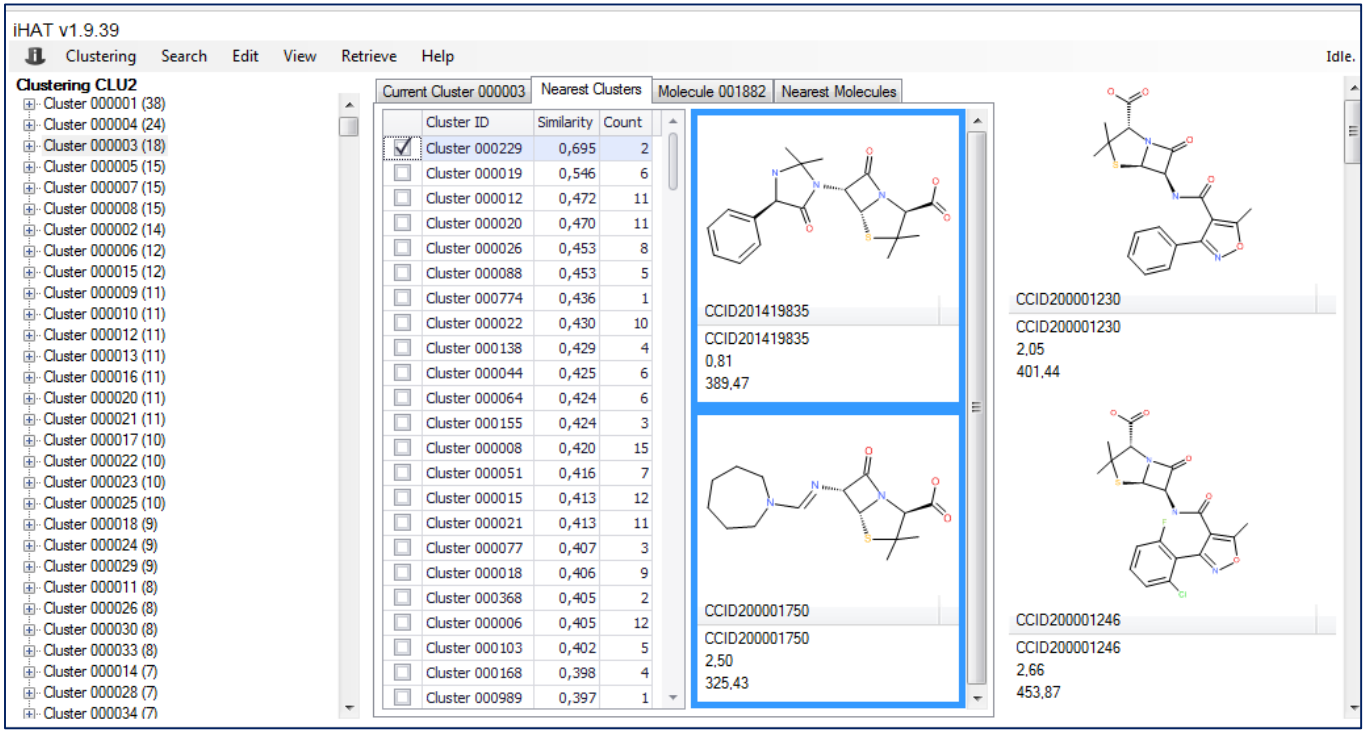
Early days

- Heavily dependent on computational chemistry resources
- Linux, scripts, static workflows, data in flat files
- Cutting, pasting and reformatting between applications
- Difficult to revisit or take over an analysis from a colleague
- Time-consuming

iHAT: An Spotfire add-in for HTS Analysis

- Leverage the powerful visualization function of Spotfire
- Annotation of compounds with in-house experimental and predicted data
- Data integration from multiple sources
- Clustering of compounds
- Visualization and manipulation of cluster tree
- NN search

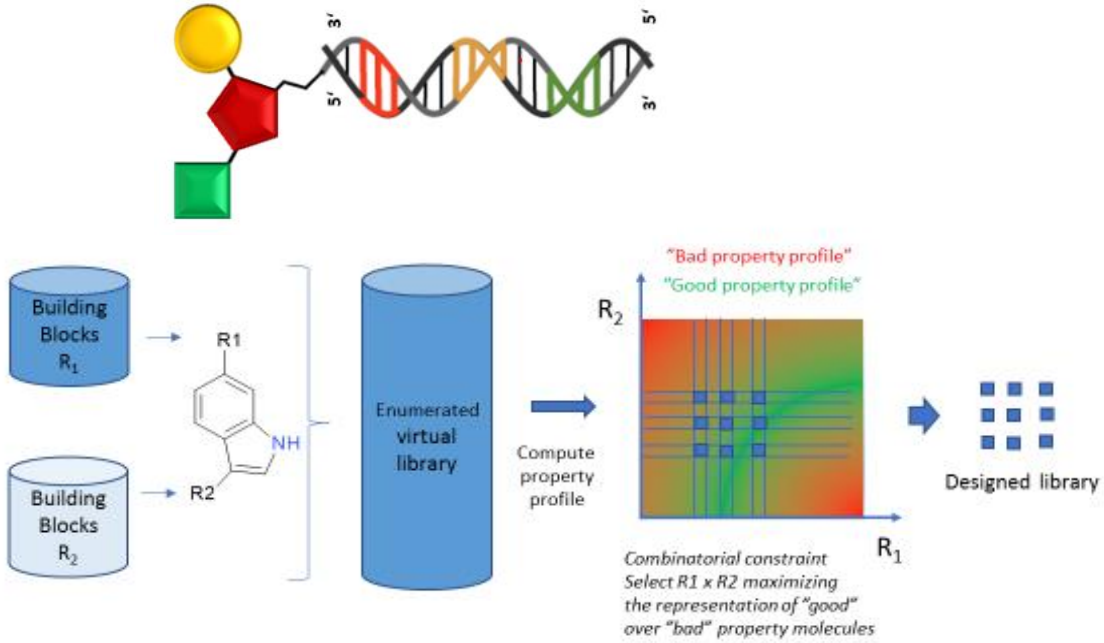
iHAT: Clustering and Reclustering

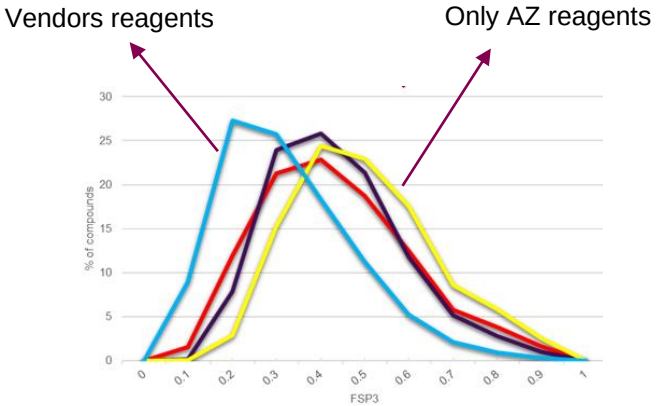


Library design @ AstraZeneca

- Diversity library is generally out of fashion
- Focused library fit for specific project need
- DNA encoded libraries become popular, but analysis is challenging, >60M to 8B library sizes

Currently, use classical library design method to reduce to 50M preferred AZ library size





Fsp3

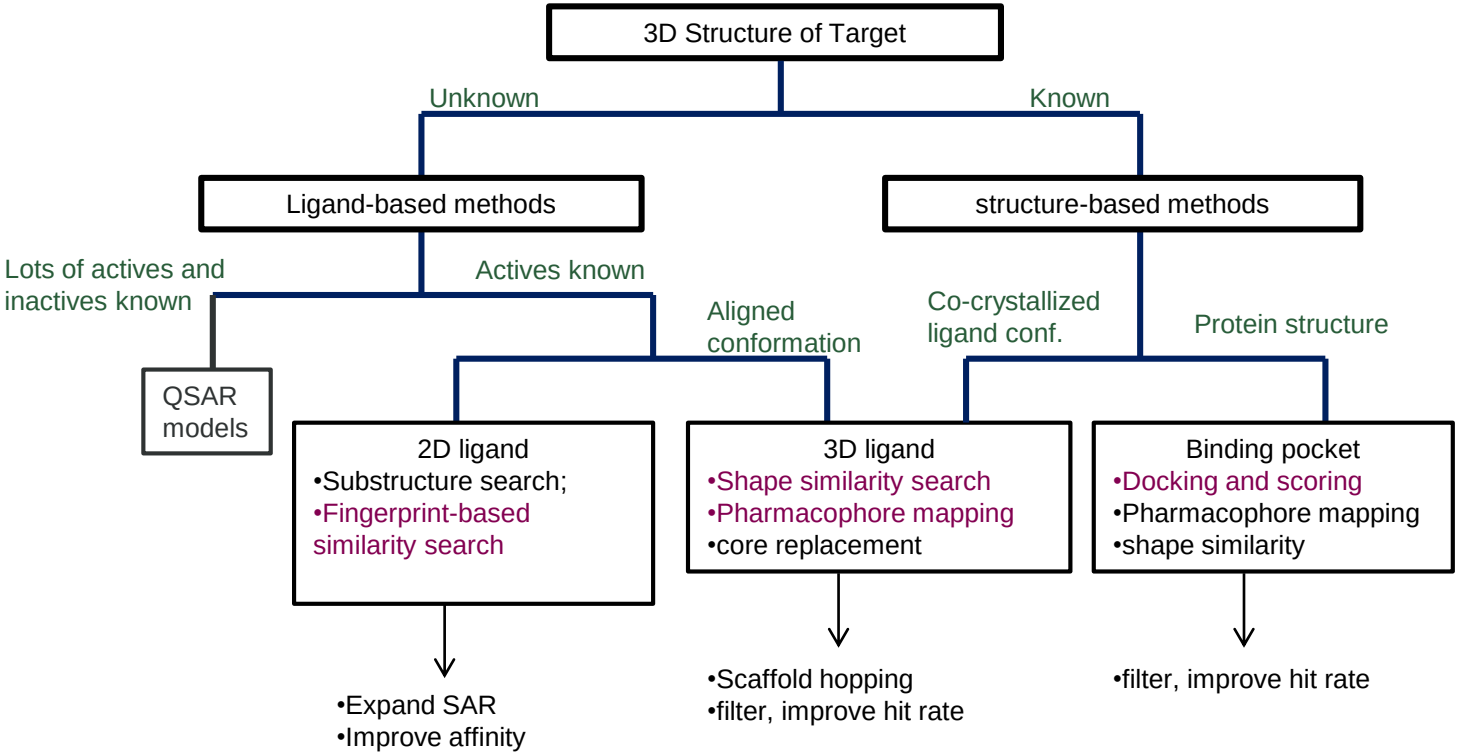
Definition of VS

- Virtual screening refers to any in-silico techniques used to search large compound databases (available chemicals or virtual libraries) to select a smaller number for biological testing

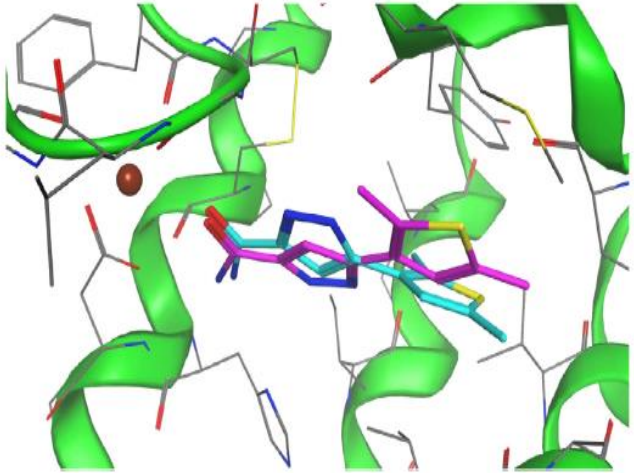
- Virtual screening can be used to
  - Select compounds for screening from in-house databases
  - Choose compounds to purchase from external suppliers
  - Select compounds from virtual libraries to be synthesized

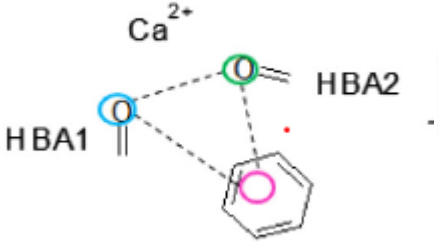
- The technique applied depends on the amount of information available about the particular disease target and the desired outcome

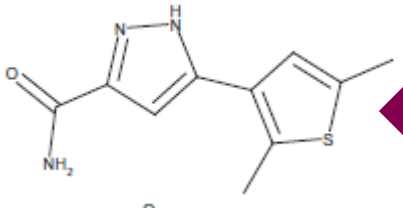
VS methods


3D Structure of Target
Unknown
Known
Ligand-based methods
structure-based methods
Lots of actives and inactives known
Actives known
Aligned conformation
Co-crystallized ligand conf.
Protein structure
QSAR models
2D ligand
•Substructure search;
•Fingerprint-based similarity search
3D ligand
•Shape similarity search
•Pharmacophore mapping
•core replacement
Binding pocket
•Docking and scoring
•Pharmacophore mapping
•shape similarity
•Expand SAR
•Improve affinity
•Scaffold hopping
•filter, improve hit rate
•filter, improve hit rate

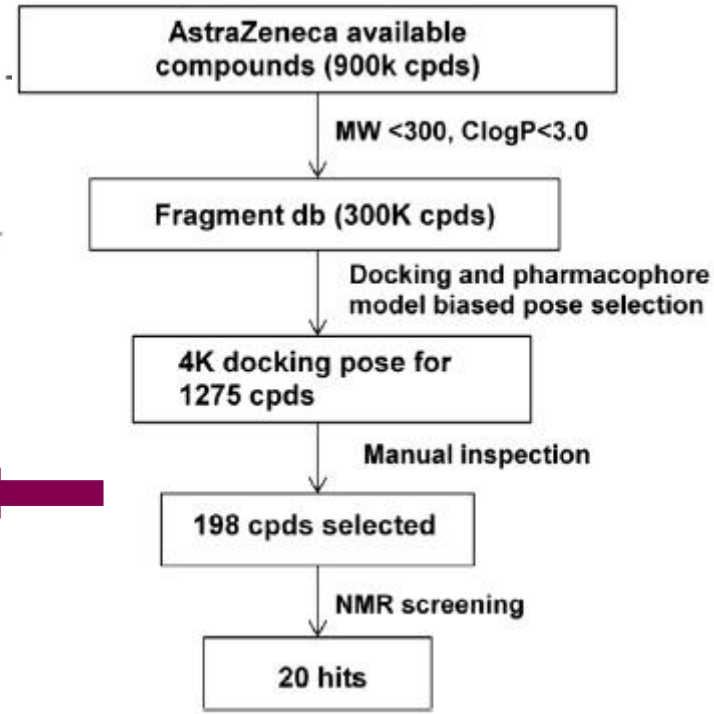
VS example







IC50 (NMR) 20uM

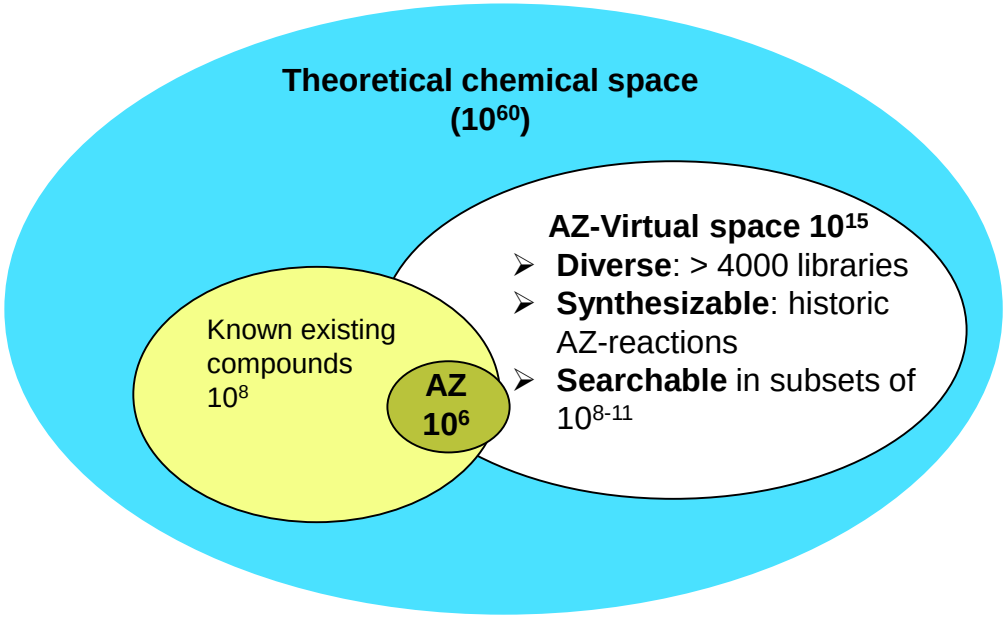


- Identification of sPLA2X inhibitors using ligand and structure based virtual screening

H. Chen et al./Bioorg. Med. Chem. Lett. 24 (2014) 5251–5255

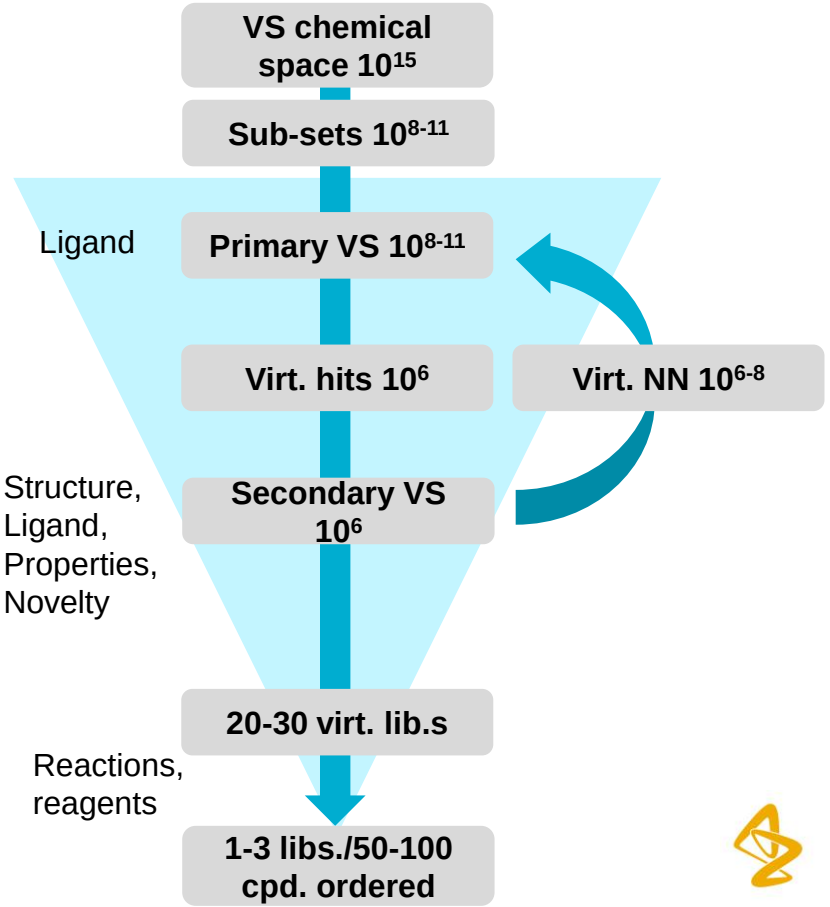
Virtual screening platform @ AZ

Chemical space

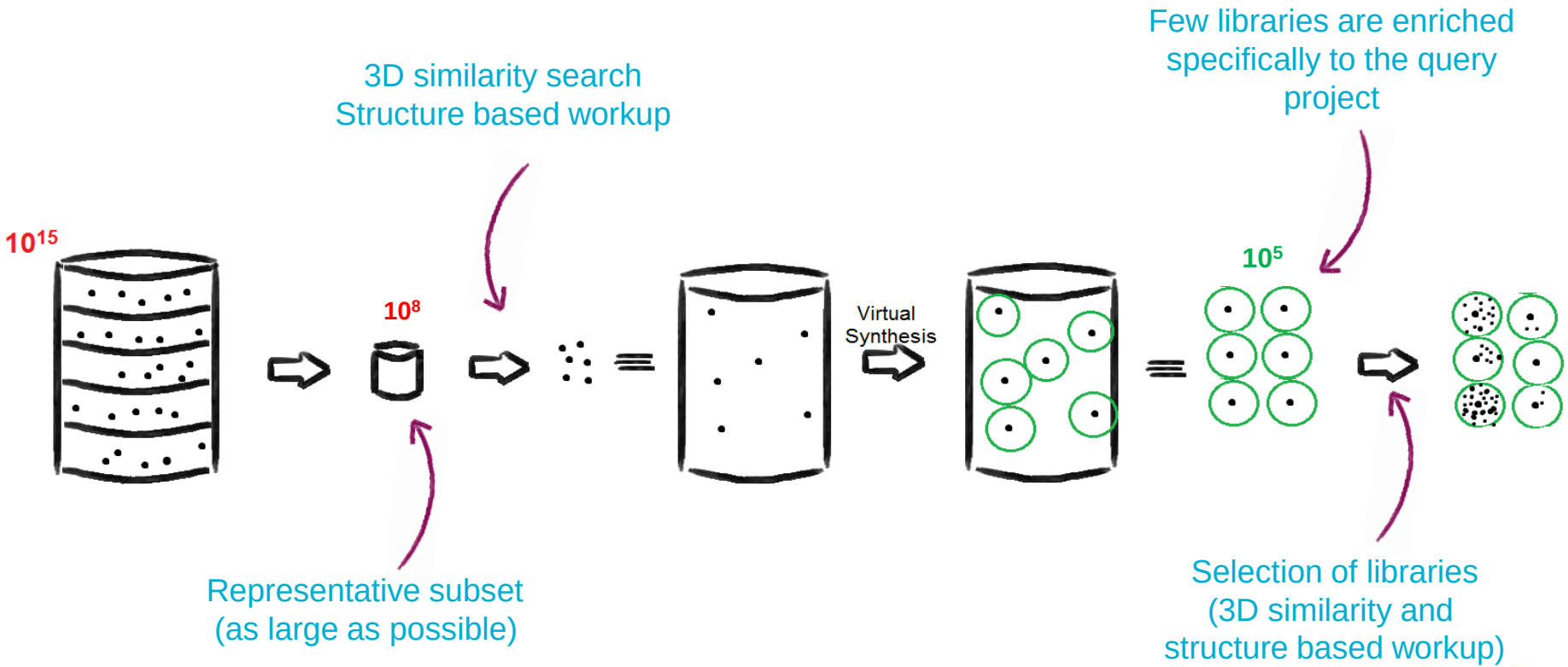


MJ Vainio *J. Chem. Inf. Model.*, **2012**, *52* (7), pp 1777–1786.

Iterative virtual screening workflow

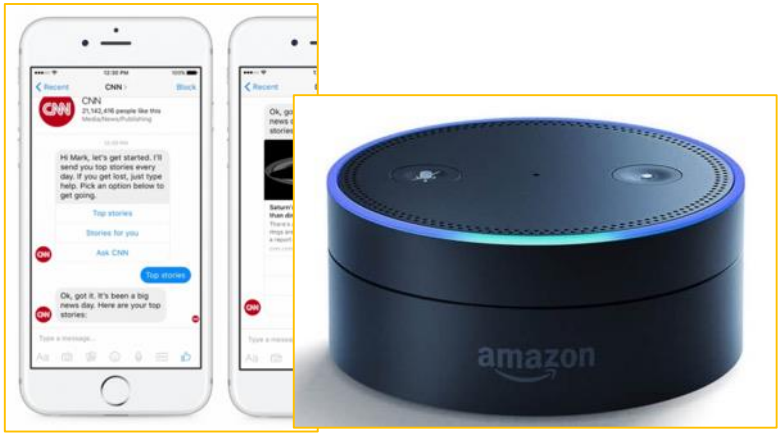


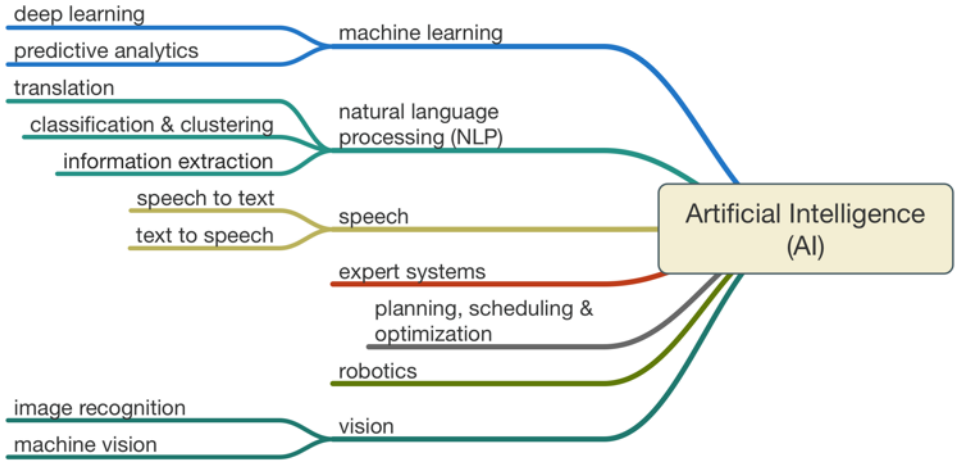
Computational strategy



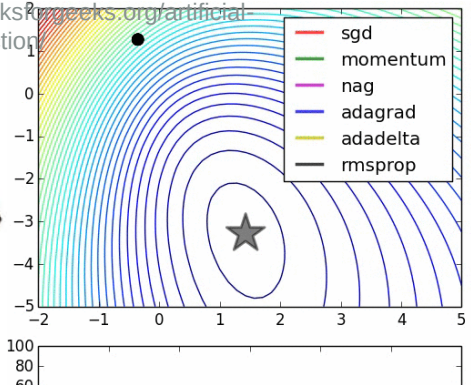
AI & Machine Learning Today

Context, Definition & Advances

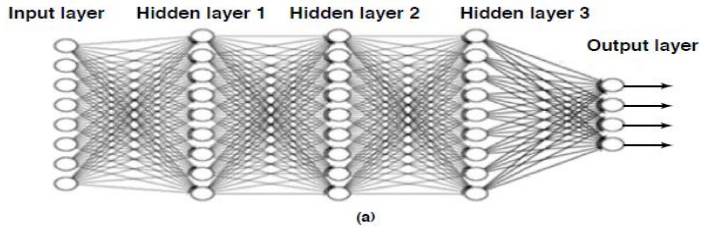


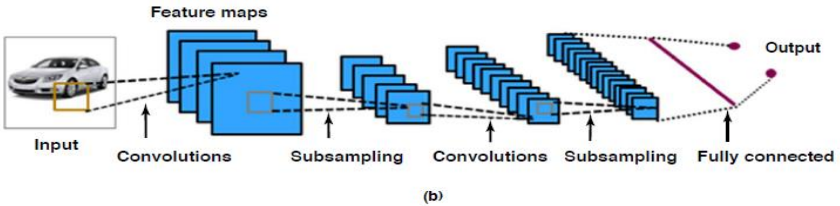


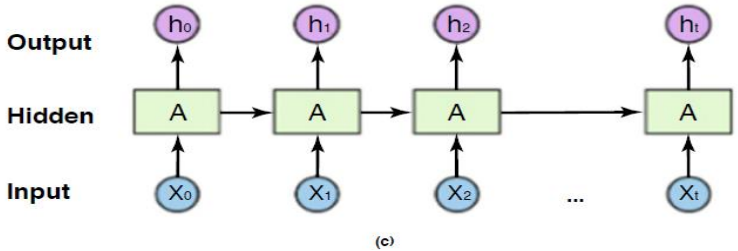
Source: http://www.geeksforgeeks.org/artificial-intelligence-an-introduction/

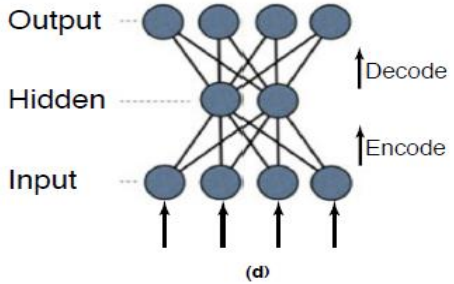


The rise of deep learning in drug discovery



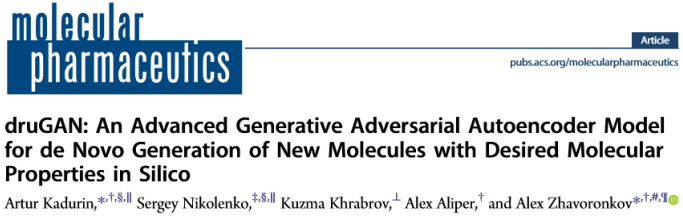






- Deep learning technologies have been adopted in drug discovery
- Various forms of NN have been applied so far

De novo molecular generation with deep learning has developed very rapidly

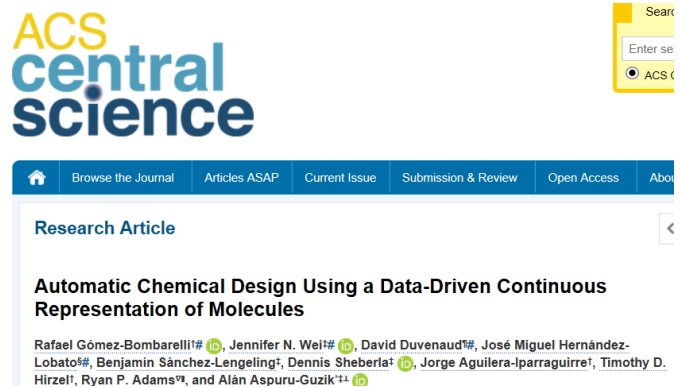

molecular pharmaceutics

Article

pubs.acs.org/molecularpharmaceutics

druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico

Artur Kadurin,[*,†,§,∥] Sergey Nikolenko,[‡,§,∥] Kuzma Khrabrov,[⊥] Alex Aliper,[†] and Alex Zhavoronkov[*,†,#,∥]


ACS central science

Search

Enter se

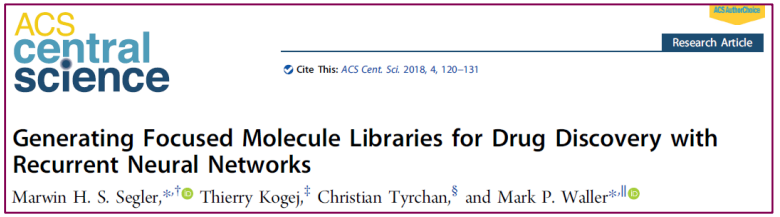
ACS

Browse the Journal | Articles ASAP | Current Issue | Submission & Review | Open Access | Abou

Research Article

Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Rafael Gómez-Bombarelli[†#], Jennifer N. Wei[‡#], David Duvenaud[¶#], José Miguel Hernández-Lobato[§#], Benjamin Sánchez-Lengeling[‡], Dennis Sheberla[‡], Jorge Aguilera-Iparraguirre[†], Timothy D. Hirzel[†], Ryan P. Adams[¶∥], and Alán Aspuru-Guzik[*†⊥]


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[*,∥]


RESEARCH

Molecular De-Novo Design through Deep Reinforcement Learning

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

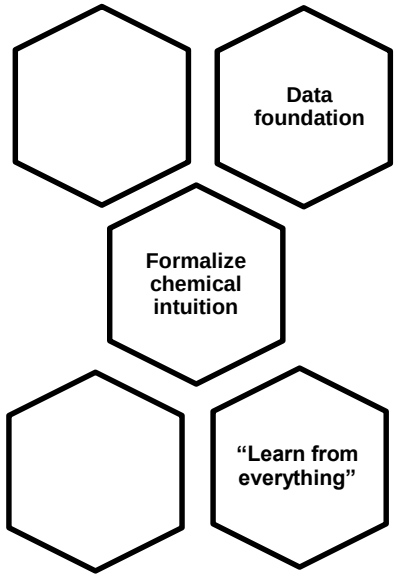

The rise of deep learning in drug discovery

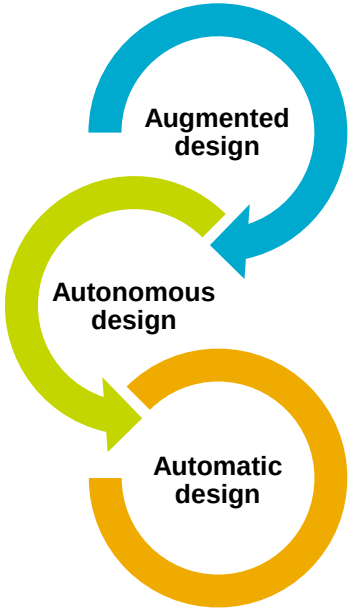
Hongming Chen[1], Ola Engkvist[1], Yinhai Wang[2], Marcus Olivecrona[1] and Thomas Blaschke[1]

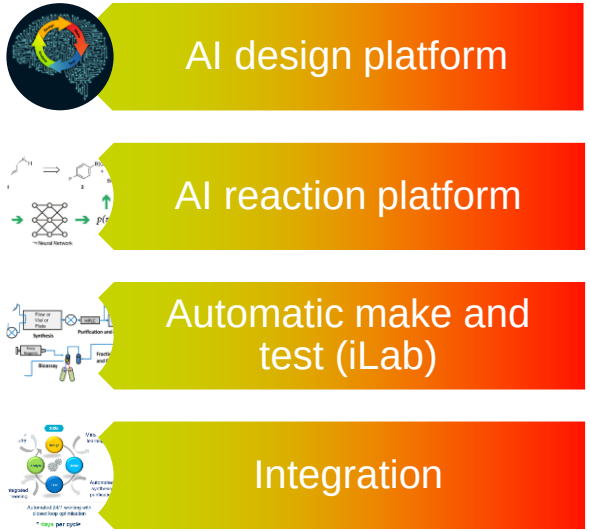
[1]Hit Discovery, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, AstraZeneca R&D Gothenburg, Mölndal 43183, Sweden
[2]Quantitative Biology, Discovery Sciences, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Unit 310, Cambridge Science Park, Milton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

Deep learning @ AstraZeneca: Vision

- Creating an integrate AI platform to impact drug discovery projects







Segler M.H.S. et al. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction, Chemistry, 2017, 23(25), 5966-5971
Segler M.H.S. et al. Planning chemical syntheses with deep neural networks and symbolic AI, Nature, 2018, 555, 604-610

Deep learning @ AZ: De Novo Molecular Augmented Design Platform (REINVENT)

Generation of novel chemical space

AZ+ Pub

Prior model

Reinforcement learning to generate project relevant compounds

Agent model

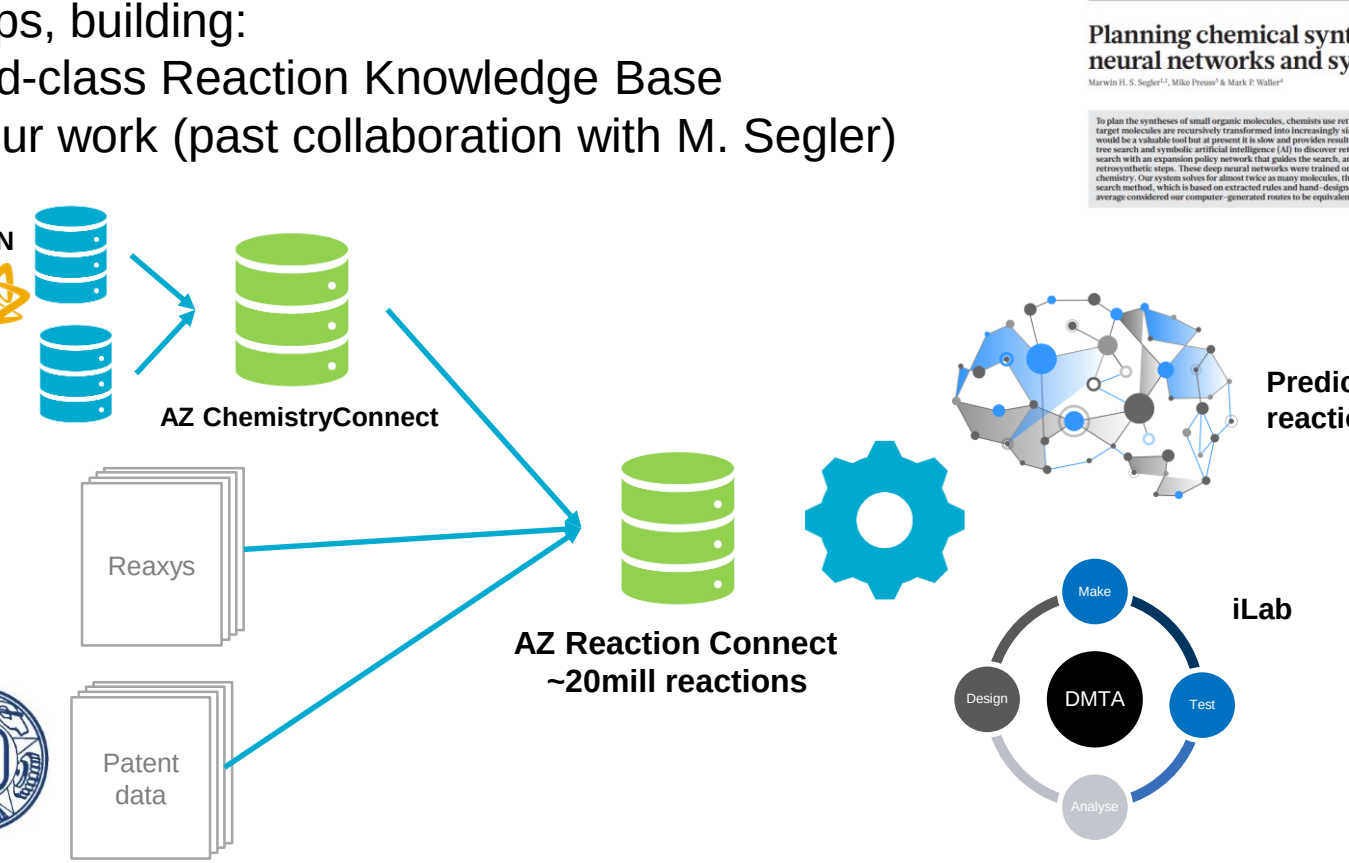
Iterations

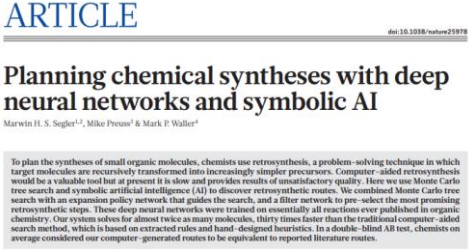
Desirability function
Σ IC50, LogP,
Novelty etc.

Iterations of design and compound synthesis

Deep learning at AstraZeneca: Reaction informatics

- First steps, building:
  - World-class Reaction Knowledge Base
  - On our work (past collaboration with M. Segler)





ARTICLE

doi:10.1038/nature25978

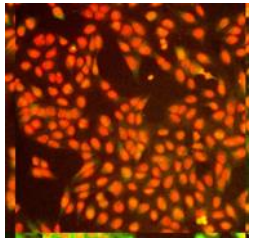
Planning chemical syntheses with deep neural networks and symbolic AI

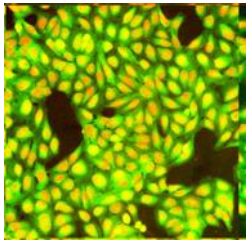
Marwin H. S. Segler[1,2], Mike Preuss[3] & Mark P. Waller[4]

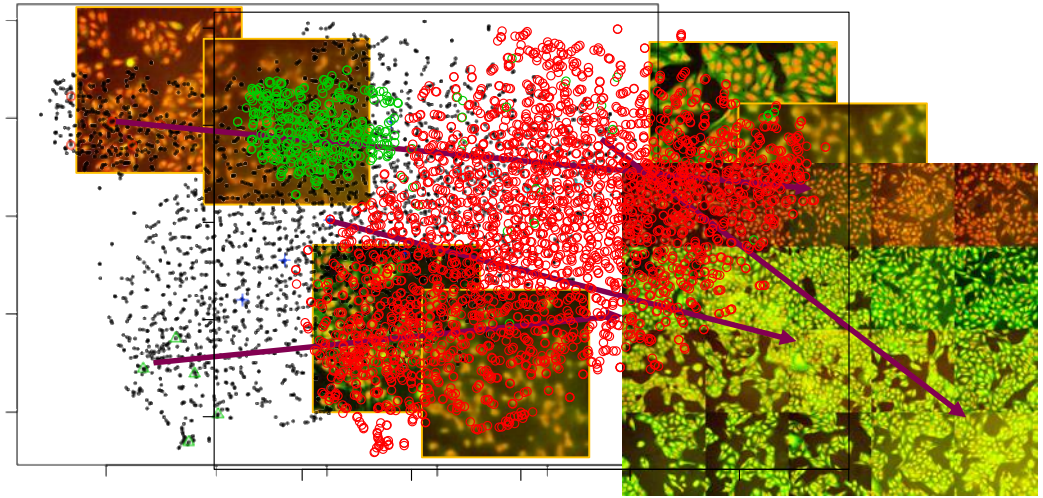
To plan the syntheses of small organic molecules, chemists use retrosynthesis, a problem-solving technique in which target molecules are recursively transformed into increasingly simpler precursors. Computer-aided retrosynthesis would be a valuable tool but at present it is slow and provides results of unsatisfactory quality. Here we use Monte Carlo tree search and symbolic artificial intelligence (AI) to discover retrosynthetic routes. We combined Monte Carlo tree search with an expansion policy network that guides the search, and a filter network to pre-select the most promising retrosynthetic steps. These deep neural networks were trained on essentially all reactions ever published in organic chemistry. Our system solves for almost twice as many molecules, thirty times faster than the traditional computer-aided search method, which is based on extracted rules and hand-designed heuristics. In a double-blind AB test, chemists on average considered our computer-generated routes to be equivalent to reported literature routes.

Becoming FASTER with AI

Through unsupervised learning for hit identification



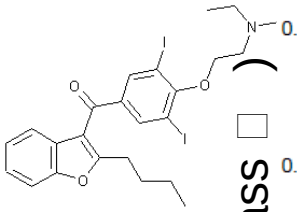




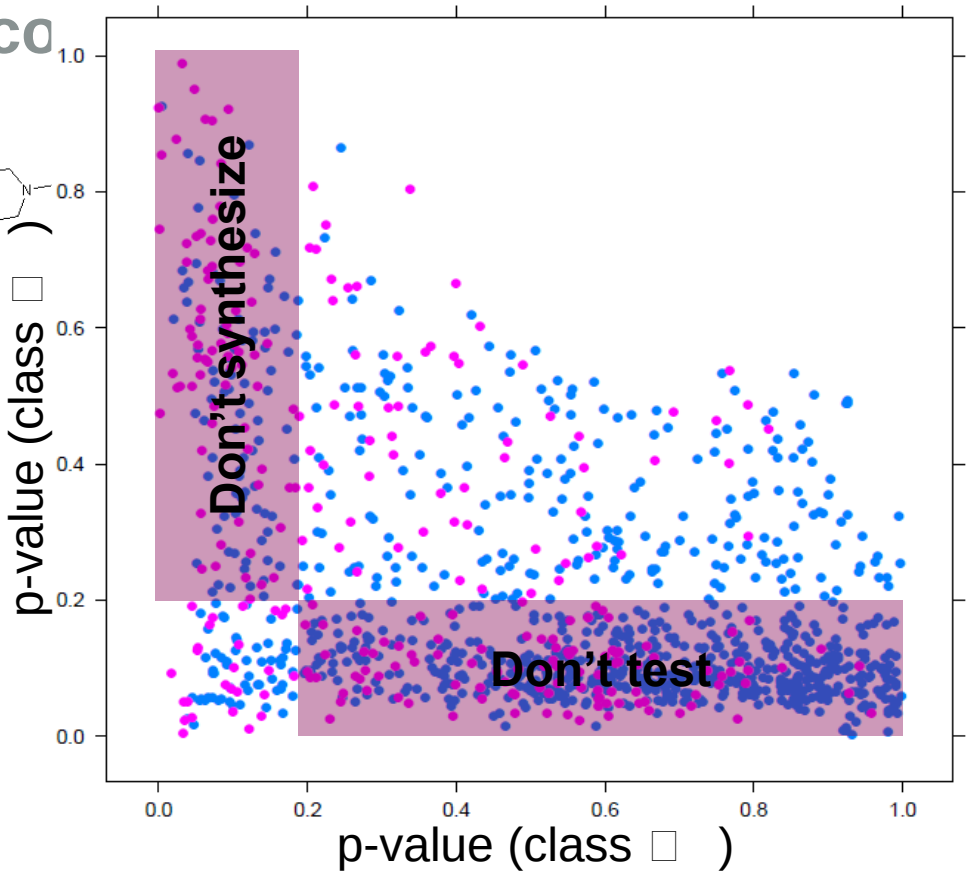
Deep CNN autoencoder → Manifold Learning (t-SNE) → Deep CNN classifier

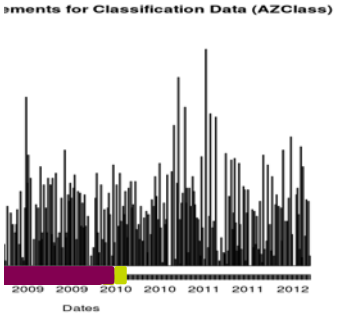
Becoming CHEAPER with ML/AI

By only co eeded



Apply ML model
ian TCP on SVM)





Becoming FASTER and CHEAPER with AI

AI augmented *de novo* molecule design

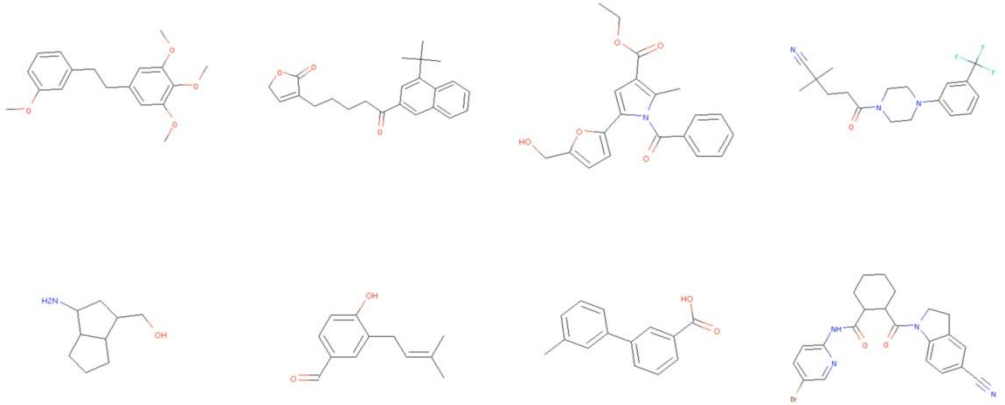
Olivecrona *et al. J Cheminform* (2017) 9:48
DOI 10.1186/s13321-017-0235-x

Journal of Cheminformatics

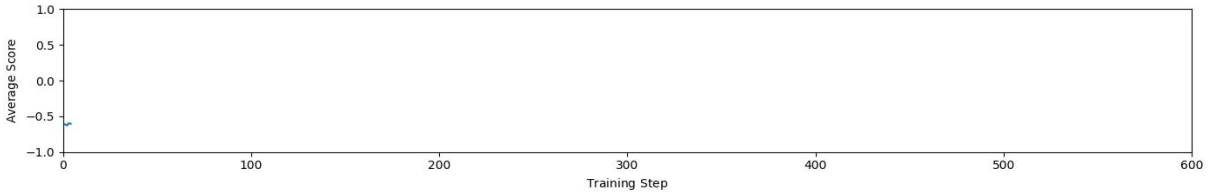
RESEARCH ARTICLE Open Access

Molecular de-novo design through deep reinforcement learning

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

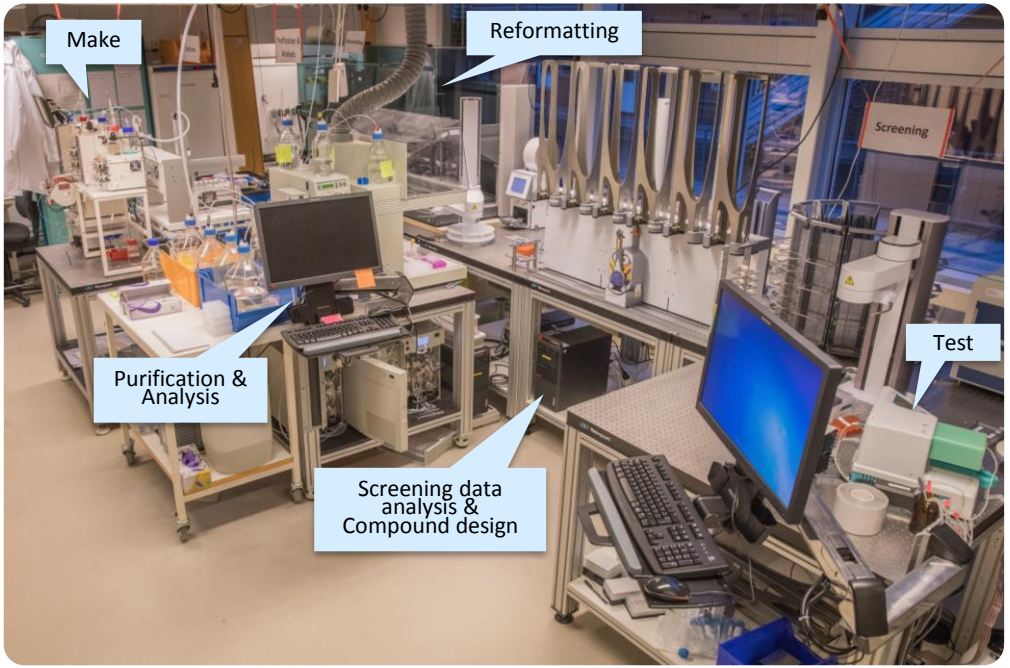


RNN + Reinforcement learning



AZ's first DMTA automation platform

- First protype built during 2017
- All DMTA steps fully integrated
- Suited for 100s of uninterrupted DMTA cycles. ML/AI module is integrated.
- Cycle times of ca. 2h
- Successfully applied in ongoing research project



Conclusions

- Cheminformatics is widely applied in Pharmaceutical industry
- Cheminformatics includes various aspects across different disciplines
- Adoption of machine learning and AI technologies will help Cheminformatics to better fit current and future research needs

Acknowledgement

- Marcus Olivercrona
- Thomas Blaschke
- Isabella Feierberg
- Christophe Grebner
- Erik Malmerberg
- Christian Tyrchan
- Garry Pairaudeau
- Clive Green
- Lars Carlsson
- Peter Varkonyi
- Michael Kossenjans

- EU Fundings

Confidentiality Notice