Deep Learning in Total Synthesis

Outline

1. **Introduction to Deep Learning**
    1. **Artificial Intelligence, Machine Learning and Deep Learning**
    2. **Deep Learning and Neural Networks**
    3. **Representation of Molecules for Machine Learning**
2. **Machine Learning for Synthesis Planning**
    1. **LHASA**
    2. **Chematica**
    3. **3N-MCTS**
3. **Machine Learning for Result Prediction**

    Newhouse, 2021

Not covered:
Evaluation of molecular complexity

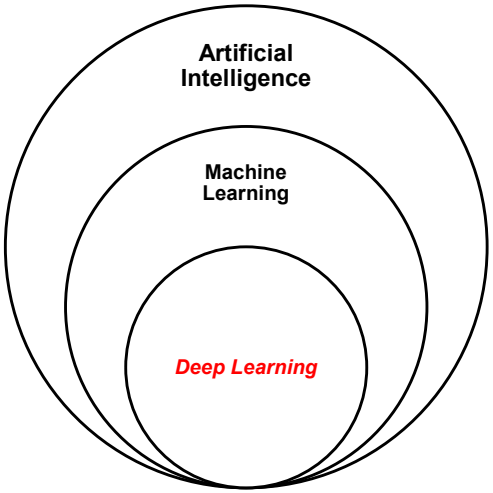
Useful Papers:

Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2016,** *55*, 5904. https://doi.org/https://doi.org/10.1002/anie.201506101

Coley, C. W.; Green, W. H.; Jensen, K. F., *Acc. Chem. Res.* **2018,** *51*, 1281. https://doi.org/10.1021/acs.accounts.8b00087

Molga, K.; Szymkuć, S.; Grzybowski, B. A., *Acc. Chem. Res.* **2021,** *54*, 1094. https://doi.org/10.1021/acs.accounts.0c00714

Artificial Intelligence, Machine Learning and Deep Learning



Artificial intelligence (AI) can be described as the effort to automate intellectual tasks normally performed by humans, which has four main views in literatures: *'acting humanly', 'thinking humanly', 'thinking rationally', 'acting rationally'.*

Machine learning (ML) focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. ML systems are trained, rather than explicitly programmed.

Classical Programming

Machine Learning

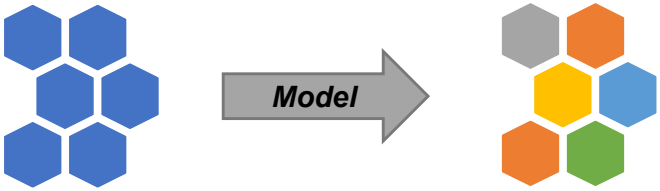
No Free Lunch Theorem (NFL)

All optimization algorithms perform equally well when their performance is averaged across all possible problems. *There is no single best optimization algorithm.*

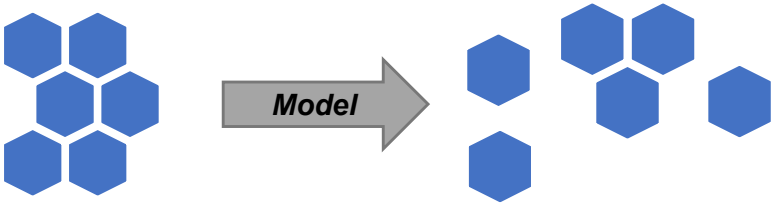
It also implies that there is no single best machine learning algorithm for predictive modeling problems.

(One classification of) types of machine learning problems

Supervised Learning (e.g. Classification):

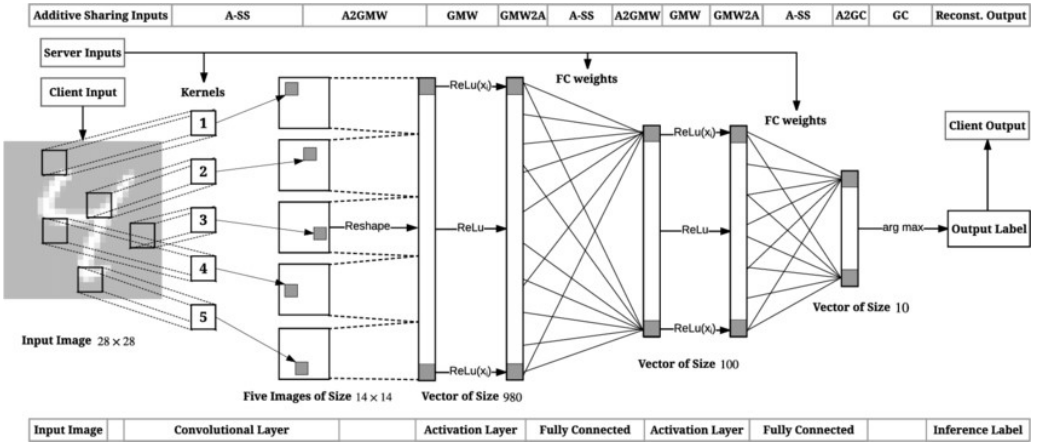


Unsupervised Learning (e.g. Clustering)



Reinforcement Learning (e.g. AlphaGo)

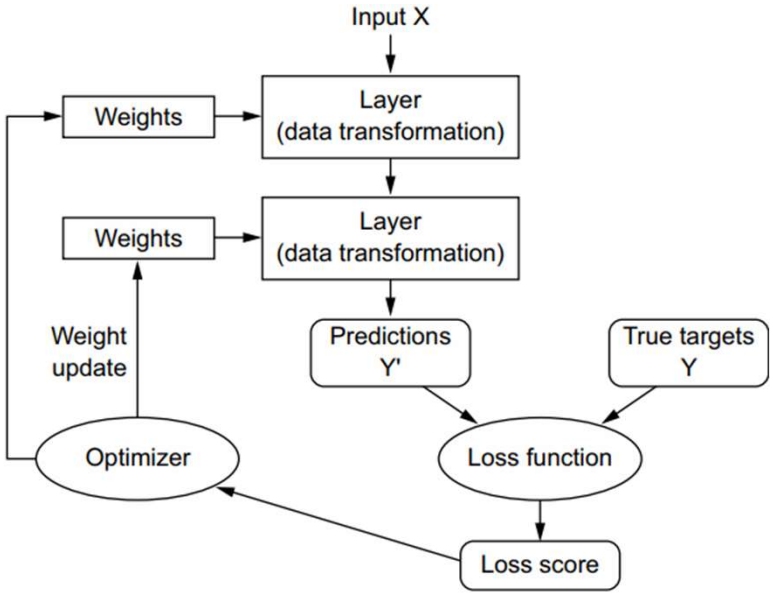
Deep Learning and Neural Networks



Input Layer Hidden Layers Output Layer

Depth

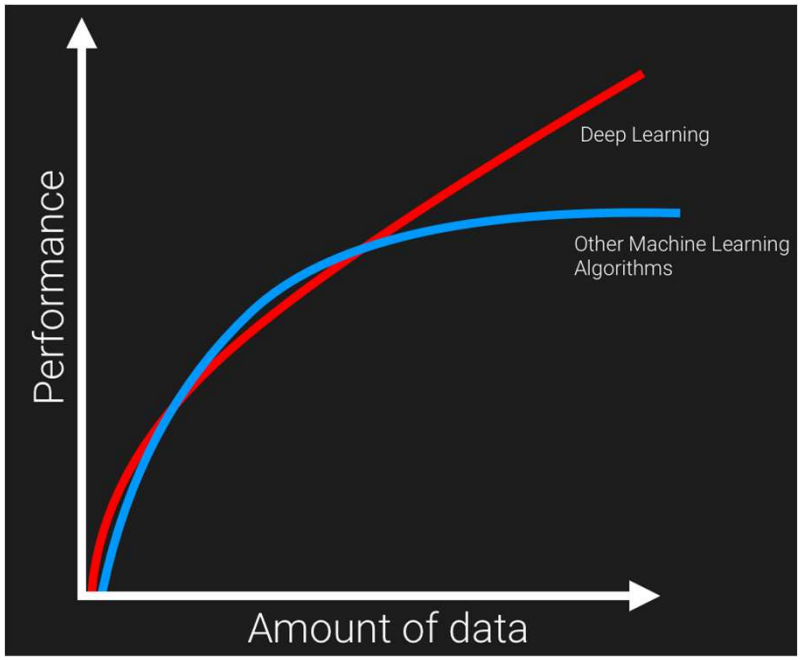
In deep learning, these layered representations are learned via models called ***neural networks***



Chollet, François. *Deep Learning with Python, Second Edition* Manning Publications Co. 2021

Deep Learning in Chemical Synthesis

Why Deep Learning



Deep Learning out perform other techniques if the data size is large.

The requirement of high end infrastructure for deep learning models to be trained in reasonable time can be satisfied.

Deep Learning is good at complex problems such as image classification, natural language processing, and speech recognition. By increasing the number of layers and neurons, neural networks can approximate almost any functions.

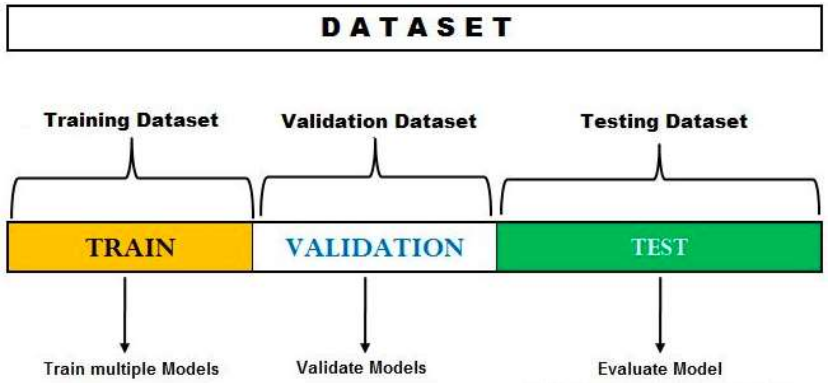
Patil, Ronil. *Is there any need of Deep Learning?* Analytic Vidhya, 2021
https://www.analyticsvidhya.com/blog/2021/05/is-there-any-need-of-deep-learning/

Hornik, K.; Stinchcombe, M.; White, H., *Neural Networks* **1989,** *2*, 359.
https://doi.org/10.1016/0893-6080(89)90020-8

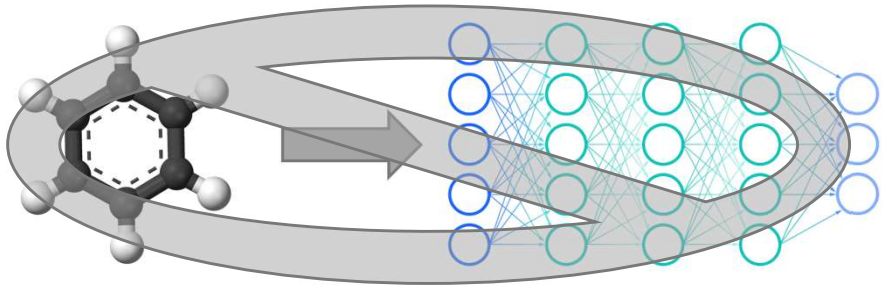
General Workflow of Machine Learning

1. Define the task
2. Develop a model
   - Prepare data
   - Choose the evaluation protocol: ***validations***
   - Find and beat a base line
   - Develop a model that overfits, then tune the model
3. Deploy the model

Validation and Testing



Representation of Molecules for Machine Learning



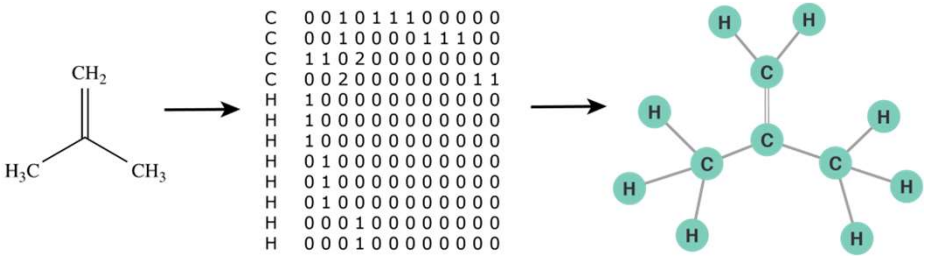
Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O., *Nat. Rev. Chem.* **2021,** *5*, 240.
https://doi.org/10.1038/s41570-021-00260-x

Kumar, Ajitesh. *Hold-out Method for Training Machine Learning Models*, Data Analytics,2022
https://vitalflux.com/hold-out-method-for-training-machine-learning-model/

Deep Learning in Chemical Synthesis

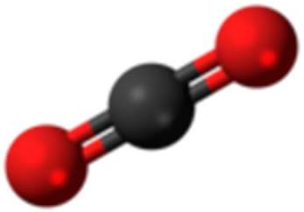
Representation of Molecules for Machine Learning

Molecular Graphs



Text-based Representations: Simplified Molecular-Input Line Entry System (SMILES)

1. Atoms indicated by atomic symbols (aromatic rings → lower case)
2. Inorganic elements are enclosed by brackets (as are formal charges)
3. Bonds represented by -,=,#, and : (single, double, triple, and aromatic); single and aromatic bonds are conventionally omitted
4. Branches are specified by enclosures in parentheses
5. Cyclic structures are indicated by breaking one bond in each ring and designating the point of opening/closure with a digit



O=C=O C(=O)=O

SMILES is not canonical

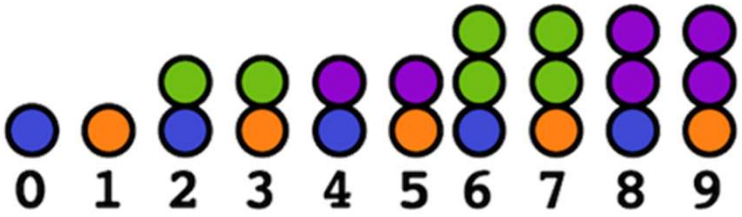
SMILES Arbitrary Target Specification (SMARTS) for Chemical Patterns

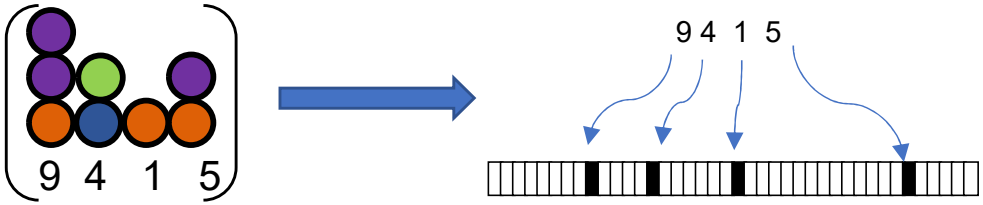
E.g. [c,n;H1] either aromatic carbon or nitrogen and exactly one hydrogen

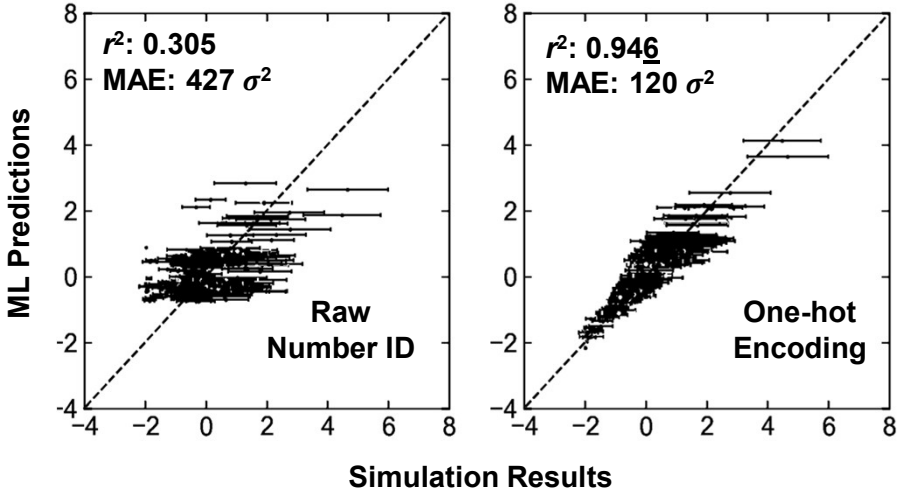
Weininger, D., *J. Chem. Inf. Comput. Sci.* **1988,** *28*, 31. https://doi.org/10.1021/ci00057a005

Weininger, D.; Weininger, A.; Weininger, J. L., *J. Chem. Inf. Comput. Sci.***1989,** *29*, 97. https://doi.org/10.1021/ci00062a008

Tokenization and One-Hot Encoding







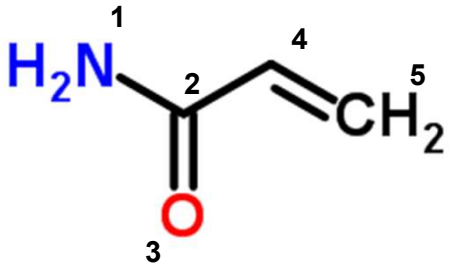
Webb, M. A.; Jackson, N. E.; Gil, P. S.; de Pablo, J. J., *Sci. Adv.* **2020,** *6*, eabc6216. https://doi.org/doi:10.1126/sciadv.abc6216

Representation of Molecules for Machine Learning

Extended Connectivity Fingerprints (ECFP)

Basic Algorithm

1. Assign each atom an identifier
2. Iteratively update identifier based on neighboring atoms
3. Remove/count duplicates
4. Fold identifiers into an *N*-bit vector



Number of nearest-neighbor heavy atoms
Atomic number
Atomic mass
Atomic charge
Number of attached hydrogens
Whether atom is part of a ring

Atom 2: hash([3, 8, 16, 0, 0, 0]) -> a unique integer identifier
E.g.
1: -9097421984
2: 5420398560
3: 1128765390
4: -0979365278
5: 2897025579

Neighbor information of atom 2:
[(1, -9097421984), (1, 1128765390), (1, - 0979365278), (2, 2897025579)]

Then HASH it ***again***, we get:

2: 12674839301029 (neighboring atoms information included)

Do the operations above iteratively, then initialize a zeros vector of a specific length and divide each identifier by the vector length and obtain the remainder. Use the remainder to set the fingerprint element to 1 or 0.

Morgan, H. L., *J. Chem. Doc.* **1965,** *5*, 107. https://doi.org/10.1021/c160017a018
1Rogers, D.; Hahn, M., *J. Chem. Inf. Model.***2010,** *50*, 742. https://doi.org/10.1021/ci100050t

Advantages: easy to generate, analogous to functional groups, flexible, robust
Disadvantages: no 3D information, not scalable to massive chemical spaces

Tanimoto similarity: is a metric for computing the inner product of two molecular fingerprint vectors. It is by far the most common similarity metric used.

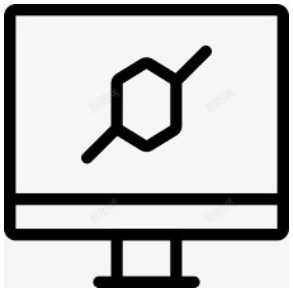
$$S_{A,B} = \frac{x_A \cdot x_B}{x_A \cdot x_A + x_B \cdot x_B - x_A \cdot x_B}$$

Mathematically, $S_{A,B}$ is the ratio of the intersection of A and B over the union of A and B.

Descriptor (Feature) Vectors

E.g. Some physiochemical features:

No. of X structure,…
log P, ASA, shape parameters, …
dipole moment, polarizability, …
electronic energy, Δhf, IP, εgap, …
simulation-derived quantities
experimental measurements



Open-Source Cheminformatics and Machine Learning
https://www.rdkit.org/

Deep Learning in Chemical Synthesis

Machine Learning for Synthesis Planning

Logic and Heuristics Applied to Synthetic Analysis (Corey, 1985)

Expert system, aka "**LHASA**", from OCSS (1969)
Database: ~1100 reactions (in 1985) -> 2100+ reactions
Step-by-step, iterative analysis with chemist

Analyzes and plans synthesis route based on logics

1. Transform-based strategies

2. Structure-goal strategies
Recognition of potential starting materials and building block.

3. Topological strategies
Identification of disconnections that can lead to major molecular simplification.

4. Stereochemical strategies
Analysis of the stereochemistry of substrates and application of stereospecific transformations

5. Functional group-oriented strategies
Reaction cascade, FGI, protection/deprotection

An implementation of *The Logic of Chemical Synthesis*

Corey, E. J.; Long, A. K.; Rubenstein, S. D., *Science* **1985,** *228*, 408. https://doi.org/doi:10.1126/science.3838594

Many other Computer-Assisted Synthesis Planning Attempts

SECS (Wipke, 1976);
SYNCHEM (Gelernter, 1977);
SYNLMA (Johnson, 1989);
SYNGEN (Hendrickson, 1989);
CHIRON (Hanessian, 1990);
IGOR (Ugi, 1993);
WODCA (Gasteiger, 1995)

Too simplified rule sets
Incompatible synthetic routes
Limited computing power

Todd, M. H., *Chem. Soc. Rev.* **2005,** *34*, 247. https://doi.org/10.1039/B104620A

How does synthetic design differ from other problems, like playing chess or Rubik's cube?

1. Much larger number of moves/rules
2. Applicability of rules is very context-dependent
3. Current "position" cannot be used to systematically plan future moves
4. Revertive search over the transformation space for global optima

Chematica / Synthia (Grzybowski, 2012 / Merck, 2017 – Present)

Expert system -> Hybrid expert–NN system
Database: >100,000 reactions (in 2020)
Close-source, commercial

Kowalik, M.; Gothard, C. M.; Drews, A. M.; Gothard, N. A.; Weckiewicz, A.; Fuller, P. E.; Grzybowski, B. A.; Bishop, K. J. M., *Angew. Chem. Int. Ed.* **2012,** *51*, 7928. https://doi.org/10.1002/anie.201202209

Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2016,** *55*, 5904. https://doi.org/10.1002/anie.201506101

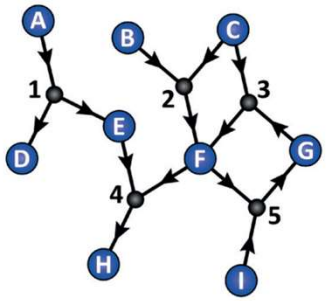
Deep Learning in Chemical Synthesis

Chematica / Synthia (Grzybowski, 2012 / Merck, 2017 – Present)

Representation of Transformations

Reactions

1) A → D + E
2) B + C → F
3) C + G → F
4) E + F → H
5) F + I → G



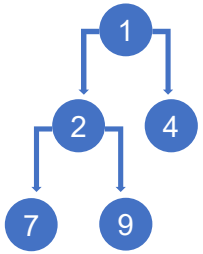
Network of Organic Chemistry (NOC)

Petri Network Representation

Search Algorithm

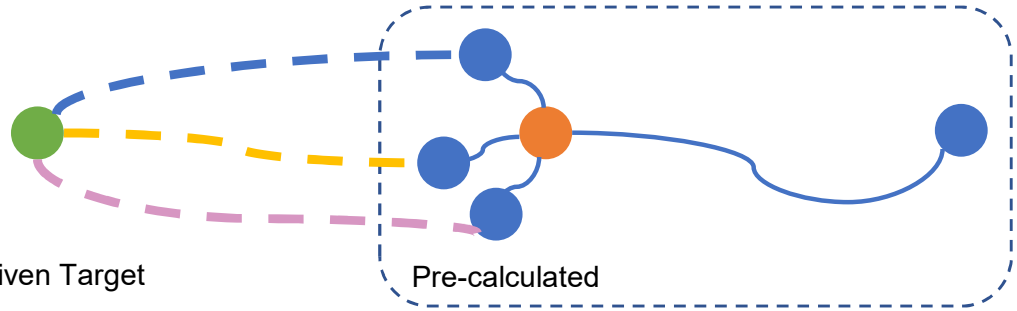
Breadth-first search (BFS) and pre-calculated intermediates

Metrics: yield, cost, popularity



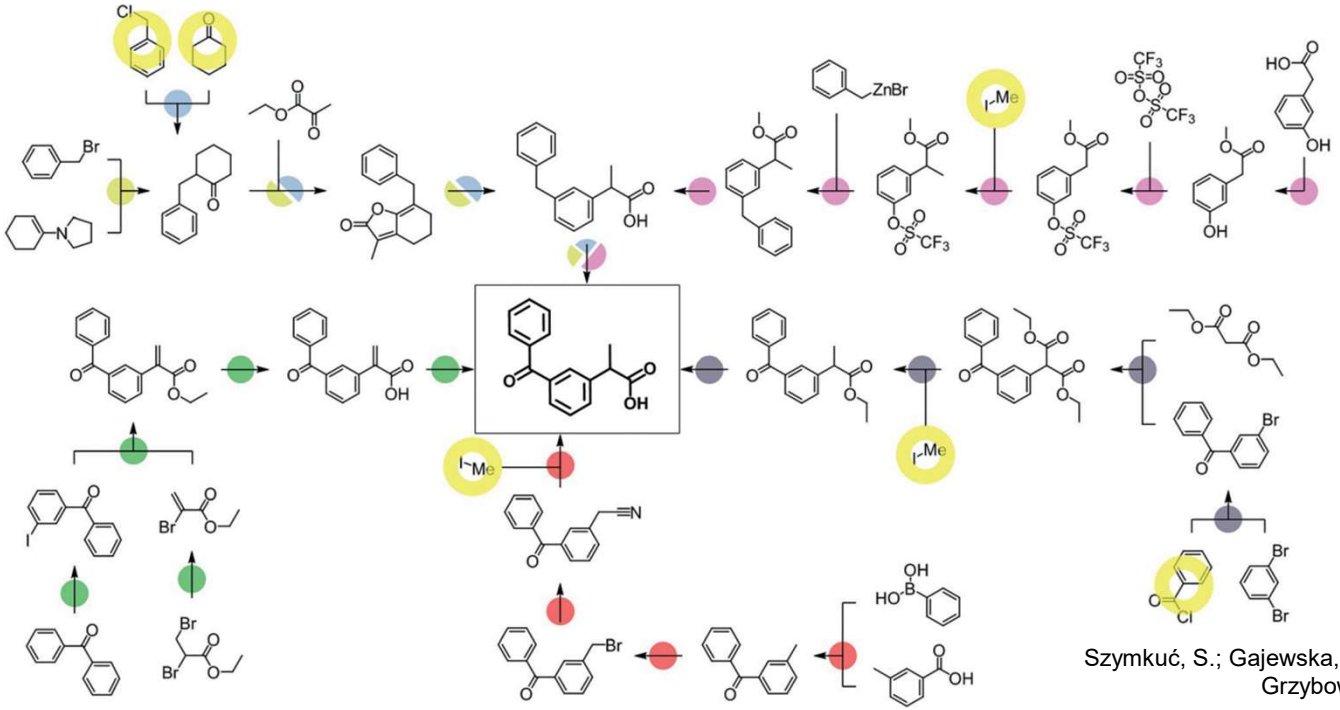
An Example of BFS:
Search Order : 1->2->4->7->9

E.g.: Popularity-defined optimality will give routes based on robust chemistries



A Given Target

Pre-calculated



Synthesis Planning under Restrictions

Red: Only one product
Green: no regulated substances
Pink: transformations after 1998
Blue: high labor-to-chemical cost ratio
Greenish-yellow: blue + no regulated substances
Grey-brownish: excluding few substances (benzaldehyde, 3-chlorobenzophenone and 3 methylbenzophenone)

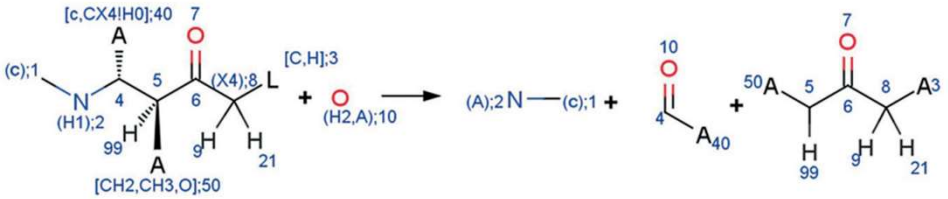
Yellow: regulated substances

Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2016,** *55*, 5904.
https://doi.org/10.1002/anie.201506101

Deep Learning in Chemical Synthesis

Chematica / Synthia (Grzybowski, 2012 / Merck, 2017 – Present)

Example: Proline-catalyzed Mannich reaction as coded



rxn_id: 8382,
name: "Proline-catalyzed Mannich Reaction",
reaction_SMARTS:[c:1][NH:2][C@H:4]([c,CX4!H0:40])[C@:5]([#1:99])([CH2,CH3,O:50])[C:6](=[O:7])[CX4:8]([#1:9])([#1:21])[#6,#1:3].[OH2:10]>>[c:1][N:2].[*:40][C:4]=[O:10].[*:50][C:5]([#1:99])[C:6](=[O:7])[C:8]([#1:9])([#1:21])[*:3]"
products:["[c][NH][C@H]([c,CX4!H0])[C@]([#1])([CH2,CH3,O])[C](=[O])[CX4]([#1])([#1])[#6,#1]", "[OH2]"]
groups to protect: ["[#6][CH]=O", "[CX4,c][NH2]", "[CX4,c][NH][CX4,c]", "[#6]C([#6])=O"]
protection_conditions_code: ["NNB1", "EA12"]
incompatible_groups: ["[#6]O[OH]", "c[N+]#[N]", "[NX2]=[NX2]", "[#6]OO[#6]", "[#6]C(=[O])OC(=[O])[#6]", "[#6]N=C=[O,S]", "[#6][N+]#[C-]", "[#6]C(=O)[Cl,Br,I]", "[CX3]=[NX2][*!O]", "[#6]C(=[SX1])[#6]", "[#6][CH]=[SX1]", "[#6][SX3](=O)[OH]", "[CX4]1[O,N][CX4]1", "[#6]=[N+]=[N-]","[CX3]=[NX2][O]"]
typical reaction conditions: "(S)-proline. Solvent, e.g., DMSO",
general references: "DOI: 10.1021/ja001923x or DOI: 10.1021/cr0684016 or DOI: 10.1021/ja0174231 or DOI: 10.1016/S0040-4020(02)00516-1"

Development of Hybrid Expert-NN System

Instytut Chemii Organicznej (**ICHO / ICHO+**); ***Semi-supervised learning-like***

Data source: reported reactions from journals and patents

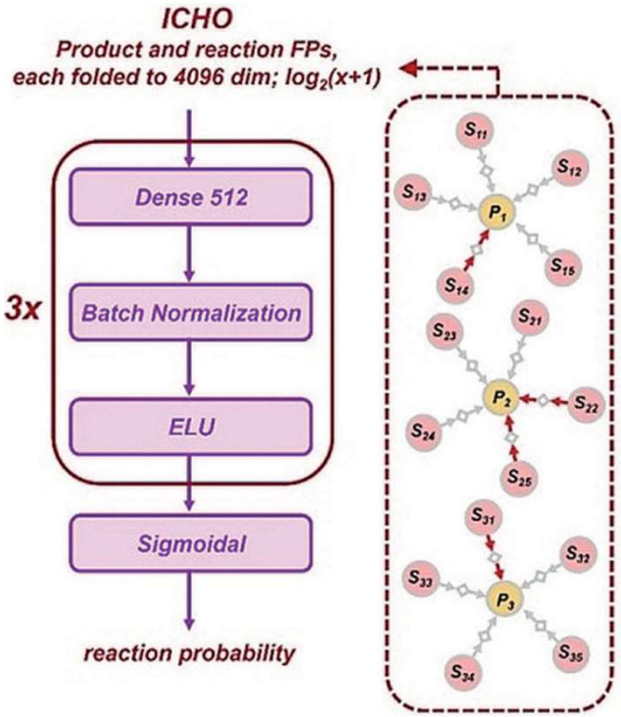
Data Filtration:
1. No protection/deprotection reactions
2. Matches at least one of Chematica's 75000 expert-coded reaction rules

Literature collection: 85 million conflict-free and high-chemical quality reactions leading to our 1.4 million products

Badowski, T.; Gajewska, E. P.; Molga, K.; Grzybowski, B. A.,
Angew. Chem. Int. Ed. **2020,** *59*, 725. https://doi.org/10.1002/anie.201912083

Input X:
(ICHO) concatenated Morgan fingerprints of a reaction and of its product (ICHO+) chemically intuitive reaction characteristics (e.g. num. of ring construction/destruction)

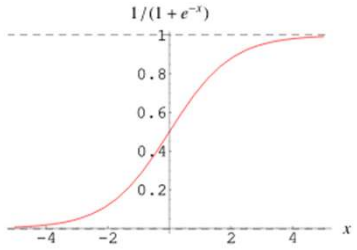
Input Y: 1 if a given conflict-free, expert reaction producing a given target is also present in our ***literature collection***, and 0 otherwise.



Red arrows (1): reported in the collection and following Chematica's reaction rules

Grey arrows (0): not reported in the collection but following Chematica's other reaction rules

Sigmoidal Function:



1. How many times an expert reaction rule with a given reaction fingerprint occurred in the literature collection
2. How many times it matched product molecules from this collection

The ratio of the answers of this two questions is the ***Synthetic Popularity***:

E.g. If an expert reaction fitted ten product molecules and this reaction type was observed in published reactions also ten times, the NN can learn that this reaction rule should be applied whenever it fits the product of interest.

Deep Learning in Chemical Synthesis

Chematica / Synthia (Grzybowski, 2012 / Merck, 2017 – Present)

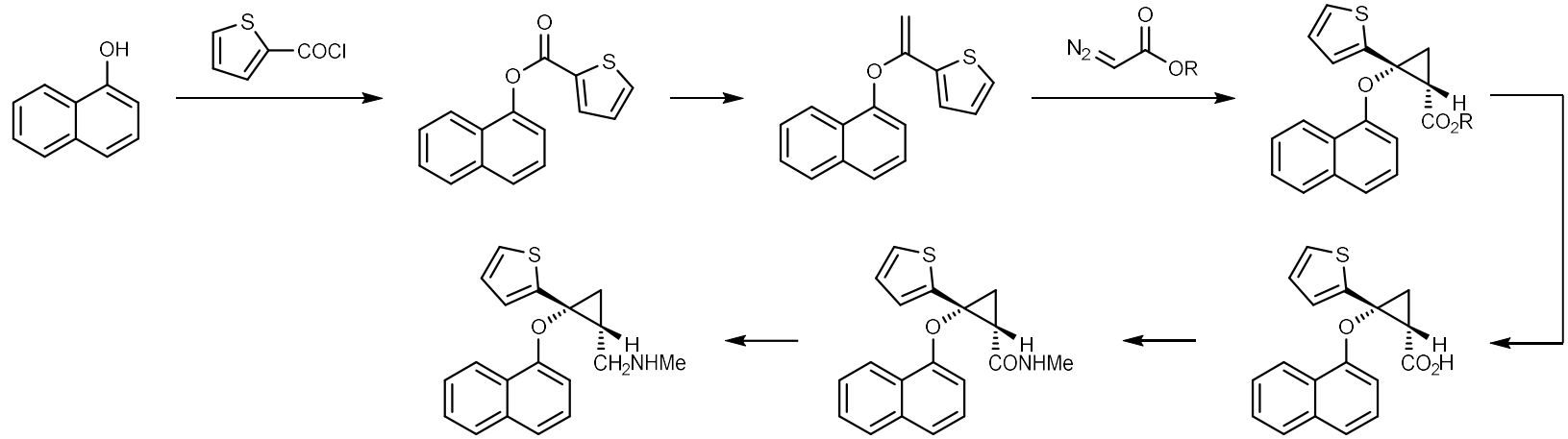
Development of Hybrid Expert-NN System

It can predict ***non-zero probabilities*** even for reactions of types ***not seen*** in the literature collection during training.

Performance Comparisons with other Models

SW2+: model by Segler and Waller, 2018

SMALLER: model that prefer shorter SMILES strings of starting materials. Sum((length of SMILES)^3 for all starting materials) in this case



	ICHO+	SW2+	SMALLER
Step 1	1	1	3
Step 2	1	1	24
Step 3	2	31	1
Step 4	2	1	10
Step 5	1	1	2
Step 6	8	1	26
Average	2.5	6	11

Values in the Table:

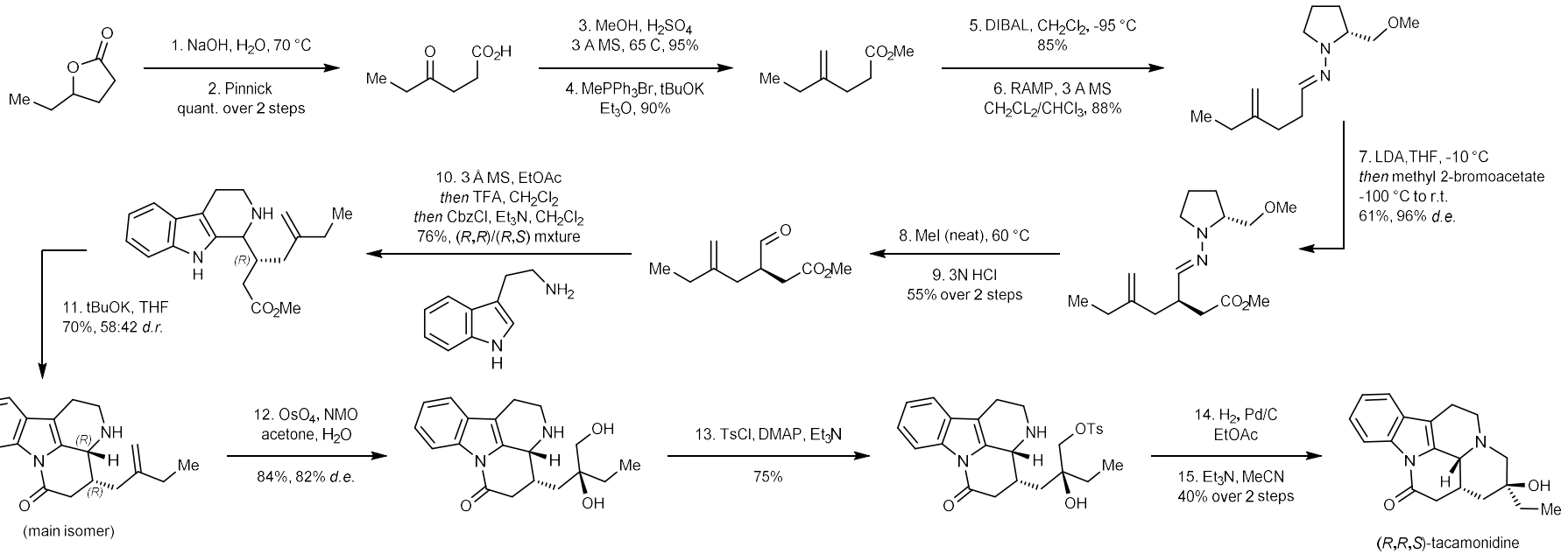
The experimental condition is the <value>-best prediction

Badowski, T.; Gajewska, E. P.; Molga, K.; Grzybowski, B. A., *Angew. Chem. Int. Ed.* **2020,** *59*, 725. https://doi.org/10.1002/anie.201912083

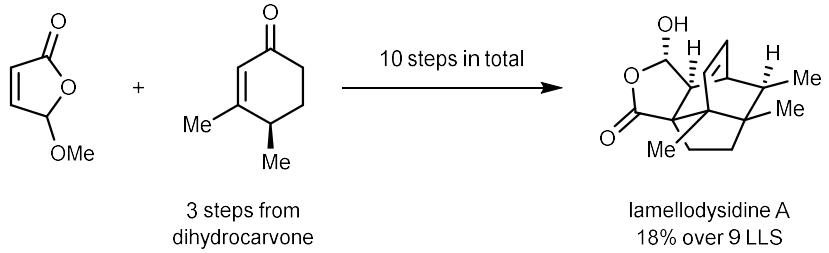
Deep Learning in Chemical Synthesis

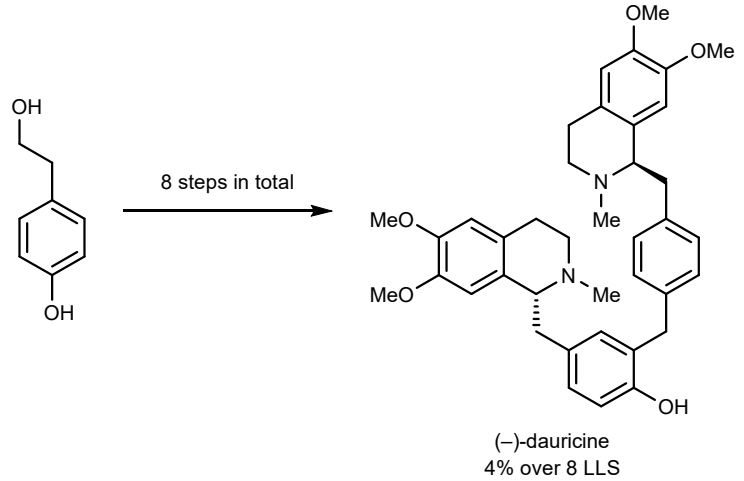
Chematica / Synthia (Grzybowski, 2012 / Merck, 2017 – Present)

Chematica for Total Synthesis



Another Two Verified Routes





Mikulak-Klucznik, B.; Gołębiowska, P.; Bayly, A. A.; Popik, O.; Klucznik, T.; Szymkuć, S.; Gajewska, E. P.; Dittwald, P.; Staszewska-Krajewska, O.; Beker, W.; Badowski, T.; Scheidt, K. A.; Molga, K.; Mlynarski, J.; Mrksich, M.; Grzybowski, B. A., *Nature* **2020,** *588*, 83. https://doi.org/10.1038/s41586-020-2855-y

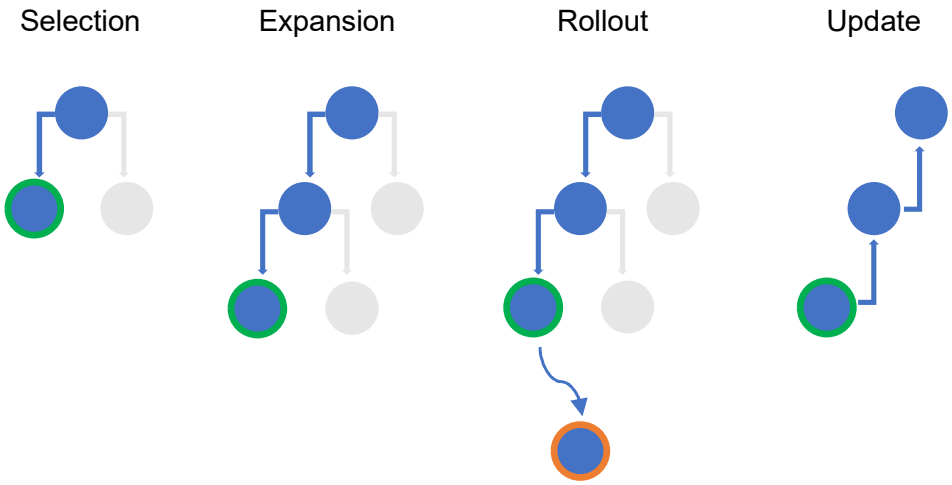
Deep Learning in Chemical Synthesis

3N-MCTS (Waller, 2018)

Deep Neural Network System & Monte Carlo tree search

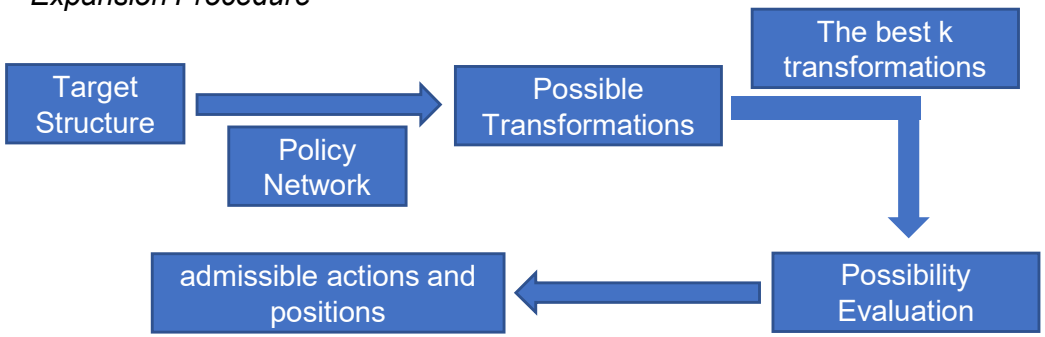
Database: 12.4 million single-step reactions from Reaxys

Monte Carlo Tree Search (MCTS)



Selection: choose the most promising position
Expansion: add new nodes by *expansion procedure* (N1 and N2)
Rollout : pick & evaluate the performance of new nodes (N3, MCTS)
Update: incorporate the position evaluation into the route

Expansion Procedure



Segler, M. H. S.; Preuss, M.; Waller, M. P., *Nature* **2018,** *555*, 604.
https://doi.org/10.1038/nature25978

Development of Expansion and Rollout Policy Networks

More data for find the candidates and less data for estimation

Expansion Policy Network: predict possible disconnections

Expansion rules: only the reaction centers was extracted. Rules occurring at least three times were kept. (301,671 rules)

Input X: structure of products
Input Y: structure of starting materials

Training set: reactions before 2015; Test set: reactions after 2015

Prediction Accuracy:
Top 1: 31%, Top 10: 63.3%, Top 50: 72.5% (max searching number)

Rollout Policy Network: evaluation

Rollout rules: contain the atoms and bonds that changed in the reaction centers and the first-degree neighboring atoms. Only rules that occurred at least 50 times in reactions published before 2015 were kept. (17,134 rules)

Development of Filter Network

Data Augmentation (100 million negative reaction generated):
1. Generated hypothetical products as negative results
2. Shuffling product-reaction pair

False positive: 1.5%, false negative 14%

Model Performance Evaluation

Double-blind AB-test, with 45 graduate-level organic chemists:
1. Chemists did not significantly prefer the literature route over our program's route
2. Chemists significantly preferred routes found by 3N-MCTS over routes generated by heuristic BFS without a policy network and an in-scope filter.

Deep Learning in Chemical Synthesis

3N-MCTS (Waller, 2018)

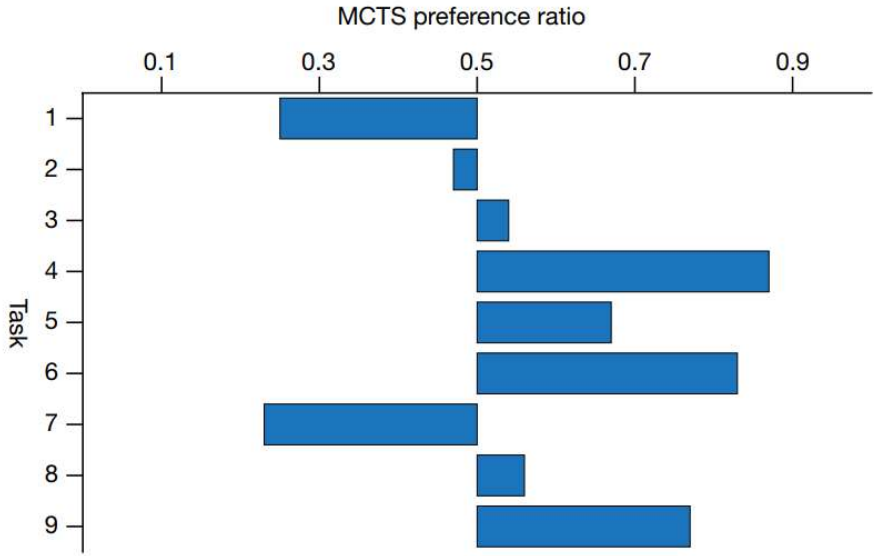
Model Performance Evaluation

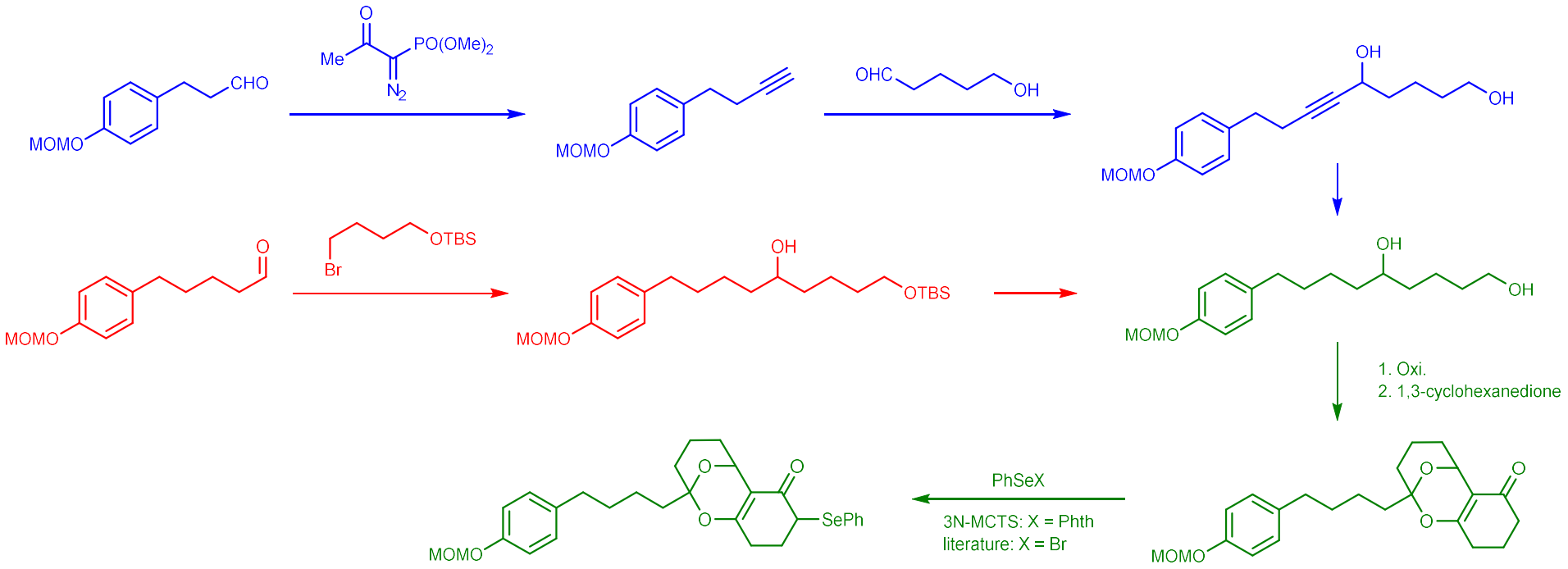
3N-MCTS versus literature routes

Right: preference ratio of 3N-MCTS

Below: comparison between proposed and reported routes

Blue: Route proposed by 3N-MCTS
Red: Route in literature
Green: Overlap of proposed route and literature



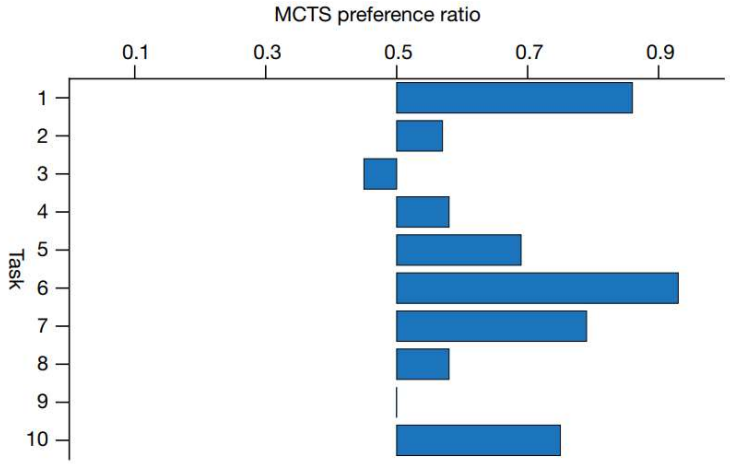


Segler, M. H. S.; Preuss, M.; Waller, M. P., *Nature* **2018,** *555*, 604.
https://doi.org/10.1038/nature25978

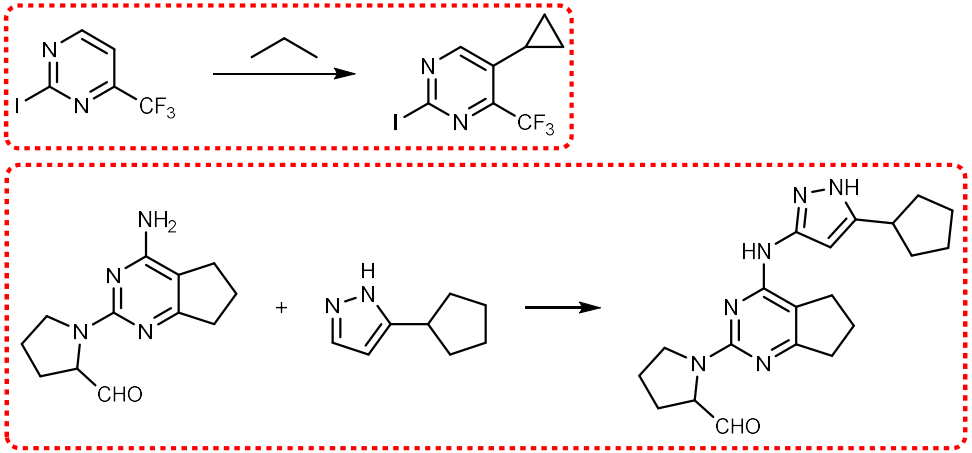
3N-MCTS (Waller, 2018)

Model Performance Evaluation

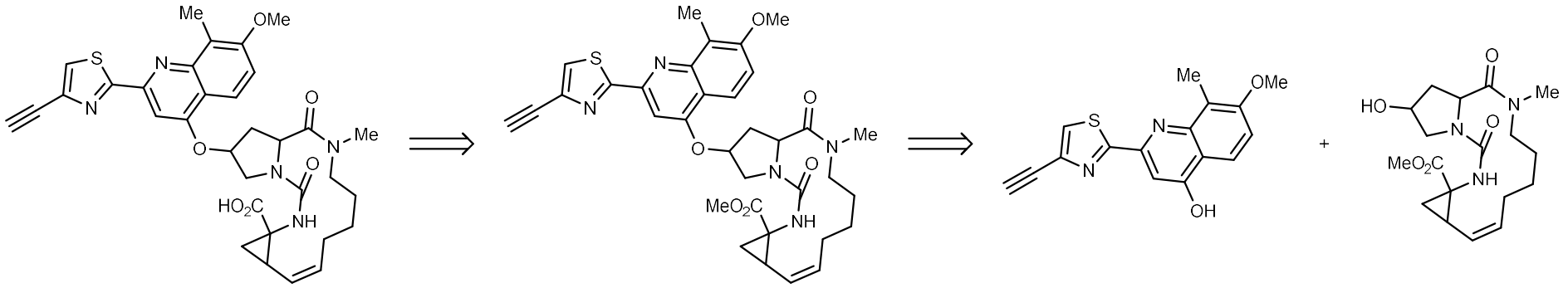
3N-MCTS versus heuristic BFS routes



Unreasonable Reactions proposed by BFS



Another retro synthesis proposed by 3N-MCTS

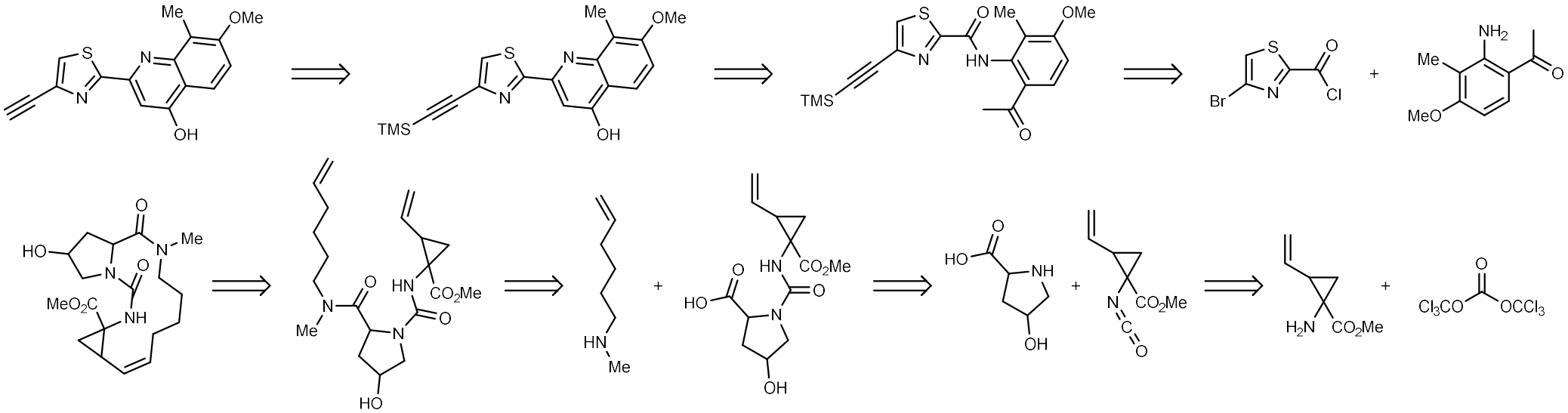


Segler, M. H. S.; Preuss, M.; Waller, M. P., *Nature* **2018,** *555*, 604. https://doi.org/10.1038/nature25978

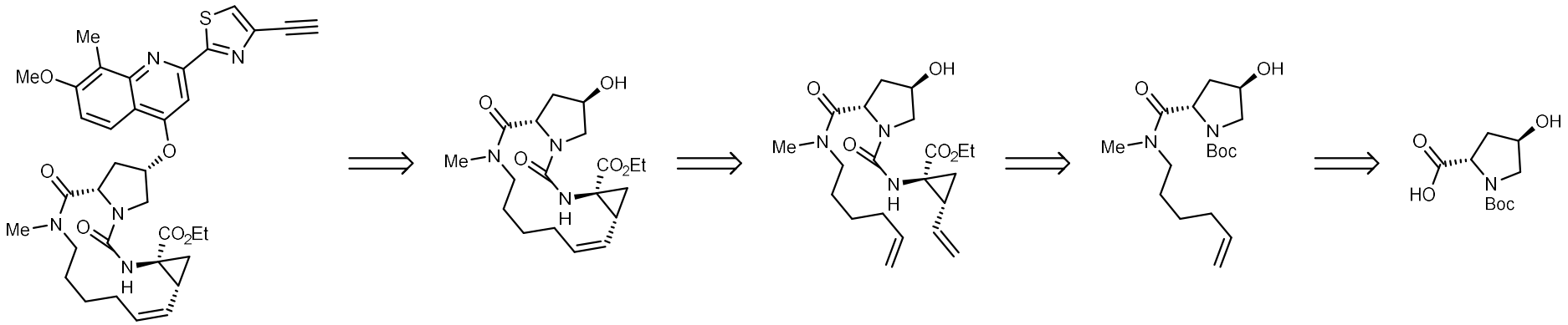
3N-MCTS (Waller, 2018)

Model Performance Evaluation

Another retro synthesis proposed by 3N-MCTS



Medicinal Chemistry Route of IDX320 Analogs (Idenix / Merck, 2015)



Parsy, C. C.; Alexandre, F.-R.; Bidau, V.; Bonnaterre, F.; Brandt, G.; Caillet, C.; Cappelle, S.; Chaves, D.; Convard, T.; Derock, M.; Gloux, D.; Griffon, Y.; Lallos, L. B.; Leroy, F.; Liuzzi, M.; Loi, A.-G.; Moulat, L.; Chiara, M.; Rahali, H.; Roques, V.; Rosinovsky, E.; Savin, S.; Seifer, M.; Standring, D.; Surleraux, D., *Bioorg. Med. Chem. Lett.* **2015,** *25*, 5427. https://doi.org/10.1016/j.bmcl.2015.09.009

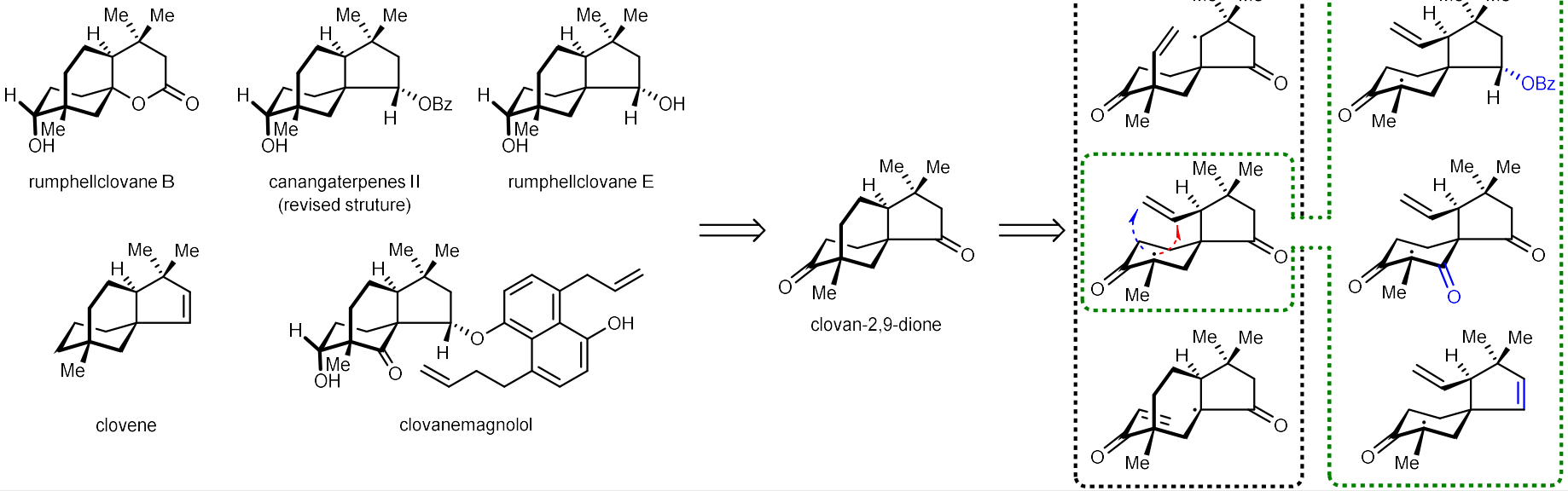
Segler, M. H. S.; Preuss, M.; Waller, M. P., *Nature* **2018,** *555*, 604. https://doi.org/10.1038/nature25978

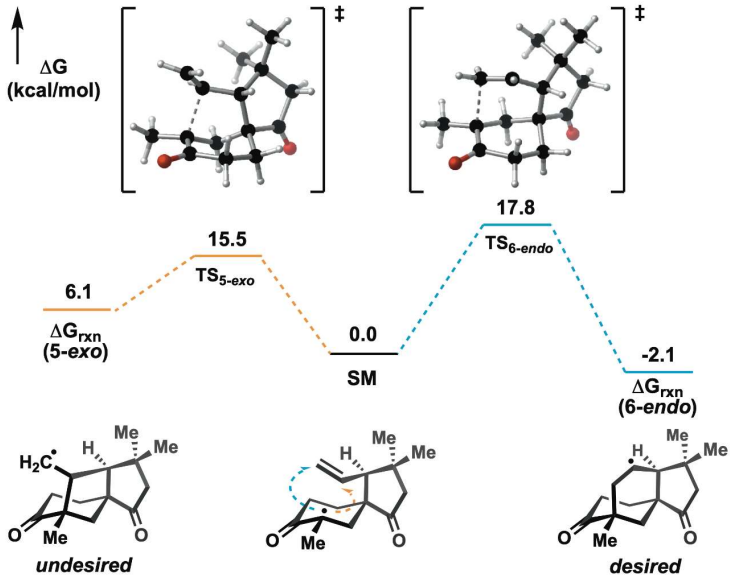
Summary: AIs for Synthesis Planning

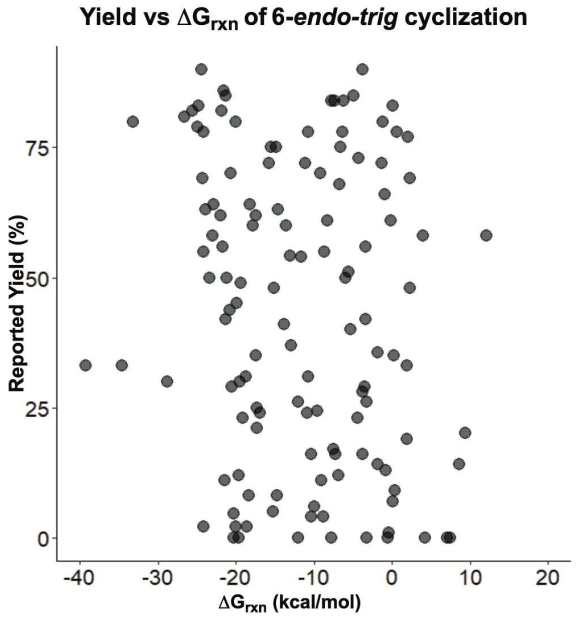
Model	LHASA	Chematica	3N-MCTS
Architecture	Expert System	Hybrid Expert-NN	Neural Network
Database	2k-Scale rules	100k-Scale encoded rules + Extractions from Reaxys	Extractions from Reaxys
Search Algorithm	BFS	BFS	MCTS
Work Flow	Step-by-step Interactive	Metric-dependent Automatic	Automatic
Scoring	Chemist	Score Function	Score Function
Data Filtration	-	Semi-supervised Learning	In-scope Filter NN
Stereochemistry?	Yes	Yes	Not Quantitatively
Natural Products?	No (Over Simplification)	Yes	No (Sparsity)
Turing Test	-	Passed	-
Other Limitation	-	Expensive	No Condition Prediction

Neural Network for the Prediction of Key Step

Newhouse, 2021







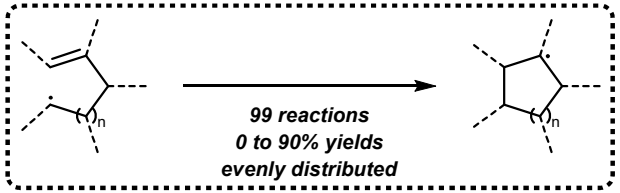
Library Preparation → Feature Extraction → Modeling → Experimental Verification

Newhouse, T.; Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. *ChemRxiv* **2021.**
https://doi.org/10.26434/chemrxiv-2021-41d5z

Deep Learning in Chemical Synthesis

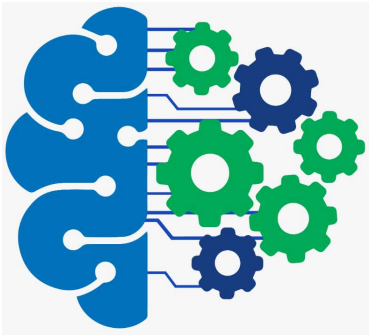
Newhouse, 2021

sp3-centered radicals
intramolecular cyclization; onto a pendant olefin

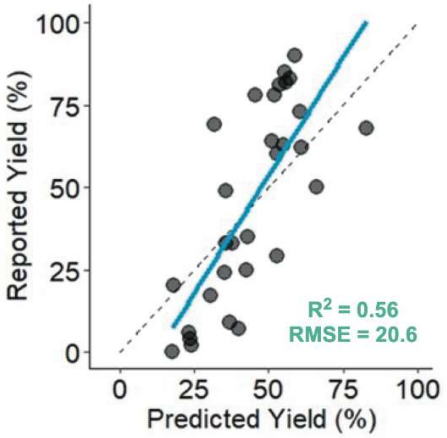


Reaxys®

DFT Calculation for
physical descriptors

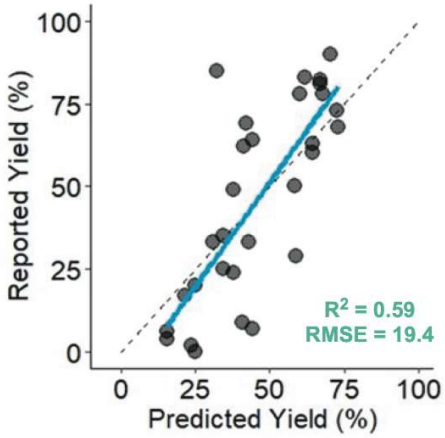


Supervised Models
Input X: descriptors of intermediates before and after cyclization
Input Y: yields



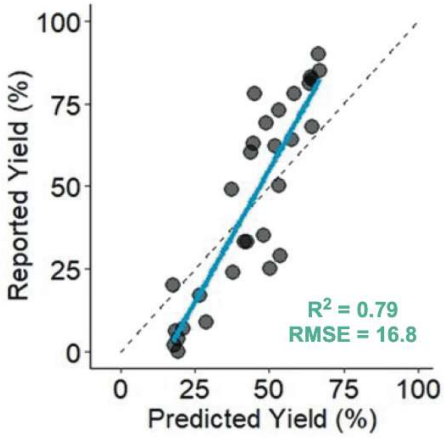
SIMPLS

Statistically Inspired Modification of the Partial Least Squares



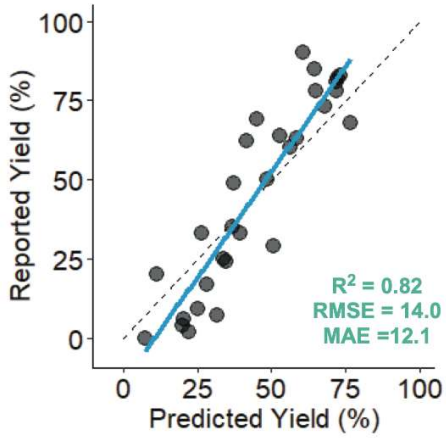
kNN

k-Nearest Neighbors



RF

Random Forest



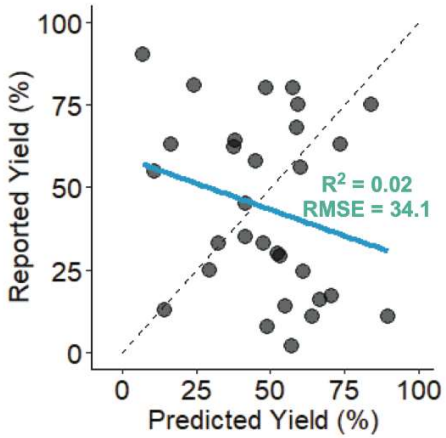
NNET

neural network

Newhouse, T.; Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. *ChemRxiv* **2021.**
https://doi.org/10.26434/chemrxiv-2021-41d5z

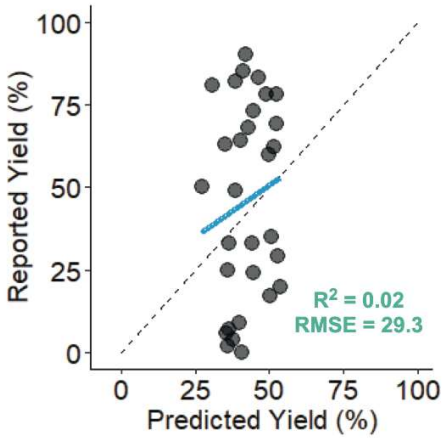
Deep Learning in Chemical Synthesis

Newhouse, 2021



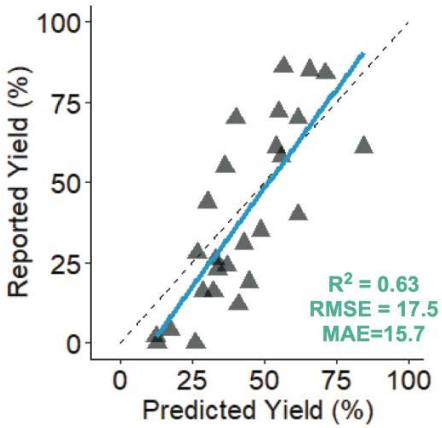
Y-Randomization test

Yields are randomly shuffled across the dataset



Random data test

Chemically meaningful descriptors are replaced with randomly generated values

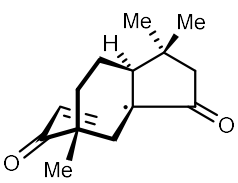


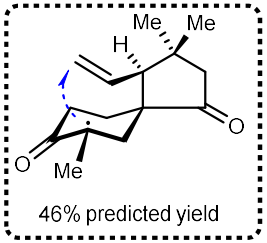
Literature Validation

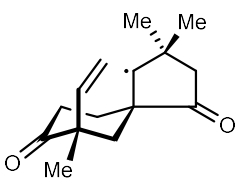
With an additional 26 new examples of 6-*endo* radical cyclization

It is like:

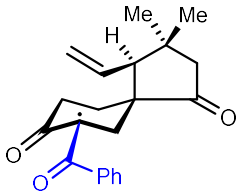
Substrate 1	Substrate 2	Catalyst	"*Anti*-yield"
X	Y	SIMPLS	20.6
X	Y	kNN	19.4
X	Y	RF	16.8
X	Y	NNET	14.0
X	-	NNET	29.3
-	Y	NNET	34.1
Substrate Scope		NNET	17.5

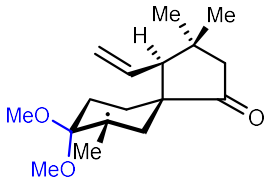

46% predicted yield

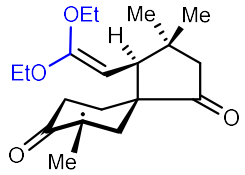

46% predicted yield

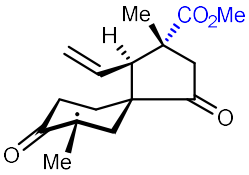

34% predicted yield

Other 6-*endo* Substrates


79% predict yield


36% predict yield

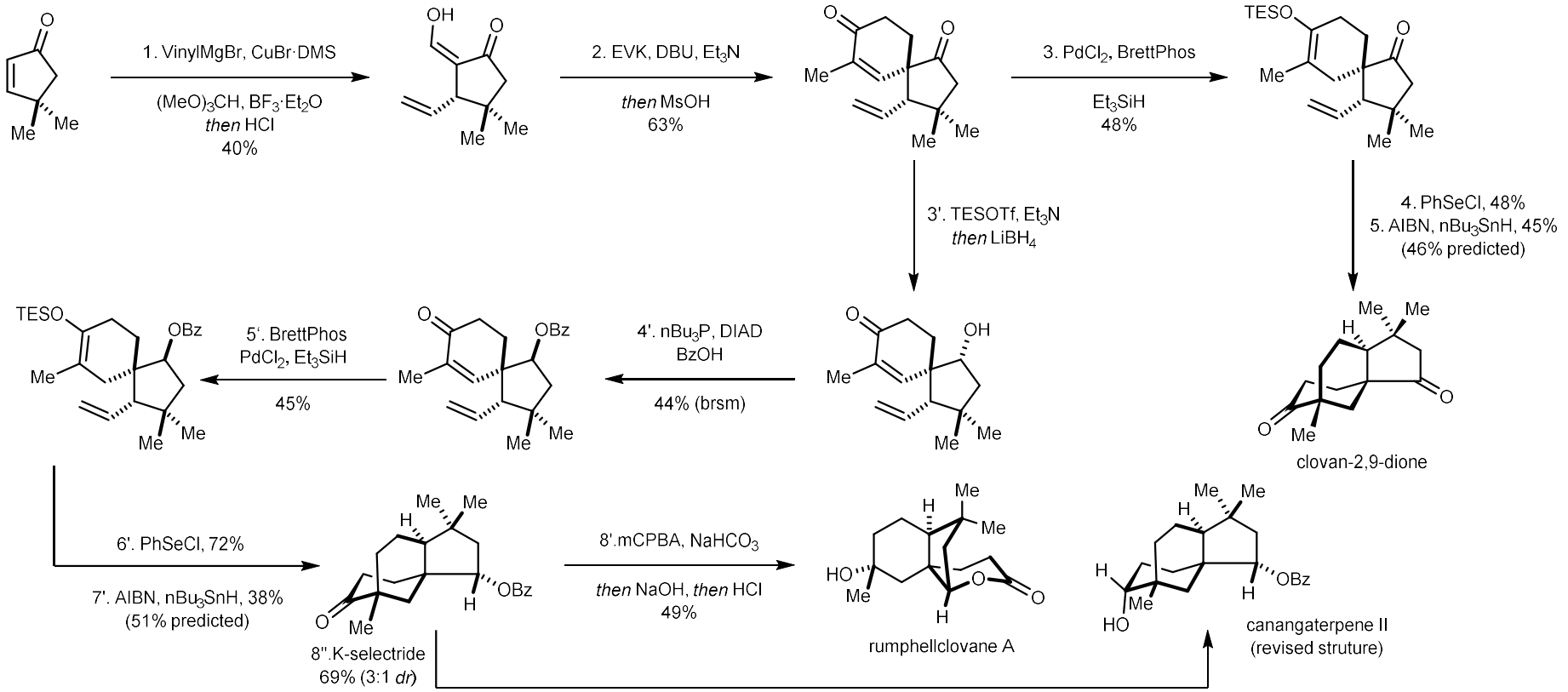

24% predict yield

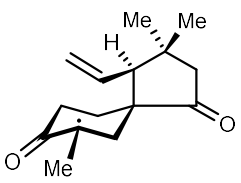

47% predict yield

Newhouse, T.; Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. *ChemRxiv* **2021.**
https://doi.org/10.26434/chemrxiv-2021-41d5z

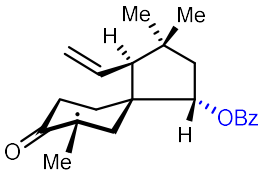
Deep Learning in Chemical Synthesis

Newhouse, 2021

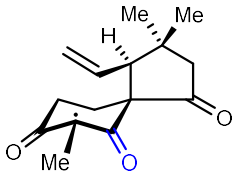




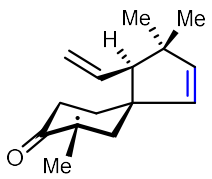
46% predicted yield
45% experimental yield



51% predicted yield
38% experimental yield



62% predicted yield
68% experimental yield



42% predicted yield
49% experimental yield

1. Transformation:
Known mechanism, intramolecular reaction, (maybe) insignificant solvent effects

2. Library Preparation:
Several structural restrictions

3. Feature Extraction:
Electronic structure-related descriptors

3. Model:
One-hidden-layer neural network

Newhouse, T.; Zhang, P.; Eun, J.; Elkin, M.; Zhao, Y.; Cantrell, R. *ChemRxiv* **2021.**
https://doi.org/10.26434/chemrxiv-2021-41d5z