Matthew O'Neill

HTE + Machine Learning

Cornella Group Meeting
1.06.2018

DISCLAIMER - Dilletennte - a person who cultivates an area of interest, such as the arts, without real commitment or knowledge

High Throuput Experimentation (HTE) is transforming chemistry, drug discovery, and biology

Machine Learning has displayed equally disruptive effects in biology and medicine but progress in chemistry has not been commensurate with these respective advances

"Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel 1959

Data is critically needed to enable machine learning, thus empowered by HTE

Included Topics

- History of development
- Common types of learning methods
- Organic methodology
- Chemical synthesis
- Drug design
- Future outlooks from the literature

Topics Not Included

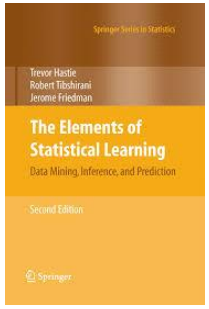
- HTE besides selected miniaturization
- Heterogenous catalysis
- Materials or property prediction
- Binary Classification ML

For popular science introductions see:

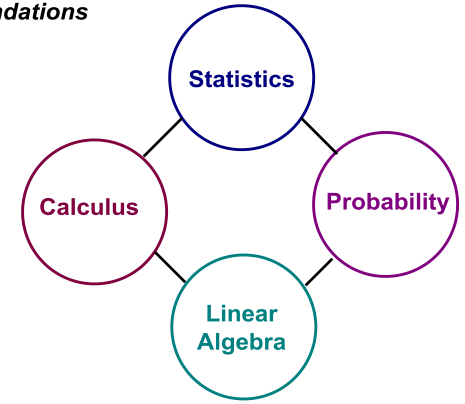
Second Machine Age - Brynjolfsson and McAfee
Life 3.0 - Max Tegmark
Superintelligence - Bostrom

Useful Technical References

Artificial Intelligence - Norwig and Russell
Elements of Statistical Learning - Friedman, Tibshirani, Hastie

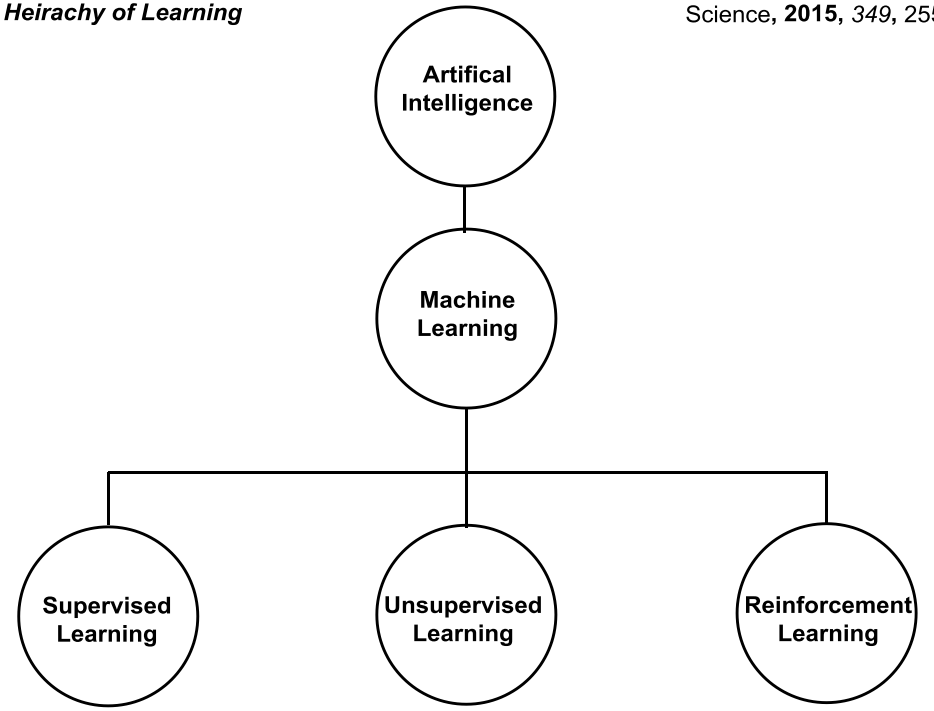


Mathematical Foundations



Heirachy of Learning

Science, **2015**, *349*, 255



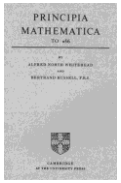
In my view, those efforts are certainly worthwhile as long as they don't come at the expense of the very field they wish to simplify. In other words, such efforts do not intimidate, threaten, or provoke fear in the hearts of any practitioner of synthesis. Promises of computational chemistry and combinatorial chemistry displacing the field were made over the years, yet we are still here. - Baran 2018

Matthew O'Neill

Selected Important Events

Cornella Group Meeting
1.06.2018

*Not to scale



1642 - Pascal invents first mechanical calculator

1913 - *Principia Mathematica* by Russell, is published, exploring Formal Logic

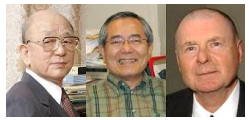
1828 - Wöhler synthesis of urea

1931 - Gödel's incompleteness theroems

1944 - Woodward and Doering complete the total synthesis of quinine - representing first 'modern' total synthesis

1986 - Hinton reports backpropagation method enabling practical ANN

1997 - Google Deepblue beats world chess master Kasparov 3.5-2.5



2010 - Nobel Prize in chemistry for palladium-catalyzed cross couplings

2015 - AlphaGo becomes first computer to beat a human player without handicaps in Go

1666-1674 - Newton and Leibniz begin laying foundation for modern calculus

1845 - Boole sets out to explore symbolic representations of reasoning, inventing Boolean Algebra

Name	Graphic Symbol	Boolean Algebra	Truth Table
AND		F = A . B Or F = AB	
OR		F = A + B	
NOT		$F = \bar{A}$	
NAND		$F = \overline{A \cdot B}$ Or $F = \overline{AB}$	
NOR		$F = \overline{A + B}$	

1932 - Sherrington shares Nobel Prize for contributions in neuroscience - discovers role of 'synapse'

1944 - McCulloch and Pitts develop algorith for Artifical Neural Networks (ANN)

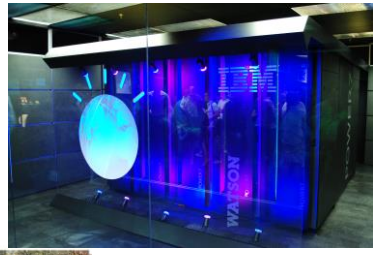
1950 - Alan Turning proposes 'Turing Test' for machine intelligence

1965 - Djerassi and Stanford colleagues begin DENDRAL - focused research on AI to transform organic synthesis

1996 - EJ Corey wins the Nobel Prize for his formalization of retrosynthesis

2003 - Human genome project is declared complete, largest collaborative biology project in world history

2011 - IBM Watson beats Jeopardy champions



Matthew O'Neill

Helpful Definitions

Cornella Group Meeting
1.06.2018

Algorithm - Process or set of rules to be followed in calculations or other problem solving operations, especially by a computer

Artifical Intelilgence - Intelligence demonstrated by machines

Artificial Neural Networks - Inspired by biological neural networks - large number of inputs into a targeted output

Backpropagation - Error attribution of coefficient weights from output layer, back to input layer

Classification - Qualitative data output - cat vs. dog image recongition

Deep Learning - Machine learning alogrithm based on multilayer neural networks Uses a series of non-linear functions combintatorially, creating multiple levels of representations for different levels of abstraction

Go - One of the world's most complex games, 10^{170} possible positions

Hueritics - An approach to problem solving, learning, or discovery that employs a practical method not guaranteed to be optimal or perfect, but likely sufficient

High-Throughput-Experimentation - Experimentation that provides large amounts of data that far exceeds traditional lab-scale preparations; examples include DEL (DNA-Encoded-Libraries), miaturization, and on-chip

Least-squares linear regression - Simplest machine learning algorithm (Hammett Plot)

Machine Learning - Study and construction of computer algorithms that can learn from data

Molecular descriptors - Describes properties of any compound by utilizing domain expertise

Monte Carlo Tree Search - Heuristic search algorithm for some kinds of decision processes, most notably those employed in game play

Neuron - Biological cell-type that transmits information in the nervous system

Processing Unit - Maps features into an output using a function - simplest unit of deep learning algorithms; artifiicial neuron

QSAR - Quantitative Structure Activity Relationship - goal to predict the biological activity of a certain compound

Random Forest Model - Machine learning algorithm that takes random samples of various decision trees, and combines their inputs to select an output

Regression - Quantitative data output - reaction yield

Reinforcement Learning - Machine learning paradigm between supervised and unsupervised; data is not labelled, but correct/incorrect response is made

RMSE - Root Mean Squared Error - Difference between values predicted and values observed; **R2** - relative error

Shallow ML - Input features provided by domain expert; template matching (does not learn representation of problem)

Support Vector Machine - Machine learning algorithm best for classification processes; utilizes 'kernels' to find boundaries between different classes

Supervised Learning - Paradigm in which input has a labelled output for training set; reaction parameters as input, reported yields as output

Test Set - Set of data that is utilized to generate a predictive algorithm, usually exploring many types of methods, i.e. linear regression, ANN, Bayes classifier, or Random Forests; train algorithm on 70% of experiments, test on the other 30%

Training Set - Set of data that is reserved to test the predictive ability of the algorithm; train algorithm on 70% of experiments, test on the other 30%

Unsupervised Learning - Paradigm where the inputs are completely unlabelled, algorithm itself comes up with classifications to attach to the data and arrives independently at outputs

Matthew O'Neill

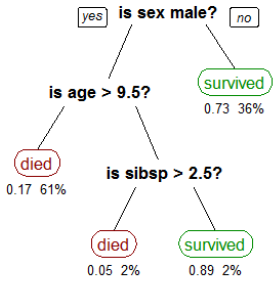
Machine Learning Techniques

Cornella Group Meeting
1.06.2018

Techniques for analyzing data set - choose based on specific problem

Similar to computational chemistry - need to choose the correct level of theory based on the problem at hand

Decision Trees

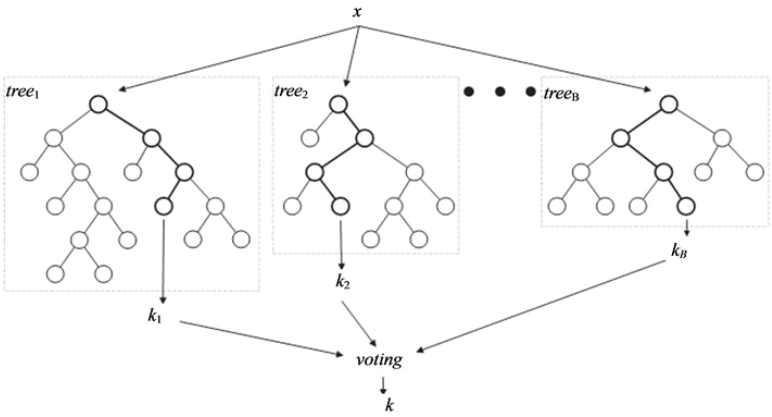


CART - Classification and Regression Trees
Node - test of an attribute
Branch - outcome of test
Leaf - class label
Low data curation
Overfitting is primary problem
Grow tree - Choose features
- Identifity splitting conditions
- Learn when to stop
- Prune tree

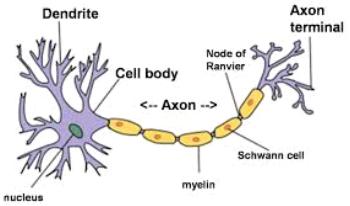
Applications - Titanic survival statistics, male or female based on height and weight, price of home

Random Forest

Grow multiple trees
More robust and accurate with multiple models
For a new object, each tree gives a vote
Output with the most 'votes' gets chosen
Avoids overfitting problem of *single* decision tree
Large data sets with high dimensionality
Disadvantage - black box
Applications - banking, epidemiology, stock behavior, reaction yield, voice classification



Artifical Neural Networks (ANN)



neuron structure

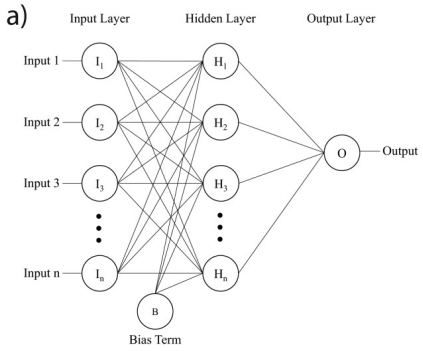
Biological inspiration
Artifical neuron - perceptron
- Multiple inputs to generate one output

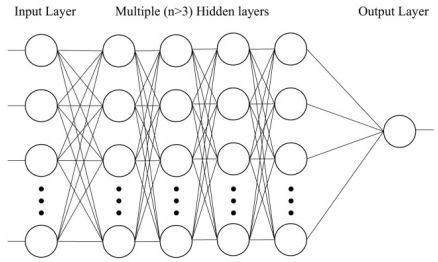
Each neuron is weighted
Distinct activation functions to decide output
- Unit step, sign, linear, logistic, etc...

Multiple types of outputs
Train either forwards or backwards
Applications - self-driving cars, image recognition

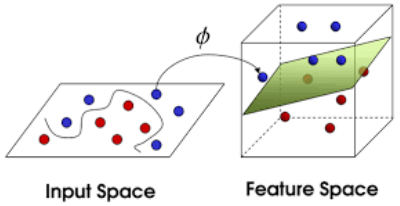
Deep Learning

ANN with multiple hidden layers
Computationally intractable until recently
Made feasible with rectified linear unit activation - transformative impact
Gradient descent for error correction
Applications - AlphaGo, retrosynthesis





Support Vector Machines



Very good for extreme cases
Support vectors brush up against the separating margin
Considers only support vectors important
Hyperplane - separates the classes in n-dimenstional space
Non linear data transformed to higher dimension
Kernel - turns vectors to dot product in feature space
- Non-trivial to choose kernel

Applications - medical imaging, air quality, medical classification, financial analysis, page ranking

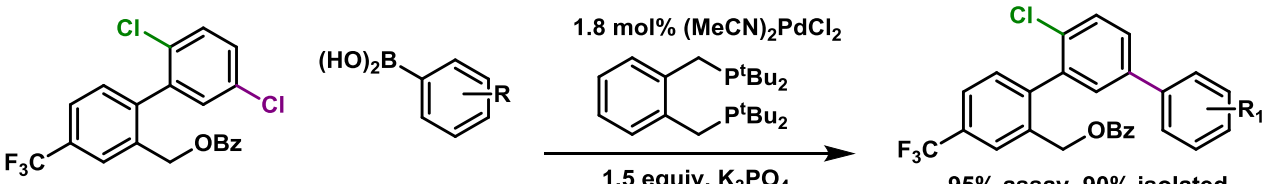
Matthew O'Neill

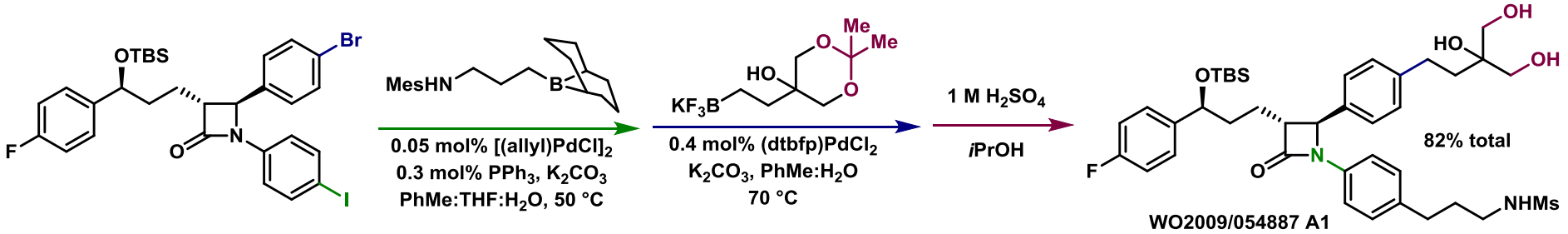
Merck: A Case Study

Cornella Group Meeting
1.06.2018

Medicinal Chemsitry	Process Chemistry
Best Molecule	Best Synthesis
Optimal Conditions	Many Reaction

Process chemistry requires extremely exacting conditions in scale up
Need quick access to a large quantity of multiple parameter permutations





Medicinal Chemistry limited primarily by time and availability of materials

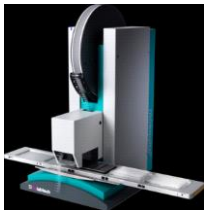
Began in house 'kits' with predispensed catalyst and base to allow rapid evaluation of conditions

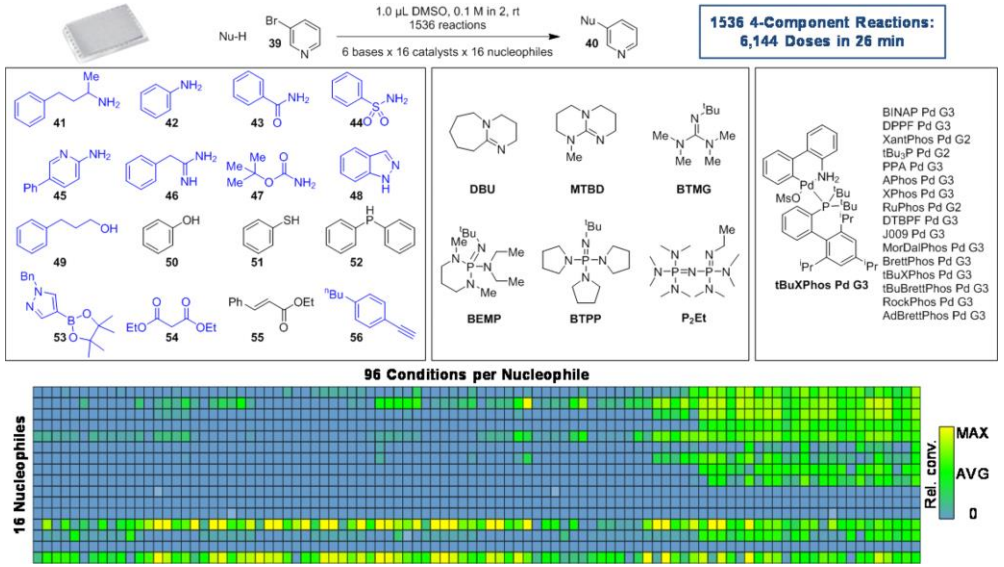
These kits enabled broader patterns to emerge from higher levels of data, leading to the Merck 'Catalyst Selection Guide'

Developed 'parallel-in-parallel' HTE - running various reaction conditions simultaneously with distinct nucleophile electrophile combinations

Bottleneck lead to development of miniaturization platform - inspired by adjacent fields, unifying robotics in biotechnology with mass spec techniques

Achieved 1500 rxns in a day, with 0.02 mg per reaction driven by TPP Mosquito robot





Acc. Chem. Res. **2017**, *50*, 2976

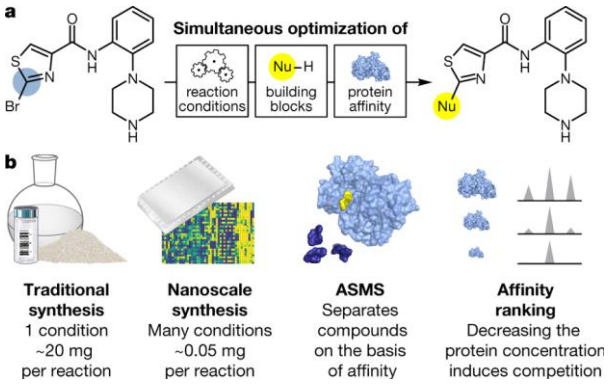
Matthew O'Neill

Merck: NanoSAR

Cornella Group Meeting
1.06.2018

HTE enables rapid evalution of potential compound leads
Greatest advances have come from a confluence of emerging technologies to enable rapid evaluation and data collection on diverse compound sets

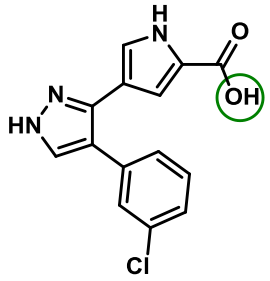
Merck developed NanoSAR for synthesis and affinity ranking using UPLC-MS (Ultra-high-Performance Liquid Chromatograph/Mass Spec) and ASMS affinity ranking (technique to detect compound bound to protein)



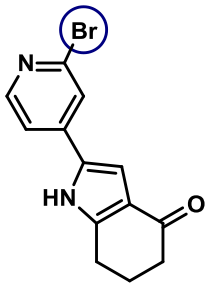
Workflow

1) Nanoliter robot dosing and reaction proceeds 20 hours
2) a. Confirm product via UPLC-MS b. info for up to 1536 rxns per run
3) Incubate protein with pool of molecules
4) a. Size exclusion column separates out unbound molecules b. HPLC-MS identifies complexes
5) Repeat ASMS assay at lower protein concentration
6) Scale up compounds and determine IC_{50} by functional assay

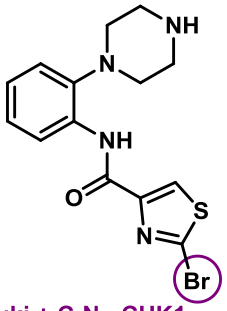
Essential Controls



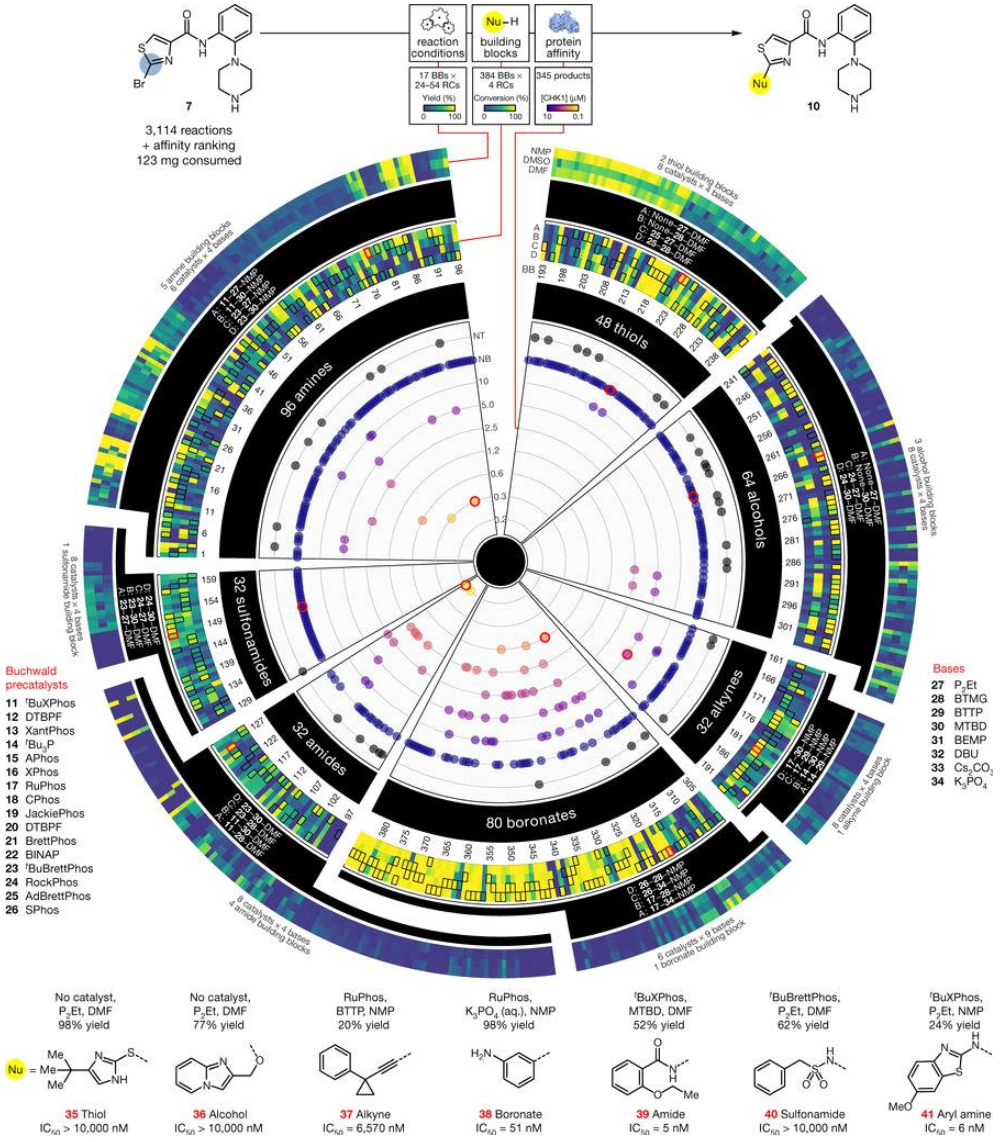
Amide - ERK2
0.035 mg / rxn
14-29 mg / rxn



Suzuki - MK2
0.035 mg / rxn
10-35 mg / rxn



Suzuki + C-N - CHK1
0.044 mg / rxn
18-22 mg / rxn



123 mg to run 3114 reactions, evaluate 384 nucleophiles, and assay material

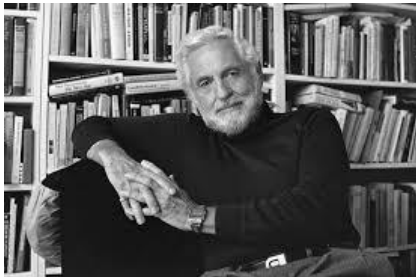
Nature, **2018**, in press

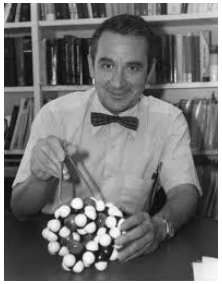
Matthew O'Neill

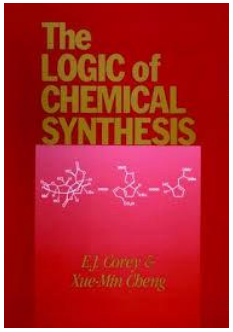
History of Pursuit

Cornella Group Meeting
1.06.2018

History of CASP - Computer-Assisted-Synthetic-Planning







History of early efforts

1963 - Vleduts proposes using computers to synthesize compounds backwards

1965 - Carl Djerassi and Stanford computer scientists begin Dendral Project
2 goals - use Artifical Intelligence to predict structure from spectra
use Artifical Intelligence to plan synthetic routes

1967 - Seminal Corey paper disclosing retrosynthesis

1969 - Corey and Wipke formalize OCSS to utilize computers for plans

1972 - Corey and Pensak formalize LHASA - first graphical platform

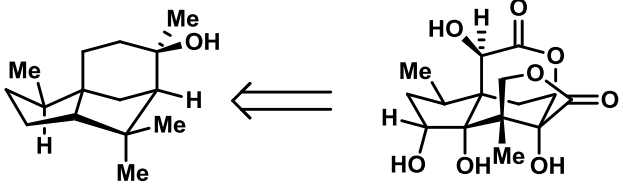
1980's - Ugi begins IGOR - represent bond electrons as matrices and reactivity as matrix transform - conceptual breakthrough from previous attempts

1990 - Hanessian applies his CHIRON system to synthetic organic planning

2012 - Chematica by Gryzbowski - uses 86,000 chemical rules

ACIE, **2016**, *55,* 5904

Recent Example of CHIRON Approach



cedrol **majucin**

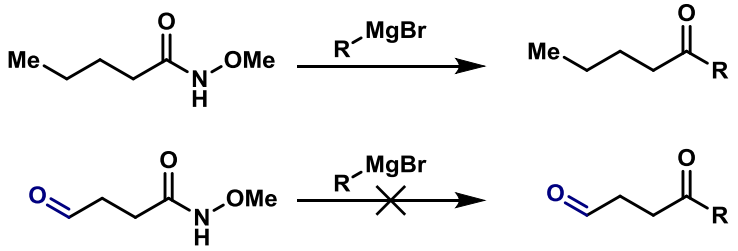
JACS, **2017**, *139*, 17783

Why so difficult (taken from Grzybowski)

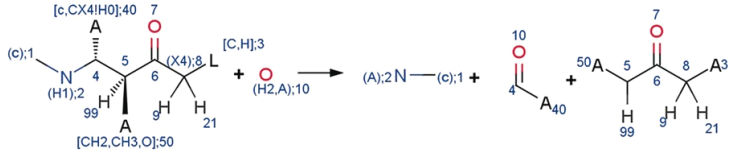
1. Excpetions to the norm abound - 'Black Swans'

2. Automation was not easy

3. Molecular context is everything!



4. Stereochemistry was hard to define (shown below in SMARTs)



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"

5. Hard to define positions - cost associated with every possible move

6. Size of search and lack of intelligent algorithms

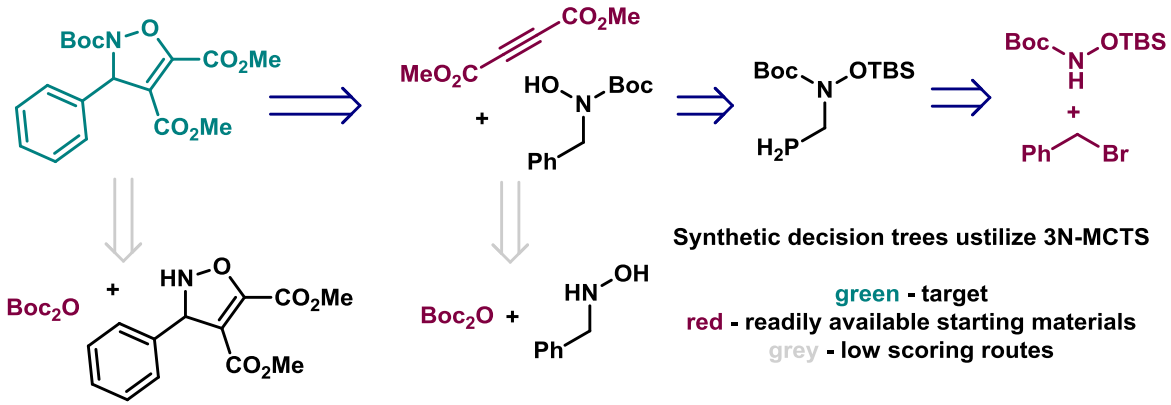
Paraphrasing Churchill's famous words after the Allies' first major victory over the Axis forces in Africa, it is not the end, it is not even the beginning of the end, but it is the end of the beginning for the computer-assisted synthesis planning. The machine is here to stay. - Gryzbowski 2016

Matthew O'Neill

Synthesis

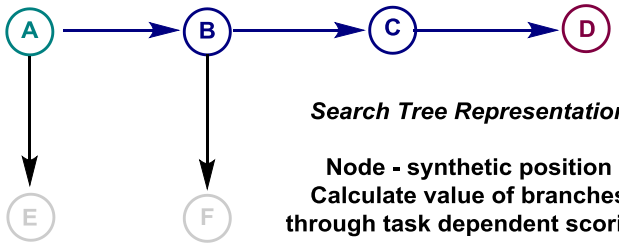
Cornella Group Meeting
1.06.2018

"These DNN were trained on essentailly all reactions ever published in organic chemistry."



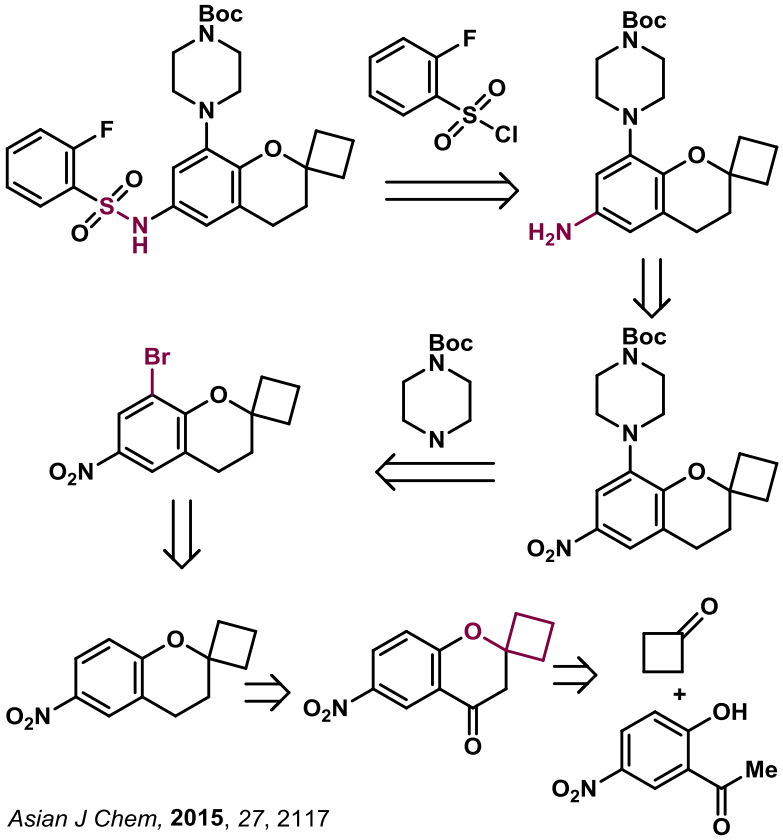
Synthetic decision trees ustilize 3N-MCTS

green - target
red - readily available starting materials
grey - low scoring routes

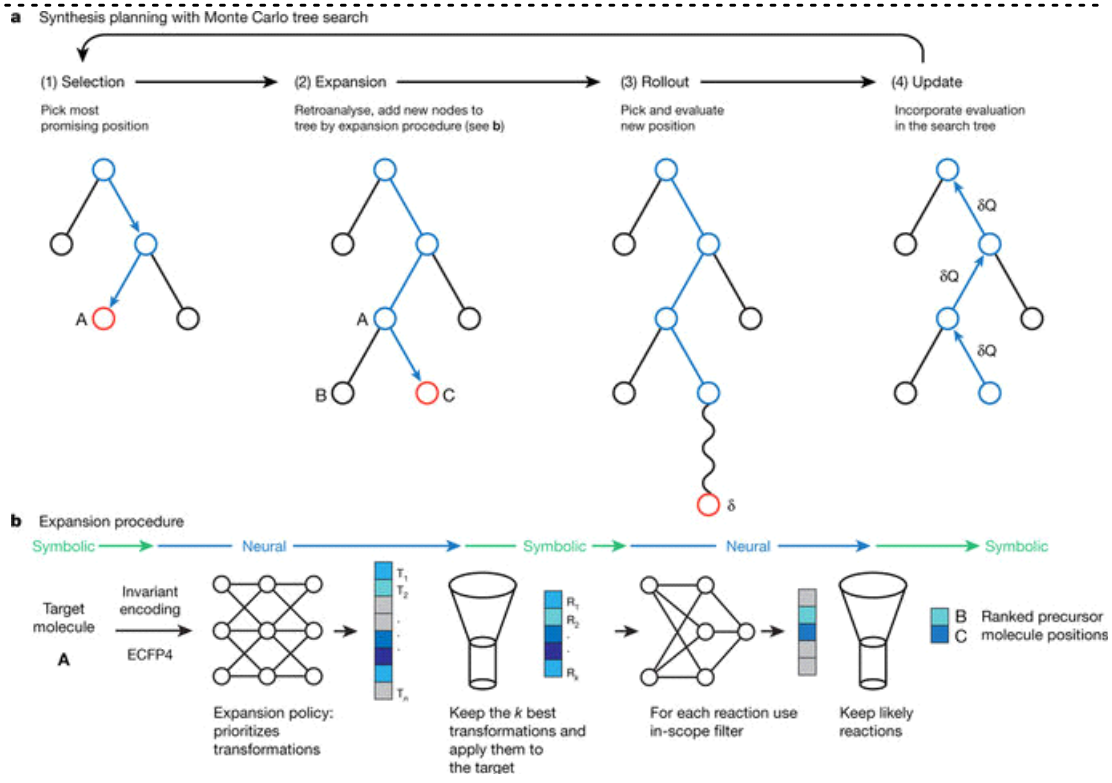


Search Tree Representation

Node - synthetic position
Calculate value of branches
through task dependent scoring



Asian J Chem, **2015**, *27*, 2117



Nature, **2018,** *555*, 604

A bigger part is the supposition by organic synthetic chemists that they are the ones best qualified to design programs for organic synthesis. Not so! - Whitesides

Matthew O'Neill

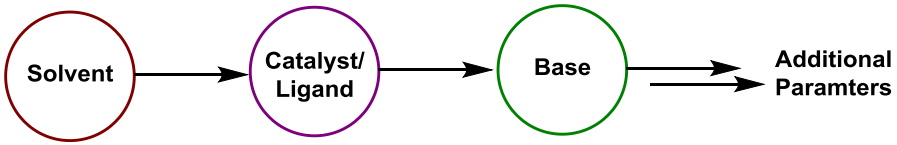
Methodology

Cornella Group Meeting
1.06.2018

Abby Doyle created first general machine learning algorithm for homogenous catalysis

Machine learning is exceptioanlly hard to amend to available data

Logistical constraints have largely precluded genuine multiparamter optimization in academia

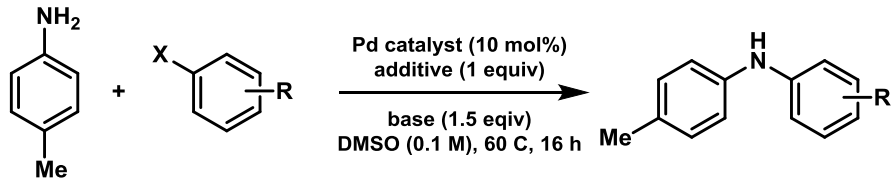


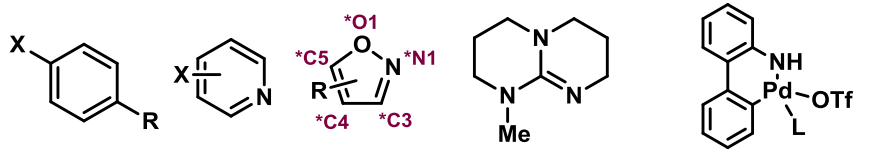
Invokes the 'curse of dimensionality' - each additional parameter *n* adds another dimension to the reactivity space

In typical methodology development, only a small fraction of this chemical space is thoroughly explored

Alternative **- Use HTE to explore large amounts of chemical space**
Can these larger data sets be used to teach algorithms how to predict yields when using unseen reagents?

Buchwald-Hartwig Amination





Aryl Halides (15) | Additives (23) | Bases (3) | Pd Catalysts (4)

1536 well plates with Mosquite robot - UPLC to determine yield - 4608 rxns

Based on Glorius 'robustness screening'
Using additive approach to model embedded isooxazole to simplify process

Nat Chem, **2014,** *6,* 859

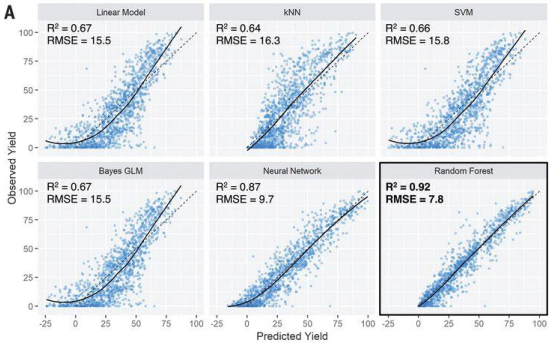
Create descriptors without invoking a mechanistic hypothesis, only look at structure
Internal consistency achieved using calculated DFT properties (B3LYP/6-31G*)

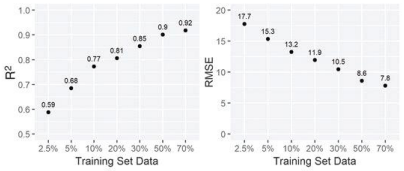
Workflow for Data Input

1) Submit molecular structure to Spartan GUI and specify reaction component
2) Calculation of molecular, vibrational, and atomic properties
3) Extraction of these features from the resulting text files
4) Generation of data table for fit - 120 total descriptors

Supervised ML Tested

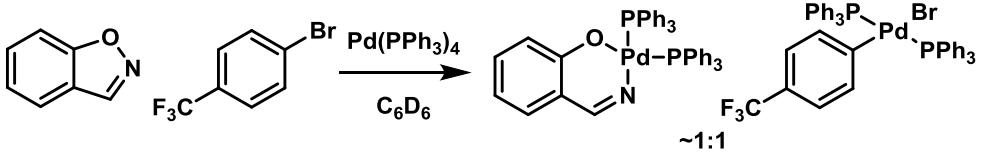
***k*-nearest neighbors, support vector machines, and Bayes generalized linear model all no better than linear regression**
Random forest model proved significantly better!
7.8% RMSE with R^2=0.96 - large amount of error attributed to experimental and analytical error
Better than any other method when trained on 5% of data vs. 70% of data for others





Examine Descriptors to Identify Most Important Qualities

Mechanistic insights from relative importance of descriptors
1) *3C NMR shift 2) LUMO energy 3) *O1 charge 4) *C5 charge



Science, **2018**, *360,* 186

Additional Methods

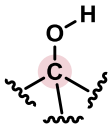
Matthew O'Neill

Cornella Group Meeting
1.06.2018

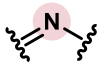
Deoxyfluorination - Second Generation work from Doyle

Fewer reactions - 640 with stock solutions and NMR quantification
No mosquito robot :(
Yields over 100% were kept to avoid bias - reveals experimental error as strong influence on subsequent prediction performance
Reaction paramters not considered - time, temp, concentration, stoich.

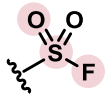
Data input different - extract 7 features of molecules, and add categorical labels for a total of 23 differenct descriptors



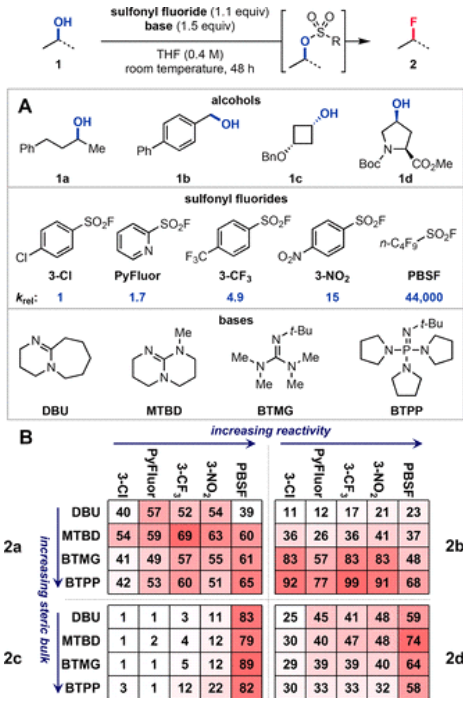
Alcohol - *C1 charge, *C1 A^2, EN



Base - *N1



Sulfonyl Fluoride - *S1 charge, *F1 charge, *O1 charge

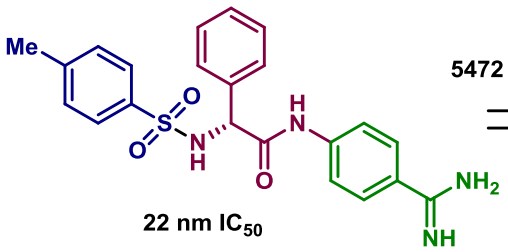


Yields for all 640 rxns determined by ^{19}F NMR

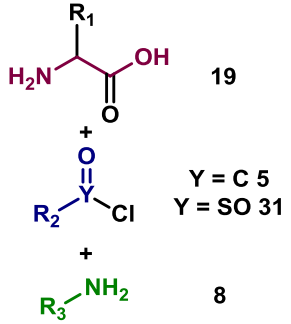
Categorical Descriptors
- Alcohol - primary
- Alcohol - secondary
- Alcohol - tertiary
- Alcohol - cyclic
- Alcohol - 4-membered ring
- Alcohol - 5-membered ring
- Alcohol - 6-membered ring
- Alcohol - 7-membered ring
- Alcohol - benzylic
- Alcohol - allylic
- Alcohol - homobenzylic
- Alcohol - homoallylic
- Alcohol - alpha-carbonyl
- Alcohol - beta-carbonyl
- Alcohol - hemiacetal
- Alcohol - amino alcohol

JACS, **2018**, *140*, 5004

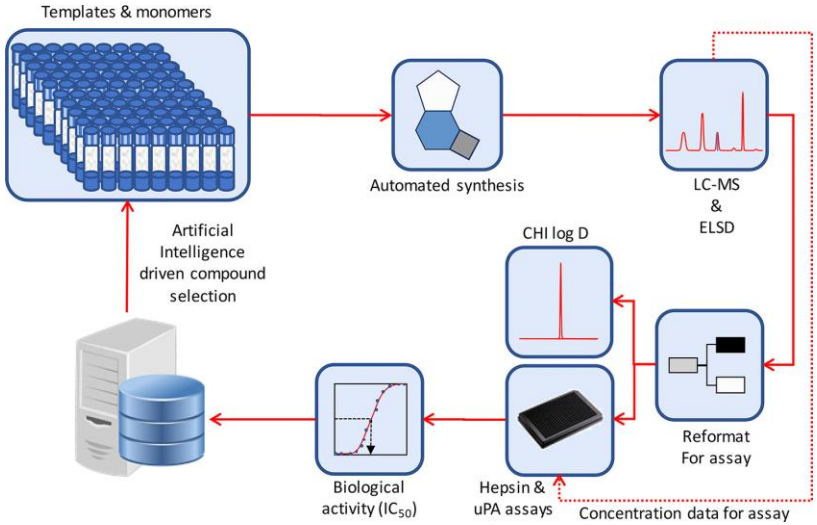
Automated Closed-Loop Optimization Platform



5472 products



Hepsin Inhibitor - clinical serine protease for carcinogenesis and metastasis via HGF signaling
Closed loop only takes 1-2 h to complete - traditional methods would take weeks
CyclOps platform for rapid synthesis vs. batch methods
BOUS - Best Objective Under-Sampled - algorithm predicts IC_{50}



J. Med. Chem. **2018**, *61*, 4335

Merck QSAR challenge in 2012

Precomputed molecular descriptors for compounds and provided experimental biological activity for 15 targets; DNN model won the competition, without a single chemist on the team

Journal of Computational Chemistry, **2017**, *38*, 1291

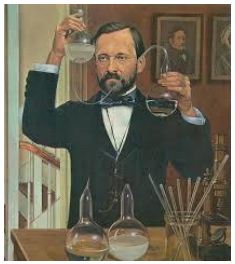
My own feelings are that any machine that takes away from me the necessity to crank out 142 amide reactions by hand is welcome to it - Derek Lowe

Matthew O'Neill

Future Outlook

Cornella Group Meeting
1.06.2018

NMR and X-Ray largely replaced degradation studies as preferred method for structure elucidation - is this bad? Led directly to the 'Golden Age of Natural Product Total Synthesis' - this beautiful field would have been precluded by technological constraints were it not for the advent of NMR and other enabling analytical techniques



"The "age of automation" thus appears to hold the potential to advance organic synthesis in a revolutionary way" - Nuno Maulide

What will this revolution look like?

Inspiration for this talk

"This evolution – guided by experts in artificial intelligence more than by experts in synthesis – raises a very important point for students: viz., for the future, will it be more important to understand AI than to be able to recall all the methods for introducing chiral centers (or other transformationof choice) from human memory? How should one balancecomputer skills and empirical synthetic skills?" - Whitesides *Isr. J. Chem.* **2018**, *58*, 142.

Coming Full Circle

What defines a practioner of synthesis?

"I did not become a "synthetic organic chemist," but almost all the research that my colleagues and I have done (and do) involves organic synthesis." - Whitesides

Satire Article about Martin Burke MD PhD Synthesis Machine - *Science*, **2015**, *347*, 1221

Urbana, IL

Ugly scenes today marred the unveiling of what may become a landmark paper, as an angry mob of organic synthesis researchers invaded the chemistry department at the University of Illinois at Urbana–Champaign before seizing and ultimately destroying a so-called "synthesis machine."

An article in the journal Science, describing the development of what is in effect a cyborg post-doc, prompted an initially peaceful protest outside the chemistry department under placards carrying the slogans KEEP NATURAL PRODUCT SYNTHESIS NATURAL, SUZUKI COUPLINGS ARE CHEATING and GIVE ME C–H ACTIVATION OR GIVE ME DEATH. However, witnesses described a marked increase in tension after the arrival of a counter-demonstration of inorganic chemists, who taunted their organic counterparts with highly charged epithets including "pot-boiler" and "column monkey".

An anonymous demonstrator later told C&EN Onion: "It all kicked off when the fucking stamp collectors showed up. The was always an undercurrent of anger, but that was when it boiled over and you became keenly aware just how many people had brought BuLi with them."

Asked to explain the motives of the inorganic counter-demonstrators, a hooded organometallic researcher said, "We're just here looking for trouble. I've got no dog in this fight, unless you're gonna tell me that thing's got an onboard SQUID magnetometer."

Anger having now reached fever pitch, a large group stormed the building, making directly for the lab housing the controversial machine. Minutes later, the helpless automaton was flung from a second floor window, landing amongst cheering protesters and breaking, ironically enough, into a number of fragments. Amid frantic shouts that the machine may have developed the capability to heal itself, clamp-stand-wielding synthetic chemists smashed what little remained. To their credit, many of them first donned appropriate personal protective equipment.

John Wiseman, a technician present during the break-in, remained sanguine as he detailed the damage to the lab. "The automated synthesis platform was what they came for, of course, but someone also found time to steal a bunch of NMR tubes and a fresh batch of DMP. You know what these people are like."

Wiseman also claimed that clashes involving armed factions of researchers were not without precedent: "You'd be surprised. There are a lot of radical chemists out there."

Taken from http://cenonion.blogspot.com/2015/03/rampaging-synthetic-chemists-smash.html

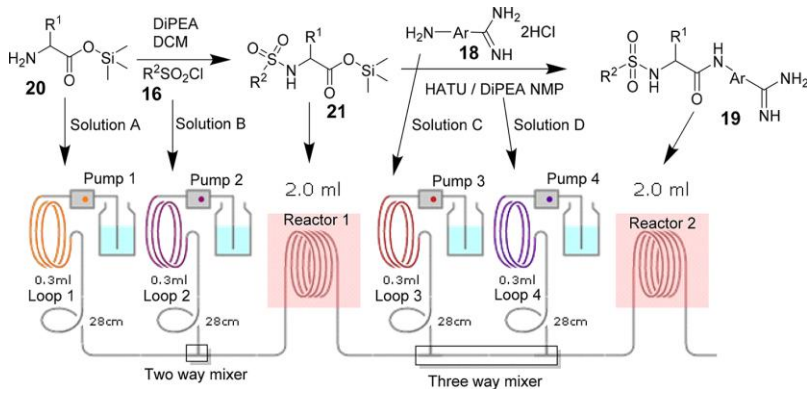
Matthew O'Neill

Appendix

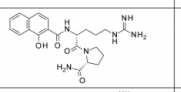
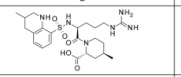
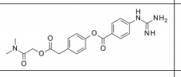
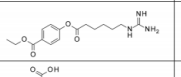
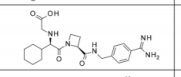
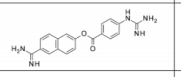
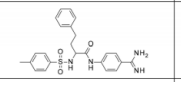
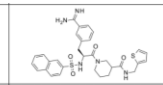
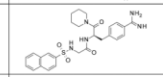
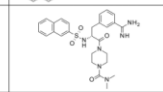
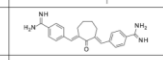
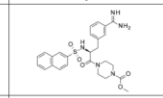
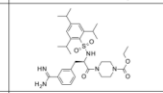
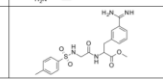
Cornella Group Meeting
1.06.2018

Closed Look Hepsin Inhibitors

Depiction of Flow Process



Commercialy available inhibitors - identify 7 as lead molecule

Compound Number	Structure	Hepsin IC_{50} (μM)	uPA IC_{50} (μM)
1		>10	ND
2		>10	ND
3		0.029	ND
4		0.383	ND
5		4.02	ND
6		0.005	ND
7		1.13	>10
8		0.676	2.43
9		4.46	ND
10		6.55	>10[a]
11		0.272	0.484
12		6.19	ND
13		3.35	0.285
14		3.14	>10

Predicting Reaction Performance in C-N couplings using ML

Examples of calculated descriptors

Additive Descriptors (n = 19)
EHOMO, ELUMO, Dipole Moment, Electronegativity, Hardness, Molecular Volume, Molecular
Weight, Ovality, Surface Area, *C3 NMR Shift, *C3 Electrostatic Charge, *C4 NMR Shift, *C4
Electrostatic Charge, *C5 NMR Shift, *C5 Electrostatic Charge, *N1 Electrostatic Charge, *O1
Electrostatic Charge, V1 Frequency, V1 Intensity

Aryl Halide Descriptors (n = 27)
EHOMO, ELUMO, Dipole Moment, Electronegativity, Hardness, Molecular Volume, Molecular
Weight, Ovality, Surface Area, *C1 NMR Shift, *C1 Electrostatic Charge, *C2 NMR Shift, *C2
Electrostatic Charge, *C3 NMR Shift, *C3 Electrostatic Charge, *C4 NMR Shift, *C4
Electrostatic Charge, *H2 NMR Shift, *H2 Electrostatic Charge, *H3 NMR Shift, *H3
Electrostatic Charge, V1 Frequency, V1 Intensity, V2 Frequency, V2 Intensity, V3 Frequency,
V3 Intensity

Base Descriptors (n = 10)
EHOMO, ELUMO, Dipole Moment, Electronegativity, Hardness, Molecular Volume, Molecular
Weight, Ovality, Surface Area, *N1 Electrostatic Charge

Ligand Descriptors (n = 64)
Dipole Moment, *C1 NMR Shift, *C1 Electrostatic Charge, *C2 NMR Shift, *C2 Electrostatic Charge, *C3 NMR Shift, *C3 Electrostatic Charge, *C4 NMR Shift, *C4 Electrostatic Charge, *C5 NMR Shift, *C5 Electrostatic Charge, *C6 NMR Shift, *C6 Electrostatic Charge, *C7 NMR Shift, *C7 Electrostatic Charge, *C8 NMR Shift, *C8 Electrostatic Charge, *C9 NMR Shift, *C9 Electrostatic Charge, *C10 NMR Shift, *C10 Electrostatic Charge, *C11 NMR Shift,
*C11 Electrostatic Charge, *C12 NMR Shift, *C12 Electrostatic Charge, *C13 NMR Shift, *C13 Electrostatic Charge, *C14 NMR Shift, *C14 Electrostatic Charge, *C15 NMR Shift, *C15 Electrostatic Charge, *C16 NMR Shift, *C16 Electrostatic Charge, *C17 NMR Shift, *C17 Electrostatic Charge, *H11 NMR Shift, *H11 Electrostatic Charge, *H3 NMR Shift, *H3 Electrostatic Charge, *H4 NMR Shift, *H4 Electrostatic Charge, *H9 NMR Shift, *H9 Electrostatic Charge, *P1 Electrostatic Charge, V1 Frequency, V1 Intensity, V2 Frequency, V2
Intensity, V3 Frequency, V3 Intensity, V4 Frequency, V4 Intensity, V5 Frequency, V5 Intensity,
V6 Frequency, V6 Intensity, V7 Frequency, V7 Intensity, V8 Frequency, V8 Intensity,
V9 Frequency, V9 Intensity, V10 Frequency, V10 Intensity