Machine Learning in Computer-aided Synthesis Planning:

Efficient Synthesis of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Lab

Ximing Li[1]

[1] Princeton University, Department of Chemistry, Princeton, NJ

PRINCETON UNIVERSITY

ABSTRACT

- An artificial synthetic organic design algorithm, which will automatically construct and manipulate robust and viable synthetic routes based on current human knowledge of organic chemistry reactivity, is a notably challenging machine learning problem without direct amenability to existing approaches.
- Expectably, equipping computers with strong, robust learning algorithms, symbolic representations, and the access to the rich history of human-developed reactions, will probably also open the door for machines to becoming masters in this discipline, which has happened recently in many other areas such as man-machine champion games (chess, Go, and some video games).
- In this work, the idea of computer-aided synthesis planning will be developed towards the goal of machines auto-learning from database and judging the most optimal methodology sequence for a given organic synthesis target. Ideally, the algorithm which this machine will be equipped with should be able to directly learn from the readily available online reaction database (SciFinder, Reaxys, Patent Office Database) which filled with reaction examples (rather than from expert-biased so-called reaction rules), read the target molecule in a symbolic representative and chemically meaningful way, and generate a library of ranked possible reaction sequences which will fulfill the given task.

BACKGROUND AND APPROACH

Highlights

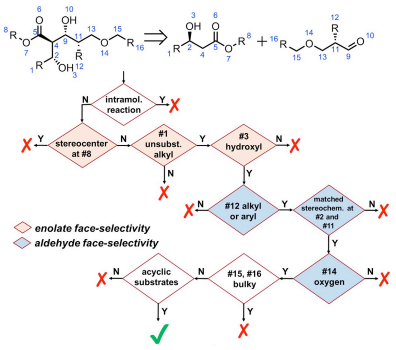
- **Computer autonomously designs chemical synthesis of medicinally relevant molecules.**
- **The syntheses are executable in the laboratory.**
- **The machine-designed routes improve on previous approaches.**

Background and Approach

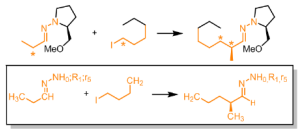
Although computers have demonstrated the ability to challenge humans in various games of strategy, their use in the automated planning of organic syntheses remains unprecedented. As a result of the impact that such a tool could have on the synthetic community, the past half century has seen numerous attempts to create in silico chemical intelligence. However, there has not been a successful demonstration of a synthetic route designed by machine and then executed in the laboratory. Here, we describe an experiment where the machine learning algorithm designed syntheses leading to commercially valuable and/or medicinally relevant targets, significantly improving on previous approaches or identified efficient routes to targets for which previous synthetic attempts had failed. These results indicate that now and in the future, chemists can finally benefit from having an "in silico colleague" that constantly learns, never forgets, and will never retire.

Reaction Networks

An example of an ideal decision tree for one of the reactions rules (double stereo-differentiating condensation of esters with aldehydes). The tree begins with a condition of the reaction being intermolecular. To ensure face selectivity of the enolate, conditions for the substituents at positions #8, #1, and #3 are considered. Conditions at positions #12, #2, #11, #14 follow and ensure proper face selectivity of the aldehyde. The last two conditions are common for both substrates. The substrates should be acyclic because cyclic structures might distort the aldehyde-titanium chelate conformation or face selectivity of the ester enolate. The other requirement concerns the consonant selectivity at both substrates that ensures the desired diastereoselectivity.



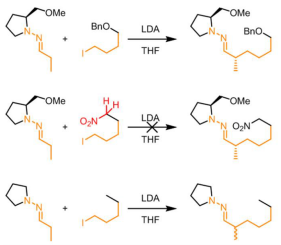
REACTION RULES: GENERAL CONSIDERATIONS



An example of a literature-reported transformation from which the "reaction core" is extracted.

The core is colored in orange, covers atoms changing their local environments (denoted), and also includes flanking atoms up to three bonds away.

Even with this extended neighborhood, the transform does not capture the influence of a distant stereo-directing group, CH2OMe.

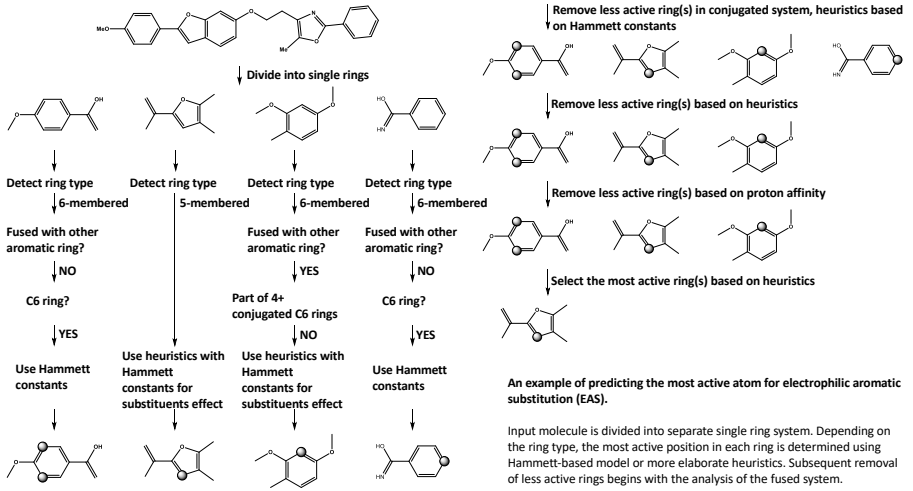


Application of an automatically-extracted reaction core to various synthetic targets.

Reaction rule extracted from a literature precedent (Figure Left) applied to Epothilone A (Figure Right, top) intermediate and middle substrate with a nitropentane side chain. In the latter case, the reaction is not feasible since the pendant nitroalkyl group is incompatible (due to the presence of acidic H's) with lithiated azaenolate formed from the hydrazone upon the initial treatment with LDA.

(bottom) In the absence of the distant stereo-directing group – not included in the reaction rule/core – the transform may still try to predict stereoselective outcome – whereas in experiment, a racemic mixture will be obtained.

EVALUATING APPLICABILITY OF TRANSFORMATIONS BEYOND RECORDS



An example of predicting the most active atom for electrophilic aromatic substitution (EAS).

Input molecule is divided into separate single ring system. Depending on the ring type, the most active position in each ring is determined using Hammett-based model or more elaborate heuristics. Subsequent removal of less active rings begins with the analysis of the fused system.

REFERENCES

1.) "Planning chemical syntheses with deep neural networks and symbolic AI." Marwin H. S. Segler, Mike Preuss, & Mark P. Waller. Nature, 2018, 555, 604
2.) "Machine learning in computer-aided synthesis planning." Connor W. Coley, William H. Green, & Klavs F. Jensen. Acc. Chem. Res., 2018, 51, 1281
3.) "Machine learning for organic synthesis: are robots replacing chemists?" Boris Maryasin, Philipp Marquetand, & Nuno Maulide. Angew. Chem. Int. Ed. (Highlight), 2018, 57, 6978
4.) "Prediction of organic reaction outcomes using machine learning." Connor W. Coley, et. al. ACS Cent. Sci., 2017, 3, 434
5.) "Computational chemical synthesis analysis and pathway design." Fan Feng, Luhua Lai, & Jianfeng Pei. Front. Chem. (Review), 2018, 6, 199
6.) "On the synthesis of machine learning and automated reasoning for an artificial synthetic organic chemist." Maneesh K. Yadav. New J. Chem. (Prospective), 2017, 41, 1411
7.) "Controlling an organic synthesis robot with machine learning to search for new reactivity." Jaroslaw M. Granda, et. al. Nature, 2018, 559, 377

ximingl@princeton.edu