

Unit 3

Amines

3.20 Special Topics

- 10/23:
- Grade cutoffs on Exam 2.
 - A-B cutoff: 80.
 - B-C cutoff: 60.
 - C-D cutoff: 45
 - Only F's were people who did not take the exam.
 - This was a significantly harder exam; y'all have been crushing it so far.
 - Remember that these grades are meant to give you a perspective for what you're on track for; they are *not* binding!
 - Notes on Steve Buchwald.
 - He's a real big-name chemist: Has his name on a ton of reactions, can make a ton of pharmaceutical drugs, does a lot of consulting for chemical companies, etc.
 - But also super kind, humble, and nice.
 - Knows a ton, but is very down-to-earth and approachable.
 - The rest of this course will be much more synthesis-heavy.
 - Feel free to continue to reach out to Prof. Elkin even though she's no longer at the blackboards!
 - Today: We'll have fun and talk about machine learning.
 - Prof. Elkin will go through Beker et al. (2018), a paper about using machine learning to predict the outcome of Diels-Alder reactions.
 - The basic idea of what the authors are saying is that if you encode the substituents, you get good prediction of the outputs!
 - Your computer doesn't know what a molecule is, so you have to encode your molecule in a way that is meaningful to a computer.
 - For example: You should not encode benzene with alternating single- and double bonds; benzene has six equivalent bonds due to resonance!
 - Nowadays, computers can predict biological activities (doesn't work perfectly yet, though great progress), solubility and crystal structures (works fine), NMR spectra (works awesome), etc.
 - Predicting optimal reaction conditions works awesome.

- Predicting reaction outcomes or yields can be hit or miss.
- There have been maybe 1 000 000 chemical reactions ever catalogued, but most of them are not that useful.
- The low-data regime of predictive modeling is the final frontier, and the especially important one for chemistry.
- Taking high-level expertise and making it algorithmically applicable can be really difficult.
- “High accuracies are achieved only if the machine is provided some chemical ‘insight’ about the reaction (in particular, information about the reaction’s core and key substituents).”
- While ML models cannot provide the generality of quantum mechanics, they work much faster.
- They trained the model with inverse electron-demand Diels-Alder reactions, Diels-Alder reactions that need to be site-selective, etc.
- The website to help you predict Diels-Alders is historical at this point, so don’t worry if you can’t access it in the paper.
- There are several classes on computational chemistry in both Course 5 and Course 10 if you’re interested!
- A problem with Reaxys: All of the reactions in the database are data-scraped from old papers, so a significant number of them are wrong or incomplete (20-30%, and worse in other databases).
- Predictive modeling really reveals how difficult it is to predict reaction outcomes: Prof. Elkin has published papers where their model can predict yield far better than even chemistry experts.
- Conclusion: ML can be useful in predicting outcomes and can generalize to unseen reactions when descriptors carrying physically relevant information are used, and the machine gets appropriately formatted information.
- Note: None of this is testable material!