PropOrb: A Frontier Molecular Orbital Interaction Proposer

Chloé-Agathe Azencott, Matthew A. Kayala, and Pierre Baldi

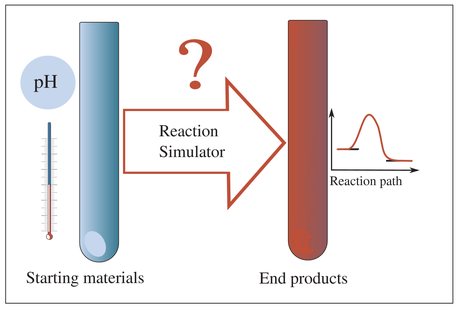
UCIrvine

Institute for Genomics and Bioinformatics
Bren School of Information and Computer Sciences

Background

Reaction Simulator

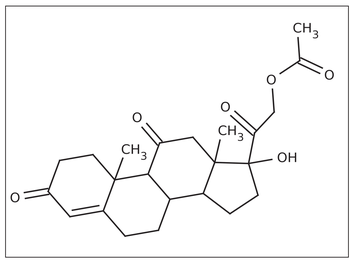
Being able to computationally simulate reactions would augment organist chemists' problem-solving abilities and enable chemical space exploration.



Molecular Orbital Interactions

We modelize elementary mechanistic steps as interactions between an occupied and an unoccupied molecular orbitals (MOs).

To determine the course of a reaction, we can list all the possible filled-unfilled MO interactions, and use the machine learning tools we concurrently develop [1] to rank them by favorability.



Cortisone alone has 116 unfilled and 186 filled MOs.
After eliminating trivially incompatible combinations, there still remains 18,875 potential intra-molecular interactions.

To circumvent the combinatorial explosion of the number of MOs to consider, we need to efficiently identify the most reactive ones.

PropOrb: Proposing Interactions

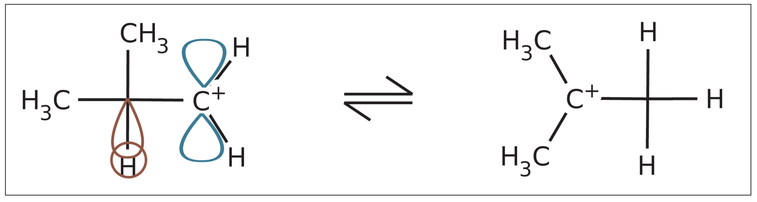
Molecular orbital type, electronegativity, and partial charges can be used to order MOs by reactivity.

We enumerate local and long-range MO interactions in order until the desired number of propositions is reached.

Local Interactions

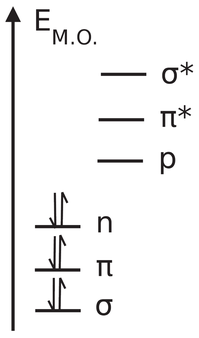
Interactions between two adjacent MOs (migratory displacements, ionizations).

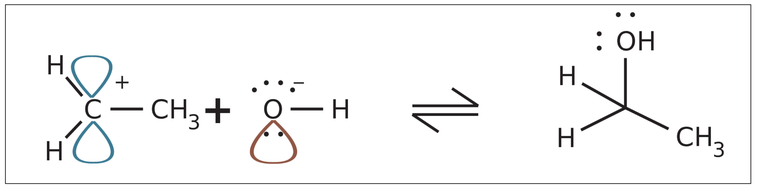
We enumerate local electrophiles (carbocations and good leaving groups) and pair them with adjacent nucleophiles.



Long-Range Interactions

We identify unoccuped MOs most likely to be the LUMO and pair them with occupied MOs most likely to be the HOMO.





Data Generation

The Reaction Explorer `http://www.reactionexplorer.org` [2] is an expert system that covers basic undergraduate chemistry. It is based on 1500+ rules covering elementary reaction steps, defined as simple electron movements from source to sink orbitals, and uses 2000+ full multistep reaction test cases.

We leverage this information to generate 3,277 systems composed of molecules made of C, H, N, O, and halide atoms, reacting under ionic conditions, and for which the most favored elementary mechanistic step is known.

Preliminary Results

Threshold: 500
Runtime: 24 minutes
Most favorable interaction included in 83% of the cases

Conclusion

PropOrb quickly identifies the most relevant MOs for reactivity.

PropOrb can be combined with the machine learning methods from [1], which are more accurate but cannot be used in practice without reducing the number of MOs to compare by order of magnitudes, ultimately leading to a complete reaction prediction system.

Future work

- refining the ordering heuristics, using for instance HSAB or calculating molecular orbital coefficients;

- expanding the coverage of the chemical space to include more diverse compounds and types of reactions.

Related Work

[1] M. A. Kayala, C.-A. Azencott, J. H. Chen and P. Baldi. ORBDB: A DATABASE OF MOLECULAR ORBITAL INTERACTIONS. ACS National Meeting, Spring 2010. San Francisco, CA.

[2] J. H. Chen and P. Baldi. NO ELECTRON LEFT BEHIND: A RULE-BASED EXPERT SYSTEM TO PREDICT CHEMICAL REACTIONS AND REACTION MECHANISMS. J. Chem. Inf. Model., 2009, 49 (9), pp 2034–2043

Further Information

Contact: `cazencot@ics.uci.edu`

Talk on Thursday, March 25th at 3:55pm
The Moscone Center - Room 206/210
CINF - General Papers

ORBDB: A DATABASE OF MOLECULAR ORBITAL INTERACTIONS (Matthew A. Kayala)
Talk on Thursday, March 25th at 2:45pm
The Moscone Center - Room 212 East
CINF - General Papers

Acknowledgments

Jonathan H. Chen and the Baldi Lab

OpenEye Scientific Software and ChemAxon academic licenses

NIH Biomedical Informatics Training grant (LM-07443-01), NSF grants EIA-0321390 and 0513376, Camille and Henry Dreyfus Foundation grant to P.B., IBM grant to C.A.