PropOrb: A Frontier Molecular Orbital Interaction Proposer

C.-A. Azencott, M. A. Kayala, and P. Baldi

Institute for Genomics and Bioinformatics Donald Bren School of Information and Computer Sciences $\underbrace{UCIRVINE}$

237th ACS Meeting

March 25, 2010

Reaction Simulation



- augment organist chemist's problem-solving abilities
- (virtual) chemical space exploration



- 5×10^6 known small molecules
- 10⁶⁰ possible small molecules
- 10²² stars in the observable universe

Reaction Explorer Jonathan H. Chen



- rule-based expert system
- elementary transformations ↔ molecular orbital interaction
- covers undergraduate chemistry
- → no ability to generalize

(日) (同) (三) (三)

OrbDB and Machine Learning joint work with Matt Kayala



- database of elementary reaction steps
- → machine learn elementary reaction favorability

- - E

-

Combinatorial explosion of the number of interactions to compare



- 116 unoccupied MOs
- 186 occupied MOs
- \rightarrow 18,875 potential non-trivial intra-molecular reactions

\rightarrow How do we efficiently reduce the number of reactions to examine?

PropOrb: Proposing Molecular Orbital Interactions



What do chemists do?

- Frontier Molecular Orbital theory
 - \rightarrow identify most reactive MOs
- Use:
 - MO type
 - Electronegativity
 - Charge / partial charge

→ Propose MO interactions in order, up to a preset threshold

Interactions between two adjacent MOs (migratory displacement, ionization...)



- Identify local electrophile:
 - carbocation p orbital
 - good leaving group σ^* orbital
- Identify adjacent nucleophile

Long Range Interactions





- Potential LUMO
 - empty p orbital
 - very acidic σ^* H-bonds
 - π* bonds
 - good leaving groups σ^* orbital
 - remaining acidic σ^* H-bonds
 - C-C σ*
- Potential HOMO
 - lone pair
 - π -bond
 - σ -bond

- from Reaction Explorer
- 4500+ test cases
- C, H, N, O and halide atoms, under ionic conditions: 3,277 systems for which the most favorable MO interaction is known

Preliminary Results



- threshold = 100
- run time = 7.5 minutes
- top interaction proposed in 76% of the cases

- threshold = 500
- run time = 24 minutes
- top interaction proposed in 83% of the cases

- Rapidly identify relevant MOs for reactivity
- Piece of a complete reaction prediction system
- Future work:
 - refine ordering heuristics aromaticity, HSAB, molecular orbital coefficients
 - expand coverage
 S and P chemistry, radicals

- Matthew A. Kayala and Jonathan H. Chen
- Prof. Pierre Baldi and the Baldi Lab
- OpenEye Academic Software License
- ChemAxon Acadedmic Software License
- NIH/NLM Biomedical Informatics Training Program
- NSF Grants 0321390 and 0513376
- The Dreyfus Foundation Special Grant Program in Chemical Sciences
- IBM PhD Fellowship