Extending SMILES to Encode Reaction Mechanisms

- Miguel Velez

Faculty Advisors: • Dr. Peter Gittins, Department of Chemistry • Dr. Jason Sawin, Department of Computer Science

This research was made possible in part by the Grants and Research Office

Introduction

The Simplified Molecular Input Line Entry System (SMILES), is a line notation that represents molecular structures using alpha-numeric characters. SMILES can also be used to represent chemical reactions, but they focus on the net rearrangement of atoms rather than the reaction mechanism. The reaction mechanism is valuable information in understanding how a reaction takes place. To address this limitation, we created the Simple Mechanism Of Reaction Encoding System (SMORES) to represent and understand the mechanisms of organic reactions.

SMORES

SMORES is an extension of the SMILES language. The grammar of SMORES language is:

| transform | molecule(s) ’>>’ mechanistic step |
| molecule(s) | SMILES |
| mechanistic steps | mtype, mechanistic steps | null |
| mtype | +b{class,class}, -b{class,class}, =b{class,class} |
| SMILES | a valid SMILES specification that uses explicit class tags |

We identified two types of electron movements in molecular reactions:

1. Heterokinetic in which a pair of electrons move together
2. Homokinetic in which two electrons move independently

These yield four types of mechanism steps:

- Heterokinetic making of bonds: molecules ’>>’ +b{class,class} | Heterokinetic breaking of bonds: molecules ’>>’ -b{class,class} 
- Homokinetic making of bonds: molecules ’>>’ =b{class,class} | Homokinetic breaking of bonds: molecules ’>>’ =b{class,class} 

We extended the Chemistry Development Kit (CDK), which is an open source Java library specific to computer science that can manipulate SMILES strings. Then, we developed the SMORES parser that applies mechanisms to the specified molecule.

- The user enters a SMORES String.
- The SMORES parser processes and creates a SMILES string and mechanism.
- The SMILES parser checks if the user entered a correct SMILES string and creates a SMORES molecule.
- The SMORES operation takes the SMORES molecule and mechanism and adjusts bonds and formal charges.
- A new SMORES molecule is created.
- The SMILES generator takes the new molecules and outputs to the user a SMILES string with the new structure of the molecules.

Future Work

- We developed a robust software that processes SMORES strings, which allows users to understand reaction mechanisms.
- It can process heterokinetic making/breaking of bonds, homokinetic making of bonds, and single movement resonance structures.
- It can also process multiple step mechanisms discretely.
- The CDK implies H atoms in molecules, which became an issue when H atoms were directly involved in a reaction or during homokinetic breaking of bonds.
- The CDK also changes the numbering of atoms during chemical reactions which does not allow us to process multi-step mechanisms in a single run.

Acknowledgements

I would like express my deep appreciation to Dr. Peter Gittins for his continuous guidance in the progress of this research work. I would also like to thank Dr. Jason Sawin for his valuable advice in the development of this work.

References