Extending SMILES to Encode Reaction Mechanisms

Miguel Velez[1], Peter Gittins[2], and Jason Sawin[3]
[1]Computer & Information Sciences, University of St. Thomas
[2]Chemistry, University of St. Thomas

Cheminformatics combines chemistry and computer science to represent and manipulate molecules in a digital environment. A standard way of representing molecules in a computer readable format is the Simplified Molecular Input Line Entry System (SMILES), a line notation that represents molecules using alpha-numeric characters. SMILES strings can be manipulated using software development kits such the Chemistry Development Kit (CDK), an open source Java library specific to computer science. These tools enable the comparison of molecules before and after a chemical reaction. However, they do not illustrate the underlying reaction mechanism, which is valuable information in understanding how a reaction takes place. We have created the Simple Mechanism Of Reaction Encoding System (SMORES), an extension of SMILES, which encodes reaction mechanism information.

Our SMORES extension consists of a SMILES string followed by an alphanumeric description of a reaction mechanism. We also extended the CDK in a Java program, to allow processing of SMORES inputs. The program allows a user to enter a SMORES string, which creates a model of the starting molecule. Then the program manipulates this molecule based on the specified mechanism. Finally, a model of the molecule produced by this mechanism is generated. This structure can be used to generate a SMILES string or a graphical representation