RULE-BASED GENERATION OF THERMOCHEMICAL ROUTES TO BIOMASS CONVERSION

Srinivas Rangarajan, Aditya Bhan†* and Prodromos Daoutidis $*
Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis

Summary

Cheminformatics and graph-theoretic techniques have been used to develop a reaction generator capable of handling homogeneous and heterogeneous reactions in the thermochemical routes to biomass conversion. This tool describes elementary reaction steps as reaction rules in a modular approach using the graph-rewriting technique. Chemical concepts describing molecules and reactions have been given a structured framework in terms of inheritance trees and other interdependencies. This tool can be used in exploring reaction pathways in biomass conversion as well as in modeling the reaction systems involved in the different stages of biofuels production.

Keywords

Reaction Path Analysis, Deployable software and databases for reaction engineering

Biorefineries are envisioned to produce both biofuels for transportation and raw materials for the organic chemicals industry in the future[1]. The design of a biorefinery, just as in the case of a petroleum refinery[2,3], will depend heavily on software tools for process analysis, design, development and optimization. Biomass conversion to fuels is a complex reaction system and can involve homogeneous reactions in the gas and liquid phase as well as catalytic chemistry[4]. Biomass contains oxygen in the ratio 1:1 with respect to carbon, in contrast to crude oil which is comprised predominantly of hydrocarbons. The high oxygen content of biomass implies that new chemical transformations have to be described for the conversion of biomass to fuels. The high variability of biomass feedstock in terms of chemical composition, results in a variable product distribution. These factors make it necessary that a rigorous as well as general tool is developed to describe reactions at the molecular level and specifically accommodate homogeneous/ heterogeneous chemistry and the chemistry of oxygen-containing molecules.

In this paper, a rule-based programming tool is presented that can generate such a complex reaction system and provide the user with the ability to handle the various types of chemistry typical in biomass conversion. This tool offers the flexibility in describing the reaction rules with varying levels of details, i.e., from a completely rule-based approach describing only the minimal chemical transformations required to a completely knowledge based approach comprising of detailed rules that describe a particular reaction. It is envisaged that this tool will assist both a chemist in exploring the reaction pathways in a reaction system to discover novel thermochemical routes in the synthesis of biofuels and a reaction engineer in modeling the different chemical processes in a biorefinery.

The automated reaction network generator developed in this research builds on the modeling strategies developed for pyrolysis[5], biochemical systems[6,7], combustion[8], hydrocarbon processing[9,10] and synthetic organic chemistry[11] and extends these by implementing algorithms from the domain of Cheminformatics to formalize the description of a reaction as a graph-rewrite system, thereby streamlining the method of reaction generation. The tool is composed of four parts - a system for molecular representation, a description of reactions, a reaction compiler to do the book-keeping and a user interface in the form of a Domain Specific Language.

The molecular representation involves the internal representation and storage of the molecule in a format amenable to the description of large reaction systems. The molecules are input, stored in memory and retrieved as a character string, employing an adapted form of the standard SMILES representation[12]. A unique string representation is obtained using the canonizing algorithm CANGEN[13] with an extended set of invariant node properties. The internal representation of a

* To whom all correspondence should be addressed, † abhan@umn.edu, $ daoutidi@cems.umn.edu

molecule is in the form of a molecular graph whose attributes are defined using the object oriented approach whereby the different chemical concepts associated with a molecule, such as Element, Atom, Fragment, etc. are given a structured framework in terms of a definite inheritance tree and other interdependencies. The nodal connectivity and the bond order are represented as an Adjacency List.

The reactions are described by a set of reaction rules, which define a reaction in three steps - determining the reactant pattern which consists of the set of atoms and bonds participating in a reaction, determining the constraints that govern the reaction, and describing the transformation operations. The reaction center is determined using a subgraph isomorphism algorithm proposed by Ullmann[14]. This algorithm finds the instances of matches of a subgraph representing the reactant pattern in the molecular graph of the reactant. The absence of matches indicates that the molecule cannot undergo the particular reaction while one or more matches imply the possibility of reactions of that molecule. The constraints describe the conditions imposed on the molecular structure, the individual atom environment and the size of the molecule. The structural constraints and the reactant pattern are represented using an adapted form of the standard SMARTS pattern representation system[15]. Complex boolean expressions of constraints can also be evaluated thereby improving the flexibility in defining constraints. The application of a reaction rule to a molecule is discarded even if one of the constraints is not satisfied. The transformation operations describe the changes to the attributes of the nodes and edges of the molecular graph. The result of the application of these transformation operations is the set of product molecules.

The generation of the reaction network is conducted by the reaction compiler, which also maintains the containers for storing all the molecules (both input and generated) and reactions involved in a reaction system. The reactions rules are applied to each one of the molecules and the new products that are obtained from a reaction are added to a list of unreacted molecules so as to generate the subsequent reactions of these molecules. Both unimolecular and bimolecular reactions are treated. In the latter case, all possible permutations of the chosen reactant molecule with every other molecule in the system is considered so as not to overlook any possible reaction.

The implementation of this reaction generation framework that mimics a graph rewriting system and that of a domain specific language (DSL) that will act as a user interface will be discussed. Specific examples from the domain of biomass conversion will be considered to illustrate the application of the reaction language and the different features of the tool developed.

References

(1) Schubert, C., Can biofuels finally take center stage? *Nature Biotechnology* **2006**, 24, (7), 777.

(2) Ghosh, P.; Hickey, K. J.; Jaffe, S. B., Development of a detailed gasoline composition-based octane model. *Industrial & Engineering Chemistry Research* **2006,** 45, 1, 337.

(3) Quann, R. J.; Jaffe, S. B., Structure oriented lumping - Describing the chemistry of complex hydrocarbon mixtures. *Industrial & Engineering Chemistry Research* **1992,** 31, (11), 2483

(4) Regalbuto, J. R., Cellulosic Biofuels-Got Gasoline? *Science* **2009**, 325, (5942), 822.

(5) Broadbelt, L. J.; Stark, S. M.; Klein, M. T., Computer-generated pyrolysis modeling -on the fly generation of species, reactions and rates. *Industrial & Engineering Chemistry Research* **1994**, 33, (4), 790.

(6) Li, C., Henry, C.S., Jankowski, M.D., Ionita, J.A., Hatzimanikatis, V. and Broadbelt, L.J., "Computational Discovery of Biochemical Routes to Specialty Chemicals", *Chem. Eng. Sci.*, **2004**, 59(22-23), 5051.

(7) Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D. and Broadbelt, L.J., "Exploring the Diversity of Complex Metabolic Networks", *Bioinformatics*, **2005**, 21(8), 1603.

(8) Warth, V.;Battin-Leclerc, F.; Fournet, R; Glaude, P. A.; Come, G. M.; Scacchi, G., Computer based generation of reaction mechanisms for gas-phase oxidation. Comput. Chem. 2000, 25, 541-560.

9(8) Prickett, S. E.; Mavrovouniotis, M. L., Construction of complex reaction systems .2. Molecule manipulation and reaction application algorithms. *Computers & Chemical Engineering* **1997,** 21, (11), 1237.

(10) Hsu, S. H.; Krishnamurthy, B.; Rao, P.; Zhao, C. H.; Jagannathan, S.; Venkatasubramanian, V., A domain-specific compiler theory based framework for automated reaction network generation. *Computers & Chemical Engineering* **2008,** 32, (10), 2455.

(11) Todd, M. H., Computer-aided organic synthesis. *Chemical Society Reviews* **2005,** 34, (3), 247.

(12) Weininger, D., SMILES, A chemical language and information systems. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **1988,** 28, (1), 31.

(13) Weininger, D.; Weininger, A.; Weininger, J. L., SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences* **1989,** 29, (2), 97.

(14) Ullmann J. R., An Algorithm for Subgraph Isomorphism. Journal of the ACM 1976, 23, 1, 31.

(15)http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html, Last Retrieved Sep13, 2009