

## Network generation and analysis of complex biomass conversion systems

Srinivas Rangarajan<sup>a</sup>, Ted Kaminski<sup>b</sup>, Eric Van Wyk<sup>b</sup>, Aditya Bhan<sup>a</sup>,  
Prodromos Daoutidis<sup>a</sup>

<sup>a</sup>*Department of Chemical Engineering and Materials Science, 421 Washington Ave.  
SE., Minneapolis, MN 55455, USA*

<sup>b</sup>*Department of Computer Science and Engineering, 200 Union Street SE  
Minneapolis, MN 55455, USA*

### Abstract

A modular computational tool for automated generation and rule-based post-processing of reaction systems in biomass conversion is presented. Cheminformatics and graph theory algorithms are used to generate chemical transformations pertaining to heterogeneous and homogeneous chemistries in the automated rule-based network generator. A domain-specific language provides a user-friendly English-like chemistry specification interface to the network generator. A rule-based pathway analysis module enables the user to extract and query pathways from the reaction network. A demonstration of the features of this tool is presented using Fructose to 5-Hydroxymethylfurfural as a case study.

**Keywords:** Network generation, pathways analysis, biomass conversion, domain specific language, Cheminformatics

### 1. Introduction

Thermochemical conversion of biomass is envisaged as a method to produce transportation fuels and chemicals from renewable resources (Huber et al. 2006). The underlying reaction network in biomass conversion systems, as in systems in conventional petrochemical refining, is large and complex, thus requiring sophisticated experimental and computational tools for analysis. Computational analysis of systems with complex reaction networks has traditionally involved the generation of a reaction network and the subsequent formulation of kinetic models based on theory, empirical methods, or parametric estimation from experimental data (Broadbelt et al. 2005). Such a methodology is required in biomass conversion too. However, biomass upgradation involves multiple potential chemistries of oxygenates – gas or aqueous phase homogeneous chemistries and heterogeneous acid, base, or metal catalysis (Huber et al. 2006). Therefore, a computational tool developed for biomass conversion should be capable of generating and analyzing these chemistries concurrently. Previously we presented Rule Input Network Generator (RING), an automated reaction network generator that can handle homogeneous and heterogeneous chemistries to generate the reaction network of a given reaction system (Rangarajan et al. 2010). In this paper, we present two additional modules – a reaction language as a user-specification interface and a post-generation pathways analysis module for extracting reaction pathways from the network. In conjunction with the reaction network generator, these two modules enable specification of user-determined chemical transformations, generation of the appropriate reaction network, and identification of desired pathways on the basis of

user-defined criteria. The conversion of Fructose to 5-Hydroxymethylfurfural (HMF) is presented as a case study for the generation and analysis of complex reaction systems using RING.

## 2. Computational tool for reaction network generation and analysis

Figure 1 shows the three-module structure of RING. Input into RING is in the form of a program written in the reaction language with English-like syntax containing the reactant information, reaction rules of the reaction system under investigation, and pathway queries. The language compiler then translates the program written in the language into a format recognized by the reaction network generator, which subsequently generates an exhaustive reaction network on the basis of the input rules. The pathway analysis module takes in the generated network to extract reaction pathways on the basis of the specified pathway queries. These three modules are discussed below.

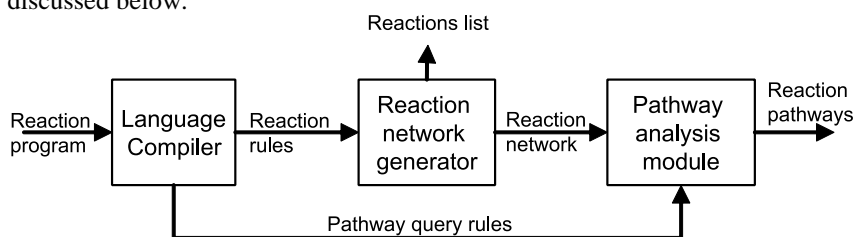


Figure 1: Overall structure of RING. Shown are the three basic modules, the inputs and outputs.

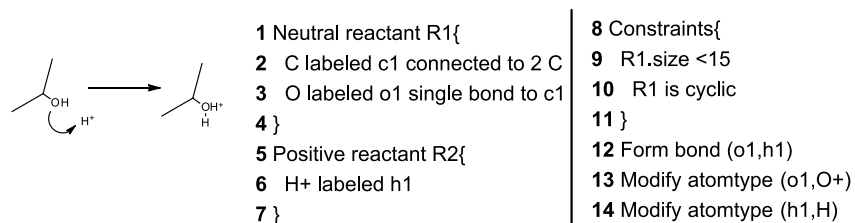


Figure 2: A reaction rule in the reaction language. A program will have multiple reaction rules

### 2.1. Reaction language compiler

RING has a new user-friendly rule input interface developed in the form of an English-like language with common organic reaction chemistry syntax, similar to the computational tools RDL (Prickett et al. 1997) and RDL++ (Hsu et al. 2008). This domain-specific language (DSL) has been developed using SILVER, a language specification tool based on attribute grammars, which permits easy modification and extension of the language's syntax (Van Wyk et al. 2010). A program written in this DSL consists of: (a) initial reactants written in SMILES (Weininger 1988) format, (b) a set of reaction rules, and (c) a set of statements for pathway extraction and querying of the reaction network. Figure 2 shows a reaction rule, pertaining to protonation of hydroxyl groups by a proton in acidic medium, written in the language. The reaction rule describes this step in a three step format consisting of: (a) identification of the reactant patterns, i.e., the atoms that participate in the step (lines 1 – 7), (b) constraints that have to be satisfied by each reactant (lines 8 – 11) to undergo the particular step, and (c) transformations on the atoms, i.e., changes in their electronic configuration and/or connectivity with each other (lines 12 – 14).

## 2.2. Reaction network generator

The working principle of the reaction network generator was described previously (Rangarajan et al. 2010). This module takes in as input the initial reactants and reaction rules pertaining to the reaction system from the language compiler and generates all possible species and reactions based on the rules using Cheminformatics and graph theory algorithms. Specifically, the reactants are converted to molecular graphs and the reaction rules are translated to graph transformation rules. A reaction is then abstracted as a graph transformation problem applied on the reactant molecular graph to generate the product graph(s). The network generator recursively applies these transformation rules on the products formed to generate additional species and reactions, until it exhaustively generates all species and reactions. The network generation feature of RING was tested on different biomass conversion chemistries including gas phase pyrolysis of fatty esters, base-catalyzed transesterification of triglycerides, and acid-catalyzed dehydration of fructose to 5-hydroxymethylfurfural (HMF). In all these cases, RING generated the pathways given in the literature, thereby demonstrating its versatility in handling a variety of organic chemical transformations.

## 2.3. Pathway analysis

The transformations occurring on a molecule to form various products can be analyzed either by extracting elementary-step acyclic pathways, or by identifying closed reaction sequences without generation/consumption of intermediates. The new pathways analysis module extracts, from the reaction network, the pathways leading from initial reactants to: (a) specified products, or (b) species satisfying certain user-specified queries in the form of size, shape, and structural features. The pathway generation algorithm employs a reverse depth-first traversal of the reaction network, without allowing directed cycles, starting from the product and traversing backwards up to the initial reactants. In addition, constraints can be prescribed on the length of a pathway, and the transformations that a pathway can/cannot have. Such features allow for extraction of specific pathways to analyze the transformations occurring in the reaction system. Section 3 presents a case study that elaborates on pathway analysis further.

## 3. Case study: Fructose to 5-Hydroxymethylfurfural (HMF)

In this section, a case study on acid catalyzed conversion of Fructose to HMF is presented. The reaction system comprises of an aqueous phase inorganic acid solution of Fructose with an organic phase to extract HMF to prevent further reactions of the HMF product (Antal et al. 1990 & Roman-Leshkov 2006). Therefore, reaction pathways leading to HMF are sought. The reaction network is generated using typical chemical transformations occurring in aqueous acidic solutions: (a) protonation of alcoholic groups, (b) dehydration of protonated alcohols, (c) hydride shifts and allylic rearrangements, and (d) deprotonation of carbenium ions to C=C or C=O groups. In addition, only the fructofuranosyl form (Antal et al. 1990) is considered. We previously presented the Fructose – HMF system as a part of a case study to demonstrate the use of RING for handling different chemistries. Herein, we analyze the pathways extracted from the reaction network.

The elementary steps discussed above were programmed in the reaction language and input into RING. Multiple simulation studies, each with different pathway extraction rules, were conducted to obtain pathways leading from fructose to HMF and the results are tabulated in Table 1. The number of pathways extracted increases exponentially with maximum pathway length. This is because longer pathways have more instances of

reversible steps (protonation/ deprotonation) and rearrangements (H shifts and allylic rearrangement) leading to isomers in a pathway. Since isomers can be generated at almost every step of the pathway, the number of extracted pathways increases exponentially. Many pathways differ from one another only because the species involved in one are isomers of those in the other or the order of occurrence of each elementary step is different. Hence, we subsequently sought distinct pathways on the basis of the frequency of occurrence of each elementary step in the pathway. Therefore, two pathways that have the same number of each elementary step are identified as identical pathways. It can be seen from Table 1 that the number of distinct pathways identified for the system is an order of magnitude lower than the total number of pathways. In addition, separate studies were conducted by imposing a limit of two on the number of occurrences of protonation of the C-C double bond, hydride shifts, and allylic rearrangements, thereby preventing unnecessary rearrangements. The number of pathways generated upon imposing these constraints is given within parentheses in Table 1. The successive application of more stringent constraints on pathway extraction resulted in the identification of five short, distinct pathways from fructofuranose form of Fructose to HMF with few rearrangements/ reversible steps. Figure 3 shows three of these distinct pathways. The pathway involving 2,5 anhydro-Mannose is the proposed scheme in the literature (Antal et al. 1990), while the other two pathways – one involving hydride shifts, and the other an allyl rearrangement – were identified using RING. The alternative pathways are of equivalent length as the pathway proposed in the literature; however, they do not have the proposed enol-aldehyde equilibrium.

Table 1. Pathway generation results for Fructose to HMF. In parentheses is the number of pathways identified when constraints limit the number of rearrangement steps in a pathway.

Pathway length	Number of pathways	Distinct pathways
$\leq 15$	947 (513)	38 (16)
$\leq 13$	144 (102)	19 (11)
$\leq 11$	22 (22)	5 (5)

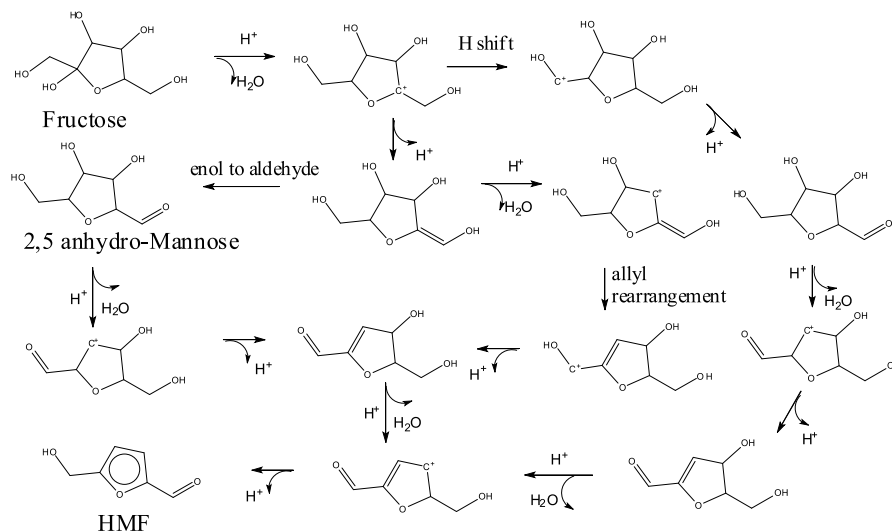


Figure 3: Reaction pathways for the conversion of Fructose to HMF. Note, not all elementary steps are shown completely.

#### 4. Discussion, conclusion, and future work

RING was used to describe the chemistry, generate the reaction network and extract reaction pathways for the Fructose-HMF system based on typical chemical transformations applicable in acid-catalyzed chemistry for aqueous solutions. RING identified the pathway proposed in the literature and generated alternative pathways. Reaction pathways that are generated by RING can be further evaluated by Density Functional Theory or ab initio calculations and semi-empirical techniques such as group contributions and linear free energy relationships. Specifically, the free energy profile, thus obtained, of a reaction pathway provides the free energy change of each step in the pathway. Such profiles can, therefore, be used to analyze plausibility of reaction steps in the system under consideration. For example, the alternative pathways identified using RING can be analyzed thermochemically by evaluating the free energy and the activation energy barriers to identify the bottleneck reactions in each pathway and/or the contribution of each pathway in the production of HMF. Thus, the problem of analyzing a large and complex reaction network can be reduced to a tractable problem of studying a relatively smaller subset of reactions that constitute the desired pathways. RING, therefore, provides a platform for the analysis of complex reaction systems. We are currently pursuing thermochemical analysis, and additional post processing features in RING, such as lumping of isomers to reduce the size of the reaction network and identifying direct mechanisms. Extension of the syntax of the language to support these additional features of RING is also being pursued.

#### Acknowledgement

This work was partially supported as part of the Catalysis Center for Energy Innovation, an Energy Frontier Research Center funded by the US Department of Energy, Office of Basic Energy Sciences under Award number DE-SC00010004. The authors also acknowledge financial support from the Initiative for Renewable Energy and the Environment (Large grant: RL-0004-09) and National Science Foundation (Award # 0905581).

#### References

- Antal, M. J. Jr., Mok, W. S. L., & Richards, G. N. (1990) Mechanism of formation of 5 hydroxymethyl- 2-furaldehyde from D-fructose and sucrose, *Carbohydrate Research*, 199, 91-109
- Broadbelt, L.J., & Pfaendtner, J. (2005) Lexicography of kinetic modeling of complex reaction networks, *AIChE J.*, 51 (8), 2112-2121
- Huber, G. W., Iborra, S., & Corma, A. (2006) Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering, *Chemical Reviews*, 106, 4044-4098
- Hsu, S.H., Krishnamurthy, B., Rao, P., Zhao, C.H., Jagannathan, S., & Venkatasubramanian, V. (2008) A domain-specific compiler theory based framework for automated reaction network generation. *Computers & Chemical Engineering*, 32 (10), 2455-2470
- Prickett S. E., & Mavrovouniotis, M. L. (1997) Construction of complex reaction systems I. Reaction description language. *Computers & Chemical Engineering*, 21, 1219 - 1235
- Rangarajan, S., Bhan, A., & Daoutidis, P. (2010) Rule-based generation of thermochemical routes to biomass conversion, *Ind.Eng.Chem.Res.*, 49 (21) 10459-10470
- Román-Leshkov, Y., Chheda, J. N., & Dumesic, J. A. (2006) Phase Modifiers Promote Efficient Production of Hydroxymethylfurfural from Fructose, *Science* 312 (5782), 1933- 1937
- Van Wyk, E., Bodin, D., Gao, J., Krishnan, L. (2010) Silver: An extensible attribute grammar system, *Science of Computer Programming*, 75, (1-2), 39-54
- Weininger, D. (1988) SMILES, A chemical language and information systems. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28 (1), 31-36