Computers and Chemical Engineering 46 (2012) 141–152

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Language-oriented rule-based reaction network generation and analysis: Applications of RING

Srinivas Rangarajan, Aditya Bhan*, Prodromos Daoutidis**

Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:
Received 28 October 2011
Received in revised form 4 April 2012
Accepted 5 June 2012
Available online 19 June 2012

Keywords:
Rule-based network generation
Reaction network analysis
Pathway analysis
Mechanism elucidation
Mechanism hypothesis testing
Synthesis routes prediction

ABSTRACT

Applications of RING in the generation and analysis of complex thermochemical reaction networks are presented. Automated generation and topological network analysis features in RING allow for: (a) constructing reaction networks exhaustively in a rule-based manner, (b) identifying dominant pathways in networks using estimates of kinetic parameters, (c) hypothesis and testing of mechanisms by comparing pathway results from RING with experimental data, and (d) predicting atom-efficient synthetic routes to valuable chemicals from known chemistries and commonly available chemicals. Case studies involving three chemical systems are used to demonstrate these features in RING: (a) acid-catalyzed propane aromatization, (b) glycerol and acetone dehydration on acid catalysts, and (c) C_4–C_9 mono-alcohols synthesis from C_2 and C_3 oxygenates on acid, base, and metal catalyzed chemistries. Through these case studies, we demonstrate that RING can be used to postulate mechanisms and predict likely products for a given system, thereby guiding experimentation and computational analysis.

1. Introduction

Large and complex reaction networks are common in chemical and biological systems. Analysis of such networks – mathematical modeling, mechanism/pathway identification, or topology identification – is challenging owing to the size and inter-connections of the network. Thus, computational techniques and tools have found significant applications in reaction network analysis, including estimation of thermochemical quantities (Broadbelt & Pfaendtner, 2005), kinetic modeling (Broadbelt & Pfaendtner, 2005; Ho, 2008), and biological network reconstruction (Feist & Palsson, 2008; Oberhardt, Palsson, & Papin, 2009). Rule-based automated network generators (Broadbelt, Stark, & Klein, 1994; Hsu et al., 2008; Prickett & Mavrovouniotis, 1997; Van Geem et al., 2006) recursively apply reaction rules to molecules, and products generated thereof, starting with initial reactants of the system. The result is an exhaustive network of all possible reactions and species that can be generated on the basis of the reaction rules. Kinetic models, typically, are then formulated (Broadbelt et al., 1994; Van Geem et al., 2006) to model the system under consideration.

In this work, we focus on generation and subsequent topological of analysis of complex reaction networks. Towards this end, we have developed a computational tool, RING (Rangarajan, Bhan, & Daoutidis, 2010; Rangarajan, Bhan, & Daoutidis, in press), which can construct the reaction network for a broad spectrum of chemistries on the basis of user-specified reaction rules and initial reactants. Further, RING can analyze the generated network in terms of identifying species and reaction lumps, mechanistic pathways, and overall reactions. RING can be used, in conjunction with experimental and computational data, to elucidate overall mechanisms, synthetic routes, and dominant reaction pathways in thermochemical reaction systems relevant for the synthesis of energy carriers and chemicals.

In the first paper of this two-part series, we presented the structure of RING, and its inputs and outputs in the context of an example involving dehydration of Fructose to HMF (Rangarajan et al., in press). In this article, we show how network generation and topological analysis features of RING can be applied together to analyze three broad classes of problems arising in complex thermochemical networks. Specifically, we use RING to: (a) analyze network characteristics of a complex reaction network (Section 2.1), (b) identify a set of plausible dominant pathways to specific products using computational and experimental data (Sections 2.2, 3.1, and 3.2), (c) hypothesize mechanisms to products in a complex network and propose potential experiments to discriminate between plausible ones (Section 3.3), and (d) identify potential atom-efficient synthetic routes to desired compounds from available starting reactants using a basis set of chemistries (Section 4).

* Corresponding author.
** Corresponding author. Tel.: +1 612 625 8818; fax: +1 612 626 7246.
E-mail addresses: abhan@umn.edu (A. Bhan), daouti001@umn.edu (P. Daoutidis).

0098-1354/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2012.06.003

Table 1
Reaction rules (elementary steps) for propane aromatization on HZSM-5.

Reactants: propane, acid site (H+)
Reaction rules
1. Adsorption of an alkane to form a carbonium ion
2. Desorption of a carbonium ion to form an alkane
3. Dehydrogenation of carbonium ions
4. Protolysis of a carbonium ion to yield a paraffin and a carbenium ion
5. Adsorption of olefins to form carbenium ions
6. Desorption of a carbenium ion to form an olefin
7. Beta scission of carbenium ions
8. Oligomerization of a carbenium ion and an olefin
9. Hydride transfer with paraffins and olefins as hydrogen donors
10. Hydride transfer with cyclic species as hydrogen donors
11. Cyclization with internal hydride shifts
12. Cyclization to form 5-carbon rings
13. Ring expansion of 5 C rings to 6
14. Alkylation of aromatics
15. Dealkylation of aromatics
Constraints (global): Forbid – vinylic carbenium ions, consecutive double bonds, linear trienes, and molecules of size ≥10 C atoms.

2. Analysis of network characteristics: propane aromatization

In this section, we show that, using a proposed set of elementary steps and their estimated kinetics relative to one another, RING can be used to: (a) reconstruct the complex network, (b) analyze the structural properties of the network such as frequency of occurrence of each reaction type in the network, (c) generate a smaller network of reactions by encapsulating a class of reactants such that information on chemical transformations is not lost, (d) identify a set of likely dominant pathways to major products, and (e) lump structural isomers that are functionally equivalent so as to reduce the overall size of the network further by several orders of magnitude. We illustrate these features for the system of acid-catalyzed aromatization of propane on ZSM-5 catalysts (Bhan et al., 2005).

Bhan et al. (2005) proposed a set of elementary steps for this system and further developed a microkinetic model for the system. Hsu et al. (2008) generated this network in RDL++, a network generator for constructing reaction networks for catalytic systems, using elementary steps from Bhan et al. (2005). We analyze this system further, in terms of methods described above. The reaction network is first constructed with the proposed elementary steps, following which pathways to benzene are identified and the network further reduced through lumping.

2.1. Network construction

The elementary steps of the system are shown in Table 1. Several of these (1–11) were considered by Hsu et al. (2008) and were also used here. Four more rules (12–15) have been added – cyclization to form five-membered rings, ring expansion to form six-membered rings, alkylation of aromatics, and dealkylation – on the basis of the elementary steps proposed by Bhan et al. (2005). The reaction rule inputs are given in the supplementary material (S1.1). For simplicity, the surface alkoxides formed upon adsorption of olefins are denoted as carbenium ions, Brønsted acid sites of the zeolite as "[H+]", and carbonium ion intermediates formed upon alkane activation are represented as "C*". These elementary step reaction rules were input into RING with propane as the initial reactant to generate the reaction network. Table 2 shows that the total number of reactions generated by RING increases four-fold upon increasing the global maximum allowed compound size constraint from five to six carbons.

Hsu et al. (2008) showed that the computational time and the number of reactions increase exponentially with the size of the largest molecule allowed and report an execution time of 48 h for generating the network with up to C_9 species that contained over one million reactions. In this section, we show how RING can be used to reduce the size of the network, yet preserve the information on transformations.

Table 2
Propane aromatization network size and reaction distribution of the four most common rules computed using RING.

Rule/network	Number of reactions (species)	
	Largest size 6	Largest size 5
Complete network	1791(199)	400(71)
Hydride transfer from rings	870	0
Hydride transfer from paraffins and olefins	429	232
Desorption	103	30
β-Scission/oligomerization	57	16

Table 2 also shows the distribution for the four most frequently occurring reaction rules in the network with a largest allowed molecule size of 6 carbon atoms. Hydride transfer steps account for a large majority of reactions (about 72%) because almost every surface intermediate (except the ones not satisfying the constraints of the reaction rule) can accept hydrogen from most paraffins, olefins, and naphthenes. The hydrogen acceptor, irrespective of the identity of the donor, undergoes the same transformation – it accepts a hydride to form a neutral gas phase species. As the size of the largest species increases, the number of hydride transfer steps naturally tends to increase exponentially to account for an even larger proportion (>70% for largest allowable molecule size greater than 6) of all reactions.

As we are interested in identifying the transformations occurring in the network, and not simulating the system using a kinetic model, hydrogen donor information can be encapsulated using a composite atom definition. A composite atom within RING is a user-defined entity used to represent: (a) an atom other than C, N, H, O, N, P, and S, or (b) a group of atoms considered together as a "unified atom". In this case, we define a "hydrogen transfer agent" (HTA), as a composite atom; the hydrogen donor then is (HTA) bonded to hydrogen, referred to as [{HTA}H] in a modified version of SMILES (Weininger, 1988). The hydride transfer rule can now be rewritten as [{HTA}H] losing a hydride ion to a carbenium ion to form [{HTA}+] (rule "HydrideTransfer1" in Scheme 1). [{HTA}H] and {HTA}+ represent all potential hydrogen donor and hydrogen acceptor species in the entire reaction network. We must therefore, write an additional rule that describes how [{HTA}+] can obtain a hydride from all potential hydrogen donors (H attached to a C in a neutral molecule) as shown in rule "HydrideTransfer2" in Scheme 1.

The original network can be restored by finding all combinations of pairs of these two rules and adding the reactions of each pair so that the common species – "[{HTA}H]" and "[{HTA}+]" – cancel out (see Section S2.1 of supporting material for more details on the method and syntax). In this sense, the encapsulation is invertible and the information on the actual transformations is not lost. Fig. 1 shows that the network size increases exponentially with the largest allowed compound size while Table 3 compares the network size and number of reactions for the case with and without encapsulation. As expected, two additional species – [{HTA}H] and [{HTA}+] – are present in the generated network in the case with encapsulation. A total of 20,000 reactions are generated when species up to C_9 are allowed using this method of encapsulation. This is considerably lower than the 1 million reactions generated using RDL++ where hydride transfer steps are explicitly enumerated. Encapsulation results in reducing the proportion of the hydride transfer reactions from about 72% to 25% of the network with a maximum allowed compound size of 6, thereby

```
//rule Hydride transfer from {HTA}H
rule HydrideTransfer1{
positive reactant r1{
   C+ labeled c1 }
neutral reactant r2{
   HTA labeled t1
   H labeled h1 single bond to t1}
modify atomtype (c1,C)
break bond (t1,h1)
form bond (h1,c1)
modify atomtype (t1,HTA+)  }
```

```
//rule Hydride transfer to {HTA}+
rule HydrideTransfer2{
positive reactant r1{
   HTA+ labeled t1}
neutral reactant r2{
   C labeled c1
   H labeled h1 single bond to c1}
break bond (c1,h1)
form bond (t1, h1)
modify atomtype (t1, HTA)
modify atomtype (c1, C+) }
```

Scheme 1. Code snippet for hydrogen transfer reaction rules with encapsulation.

Table 3
Network size, number of reactions, and execution time of propane aromatization network with/without encapsulation.

Max. compound size	With encapsulation				No encapsulation	
	Species	Reactions	Time (s)	% H-transfer steps	Time (s)	% H-transfer steps
5	73	221	20.6	24.0	24.9	58.0
6	201	658	140.6	25.2	190.4	72.5

accounting for lesser number of reactions generated as other reactions of the system are not affected. Further, for this network, there is a savings of 25% in execution time, which can potentially be much higher for larger networks. Thus, the definition of composite atoms within RING can reduce the network size and speed up the network generation significantly, and concurrently preserve the information regarding transformations.

2.2. Dominant benzene production pathways

Benzene, the major aromatic product in this system can be produced in several ways. A query, for instance, for the shortest (and up to one step longer) distinct (Rangarajan et al., in press) pathways to benzene identified more than 100 possible pathways. Not all of them are equally likely, and so it is desirable to identify the set of dominant ones. We can use experimental and computational data to provide RING with additional logical constraints that can prune down the set of pathways identified by RING. These constraints will determine if a particular pathway is to be explored or discarded.

Bhan et al. (2005) calculated activation energies and pre-exponential factors for elementary steps of the system using experimental data and computations for a simplified model of the system comprising of about 300 reactions. The kinetic parameters were obtained by Bhan et al. from: (a) parametric fitting to their experimental data, (b) computational chemistry calculations for certain reaction rules such as cyclization (Joshi, Bhan, & Thomson, 2004; Joshi & Thomson, 2005) and hydride transfer (Kazansky, Frash, & Van Santen, 1997a; Kazansky, Frash, & Van Santen, 1997b), and (c) parametric fitting to experimental data in the literature for rules such as β-scission/oligomerization (Buchanan, Santiesteban, & Haag, 1996). From this information, therefore, for any molecule, we can compare the rates of all possible reactions it can undergo with knowledge of rate constants and the concentration of any co-reactants at hand. We further use this to identify the dominant transformations of each species – molecule or intermediate, and therefore, can construct a set of dominant pathways leading to benzene. In this section, parameters as listed in Bhan et al. (2005) have been directly used.

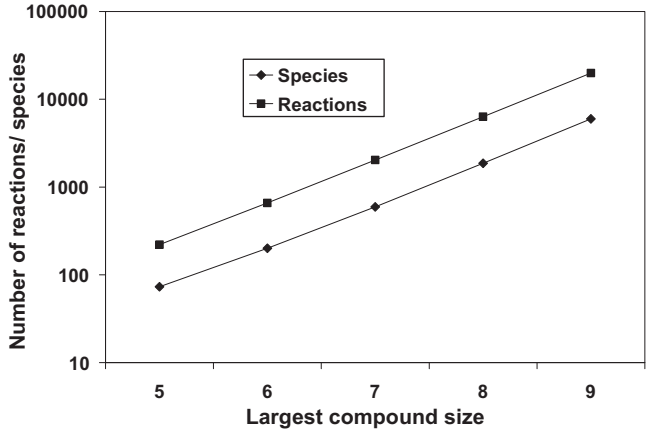


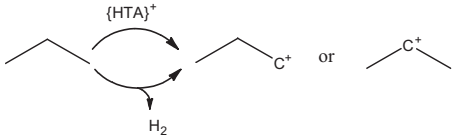
Fig. 1. Reactions and species count for acid-catalyzed propane aromatization with hydride transfer encapsulation as a function of largest allowed species size.

Table 4 shows the ratios of reaction rates of different reaction steps (at specific conditions) calculated with the kinetics data reported by Bhan et al. (2005). Carbenium ions (or surface alkoxides) can be formed from alkanes either by hydride transfer steps or by alkane adsorption followed by dehydrogenation. Table 4(i) reports the ratio of rates of these two steps for propane and hexane. It can be seen that the rate of hydride transfer to form a secondary carbenium ion can be up to an order of magnitude larger than dehydrogenation; the relative rate of hydride transfer to form a primary carbenium ion, however, is much smaller. Indeed, hydride transfer to form tertiary and secondary carbenium ions is several orders of magnitude faster than that to form a primary carbenium ion as seen in Table 4(iv). We assume an approximate coverage of 1.5% based on results of Bhan et al. (2005) and use exit concentrations in our calculations; in reality, however, the concentrations of species, including acid site coverage, vary along the bed and local concentrations are required to calculate relative rates at a specific location of the reactor. As a first approximation, nevertheless, we assume hydride transfer as the predominant step for the formation of secondary/tertiary carbenium ions, while primary carbenium ions are formed by alkane activation and dehydrogenation.

Alkyl carbenium ions can grow in chain size either by alkylation or by oligomerization. Table 4(ii) shows that oligomerization with ethylene and propylene (to form secondary carbenium ions) clearly dominates over alkylation. Alkylation with propane is higher compared to other paraffins (C_2H_6:C_3H_8:C_4H_{10} = 0.015:1:0.005) because of the higher partial pressure of the propane reactant. Therefore, the pathway query restricted alkylation steps to necessarily have propane as a co-reactant. An alkyl carbenium ion can:

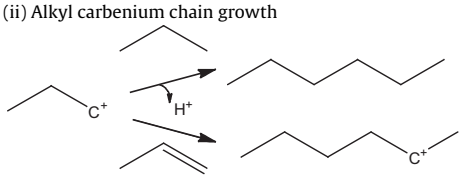
Table 4
Reaction rate ratios for different species in propane aromatization.[a]

(i) Alkanes → alkyl carbenium ion
Hydride transfer to form 2° (to form 1°): dehydrogenation[b]



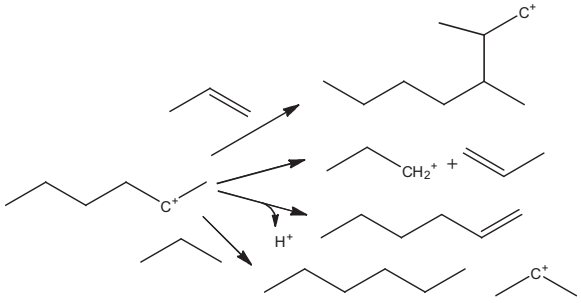
Propane – 6.68:(0.03:1)
Hexane – 12.45:1 (0.05:1)

(ii) Alkyl carbenium chain growth



Alkylation (with alkanes):oligomerization (with alkenes)	C_2H_6:C_3H_8:C_4H_{10}:C_2H_4:C_3H_6:C_4H_8
Propyl	0.015:1:0.005:3.5:6.54 (0.33):0.53 (0.026)[c]

(iii) Fate of alkyl carbenium ion



	Oligomerization (to 1°):β-scission (to 1°):desorption:hydride transfer
Hexyl	1:16.32:0.17:0.22

(iv) Reaction rate ratio of different steps

	Any → 1°:→2°:→3°[d]
Hydride transfer	0.004:1:12.20
Oligomerization/β-scission	1:20.00:60.59

[a] All values are calculated on the basis of kinetics information presented in Bhan et al. (2005). Reference temperature is 803 K, W/F ~ 2 g-cat h/mol.
[b] For the case of dehydrogenation, we assume that the alkane forms a carbonium-ion like species upon adsorption in a quasi-equilibrated step, the concentration of acid sites as calculated for Si/Al = 16 with 98.5% of sites vacant (reported in Bhan et al., 2005).
[c] Values in parentheses are relative oligomerization rates leading to primary carbenium ions.
[d] All ratios are reported on the basis of a secondary carbenium ion as a starting reactant.

(a) grow in chain size (by oligomerization primarily), (b) crack to form smaller molecules, (c) desorb to form an olefin, or (d) undergo hydrogen transfer. Table 4(iii) shows the relative rates of these four reactions for hexyl carbenium ions; β-scission is the dominant reaction. However, as expected, desorption is essential to all pathways and therefore, we do not discard the reaction rule. It can be argued that smaller carbenium ions (such as secondary propyl carbenium ions) cannot undergo β-scission and will likely oligomerize to form a larger carbenium ion or desorb to form the corresponding olefin. Hence, the rate of scission of hexyl carbenium ions can be compensated by the rate of oligomerization of smaller carbenium ions. Bhan et al. (2005) calculate a first order cyclization rate constant to be $9.7 \times 10^7\ s^{-1}$ while that of reference ($2^\circ \rightarrow 1^\circ$) β-scission rate constant is $516.4\ s^{-1}$. Therefore, if a molecule can cyclize (a carbenium ion with a carbon–carbon double bond three or four carbons away), it will do so rapidly. Therefore, we constrain the pathway detection algorithm to avoid oligomerization or β-scission of carbenium ions that can cyclize. Oligomerization and β-scission rates to form secondary and tertiary carbenium ions are larger than those to form primary carbenium ions as shown in Table 5(iv). Therefore, if a molecule can crack/oligomerize to either form a primary or a secondary carbenium ion, it would prefer to form the latter. This constraint is also considered for pathway analysis. Finally, dealkylation of toluene is reported to have an activation energy 30 kJ/mol higher (leading to a rate constant 90 times lower) than dealkylation of C_8 and C_9 aromatics (Bhan et al., 2005). Selectivity to toluene is also experimentally found to be higher than selectivity to xylene (Bhan et al., 2005; Lukyanov, Gnep, & Guisnet, 1995). Therefore, pathways to benzene through toluene were avoided. Three major classes of pathways to

Table 5
Pathways statistics for acid-catalyzed acrolein production from glycerol.

Length	Number of pathways			
	No constraints		With additional constraints	
	All	Distinct	All	Distinct
12	37,203	1385	156	16
10	909	161	45	8
8	22	12	6	3

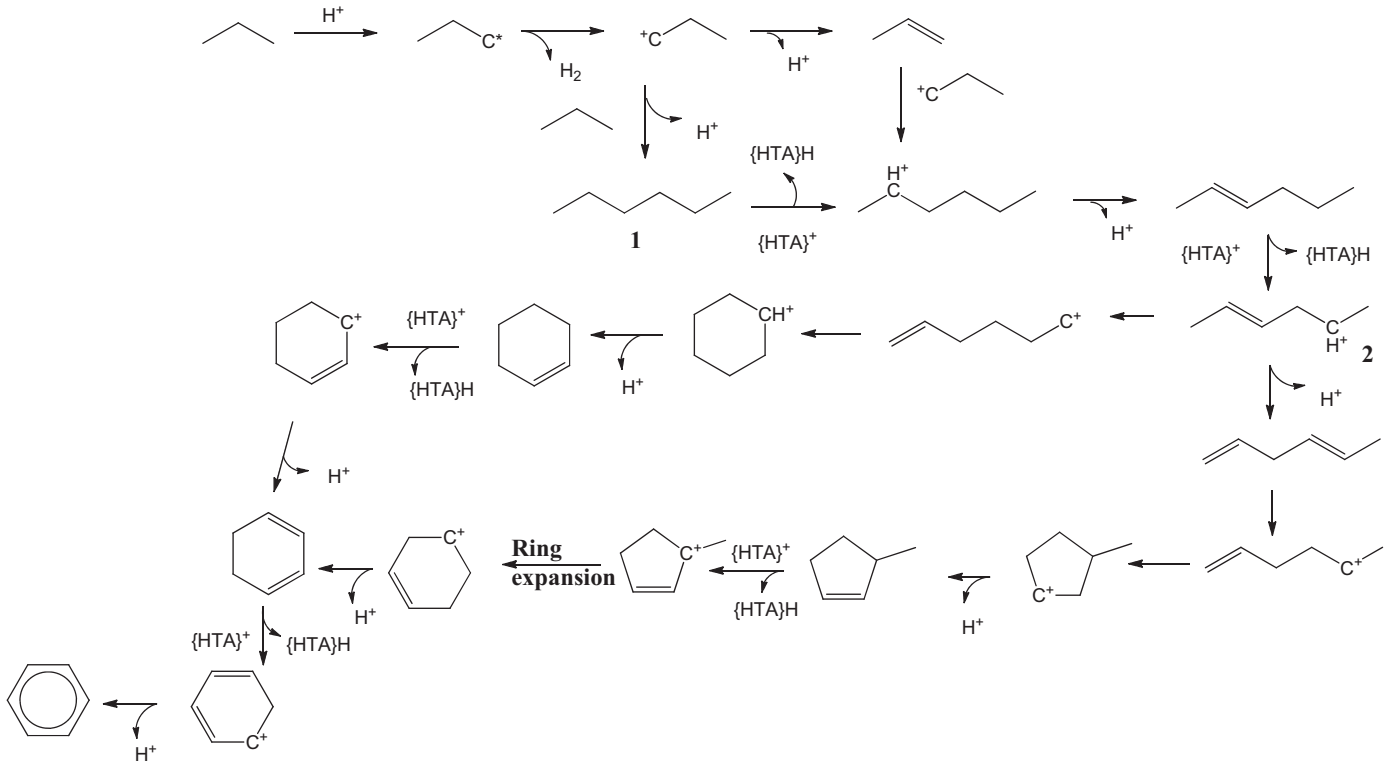


Fig. 2. Acid-catalyzed propane aromatization to form benzene involving cyclization of a C_6 species.

benzene formation exist depending upon the largest compound size in the pathway. Fig. 2 depicts pathways involving cyclization of C_6 species, which formed either by alkylation or by oligomerization, ultimately leading to species 2. Given that the rate of oligomerization is about seven times higher than the rate of alkylation and both routes lead to the same intermediate, we can conclude that species 2 is almost exclusively formed via oligomerization and subsequent hydride transfer. Cyclization, as shown, can happen by forming five- or six-membered rings. Species 1 or 2 (Fig. 2) can also lead to the formation of C_9 species, which ultimately undergoes cyclization and aromatization to form ethyl-methyl-benzenes (Supporting material S2.2). These C_9 aromatics can then undergo dealkylation to form benzene. A third route is through the formation of C_8 compounds from butene, which alternatively could also be formed from propene and ethane (Supporting material S2.2). The application of kinetic information as constraints resulted in identifying 10 dominant pathways from an initial set of over 100 pathways, leading to an order of magnitude reduction. Longer pathways could also be significant, and the exact contribution of each pathway can be found if the reaction fluxes are available from kinetic modeling.

2.3. Network lumping

The propane aromatization network, even upon encapsulation, contains about 20,000 reactions and close to 6000 species (Fig. 1). However, many species – both stable and ionic species – are functionally equivalent to each other; they have the same molecular formula and are composed of the same number and type of functional groups. These molecules have similar chemical properties owing to their chemical similarity and can, therefore, be grouped together into "lumps" of molecules. Lumps include both stable species and ionic intermediates. The number of lumps is smaller than the total species count which enables us to get a compact representation of the network because reactions can also be lumped by assuming them to be composed of lump(s) reacting to form a set of product lumps. The number of "lumped reactions" can be orders of magnitude lower than the original network size. In general, lumping requires taking into consideration functional group information and becomes more involved when several types of functional groups, such as in oxygenates, exist. We have developed a new lumping technique that identifies such functional equivalence (Rangarajan et al., in press). We further lump paraffins, olefins, naphthenics, and aromatics on the basis of molecular formula using an additional lumping instruction (Rangarajan et al., in press) as this is a common classification for hydrocarbons. In Rangarajan et al. (in press), we demonstrated how lumping reduced the network size by a factor of 2 for the fructose-to-HMF system, which contains significant amount of oxygenates. For the case of hydrocarbons, this reduction is even higher as discussed below.

Fig. 3 shows the size of the lumped network as a function of the largest allowed compound size. The number of reactions is of the order of a few hundred reactions, which is two orders of magnitude lower than the original network with encapsulation, and four orders of magnitude lower than the network without encapsulation based on the numbers reported by Hsu et al. (2008). This 100-fold reduction in network size leads to a lumped network that is more amenable to kinetic modeling. A lumped network could also potentially be a starting point for a corresponding pathway analysis with lumped reactions.

This example shows how, using RING, (a) a complex network can be constructed from elementary steps, (b) the network can be collapsed into one of a smaller size using encapsulation and inverted back to the original network, (c) the network can be grouped to obtain a lumped network of size that is two or more orders of magnitude smaller, and (d) likely shortest dominant pathways can be obtained from relative ratios of estimated kinetic parameters.

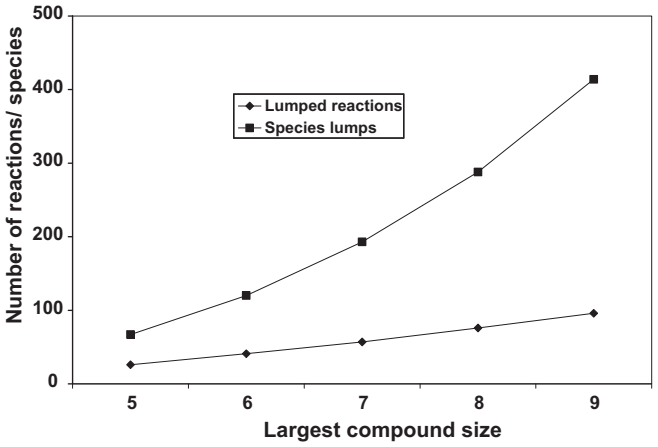


Fig. 3. A count of species lumps and reaction lumps with hydride-transfer encapsulation as a function of largest allowed species size for acid-catalyzed propane aromatization.

3. Mechanism hypothesis based on experimental data: glycerol dehydration

In this section, we show that RING can be used for testing postulated elementary steps and further propose new mechanisms for complex reaction systems. We use the system of acid-catalyzed glycerol dehydration for illustration. Corma, Huber, Sauvanaud, and O'Connor (2008) studied the conversion of glycerol–water mixtures to acrolein on ZSM-5 based zeolite catalysts. A complex spectrum of products, including acrolein, acetaldehyde, and a variety of acids, was experimentally observed. On the basis of reported mechanisms for oxygenates on solid acid catalysts, the authors suggest plausible pathways to certain products. Using RING, in Sections 3.1 and 3.2, we identify pathways to acrolein and acetone exhaustively on the basis of a set of postulated elementary steps, and further test their plausibility by noting if a predicted unique intermediate molecule is actually observed in experiments. Further, in Section 3.3 we hypothesize a mechanism for acetone conversion under the same reaction conditions using the same set of elementary steps.

In heterogeneous Brønsted-acid catalyzed conversion of oxygenates, the following elementary steps are commonly observed: alcohol adsorption–desorption, dehydration, alkene adsorption–desorption, hydride transfer, β-scission, oligomerization, 1,2-hydride shift, cyclization, etherification, and carbonyl adsorption–desorption. None of the experiments discussed by Corma et al. (2008) report the production of ethers. This is perhaps because ethers can decompose at the temperature of operation (300–500 °C) by the microscopic reverse of the etherification step. We, therefore, input the aforementioned steps into RING, with the exception of the etherification step. Carbon monoxide is observed in these experiments, indicating decarbonylation steps at these high temperatures. A microscopic reverse of the carbonylation step (Cheung, Bhan, Sunley, Law, & Iglesia, 2007) was, therefore, included. Water concentration used in the experiments by Corma et al. (2008) was considerably high (20–85 glycerol wt% feed); at these conditions, hydration is likely and is considered as a possible elementary step. We model this system such that surface species are represented as carbenium ions and acid sites are represented as [H+]. The hydride transfer steps are modeled by encapsulation, as discussed in Section 3. The adsorption of carbonyl compounds has been known to lead to acylium-like intermediates (Kresnawahjuesa, Gorte, & White, 2004); therefore, intermediates of carbonyl compounds are modeled to form carbenium ions. The inputs into RING are given in the supplementary material (S1.2).

The global constraints for this system are: (a) largest allowed molecule is 8 atoms (number of carbon and oxygen atoms), (b) consecutive double bonds are prohibited, (c) C=C+ is prohibited, and (d) gem-diols cannot be formed because they are highly reactive and form the respective ketone or aldehyde (Gomez-Bombarelli, Gonzalez-Perez, Perez-Prior, Calle, and Casado (2009) report that $\log K_{eq}$ of hydration of C_2–C_4 aldehydes and ketones vary from 0 to −2.7).

3.1. Acrolein production pathways

Acrolein is the major product of glycerol conversion. Pathways to acrolein from glycerol were queried in RING, the results of which are presented in Table 5. Initially, a query is set such that pathways of length up to 12 steps are sought with no further constraints. Several thousand pathways were identified by RING as reported in the table. The number of pathways was reduced to several hundred upon: (a) setting reversible rearrangements (e.g., Hshift) to less than two instances in the pathway so as to prevent several rearrangements in a pathway, and (b) preventing the cleavage of C—C bonds so that the carbon backbone of glycerol is preserved. Table 5 also lists the distinct pathways identified by RING for the cases with/without additional constraints. By imposing more stringent constraints, we progressively identify fewer pathways and obtain three pathways that can be analyzed manually.

Fig. 4 shows the distinct pathways to acrolein. All pathways involve 3-hydroxypropionaldehyde (O=CCCO), which suggests that it is a necessary intermediate and a primary product upon glycerol dehydration. Corma et al. (2008) report that acetol and 3-hydroxypropionaldehyde are observed as products at low conversions. They also infer that 3-hydroxypropionaldehyde is an intermediate in the formation of acrolein since a separate experiment with acetol as feed resulted in negligible acrolein production. Thus, our inference of a unique and necessary intermediate in acrolein synthesis pathways starting from glycerol, made using the exhaustive set of pathways obtained through RING corroborates experimental observation of that species.

3.2. Pathways to acetone

Corma et al. (2008) report significant acetone production when acetol was fed to the reactor. They also observed that temperature and conversion did not affect the selectivity of acrolein but affected all other products such as acetone, acetaldehyde, and higher oxygenates. They also found that when 1,2-propanediol was used as feed, it resulted in 60% carbon selectivity to acetone. On the basis of these observations, Corma et al. (2008), conjectured that acrolein and acetone are formed in parallel pathways.

A pathway query to acetone was input into RING with the constraint that the carbon backbone of glycerol is preserved (by ensuring no C—C scission). This resulted in less than 10 distinct pathways, some of which are shown in Fig. 5. Each of these pathways involved either acetol (hydroxyacetone), 1,2-propanediol, or their corresponding adsorbed surface intermediates. Thus, analysis through RING also shows that acetol or propanediol are primary products obtained upon dehydration of the terminal hydroxy group, while dehydration of the secondary hydroxyl group results in the formation of 3-hydroxypropionaldehyde hydroxypropionaldehyde and subsequently acrolein (Fig. 4).

The above analyses for acrolein and acetone formation (and that in Section S3.1 for pathways to acetaldehyde given in the supporting material) clearly show that pathways identified from the postulated elementary steps involve intermediates observed experimentally. This, therefore, suggests that the elementary steps considered for this study are adequate to describe the chemistry of the system. We can now postulate mechanisms to observed

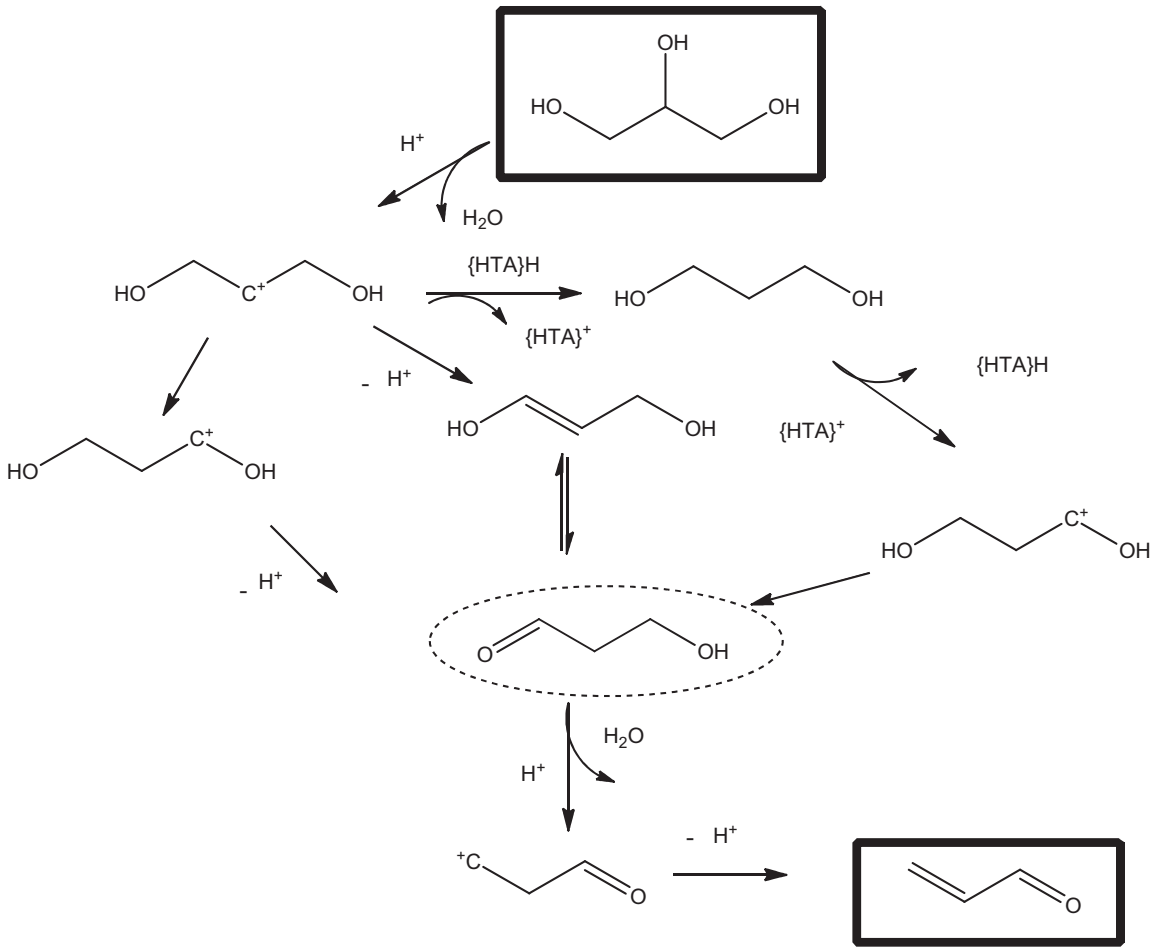


Fig. 4. Acrolein production pathways over solid acid catalysts. The 3-hydropropionaldehyde precursor to acrolein is highlighted by a dashed oval.

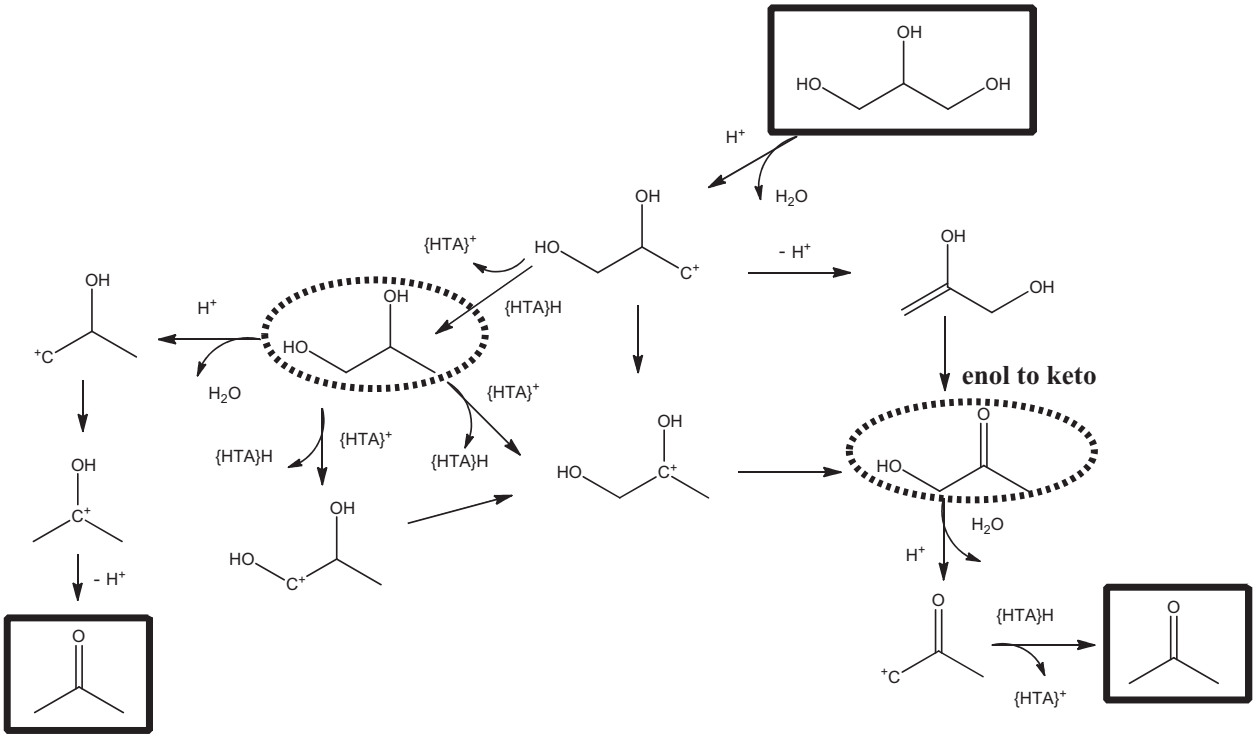


Fig. 5. Acetone production pathways from glycerol over solid acid catalysts (note all pathways require either hydroxyacetone or 1,2-propanediol).

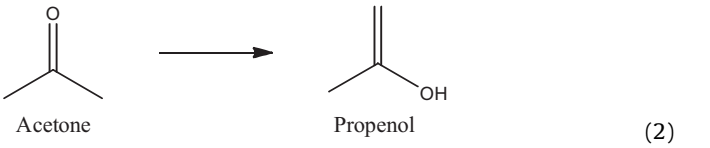
products for other oxygenate conversion systems operating under the same conditions as glycerol dehydration.

3.3. Potential mechanisms that result in acetic acid production from acetone

In a separate experiment, Corma et al. (2008) used an aqueous solution of acetone as feed at similar temperatures and WHSV as that of the glycerol conversion experiments. They reported butenes and acetic acid as the major products, leading them to postulate a stoichiometric reaction (1) to account for these two products:

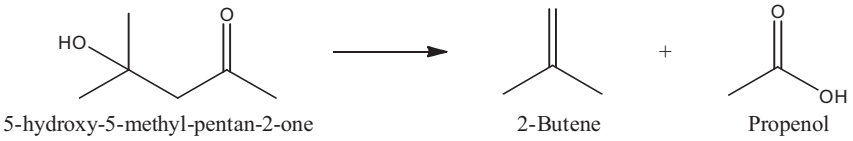
$$2\,\text{Acetone} \rightarrow \text{Butene} + \text{Acetic acid} \quad (1)$$

As the experimental conditions were similar, we generated a network for acetone conversion using RING with the same elementary steps as used for glycerol dehydration, and identified pathways to acetic acid from acetone (Fig. 6). Several routes exist for acetic acid production, all of which require the hydration of a surface carbonyl species and a larger C_6 molecule formed upon the addition of acetone with another acetone tautomer. Only one pathway (involving addition product '1') results in butene (2 methyl propene) as a by-product. This pathway involves dehydration and a hydration step, with no net additions. The catalytic cycles (2)–(4) and overall reaction (5) represent the transformation of acetone to acetic acid and iso-butene:

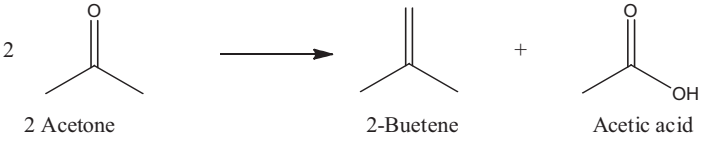


(2)

(3)



(4)



(5)

We note that the overall stoichiometric reaction based on our postulated sequence of elementary steps formulated within RING matches that proposed by Corma et al. (2008). The mechanisms involving hydrogen transfer involve the formation of butan-3-en-2-ol, which could further dehydrate to form butadiene. No alcohols or dienes are, however, reported suggesting that the mechanism without hydrogen transfer is the most likely one. Several experiments can be devised to test the hypothesis that the proposed mechanism is indeed the actual route to form butene. It is evident from the pathway shown in Fig. 6 that at low conversions the C_6 addition product should be observed and the rate of acetone conversion would be second order with respect to acetone. In addition to this, as we show below, we can simulate isotope tracking with RING to predict possible observations if isotope labeling experiments are conducted.

The acetone system was generated in RING with: (a) unlabeled water, (b) unlabeled acetone, (c) completely ^{13}C- and ^{18}O-labeled acetone, and (d) ^{18}O-labeled water; Fig. 7 shows the results of this simulation. The pathway marked "Pathway 1" shows the condensation of two unlabeled acetone molecules, followed by hydration with $^{18}OH_2$ leading to an ^{18}O-acid. This shows that an experiment with unlabeled acetone in ^{18}O-water solution should lead to ^{18}O incorporation into the acid product. Pathways marked "2" and "3", on the other hand, form a completely labeled acetic acid molecule with one or all atoms of butene being labeled. "Pathway 4" forms butene that has 3 carbons labeled. Therefore, if the mechanism proceeds with condensation, dehydration, scission, and hydration, we should expect that upon co-feeding equimolar concentrations of completely labeled and unlabeled acetone, the butenes formed should have equal distribution of zero, one, three, and four ^{13}C-labeled carbons. Further, Fig. 7 shows acetic acid is either completely labeled or unlabeled with respect to carbons, and will have ^{18}O incorporation when $^{18}OH_2$ is used.

Thus, RING has been used to: (a) propose and test a set of elementary steps to identify plausible pathways to observed products, by comparing pathway predictions of intermediates with that of experimentally observed ones, and (b) hypothesize mechanisms for product formation in related systems as well as suggest suitable experiments, especially isotope labeling studies, to validate it. In general, manual enumeration of possible routes to observed products is tedious for systems having a complex reaction network. RING, in such cases, can be used to identify these routes exhaustively, and can be a starting point for proposing and testing mechanisms.

4. Synthetic routes identification: alcohols from small oxygenates

A chemical compound can be synthesized in many ways depending upon the chemistries and starting reactant(s) used. This large number of possibilities could lead to a complex network of synthetic routes. Such a network can also be constructed using RING given a set of starting chemicals and a basis set of chemistry rules. Subsequently, potential products that can be synthesized as well as the corresponding synthesis routes can be identified by RING. We illustrate this through an example for synthesis of alcohols (butanols to nonanols) from C_2 and C_3 oxygenates (ethanol, acetaldehyde, acetic acid, propanal, acetone, 1-propanol, 2-propanol, and propanoic acid). A representative basis set of chemistries includes several acid catalyzed (dehydration, esterification), base catalyzed (esterification, aldol condensation, ketonization), and metal catalyzed (hydrogenation

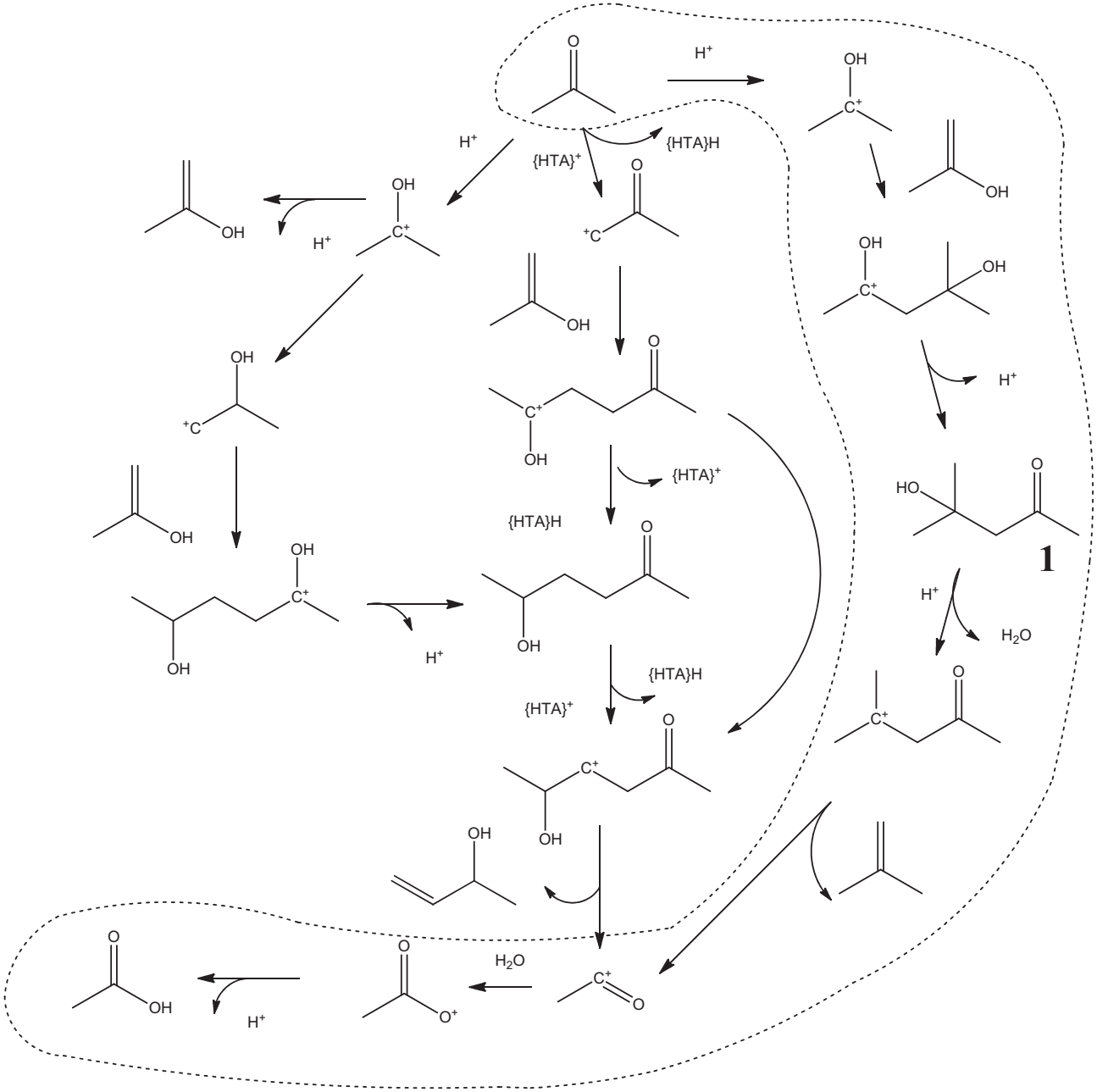


Fig. 6. Pathways identified by RING from acetone to acetic acid over solid acid catalysts. The proposed mechanism is enclosed within the dashed curve.

and dehydrogenation) reactions. We use RING to identify: (a) the potential spectrum of higher alcohols (primary/secondary/tertiary, linear/branched), for example, butanols through nonanols, which can be formed from these oxygenates, and (b) the most atom-efficient sequence of synthesizing these alcohols from the feed oxygenates.

The reaction rule inputs into RING are the overall stoichiometric reaction steps for each of the chemistries in the basis set (see supplementary material for the rules). For example, hydrogenation of ketones and aldehydes involves the carbonyl group and molecular hydrogen as reactants, and the transformations involve forming C—H and O—H bonds, breaking the H—H bond, and reducing the bond order of the carbonyl bond to a single bond. Thus, each reaction is potentially an independent reaction system that could be run in a separate reactor using independent temperature and pressure conditions. Such a network differs from the examples considered in Sections 2–4 in that the synthesis network is composed of overall stoichiometric reactions in contrast with elementary reaction steps considered so far.

4.1. Product spectrum

The generated network, comprising of about 4000 reactions, was queried for all alcohols produced. Table 6 shows the distribution of alcohols that were formed based on the given set of chemistries specified. It is evident that a large number of alcohols can potentially be formed from the given smaller oxygenates through the given set of chemistries. Furthermore, the spectrum of products is broad – from linear primary alcohols to branched tertiary alcohols – implying that the network is 'rich' and can be used as a starting

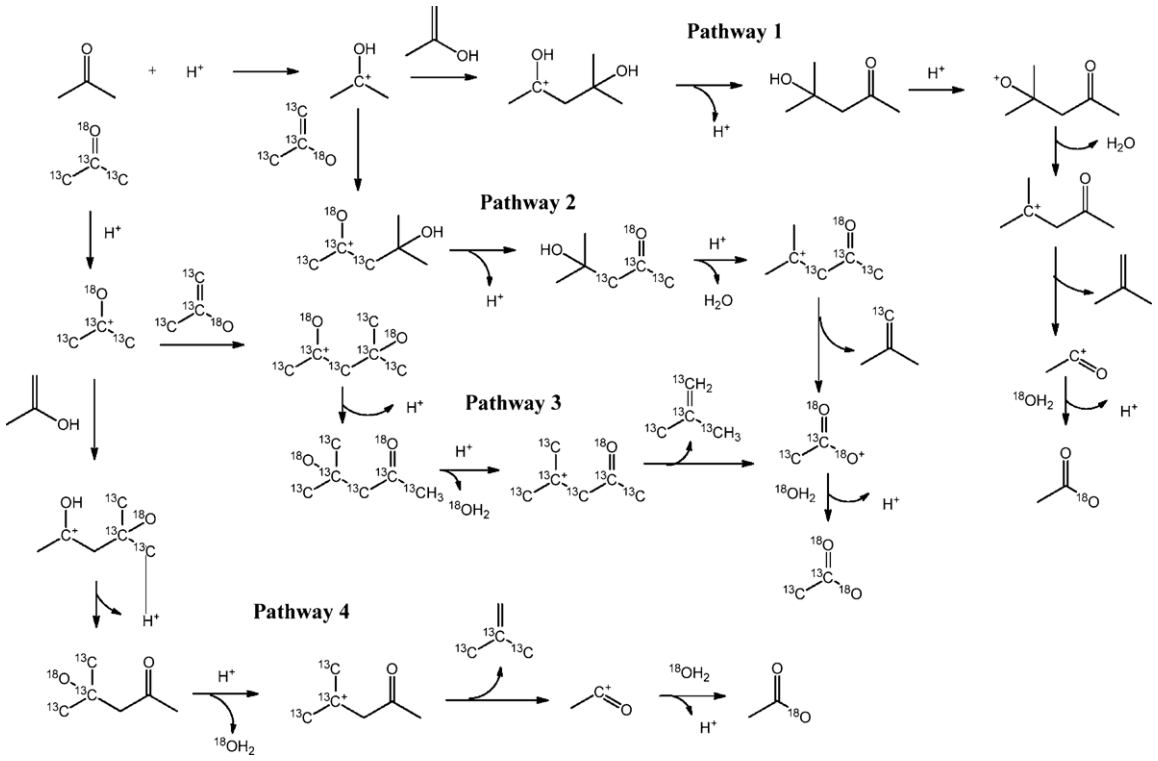


Fig. 7. ^{13}C- and ^{18}O-labeling simulations using RING for the conversion of aqueous acetone solution on zeolites.

point to find optimal synthetic routes to desired alcohols, within the specified chemistry basis set.

4.2. Atom-efficient synthesis steps to form mono-alcohols

The network generated using RING was queried for overall mechanisms to linear C_4–C_9 homologous alcohols up to 10 steps long. For each homologue, the overall reaction with highest atom efficiency to the desired alcohol was identified to select those synthetic steps that result in minimal by-products. The atom-efficiency value is easily obtained from overall mechanisms, but not from pathways, because the latter identify how a product is formed but not what the by-products are. Since RING generates all mechanisms (up to 10 steps in length), we can obtain the most atom-efficient synthesis routes by looking up the list of mechanisms output from RING and calculating the atom efficiency for each of them. Fig. 8 shows such routes to C_4–C_9 primary linear mono-alcohols. It is evident from the figure that all these products can be obtained starting from acetaldehyde upon multiple steps of aldol condensation and hydrogenation. Alcohols with odd number of carbons require aldol condensation with propanal once, while those with even number of carbons can be formed with acetaldehyde as the only reactant. This observation is expected given that no carbon is lost in aldol condensation. It is also evident that successive aldol condensation with acetaldehyde/propanal is required to maintain the aldehyde group at each step. The molar ratio of hydrogen to aldehydes is 1:1, implying that if aldehydes are derived from their respective alcohols (ethanol and propanol) the net reaction for synthesis of the C_4–C_9 linear primary alcohol involves no net hydrogen consumption. Since hydrogenation/dehydrogenation can be carried out on metal catalysts, and aldol condensation reactions occur over metal oxide catalysts, a series of steps with metal and metal oxide catalysts can achieve the synthesis of the C_4–C_9 linear primary alcohol homologues.

Similarly, atom-efficient synthetic pathways to other types of desirable products can be identified. For example, synthetic routes to secondary alcohols and those alcohols that can be formed through self-condensation alone (and hence in a higher selectivity) are shown in the supporting material (Section S4). The generated network is restricted by the input rules provided by the user and is not necessarily exhaustive in terms of including all plausible reactions. Alternative pathways involving other chemistries and having better selectivity/atom-efficiency for a given product could exist. However, since the network generated by RING is exhaustive for given inputs, this limitation can be addressed by having a comprehensive list of likely reaction rules. Thus, it would be possible to

Table 6
Spectrum of mono-alcohols formed from C_2 and C_3 oxygenates using RING.

Size	Primary	Secondary	Tertiary	Linear	Branched	Cyclic
5	1	1	0	2	0	0
6	3	2	0	3	2	0
7	7	5	0	3	7	0
8	19	14	2	6	23	0
9	48	35	9	11	67	1
10	37	39	0	6	70	0
	115	**96**	**11**	**31**	**169**	**1**

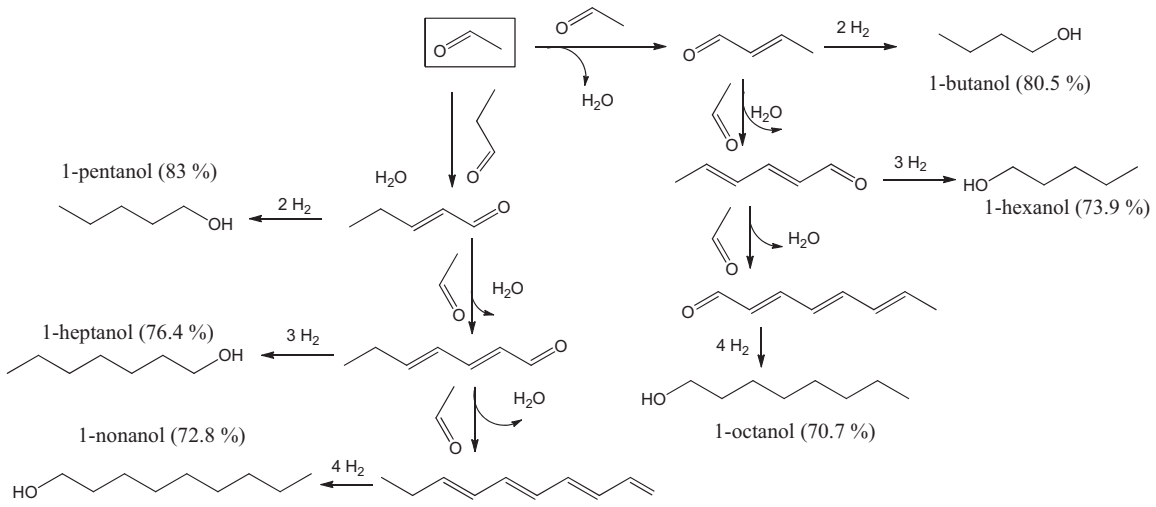


Fig. 8. Synthesis routes to form primary linear mono-alcohols. Atom efficiency values are given in parenthesis. Initial reactants are identified by enclosing them in a box.

determine all products that can be obtained from the given starting compounds and chemistries and how they can be synthesized in an optimal manner (optimality being any qualitatively defined parameter, such as allowing only self-condensations, etc.).

5. Discussion and conclusion

The three case studies illustrate the different classes of network analysis problems that can be addressed using RING. In the first example, propane aromatization, the elementary steps and kinetics are already established and we infer the characteristics of the network such as reaction distribution, combinatorial complexity, most abundant lumps, and likely dominant pathways to benzene. In the second example, we hypothesize a set of elementary steps for glycerol dehydration and compare pathways enumerated by RING with experimental data to validate these elementary steps. Further, we use the elementary steps to propose a mechanism for acetone dehydration. Therefore, these two cases are contrasting examples; however, both case studies illustrate that RING can be used in conjunction with additional data to elucidate the transformations to desired products occurring in complex reaction networks. Thus, we postulate that RING can be used along with experimental and computational studies so that data can be directly input as constraints in network generation or analysis.

Four characteristics of RING enable it to be applied in network analysis: (a) flexibility and versatility – a broad spectrum of chemistries can be handled as discussed previously (Rangarajan et al., 2010), several types of network analysis features (pathways, lumping, and mechanisms) are available, and a variety of options are available in the form of constraints on rules and post-processing steps that allow for including expert knowledge; (b) rule-based – the entire process of network generation and analysis is governed by rules, viz., reaction rules and post-processing instructions written in an English-language interface (Rangarajan et al., in press); (c) exhaustiveness – the generated network is complete and correct for the given inputs, and the pathways and mechanism algorithms traverse the network completely; and (d) speed – generation of a network takes seconds to hours depending on the nature of the reaction system and system specific modifications, such as the encapsulation discussed in Section 2, allows for further reducing the execution time. RING, as shown in the three case studies, can therefore be applied in several types of complex network analysis problems.

Acknowledgments

The authors would like to thank Prof. Eric Van Wyk and Ted Kaminski, Department of Computer Science and Engineering, University of Minnesota, for helpful suggestions on computer science algorithms and collaboration on developing the reaction language for RING. Financial support from the Initiative for Renewable Energy (Large Grant: RL-0004-09) at the University of Minnesota, the National Science Foundation Emerging Frontiers in Research and Innovation program, Grant #0937706 is gratefully acknowledged. AB was supported as part of the Catalysis Center for Energy Innovation, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Award number DE-SC0001004.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compchemeng.2012.06.003.

References

Bhan, A., Hsu, S.-H., Blau, G., Caruthers, J. M., Venkatasubramanian, V., & Delgass, W. N. (2005). Microkinetic modeling of propane aromatization over HZSM-5. *Journal of Catalysis*, *235*, 35–51.

Broadbelt, L. J., & Pfaendtner, J. (2005). Lexicography of kinetic modeling of complex reaction networks. *AIChE Journal*, *51*(8), 2112–2121.

Broadbelt, L. J., Stark, S. M., & Klein, M. T. (1994). Computer-generated pyrolysis modeling – On the fly generation of species, reactions and rates. *Industrial & Engineering Chemistry Research*, *33*(4), 790–799.

Buchanan, J. S., Santiesteban, J. G., & Haag, W. O. (1996). Mechanistic considerations in acid-catalyzed cracking of olefins. *Journal of Catalysis*, *158*(1), 279–287.

Cheung, P., Bhan, A., Sunley, G. J., Law, D. J., & Iglesia, E. (2007). Site requirements and elementary steps in dimethyl ether carbonylation by acidic zeolites. *Journal of Catalysis*, *245*, 110–123.

Corma, A., Huber, G. W., Sauvanaud, L., & O'Connor, P. (2008). Biomass to chemicals: Catalytic conversion of glycerol/water mixtures into acrolein, reaction network. *Journal of Catalysis*, *257*, 163–171.

Feist, A. M., & Palsson, B. O. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, *26*, 659–667.

Gomez-Bombarelli, R., Gonzalez-Perez, M., Perez-Prior, M. R., Calle, E., & Casado, J. (2009). Computational calculation of equilibrium constants: Addition to carbonyl compounds. *Journal of Physical Chemistry A*, *113*, 11423–11428.

Ho, T. C. (2008). Kinetic modeling of large-scale reaction systems. *Catalysis Reviews*, *50*(3), 287–378.

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.

Joshi, Y. V., Bhan, A., & Thomson, K. T. (2004). DFT-based reaction pathway analysis of hexadiene cyclization via carbenium ion interemediates: Mechanistic study of light alkane aromatization catalysis. *Journal of Physical Chemistry B*, *108*(3), 971–980.

Joshi, Y. V., & Thomson, K. T. (2005). Embedded cluster (QM/MM) investigation of C_6 diene cyclization in HZSM-5. *Journal of Catalysis*, *230*(2), 440–463.

Kazansky, V. B., Frash, M. V., & Van Santen, R. A. (1997a). A quantum-chemical study of hydride transfer in catalytic transformations of paraffins on zeolites. Pathways through adsorbed nonclassical carbenium ions. *Catalysis Letters*, *48*(1–2), 61–67.

Kazansky, V. B., Frash, M. V., & Van Santen, R. A. (1997b). A quantum-chemical study of hydride transfer in catalytic transformations of paraffins on zeolites. *Studies in Surface Science and Catalysis*, *105*, 2283–2290.

Kresnawahjuesa, O., Gorte, R. J., & White, D. (2004). Characterization of acylating intermediates formed on H-ZSM-5. *Journal of Molecular Catalysis A: Chemical*, *208*, 175–185.

Lukyanov, D. B., Gnep, N. S., & Guisnet, M. S. (1995). Kinetic modeling of propane aromatization reaction over HZSM-5 and GaHZSM-5. *Industrial & Engineering Chemistry Research*, *34*, 516–523.

Oberhardt, M. A., Palsson, B. O., & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, *5*. Article No. 320

Prickett, S. E., & Mavrovouniotis, M. L. (1997). Construction of complex reaction systems. 2. Molecule manipulation and reaction application algorithms. *Computers & Chemical Engineering*, *21*(11), 1237–1254.

Rangarajan, S., Bhan, A., & Daoutidis, P. (2010). Rule-based generation of thermochemical routes to biomass conversion. *Industrial & Engineering Chemistry Research*, *49*(21), 10459–10470.

Rangarajan, S., Bhan, A., & Daoutidis, P. Language-oriented rule-based network generation and analysis: Description of RING, *Computers & Chemical Engineering*, http://dx.doi.org/10.1016/j.compchemeng.2012.06.008, in press.

Van Geem, K. M., Reyniers, M.-F., Marin, G. B., Song, J., Green, W. H., & Matheu, D. M. (2006). Automatic reaction network generation using RMG for steam cracking of *n*-hexane. *AIChE Journal*, *52*(2), 718–730.

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.