JCTC Journal of Chemical Theory and Computation
Letter
pubs.acs.org/JCTC

Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry

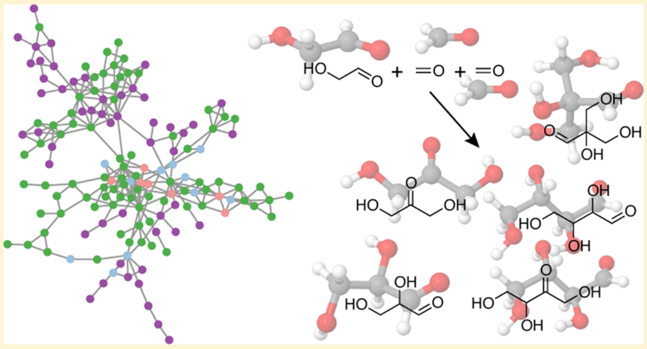
Dmitrij Rappoport,*[†] Cooper J. Galvin,[‡] Dmitry Yu. Zubarev,[†] and Alán Aspuru-Guzik*[†]

[†]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

[‡]Pomona College, 333 North College Way, Claremont, California 91711, United States

Supporting Information

ABSTRACT: While structures and reactivities of many small molecules can be computed efficiently and accurately using quantum chemical methods, heuristic approaches remain essential for modeling complex structures and large-scale chemical systems. Here, we present a heuristics-aided quantum chemical methodology applicable to complex chemical reaction networks such as those arising in cell metabolism and prebiotic chemistry. Chemical heuristics offer an expedient way of traversing high-dimensional reactive potential energy surfaces and are combined here with quantum chemical structure optimizations, which yield the structures and energies of the reaction intermediates and products. Application of heuristics-aided quantum chemical methodology to the formose reaction reproduces the experimentally observed reaction products, major reaction pathways, and autocatalytic cycles.



1. INTRODUCTION

Complex reaction mechanisms, in which many competing reaction steps combine to form a network of chemical reactions, are increasingly recognized as a common pattern in chemistry.[1,2] Characteristic features of complex reactions include branching and interference of reaction pathways, autocatalysis, and product inhibition and are observed in systems as varied as transition-metal catalysis,[3] cell metabolism,[4−6] and polymerization.[1,7,8] These features arise from the interplay between the properties of the elementary intermolecular reactions and the global topology of the reaction network. A better understanding of the network effects in complex reactions offers means for influencing their dynamics and product composition.[2,9,10] Studies of topological properties of reaction networks help characterize the chemical and functional characteristics of chemistries underlying cell metabolism and prebiotic chemistry.[4−6,11−14] A significant contribution to these efforts can be expected from theoretical works combining quantum chemical methods and network theory. Theory and computation of kinetics of elementary reactions from first principles have made enormous progress.[15−17] The topology of the network of cell metabolism has been extensively studied,[4−6] and specific network topologies have been associated with self-replication and origin of life.[13,14] However, our knowledge of topological structures occurring in chemical reaction networks is still very incomplete.

In the language of quantum chemistry, a complex chemical reaction is represented by its reactive potential energy surface, the minima of which correspond to reactants, intermediates, and products, while the first-order saddle points designate reaction barriers.[18] Exploration of energy landscapes has provided both a convenient conceptual framework and a common tool set for studying conformational dynamics, protein folding, and molecular rearrangements.[18−26] In particular, the mapping of the continuous potential energy onto a network representation allows analyses of topological and kinetic properties of these chemical systems to be carried out in an efficient manner. The goal of this work is to develop computational tools to explore potential energy surfaces involving unimolecular and bimolecular reactions and to study the topological properties of the corresponding reaction networks. The topography of the reactive potential energy surfaces must reflect the considerable structural changes associated with bond formation and bond breaking but also the very soft modes of relative molecular motion. Therefore, multiscale dynamical modeling would be required to encapsulate the wide range of characteristic time scales present in the reactive system.[27] Applications of multiscale models to reactive dynamics of systems relevant to cell metabolism and prebiotic evolution are computationally prohibitive at present.

On the other hand, chemical intuition suggests a conceptually quite different approach to describing reactive potential energy surfaces. It is well-known that chemical reactivity of organic compounds is well captured by a set of heuristic rules that represent chemical transformations as flows

Received: November 21, 2013
Published: February 7, 2014

of electron pairs ("arrow pushing").[28] The heuristic "arrow pushing" rules provide an expressive language for describing organic reaction mechanisms, especially of polar organic reactions and pericyclic reactions, which can be considered as a coarse-grained description of the reactive channels available on the potential energy surface. Quantum chemical optimizations of transition states support the reactivity patterns suggested by the heuristic rules.[29,30] Rule-based systems have a long tradition in organic synthesis since the pioneering work of Corey and Wipke over 40 years ago[9,31−37] and have been recently developed into a broad-spectrum synthetic tool by Grzybowski and co-workers.[38,39]

Reaction mechanism generation also has great utility in the field of combustion chemistry, see refs 40 and 41 and references cited therein. Detailed kinetic models of gas-phase combustion processes consisting of thousands of elementary reactions have been created using experimental data and, increasingly, quantum chemical results.[9,42−44] The major steps in the automatic reaction generation involve the enumeration of feasible reaction pathways based on heuristic rules[42,45−49] and the prediction of reaction rates using transition-state theory and quantum chemical calculations of activation barriers.[9,44,50] These developments are available via software development projects such as RMG[51] or EXGAS.[46]

In this work, we are primarily interested in topological properties of chemical reaction networks in solution, especially of the reaction networks related to cell metabolism and prebiotic chemistry. To investigate these systems, we developed a computational framework of heuristics-aided quantum chemistry (HAQC) that is suitable for efficient explorations of their reactive potential energy surfaces. The proposed methodology employs a combination of constructive heuristic rules, which are used to generate tentative reactive channels, and quantum chemical calculations to obtain the energies and structures corresponding to stable reactants, intermediates, and products. The constructive heuristic rules do not rely on known reaction mechanisms and are chosen to be generic representations of polar organic reactions as combinations of bond breaking and bond formation steps. Subsequently, the results of the quantum chemical calculations provide the energetic criteria for selecting the feasible reactive channels.

The reactive potential energy surface describes a closed thermodynamic system, which conserves the numbers of atoms but can vary its energy. The minima on the potential energy surface define the available system states and can thus be represented as nodes in a network with edges given by reactive transitions.[18] The resulting network representation is analogous to kinetic transition networks (TN), which were developed in the context of protein folding and conformational equilibria.[18−20,26] However, an individual energy minimum generally represents a collection of molecules in our case. Under the additional assumption of Markovian transition dynamics, one obtains a Markovian state model,[21−25] which lends itself to studies of global system dynamics and stationary states. In protein folding, the network nodes are obtained from dynamical trajectory data using clustering approaches, and the transition probabilities can be estimated from hopping statistics.

In chemical reaction networks, the transition probabilities follow from transition state theory and are defined by the corresponding reaction energy barriers. If the reaction mechanism is known, the barrier heights can be obtained by transition-state optimization procedures[17] or chain-of-states approaches.[16] These techniques have been recently used for reaction mechanism generation.[50] Discrete path sampling provides an efficient way of mapping out complex energy landscapes.[52−54] For simplicity, we model reaction barrier heights in this work by heuristic functions of the energies of the reactants, intermediates, and products along the reaction path. Our approach is motivated by Hammond's postulate, which holds that transitions states of reactions involving unstable intermediates resemble the intermediates themselves[55] or, alternatively, that the reaction energy and the height of the activation barrier are correlated with each other.[56] We show below that even simple heuristic kinetic parameters lead to useful predictions of reaction products and pathways.

We apply the HAQC methodology to the reaction network of the *formose* reaction, a well-studied organic reaction occurring in alkaline solutions of formaldehyde and resulting in a complex mixture of aldose and ketose sugars.[57,58] More than 40 compounds were experimentally identified as products of the formose reaction,[59,60] and major pathways are known;[58,61] however, many mechanistic details remain obscure. The formose reaction is one of the simplest organic reaction exhibiting autocatalysis[62] and was early conjectured as a potential route to sugars in the course of prebiotic evolution.[13,61,63,64] We present models of the formose reaction in different stoichiometries obtained using a combination of chemical heuristics and semiempirical quantum chemistry. Without *a priori* input of known reaction mechanisms, the formose reaction models predict formose sugars up to C_5 known from experiments[59,60] and major reaction pathways postulated in the literature.[56,62] The analyses of the network topology and energetics of the resulting formose reaction networks are detailed in our companion publication.[65]

This paper is organized as follows: Section 2 develops the framework of the HAQC methodology and heuristic thermodynamic and kinetic reaction feasibility criteria. Models of the formose reaction network in different stoichiometries are constructed and their chemical compositions are analyzed in section 3. A discussion and outlook are given in section 4.

2. CHEMICAL HEURISTICS FOR COMPLEX REACTION MECHANISMS

Many, if not most, hard problems in chemical structure and reactivity may be traced back to the high dimensionality of the quantum chemical models for electrons and nuclei. This is particularly true for complex reaction networks, which are characterized by having complicated potential energy surfaces with numerous energy minima. Furthermore, the reactive events on these surfaces are very rare, with energy barriers for bond breaking and bond formation being typically on the order of 10−50 kJ/mol for reactions in solution. The kinetic time scales for the formose reaction in alkaline solutions range from 10 min to few hours, depending on concentrations, temperature, and catalyst.[66] It is assumed that the prebiotic sugar formation via formose reaction took place on the surfaces of minerals over thousands or millions of years.[61,67,68] These time scales are presently out of reach even with the most powerful molecular dynamics methods.

Our goal in this work is to identify the reactions possible in the complex reaction mechanism and to analyze the topological structure of the corresponding reaction network. In order to circumvent the time scale problem, we make use of chemical heuristics to simulate motion along reactive trajectories, coupled with quantum chemical structure optimizations that enumerate the structures and energies of the reactants, intermediates, and products. We refer to the proposed methodology as heuristics-aided quantum chemistry (HAQC) in the following. The central assumptions of the HAQC approach are summarized and discussed below.

1. Reaction products and pathways are obtained by a set of heuristic *transformation rules*, which are recursively applied to structure formulas of molecules. We encode molecular structures by their simplified molecular-input line entry system (SMILES) representations.[69] The transformation rules used in this work are given in Scheme 1, where X, Y, and Z represent arbitrary atoms.

Scheme 1. Heuristic Transformation Rules for Polar Reactions Used in This Work

(a)	$X{=}Y + Z^+ \rightarrow X^+{-}Y{-}Z$	Electrophilic addition
(b)	$X{=}Y + Z^- \rightarrow X^-{-}Y{-}Z$	Nucleophilic addition
(c)	$X^+{-}Y^- \rightarrow X{=}Y$	Double bond depolarization
(d)	$X{-}Y \rightarrow X^+ + Y^-$	Single bond breaking
(e)	$X^+ + Y^- \rightarrow X{-}Y$	Single bond formation
(f)	$X^+ \smile Y^- \rightarrow \underset{\smile}{X{-}Y}$	Ring closure

We wish to stress that these primitive transformations are not required to describe genuine elementary reactions. Rather, they provide a simple device for constructing elementary reactions in an unbiased fashion and should capture the electron flow in polar organic reactions in aqueous solutions. The primitive transformations (a), (b), and (d)−(f) correspond to actual elementary reactions, while depolarization of multiple bonds (c) does not have an equivalent in quantum chemistry and is energy neutral.

2. The SMILES representations of the reaction intermediates and products obtained by way of heuristic transformations are mapped onto the corresponding three-dimensional structures and are subject to quantum chemical structure optimizations. In order to obtain a consistent description of the chemical structures that are part of the complex reaction network, a robust equivalence should be enforced between the structure formulas (given by SMILES) and the three-dimensional optimized structures from quantum chemistry. Therefore, we exclude all molecules, for which structure optimization does not preserve heavy-atom connectivity.
3. The heuristic transformation rules operate on molecular collections, which we refer to as *flasks* $\mathcal{F}_K = \{M_{K1},...,M_{Km_K}\}$ in the following. K is the flask index and M_{Kk} denotes the constituent molecules of flask K. We consider the molecular collection as a closed system and keep the numbers of atoms constant across flasks. As a consequence, flask energies are directly comparable to each other. Further, we assume that interactions between the molecules are negligible and thus flask energies are well approximated by sums of the energies of its constituent molecules $E_K = \Sigma_k \varepsilon_{Kk}$, which may be computed using any suitable quantum chemical method.
4. We distinguish between neutral and charged constituent molecules and label the flasks containing only neutral constituent molecules as *product flasks*. Assuming that the overall flask stoichiometry is conserved and the total charge is zero, we can expect the neutral forms of all constituent molecules to form in a sufficiently large number of transformation steps. Therefore, we may represent all stable reaction products as constituent molecules of product flasks without limiting the generality of the procedure. We utilize that polar reactions involve movement of electric charges between reaction participants, producing charged compounds as intermediates, and following Hammond's postulate, we make the additional assumption that the sequence of flasks containing one or more charged constituent molecules (*intermediate flasks*) may be considered as approximations to the instantaneous configurations along the reaction trajectory.
5. The recursive application of heuristic rules produces an auxiliary network representation containing both product flasks and intermediate flasks. (Figure 1a) The root node of the network is the initial flask $\mathcal{F}_1$, which is referred to as *generation* 0 of the network, and the *generation* $g > 0$ is obtained by combinatorially applying heuristic rules of Scheme 1 to all flasks of generation $g - 1$. Since multiple

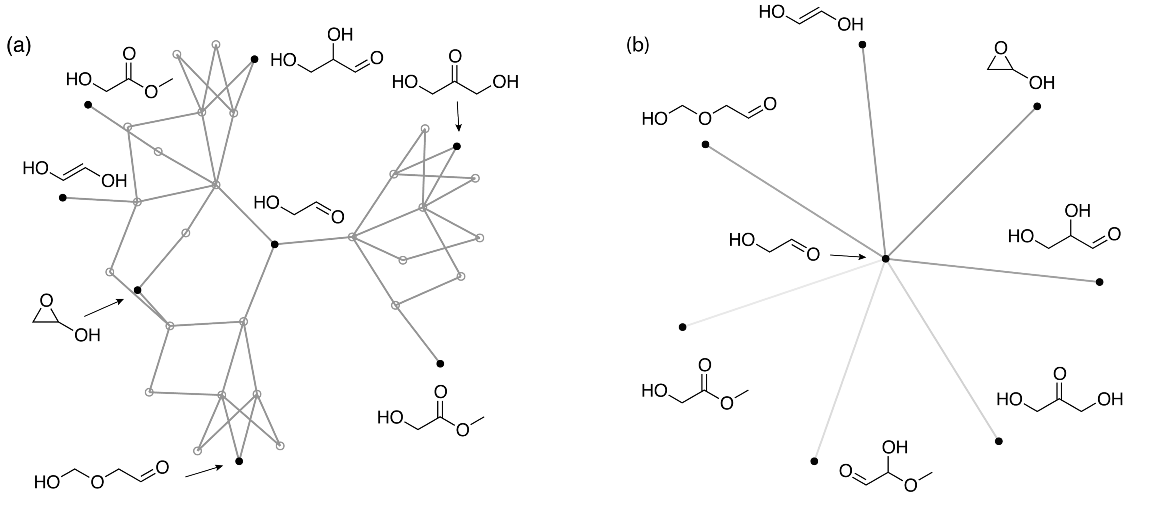


Figure 1. (a) Auxiliary network and (b) reaction network T_3 of formose reaction after three generations. Neutral flasks are indicated by black solid circles, intermediate flasks are shown by open circles. Chemical formulas denote the largest constituent molecule of each flask. Line intensities signify kinetic arc parameters of individual reaction steps; smaller arc values (more feasible reactions) are denoted by darker lines.

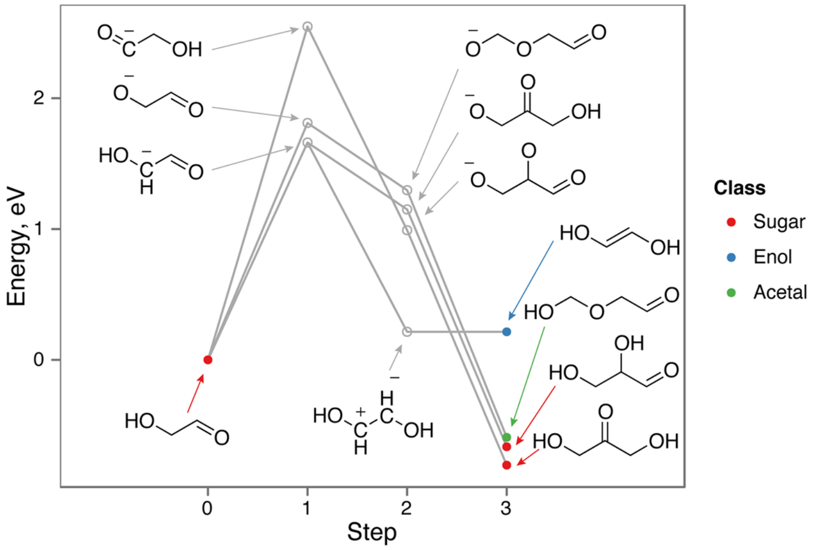


Figure 2. Selected energy profiles along 3-step chemical reactions between glycolaldehyde and formaldehyde ($\mathcal{F}_1 = \{O{=}CHCH_2OH, CH_2{=}O\}$). Product flasks are represented by solid circles, intermediate flasks by empty circles. Color coding and chemical formulas denote the largest constituent molecule of each flask (see legend).

paths may lead to the same flask, the auxiliary network representation is not a true tree graph. The *reaction network* is obtained from the auxiliary network representation by retaining only product flasks as *network nodes* and adding *network edges* based on the threshold criteria for thermodynamic and kinetic reaction parameters developed below. (Figure 1b).

6. We employ flask energies of product flasks and intermediate flasks to define thermodynamic and kinetic reaction parameters for transformations between product flasks $\mathcal{F}_K \rightarrow \mathcal{F}_L$. Energy differences between initial and final product flasks, $\Delta E_{K\rightarrow L} = E_L - E_K$, are natural choices for thermodynamic parameters and are independent of possible multiple pathways between $\mathcal{F}_K$ and $\mathcal{F}_L$. In addition, we develop heuristic kinetic reaction parameters, which take into account the flask energies of initial and final product flasks as well as the flask energies of the intermediate flasks connecting them. The heuristic kinetic reaction parameter $W_{K\rightarrow L}$ of a N-step transformation $\mathcal{F}_K \rightarrow \mathcal{F}_L$ should be a function of the flask energies $\{E_{K_i}, i = 0,...,N\}$ of the sequence of flasks $\{\mathcal{F}_{K_0} = \mathcal{F}_K, \mathcal{F}_{K_1},...\mathcal{F}_{K_N} = \mathcal{F}_L\}$, which is non-negative, additive for concatenated reaction sequences, and physically reasonable. We suggest *climb* parameter $W_{c,\,K\rightarrow L}$ and *arc* parameter $W_{a,\,K\rightarrow L}$ as heuristic kinetic parameters and assess their performance below. If multiple paths exist for the transformation $\mathcal{F}_K \rightarrow \mathcal{F}_L$ are present, we choose the *most feasible* path among them, which is defined by having the smallest heuristic kinetic reaction parameter.
7. Simple threshold criteria serve to determine thermodynamic and kinetic feasibility of transformations between product flasks. Only transformations $\mathcal{F}_K \rightarrow \mathcal{F}_L$ with $\Delta E_{K\rightarrow L} \le \Delta E^{\max}$ and $W_{K\rightarrow L} \le W^{\max}$ are added as network edges to the reaction network, where $\Delta E^{\max}$ and $W^{\max}$ are the thermodynamic and kinetic threshold constants, respectively.

The heuristic kinetic reaction parameters are motivated by Hammond's postulate and are designed to approximate reaction activation barriers. In the framework of transition state theory,[15] the activation barrier of a reaction is given by the energy of the transition structure, that is, the highest point along the reaction energy profile relative to the preceding energy minimum. Following Hammond's postulate, the highest-energy reaction intermediate is assumed to approximate the transition structure. For multistep reactions, the elementary reaction with the highest activation barrier determines the overall kinetics as the rate-limiting step. A convenient functional form for heuristic kinetic parameters is suggested by the following analogy: In thermal equilibrium, the abundance of flask $\mathcal{F}_K$ is given by the Boltzmann distribution, $c_K \propto \exp(-\beta E_K)$, in which $\beta = 1/(k_B T)$ with Boltzmann constant k_B and absolute temperature T. By analogy, we define heuristic kinetic parameters $W_{K\rightarrow L}$ for the reaction $\mathcal{F}_K \rightarrow \mathcal{F}_L$ in such a manner that the corresponding reaction rate may be represented as $k_{W\rightarrow L} \propto \exp(-\beta E_{K\rightarrow L})$.

The replacement of transition state calculations by heuristic kinetic reaction criteria based on Hammond's postulate is the most sensitive approximation in the proposed methodology. However, Hammond's postulate is based on a rich body of empirical data across different reaction types.[70,71] A theoretical justification for Hammond's postulate for elementary reaction steps can be obtained from catastrophe theory in the limiting case of small structure changes.[72]

The simplest approximation for the kinetic reaction parameter follows if we assume that the energy of the highest-energy intermediate flask approximates the activation barrier of the rate-limiting step. The corresponding kinetic *climb* parameter $W_{c,\,K\rightarrow L}$ for the N-step reaction $\mathcal{F}_K \rightarrow \mathcal{F}_L$ is given by

$$W_{c,K\rightarrow L} = \sum_{i=0,...,N-1} \max(E_{K_{i+1}} - E_{K_i}, 0) \qquad (1)$$

where we use the flask energies $\{E_{K_i}, i = 0,...,N\}$ as defined above. By definition, $W_{c,K\rightarrow L}$ yields the highest activation barrier of a multistep reaction relative to the initial flask $\mathcal{F}_K$.

In complex reaction mechanisms, a further consideration are branching points in reactive trajectories, which reduce the yield of each individual reaction product. Assuming that trajectory

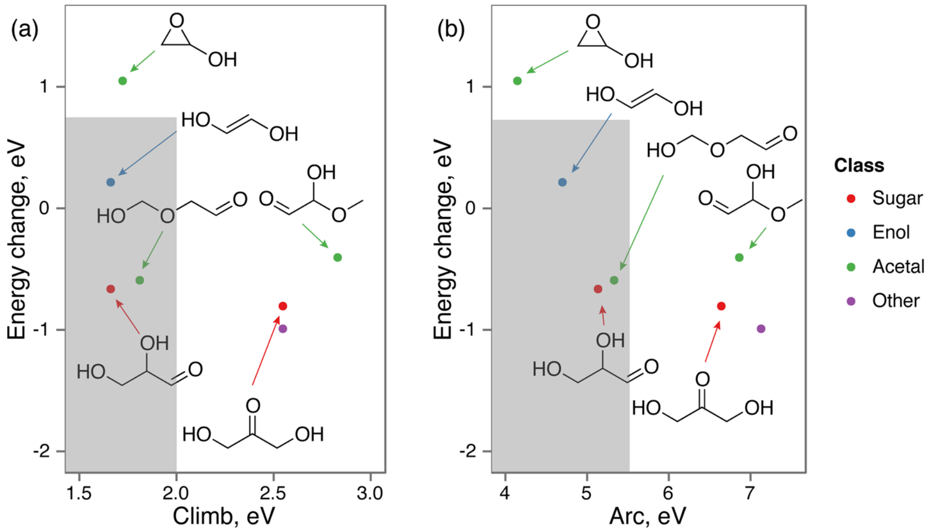


Figure 3. Thermodynamic and kinetic reaction parameters for formose reaction products in the T_3 network using (a) kinetic climb parameter W_c, (b) kinetic arc parameter W_a. Filled circles represent product flasks; color coding and chemical formulas denote the largest constituent molecule of each flask (see legend). The dark shaded areas depict the range of feasible reactions given by the threshold criteria for thermodynamic and kinetic reaction parameters.

bifurcations occur with a constant rate at each intermediate flask, the probability of reaching a given product flask decreases exponentially with the number of steps. Hence, it appears reasonable to use an energetic parameter that increases roughly linearly with the number of transformation steps, and we are led to define the kinetic *arc* parameter $W_{a,\,K\to L}$ for the reaction $\mathcal{F}_K \to \mathcal{F}_L$ as

$$W_{a,K\to L} = \sum_{i=0}^{N-1} ((E_{K_{i+1}} - E_{K_i})^2 + \alpha^2)^{1/2} \qquad (2)$$

where α is an empirical parameter and the flask energies $\{E_{K_i}, i = 0,...,N\}$ are defined as above. We can consider α as a penalty factor for long paths and set $\alpha = 1$ eV for the purposes of the following discussion. Finally, we note that by introducing heuristic kinetic parameters $W_{K\to L}$ we have made a tacit assumption of a constant pre-exponential factor A for all feasible reactions. The value of A sets the overall time scale of the reaction kinetics.

We can calibrate the kinetic climb and arc parameters and assess their performance using experimental knowledge of constituent processes of the formose reaction. We employ the heuristic rule set of Scheme 1 and use the OpenBabel structure builder to convert SMILES strings to three-dimensional models.[73−75] The energies are determined throughout this work by structure optimizations using the PM7 semiempirical method within the MOPAC package.[76] Solvation effects in water are included using the conductor-like solvation model (COSMO)[77] with an effective dielectric constant of $\varepsilon = 78.4$. COSMO is a widely used implicit solvation model that models the electrostatic effects of the solvent by a dielectric cavity and is closely related to polarizable continuum models (PCM).[78] Alternative methods for representing molecules in solution include explicit solvation models, quantum mechanics/molecular mechanics (QM/MM) approaches,[79] as well as empirical solvation models.[80]

We consider reactions involving one molecule glycolaldehyde and one formaldehyde molecule ($\mathcal{F}_1$ ={O=CHCH$_2$OH, CH$_2$=O}, Figure 2). The reaction mechanisms predicted on the basis of generic heuristics for polar organic reactions correspond to well-established reaction routes: (i) enolization of glycolaldehyde to ethene-1,2-diol (product indicated by blue circle in Figure 2),[81,82] (ii) aldol addition of glycolaldehyde and formaldehyde to form glyceraldehyde[83,84] (product in red), and (iii) hemiacetal formation (product in green).[85] As suggested by Hammond's postulate, the intermediate flasks, shown by empty circles in Figure 2, trace the movement of charge in reactions i−iii in fairly good approximation. The last step of the enolization describes a fictitious depolarization of the C=C double bond and is energy neutral. Reaction paths i and ii share the enolate anion as the highest-energy intermediate flask and thus have the same climb parameter $W_c = 1.66$ eV, while their arc parameters are different: $W_a = 4.70$ eV (enolization) and $W_a = 5.13$ eV (aldol addition). An additional reaction, (iv) a C−C coupling reaction via an aldehyde anion is predicted to occur at larger values of kinetic parameters ($W_c = 2.75$ eV, $W_a = 7.45$ eV). While the reaction product of iv, dihydroxyacetone (shown in red in Figure 2), is more stable than the products of reactions i−iii, the larger values of kinetic parameters are consistent with the experimental finding that deprotonation of an aldehydic proton is unfavorable and requires *umpolung* techniques.[86] In contrast, we expect the enolate-based reactions (i and ii) as well as the hemiacetal formation (iii) ($W_c = 1.81$ eV, $W_a = 5.33$ eV) to be feasible in aqueous solution.

In order to investigate the performance of kinetic climb and arc parameters in more detail, we consider the predicted formose reaction products after 3 and 6 generations starting from the flask $\mathcal{F}_1$ = {O=CHCH$_2$OH, CH$_2$=O, CH$_2$=O} (tetrose stoichiometry). We denote the resulting reaction networks as T_3 and T_6, respectively. Using suitable threshold values for either kinetic climb parameter ($\Delta E^{\max} = 0.75$ eV, $W_c^{\max} = 2.00$ eV) or kinetic arc parameter ($\Delta E^{\max} = 0.75$ eV, $W_a^{\max} = 5.50$ eV), we are able to generate a classification of feasible/unfeasible reactions for the T_3 network that agrees with empirical expectations outlined above (Figure 3).

The differences between the threshold criteria based on the kinetic climb and arc parameters are noticeable in the T_6 network (Figure 4). A large number of compounds can be

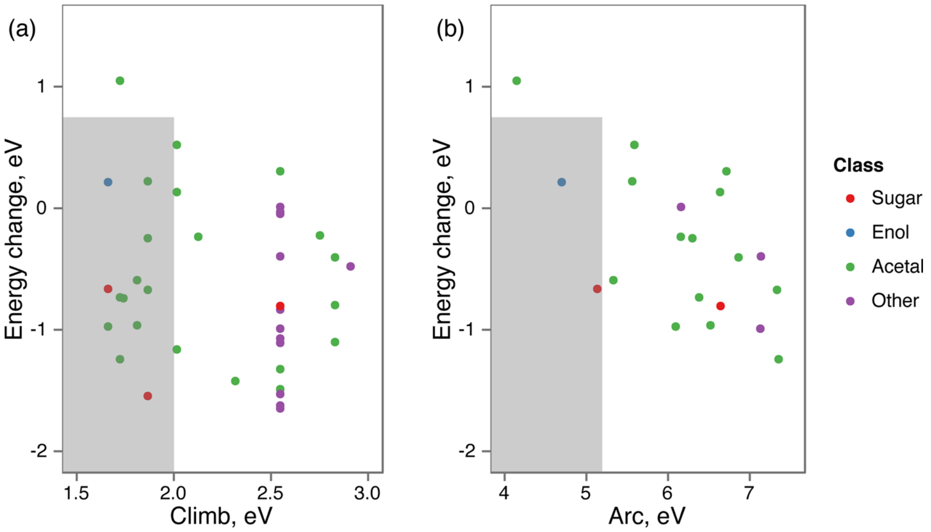


Figure 4. Thermodynamic and kinetic reaction parameters for formose reaction products in the T_6 network using (a) kinetic climb parameter W_c, (b) kinetic arc parameter W_a. See Figure 3 for details.

reached directly from the initial flask via a comparatively low barrier (small values of W_c); however, for many of them this is possible only by way of a long sequence of intermediate flasks. A simple threshold criterion using kinetic climb parameter does not treat this problem adequately and therefore does not permit a simple feasible/unfeasible classification for multistep reactions. The kinetic arc parameter exhibits a more desirable behavior: By penalizing long transformation sequences, it spreads the parameter distribution such that a consistent set of threshold criteria ($\Delta E^{\max}$ = 0.75 eV, $W_a^{\max}$ = 5.50 eV) remains useful over longer sequences of steps. These criteria are used throughout this work. The detailed results are given in section S1 of the Supporting Information.

A potential weakness of the kinetic arc parameter is that it does not distinguish between the forward and backward reactions. However, since we only apply simple threshold selection criteria, this is unlikely to significantly affect our conclusions. Nevertheless, it is desirable to develop kinetic parameters that are irreversible and show linear increase with number of steps. This effort will require considering a wider range of chemical reactions and is reserved for future work. The accuracy of our predictions are limited by the choice of the heuristic kinetic parameters as well as systematic errors of quantum chemical calculations. Furthermore, we disregard the stereochemistry and conformation equilibria of the formose products in this work. We expect the effects of the latter approximations to be small compared to the errors related to heuristic kinetic parameters and simple threshold criteria. The deviation from experimental results due to these challenges will be addressed in future work.

3. PROBING THE CHEMISTRY OF THE FORMOSE REACTION NETWORK

The formose reaction is a self-condensation of formaldehyde in alkaline solutions[57,58] and at surfaces of various minerals.[61,67,68] The presence of autocatalytic cycles[62] and the mechanistic parallels to sugar metabolism led to conjectures that it played an important role in the prebiotic formation of sugars.[13,63,64] The product mixture of the formose reaction was analyzed by multiple groups and more than 40 reaction products were identified to date.[58−60]

We investigated the structures and properties of formose reaction networks obtained after 9 generations starting from the initial flasks $\mathcal{F}_1$ = {O═CHCH$_2$OH, CH$_2$═O, CH$_2$═O} (tetrose stoichiometry, denoted by T_9) and $\mathcal{F}_1$ = {O═CHCH$_2$OH, CH$_2$═O, CH$_2$═O, CH$_2$═O} (pentose stoichiometry, P_9). Since the heuristic heuristic transformation rules used in the network construction preserve flask stoichiometry (Scheme 1), the nodes of the resulting network representation correspond to the possible *states* of the reactive system with fixed stoichiometry. The resulting TN representation is analogous to kinetic transition networks introduced previously in the contexts of conformational dynamics and protein folding.[20,23−26] It should be contrasted with the commonly used representations of metabolic networks as *interaction networks*, in which individual metabolites are network nodes and network edges connect all reaction participants with each other.[5,6,87,88] The TN representation of the reaction network is a directed network with edge weights given by thermodynamic and/or kinetic parameter values. In the following, we only consider the out-component of the reaction network reachable from the initial flask $\mathcal{F}_1$.

The T_9 network contained a total of 149 nodes, including 146 distinct neutral molecules, and 445 edges. The chemical composition of the T_9 network is shown in Table 1. The graphic representation of the network was created by the open-source Cytoscape program[89] using a force-directed algorithm followed by minimal manual adjustments (Figure 5). The product flasks were characterized by the chemical class of the largest constituent molecule as sugars, enols/enediols, acetals/hemiacetals, or other. (Table 1) The predicted formose products included 2 trioses (glyceraldehyde and dihydrox-

Table 1. Chemical Composition of T_9 and P_9 Networks

	T_9	P_9
sugars	6	11
acetals	78	235
enols	9	9
other	53	99
total	146	354

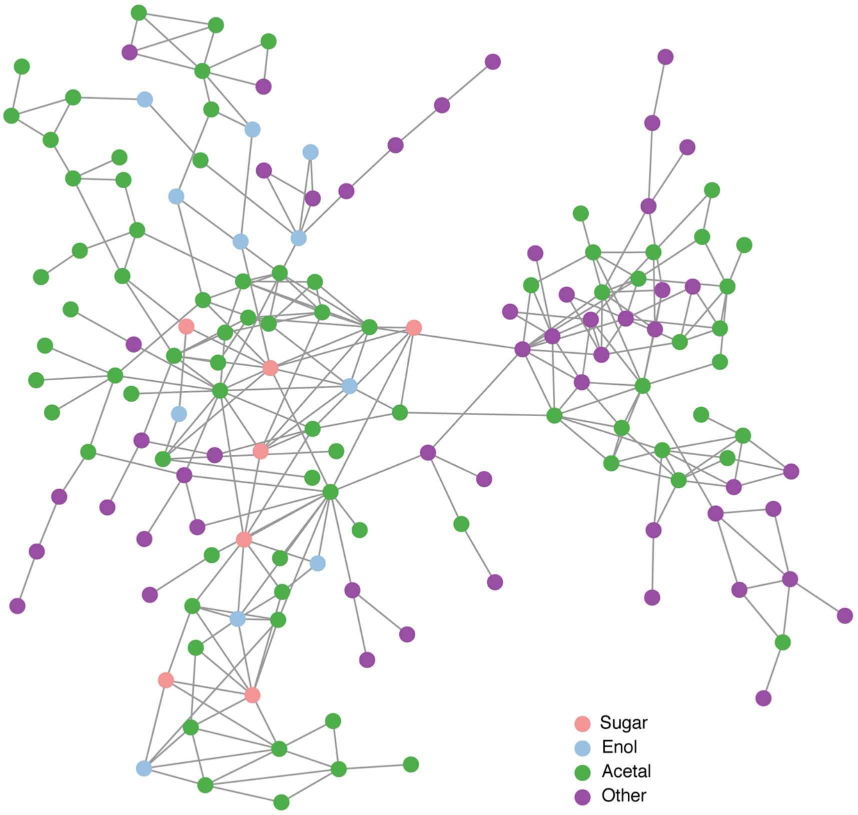


Figure 5. TN representation of the T_9 network. Filled circles represent product flasks. Color coding and chemical formulas denote the largest constituent molecule of the respective flask (see legend). Black solid lines indicate major pathways of sugar formation.

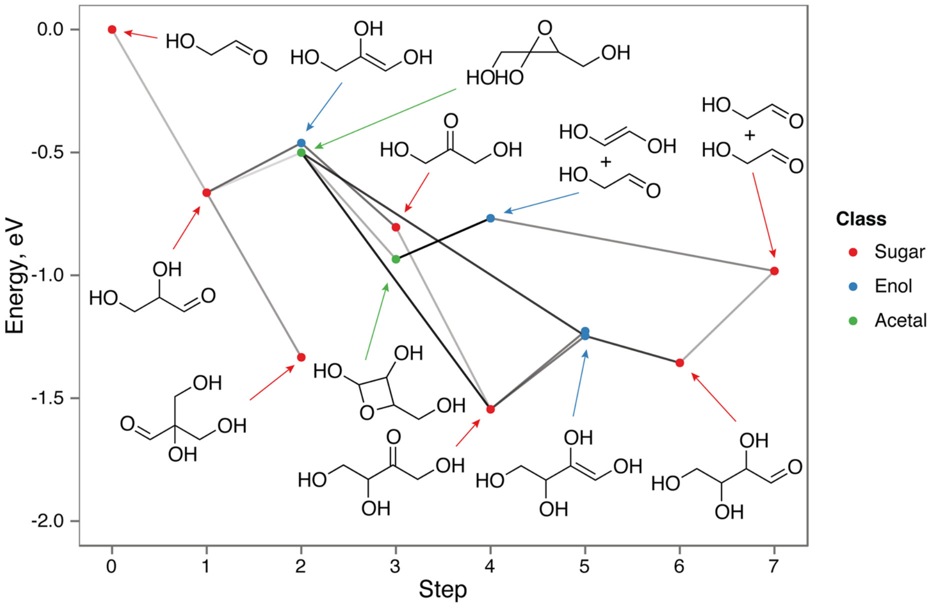


Figure 6. Major reaction pathways of sugar formation in the T_9 network. See Figure 5 for details. Line intensities signify kinetic arc parameters of individual reaction steps; smaller arc values (more feasible reactions) are denoted by darker lines.

yacetone) and 3 tetroses (aldotetrose, ketotetrose, and the branched tetrose 2,3-dihydroxy-2-(hydroxymethyl)propanal), which were experimentally identified in the formose reaction mixture.[59,60]

Major reaction pathways of sugar formation were determined by minimizing the sum of the kinetic arc parameters along the path from $\mathcal{F}_1$ to the sugar-containing product flasks (Figure 5). Pathways predicted in this way included mechanisms previously postulated for the formose reaction.[58,61] The central pathway of carbon-chain elongation was found to involve sequences of aldol additions[83,84] and aldose−ketose isomerizations[82] (Figure 6). As discussed above, glyceraldehyde was formed by aldol condensation of glycolaldehyde and formaldehyde, while subsequent aldol condensation with another molecule of formaldehyde yielded the branched tetrose 2,3-dihydroxy-2-

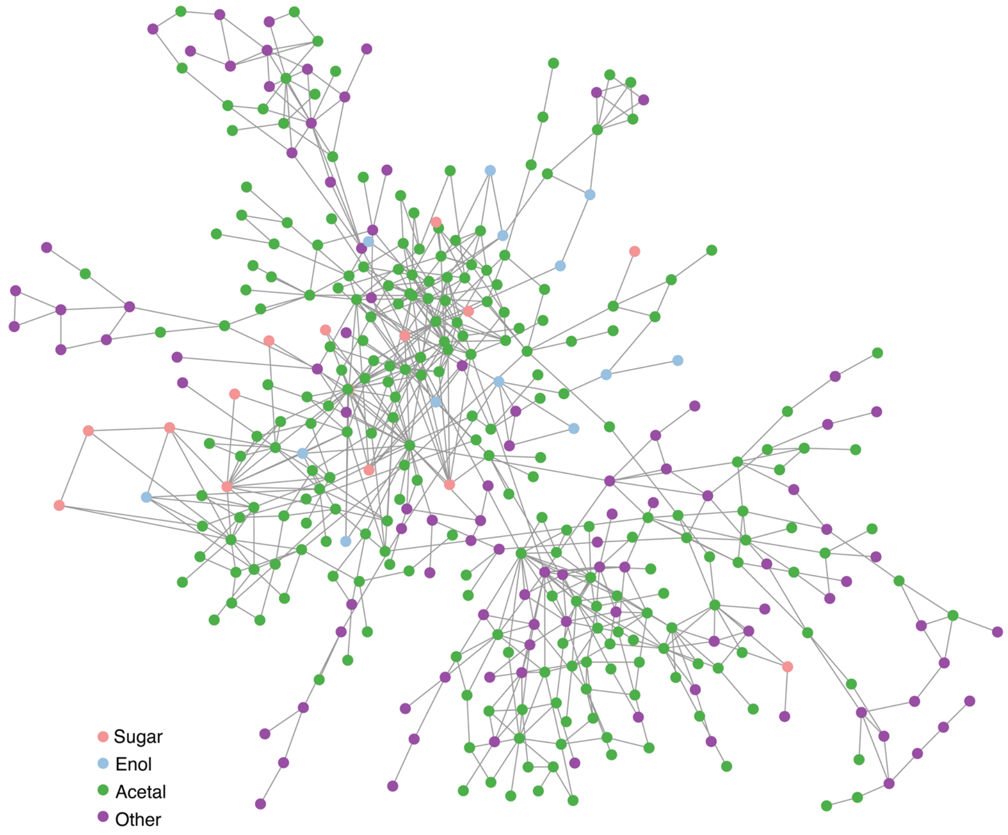


Figure 7. TN representation of the P_9 network. See Figure 5 for details.

(hydroxymethyl)propanal. Unbranched carbon−chain elongation involved an isomerization of glyceraldehyde to dihydroxyacetone via an enediol intermediate (Lobry de Bruyn−van Ekenstein isomerization),[82] followed by another aldol condendation reaction, which produced ketotetrose. The isomerization of ketotetrose via an enediol intermediate produced aldotetrose. Notably, the aldose−ketose isomerizations involve endothermic steps and appear as the slow steps of sugar formation. (Figure 5)

In addition, several unexpected reaction pathways were obtained involving three-membered and four-membered cyclic tetrose hemiacetals (Figure 6). These reaction pathways involve fewer reactions and appear to provide a shortcut to tetrose sugars. However, the strained three-membered and four-membered hemiacetal structures have not been experimentally characterized, and it is undetermined if they occur as reaction intermediates in aqueous solutions. The favorable flask energies associated with these structures are possibly an artifact of the semiempirical PM7 method and may be corrected by more accurate quantum chemical methods. The full list of reaction products of the T_9 network is given in section S2 of the Supporting Information. The details of the sugar formation pathways can be found in section S3 of the Supporting Information.

The T_9 network contained the prominent autocatalysis feature of the formose reaction suggested by Breslow.[62] Breslow's mechanism includes the formation of aldotetrose via a sequence of aldol additions and isomerizations, followed by the retroaldol cleavage of aldotetrose into two glycolaldehyde molecules.[62] Note that in the TN representation of the reaction network, the initial and the final flasks of autocatalytic processes are not identical and thus do not form closed cycles. Instead, product flasks arising from autocatalytic processes can be recognized by the doubling of the number of glycolaldehyde molecules per flask (Figure 5). In addition, autocatalytic cycles involving strained three- and four-membered hemiacetals were found in the T_9 network and were favored by shorter reaction sequences (Figure 6). The key step of these pathways involved an oxetane ring cleavage to glycolaldehyde and ethene-1,2-diol. Along with the four-membered aldotetrose hemiacetal, this mechanism might be rejected on thermodynamic grounds by more accurate quantum chemical methods.

The P_9 network consisted of 371 neutral flasks (354 distinct molecules) connected by 1114 reactions (Figure 7). The reaction mixture contained 11 sugars including 3 pentoses: 3-ketopentose, 2,3,4-trihydroxy-2-(hydroxymethyl)butanal, and 1,3,4-trihydroxy-3-(hydroxymethyl)butan-2-one. The formation of 2-ketopentose and aldopentose is expected after 12 and 15 generations, respectively. The subgraph of the P_9 network containing sugars, enols, and enediols was found to be qualitatively similar to that of the T_9 network but exhibited a larger set of concurrent reaction pathways as well as several Breslow-type autocatalytic processes involving higher sugars (e.g., dihydroxyacetone) as catalysts for condensation of formaldehyde (Figure 7). The full list of reaction products of the P_9 network is given in section S2 of the Supporting Information.

4. DISCUSSION AND OUTLOOK

The HAQC approach developed in this work provides a simple and efficient framework for exploring the structures of complex chemical reactions. It seeks to address the challenges of high dimensionality of reactive potential energy surfaces and the wide range of relevant time scales by combining constructive chemical heuristics and quantum chemical structure optimizations. The TN representations of reaction networks (Figures 6 and 7) generated by the HAQC approach yield a convenient starting point for studying the structures and dynamics of reactive potential energy surfaces. We present analyses of topological structure, global energetics, and network properties in our companion publication.[65] Moreover, the TN representation helps to link the networks of complex chemical reactions with the numerous analysis and dynamics methods developed for kinetic transition networks of protein folding and molecular rearrangements.[20,23−26] The applications of these methods to the reaction networks of cell metabolism and prebiotic chemistry will provide deeper insight into qualitative and quantitative aspects of chemistry of life and its origins.

As was previously shown in the context of protein folding models,[20,23−26] the TN network representation, in particular its Markovian state model variant, can be substituted for the full atomistic model in studies of both structural and kinetic properties of chemical systems. The structural properties can be derived from the eigenvectors of the transition matrix, which gives rise to spectral clustering and coarse-graining techniques.[20,23−26] Since the transitions between the network nodes obey first-order kinetics, the methods of kinetic Monte Carlo[90] are immediately applicable to TN representations of chemical systems. The kinetics of Markovian state models was found to reproduce the full system kinetics very well in previous studies.[23,25]

Once the feasible reactive channels on the potential energy surfaces are mapped out using the HAQC approach, many further refinements become possible. Perhaps most importantly, conventional methods of transition state optimizations can be used to provide accurate estimates of reaction barriers and transition probabilities.[16,17] With this information, the TN representation can be transformed into a continuous kinetic model that can be numerically solved using the established algorithms and codes.[46,51] The TN representation can also be used to construct flux balance analysis models of the reaction networks.[4] Work along these lines in currently in progress.

As we have shown in this work, simple constructive heuristics for polar organic reactions (Scheme 1) and semiempirical quantum chemistry predict the sugars up to C_5 and the major reaction pathways in the formose reaction network in line with expectations from experiment. However, the presence of strained three- and four-membered cyclic hemiacetals (Figure 6) indicates that a number of improvements are desirable. These further developments include (i) more accurate quantum chemical methods than the PM7 semiempirical method and COSMO solvation model used in this work (the mean unsigned error of the PM7 method for reaction energies of simple organic reactions is 4 kcal/mol[76]); (ii) improvements in thermodynamic and kinetic reaction parameters and more sophisticated classification approaches for feasible/unfeasible reactions; (iii) mechanism filtering and efficient sampling techniques[52,53] to reduce the number of quantum chemical calculations performed and the scaling with system size; and (iv) refinement and extension of rules of chemical transformation beyond "arrow pushing" rules of polar organic reactions. An important extension is the development of heuristic rule sets for more challenging classes of chemical reactions such as radical reactions, photochemical processes, and reactions involving organometallic compounds. Methods of statistical inference may help in *deriving* new rule sets specific to these domains from the existing body of experimental data or quantum chemical calculations.

The formose reaction is a useful testbed for the HAQC approach, since many formose products have been identified and mechanistic proposals for major reaction pathways exist. A host of other complex reaction networks have been described, but little is known about their product compositions and mechanisms. Complex chemical reactions of relevance to prebiotic chemistry include selective formose reactions catalyzed by phosphate,[91] borate,[61] or silicate;[92] condensations of hydrogen cyanide and formamide to nitrogen heterocycles;[93] the triose−ammonia reaction;[94,95] and the nucleoside synthesis recently suggested by Sutherland and co-workers.[96,97] Detailed studies of these and other abiotic reaction networks may help to elucidate common properties of reaction networks and differences from networks formed by evolution.[65]

Finally, the combination of heuristic rules and quantum chemical calculations might be viewed as an expedient tool for exploring chemically accessible regions of chemical space.[18,98−100] Coupled with efficient quantum chemical methodology and high-throughput computation, it holds promise for novel approaches for molecular design and optimization.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Kinetic selection criteria for formose reaction; reaction products of the formose reaction; major sugar formation pathways of the formose reaction. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: rappoport@chemistry.harvard.edu.
*E-mail: aspuru@chemistry.harvard.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported the Cyberdiscovery Initiative Type II (CDI^2) grant of the National Science Foundation (NSF), grant No. OIA-1125087. C.J.G. was supported by the Research Experience for Undergraduates (REU) summer research program of the NSF.

■ REFERENCES

(1) Helfferich, F. G. *Kinetics of Multistep Reactions*, 2nd ed.; Comprehensive Chemical Kinetics; Elsevier: Amsterdam, 2004; Vol. *40*.

(2) Vinu, R.; Broadbelt, L. J. *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3*, 29−54.

(3) Hartwig, J. F. *Organotransition Metal Chemistry. From Bonding to Catalysis*; University Science Books: Sausalito, CA, 2010.

(4) Palsson, B. *Systems Biology: Simulation of Dynamic Network States*; Cambridge University Press: Cambridge, U.K., 2011.

(5) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabási, A. L. *Nature* **2000**, *407*, 651−654.

(6) Barabási, A.-L.; Oltvai, Z. N. *Nature Rev. Gen.* **2004**, *5*, 101−113.

(7) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2007**, *37*, 101.

(8) Li, J.; Nowak, P.; Otto, S. *J. Am. Chem. Soc.* **2013**, *135*, 9222−9239.

(9) Broadbelt, L. J.; Pfaendtner, J. *AIChE J.* **2005**, *51*, 2112−2121.

(10) Green, W. H., Jr. In *Advances in Chemical Engineering*; Marin, G. B., Ed.; Elsevier: Amsterdam, 2007; Vol. *32*; pp 1−50.

(11) Noor, E.; Eden, E.; Milo, R.; Alon, U. *Mol. Cell* **2010**, *39*, 809−820.

(12) Bar-Even, A.; Flamholz, A.; Noor, E.; Milo, R. *Nature Chem. Biol.* **2012**, *8*, 509−517.

(13) Cairns-Smith, A. G.; Walker, G. L. *Biosystems* **1974**, *5*, 173−186.

(14) Wächtershäuser, G. *Proc. Nat. Acad. Sci.* **1990**, *87*, 200−204.

(15) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771−12800.

(16) Henkelman, G.; Jóhannesson, G.; Jónsson, H. In *Theoretical Methods in Condensed Phase Chemistry*; Schwartz, S. D., Ed.; Kluwer: Dordrecht, 2002; pp 269−302.

(17) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.

(18) Wales, D. J. *Energy Landscapes. Applications to Clusters, Biomolecules, and Glasses*; Cambridge University Press: Cambridge, U.K., 2003.

(19) Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. *J. Chem. Theory Comput.* **2006**, *2*, 840−857.

(20) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.

(21) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571−6581.

(22) Singhal, N.; Snow, C. D.; Pande, V. S. *J. Chem. Phys.* **2004**, *121*, 415−425.

(23) Pande, V. S.; Beauchamp, K.; Bowman, G. R. *Methods* **2010**, *52*, 99−105.

(24) Bowman, G. R.; Huang, X.; Pande, V. S. *Cell Res.* **2010**, *20*, 622−630.

(25) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.

(26) Wales, D. J. *Curr. Op. Struct. Biol.* **2010**, *20*, 3−10.

(27) Tozzini, V. *Acc. Chem. Res.* **2010**, *43*, 220−230.

(28) Levy, D. E. *Arrow-Pushing in Organic Chemistry. An Easy Approach to Understanding Reaction Mechanisms*; Wiley: Hoboken, NJ, 2008.

(29) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936−2941.

(30) Ess, D. H.; Houk, K. N. *J. Phys. Chem. A* **2005**, *109*, 9542−9553.

(31) Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178−192.

(32) Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science* **1985**, *228*, 408−418.

(33) Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz, A.; Fontain, E.; Gruber, B.; Herges, R.; Knauer, M.; Reitsam, K.; Stein, N. *Angew. Chem., Int. Ed.* **1993**, *32*, 201−227.

(34) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. *Pure Appl. Chem.* **1990**, *62*, 1921−1932.

(35) Todd, M. H. *Chem. Soc. Rev.* **2005**, *34*, 247.

(36) Ihlenfeldt, W.-D.; Gasteiger, J. *Angew. Chem., Int. Ed.* **1996**, *34*, 2613−2633.

(37) Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B.; Song, J.; Green, W. H., Jr; Matheu, D. M. *AIChE J.* **2006**, *52*, 718−730.

(38) Gothard, C. M.; Soh, S.; Gothard, N. A.; Kowalczyk, B.; Wei, Y.; Baytekin, B.; Grzybowski, B. A. *Angew. Chem., Int. Ed.* **2012**, *51*, 7922−7927.

(39) Kowalik, M.; Gothard, C. M.; Drews, A. M.; Gothard, N. A.; Weckiewicz, A.; Fuller, P. E.; Grzybowski, B. A.; Bishop, K. J. M. *Angew. Chem., Int. Ed.* **2012**, *51*, 7928−7932.

(40) *Cleaner Combustion*; Green Energy and Technology; Battin-Leclerc, F., Simmie, J. M., Blurock, E., Eds.; Springer: London, 2013.

(41) Tomlin, A. S.; Turányi, T.; Pilling, M. J. In *Comprehensive Chemical Kinetics*; Pilling, M. J., Ed.; Elsevier: Amsterdam, 1997; Vol. 35; pp 293−437.

(42) Susnow, R. G.; Dean, A. M.; Green, W. H., Jr; Peczak, P.; Broadbelt, L. J. *J. Phys. Chem. A* **1997**, *101*, 3731−3740.

(43) Green, W. H.; Barton, P. I.; Bhattacharjee, B.; Matheu, D. M.; Schwer, D. A.; Song, J.; Sumathi, R.; Carstensen, H. H.; Dean, A. M.; Grenda, J. M. *Ind. Eng. Chem. Res.* **2001**, *40*, 5362−5370.

(44) Sumathi, R.; Green, W. H., Jr. *Theor. Chem. Acc.* **2002**, *108*, 187−213.

(45) Curran, H. J.; Gaffuri, P.; Pitz, W. J.; Westbrook, C. K. *Combust. Flame* **1998**, *114*, 149−177.

(46) Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P. A.; Côme, G. M.; Scacchi, G. *Comput. Chem. (Oxford, U.K.)* **2000**, *24*, 541−560.

(47) Bournez, O.; Côme, G.-M.; Conraud, V.; Kirchner, H.; Ibānescu, L. In *Rewriting Techniques and Applications*; Springer: Berlin, 2003; pp 30−45.

(48) Ratkiewicz, A.; Truong, T. N. *Int. J. Quantum Chem.* **2006**, *106*, 244−255.

(49) Muharam, Y.; Warnatz, J. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4218−4229.

(50) Magoon, G. R.; Green, W. H., Jr. *Comput. Chem. Eng.* **2013**, *52*, 35−45.

(51) Green, W. H. et al. *RMG: Reaction Mechanism Generator*, version 4.0.1. 2013; http://rmg.sourceforge.net (accessed Jan. 27, 2014).

(52) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285−3305.

(53) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891−908.

(54) Prentiss, M. C.; Wales, D. J.; Wolynes, P. G. *PLoS Comput. Biol.* **2010**, *6*, e1000835.

(55) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.

(56) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1938**, *34*, 11.

(57) Boutlerow, A. *C. R. Acad. Sci.* **1861**, *53*, 145−147.

(58) Mizuno, T.; Weiss, A. H. In *Adv. Carbohydr. Chem. Biochem.*; Tipson, R. S., Horton, D., Eds.; Academic Press: New York, 1974; Vol. *29*; pp 173−227.

(59) Decker, P.; Schweer, H.; Pohlmann, R. *J. Chromatogr.* **1982**, *244*, 281−291.

(60) Zweckmair, T.; Böhmdorfer, S.; Bogolitsyna, A.; Rosenau, T.; Potthast, A.; Novalin, S. *J. Chromatogr. Sci.* **2014**, *52*, 169−175.

(61) Kim, H.-J.; Ricardo, A.; Illangkoon, H. I.; Kim, M. J.; Carrigan, M. A.; Frye, F.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133*, 9457−9468.

(62) Breslow, R. *Tetrahedron Lett.* **1959**, *1*, 22−26.

(63) Oparin, A. I. *The Origin of Life on the Earth*, 3rd ed.; Academic Press: New York, 1957.

(64) Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99−123.

(65) Rappoport, D.; Zubarev, D. Y.; Galvin, C. J.; Aspuru-Guzik, A. Network Properties of Abiotic Chemical Reaction Networks. **2014**; In preparation.

(66) Pfeil, E.; Schroth, G. *Ber. Dtsch. Chem. Ges.* **1952**, *85*, 293−307.

(67) Gabel, N. W.; Ponnamperuma, C. *Nature* **1967**, *216*, 453−455.

(68) Schwartz, A. W.; de Graaf, R. M. *J. Mol. Evol.* **1993**, *36*, 101−106.

(69) Weininger, D. *J. Chem. Inf. Model.* **1988**, *28*, 31−36.

(70) Jencks, W. P. *Chem. Rev.* **1985**, *85*, 511−527.

(71) Williams, I. H. *Chem. Soc. Rev.* **1993**, *22*, 277−283.

(72) Wales, D. J. *Science* **2001**, *293*, 2067−2070.

(73) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminf.* **2011**, *3*, 33.

(74) OpenBabel package, version 2.3.1. 2013; http://openbabel.org (accessed Jan. 27, 2104).

(75) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. *Chem. Cent. J.* **2008**, *2*, 5.

(76) Stewart, J. J. P. *J. Mol. Model.* **2013**, *19*, 1−32.

(77) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799−805.

(78) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.

(79) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.

(80) Jalan, A.; West, R. H.; Green, W. H., Jr *J. Phys. Chem. B* **2013**, *117*, 2955−2970.

(81) Keeffe, J. R.; Kresge, A. J. In *The Chemistry of Enols*; Rappoport, Z., Ed.; Wiley: Chichester, U.K., 1990; Chapter 7, pp 399−480.

(82) Angyal, S. J. *Top. Curr. Chem.* **2001**, *215*, 1−14.

(83) Heathcock, C. H. In *Comprehensive Organic Synthesis*; Trost, B. M., Fleming, I., Eds.; Pergamon Press: Oxford, U.K., 1991; Vol. *2*; Chapter 1.5, pp 133−179.

(84) Braun, M. In *Modern Aldol Reactions*; Mahrwald, R., Ed.; Wiley-VCH: Weinheim, 2004; Vol. *1*; Chapter 1, pp 19−61.

(85) Schmitz, E.; Eichhorn, I. In *The Ether Linkage*; Patai, S., Ed.; Wiley: Chichester, U.K., 1967; pp 309−351.

(86) Seebach, D. *Angew. Chem., Int. Ed.* **1979**, *18*, 239−258.

(87) Ravasz, E.; Somera, A. L.; Mongru, D. A.; Oltvai, Z. N.; Barabási, A.-L. *Science* **2002**, *297*, 1551−1555.

(88) Newman, M. E. J. *Networks. An Introduction*; Oxford University Press: Oxford, U.K., 2009.

(89) Shannon, P. *Genome Res.* **2003**, *13*, 2498−2504.

(90) Voter, A. F. In *Radiation Effects in Solids*; Springer: Dordrecht, 2007; pp 1−23.

(91) Müller, D.; Pitsch, S.; Kittaka, A.; Wagner, E.; Wintner, C. E.; Eschenmoser, A. *Helvet. Chim. Acta* **1990**, *73*, 1410−1468.

(92) Lambert, J. B.; Gurusamy-Thangavelu, S. A.; Ma, K. *Science* **2010**, *327*, 984−986.

(93) Roy, D.; Najafian, K.; von Ragué Schleyer, P. *Proc. Nat. Acad. Sci.* **2007**, *104*, 17272−17277.

(94) Weber, A. L. *Orig. Life Evol. Biosph.* **2007**, *37*, 105−111.

(95) Eschenmoser, A. *Chem. Biodivers.* **2007**, *4*, 554−573.

(96) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459*, 239−242.

(97) Powner, M. W.; Sutherland, J. D.; Szostak, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 16677−16688.

(98) Blum, L. C.; Reymond, J.-L. *J. Am. Chem. Soc.* **2009**, 8732−8733.

(99) von Lilienfeld, O. A. *Int. J. Quantum Chem.* **2013**, *113*, 1676−1689.

(100) Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. *J. Am. Chem. Soc.* **2013**, *135*, 7296−7303.