





Review

# Automated Exploration of Prebiotic Chemical Reaction Space: Progress and Perspectives

Siddhant Sharma <sup>1,2,3</sup> , Aayush Arya <sup>1,4,†</sup> , Romulo Cruz <sup>1,5,†</sup>  and Henderson James Cleaves II <sup>1,6,\*</sup> 

<sup>1</sup> Blue Marble Space Institute of Science, Seattle, WA 98154, USA; sharmasi@chalmers.se (S.S.); aayush.11912610@lpu.in (A.A.); romulo.cruz.s@uni.pe (R.C.)

<sup>2</sup> Department of Biochemistry, Deshbandhu College, University of Delhi, New Delhi 110019, India

<sup>3</sup> Department of Chemistry and Chemical Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

<sup>4</sup> Department of Physics, Lovely Professional University, Jalandhar-Delhi GT Road, Phagwara 144001, India

<sup>5</sup> Big Data Laboratory, Information and Communications Technology Center (CTIC), National University of Engineering, Amaru 210, Lima 15333, Peru

<sup>6</sup> Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan

\* Correspondence: hcleaves@elsi.jp

† These authors contributed equally to this work.

**Abstract:** Prebiotic chemistry often involves the study of complex systems of chemical reactions that form large networks with a large number of diverse species. Such complex systems may have given rise to emergent phenomena that ultimately led to the origin of life on Earth. The environmental conditions and processes involved in this emergence may not be fully recapitulable, making it difficult for experimentalists to study prebiotic systems in laboratory simulations. Computational chemistry offers efficient ways to study such chemical systems and identify the ones most likely to display complex properties associated with life. Here, we review tools and techniques for modelling prebiotic chemical reaction networks and outline possible ways to identify self-replicating features that are central to many origin-of-life models.

**Keywords:** prebiotic chemistry; automated chemical space searches; chemical reaction networks; computational modelling; network autocatalysis; self-replicating structures



**Citation:** Sharma, S.; Arya, A.; Cruz, R.; Cleaves II, H.J. Automated Exploration of Prebiotic Chemical Reaction Space: Progress and Perspectives. *Life* **2021**, *11*, 1140. <https://doi.org/10.3390/life11111140>

Academic Editor: Michael Russell

Received: 7 September 2021

Accepted: 18 October 2021

Published: 26 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

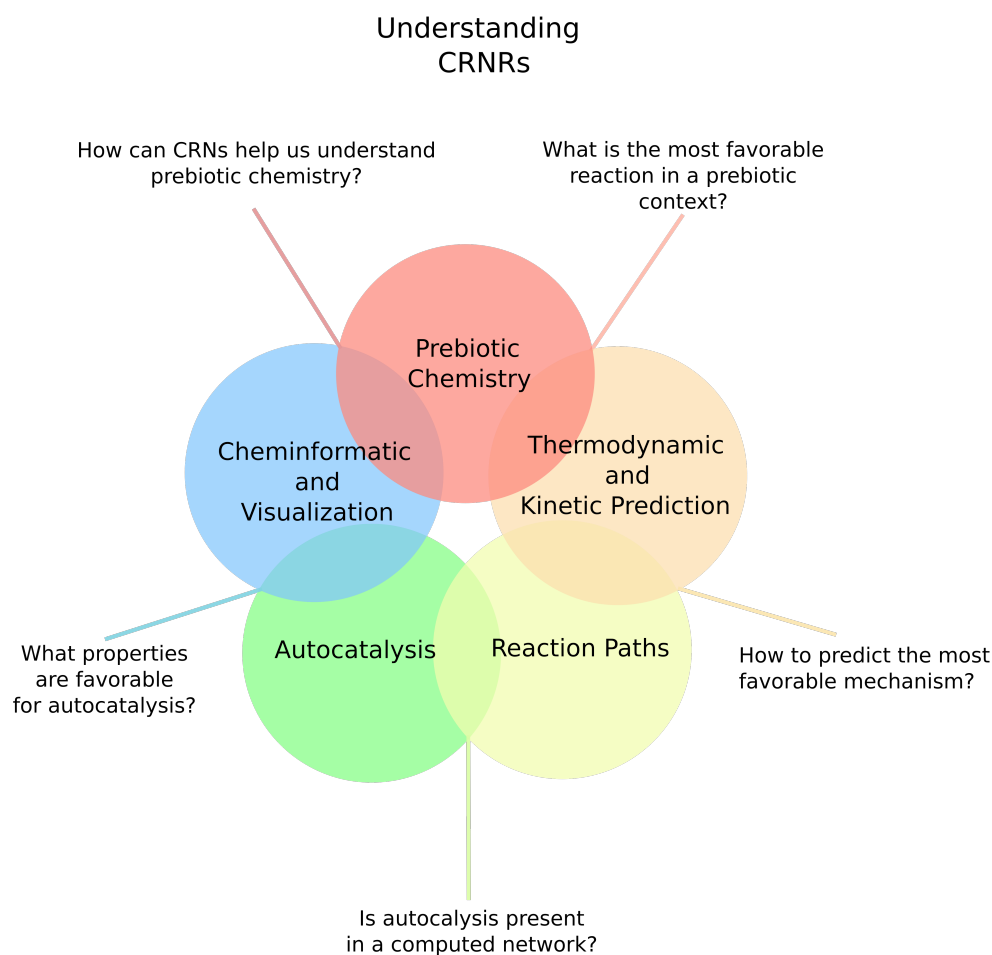
## 1. Introduction

The study of prebiotic chemistry requires understanding complex phenomena involving the interplay of highly variable and as-yet uncertain primitive environmental conditions, often in the context of diversity-generating chemical reactions [1]. These reactions may have together produced large and diverse sets of products that can differ subtly or dramatically under variable conditions, e.g., [2–5]. This interplay has been speculated to have produced the emergent chemical systems which gave rise to life. However, the specific environmental conditions and chemical processes which gave rise to life have now been lost to Earth's dynamic geological history, and it is difficult for experimentalists to recreate all possible combinations of conditions that may have been present on primitive Earth in the laboratory, or analyze the complex products which often result from such lab simulations [6].

Computational approaches (see for example [7]) offer efficient ways for chemists to study chemical systems which may display complex properties conducive to the emergence of chemical systems with life-like properties. Here we name the putative collection of chemical processes, their interplay with environmental parameters, and the resultant chemical diversity that appears in such a computational model as a chemical reaction network representation (CRNR). CRNRs may be thought of as idealized collective material flows

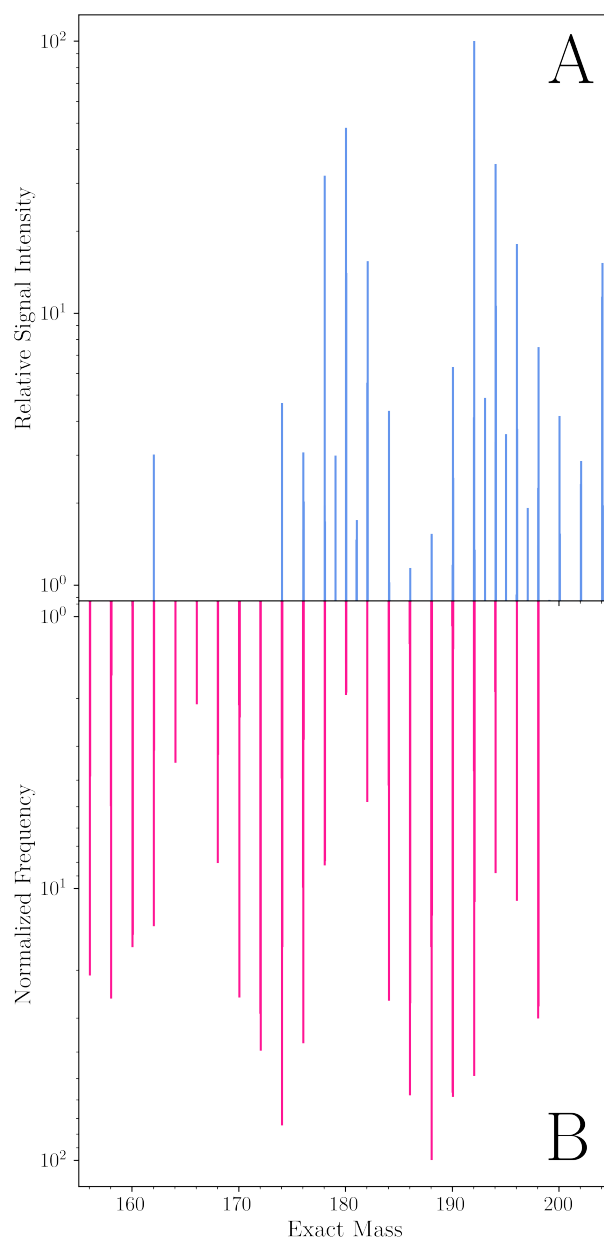
through allowed reactions channels, which can vary as a function of reaction conditions, including temperature, pH, concentration, molecularity, etc. (e.g., [8–10]).

These representations may not necessarily be complete descriptions of complex reaction systems, but may nonetheless offer roadmaps for understanding complicated fluxes through chemical systems. Many aspects of real chemical reaction network (CRN) chemistry may not be realized in CRNRs. Understanding why CRNRs fail to accurately mimic real CRN outcomes is a central challenge for computational chemistry to help understand prebiotic chemistry and the origins of life, and offers a route to improve the use of CRNs as guides for such purposes. Many important questions remain as to how real CRNs could have become capable of Darwinian selection [11,12]. Some authors have suggested the emergence of complexity due to network properties may be as important as the nature of the chemical reactions involved in CRNs [13,14]. An overview of current questions and methods aiding in the exploration of prebiotic chemical reaction space is depicted in Figure 1.



**Figure 1.** Some of the current questions and methods for the exploration of prebiotic CRNRs.

Comparing computational and experimental investigations of CRN reactions, real-world reactions may produce hundreds to millions of products, and their identification is limited by analytical method detection limits. The size of the product space that can be practically computed, if not carefully informed by undetermined variables associated with kinetic parameters, may also grow exponentially (e.g., [15]). Locally variable environmental factors, such as the presence of certain minerals (e.g., [2,16]), may also alter the course of reactions and steer product distributions (e.g., [17]). Figure 2 illustrates one example of how CRNRs attempt to predict the outcomes of real-world CRNs.



**Figure 2.** CRNRs may accurately predict the many features of CRNs. **(A)** The 150–210 amu region of the mass spectrum of the products of a laboratory formose reaction measured using high-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) in negative ESI mode. **(B)** The predicted mass distribution after six generations of the products of the same reaction was generated using CRNR methods. In this formose reaction, 2 M paraformaldehyde, 1 M glycolaldehyde, and 0.05 M  $\text{Ca}(\text{OH})_2$  were heated in aqueous solution in sealed glass ampoules under nitrogen at  $85^\circ$  for eight days. The code for generating this figure is described in Supplementary Materials.

Because the salient features of prebiotic chemical systems may be difficult to measure directly, CRNR methods [18,19], *in silico* exploration of high-dimensional chemical spaces (e.g., [15,20,21]), and network theory offer promising tools to explore origins questions (e.g., [12,22]). As Smith et al. [23] have pointed out, network theory may be useful in the study of collective chemical behavior. It is crucial that collective systemic behavior be understood to fully characterize a chemical or biological system, as the study of only one or a few components of a complex product suite is likely inadequate to infer higher-order properties of chemical systems.

A large variety of computational tools can be pipelined in CRNRs to explore prebiotic chemical reaction space, ranging, for example, from *ab initio* molecular dynamics (MD) simulations (e.g., [24,25]) to chemical assembly theory (e.g., [26]), and molecular assembly trees (e.g., [27]). Informatics has also opened many new avenues of study in biology and chemistry [28]. Chemoinformatics has rapidly become a routine discovery tool [29], with ever more powerful open-source resources becoming available [30,31]. Likewise, the types of chemical systems of prebiotic relevance that CRNRs can be applied to are diverse, ranging from experimental methods for life detection (e.g., [32]) to the simulation of primitive planetary atmospheric chemistry (e.g., [33]).

In the present article, we briefly review the developments in computational chemistry that can assist in the application of CRNR computation and analysis to understanding problems of astrobiological relevance, especially prebiotic chemistry.

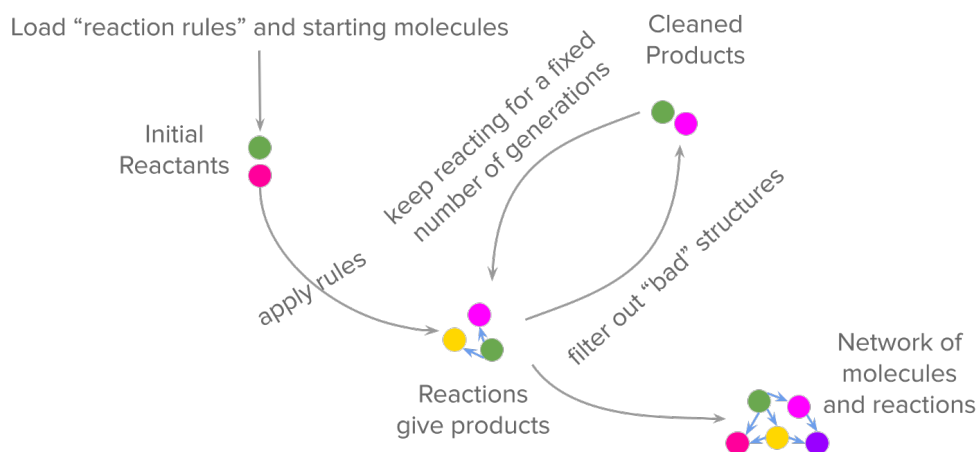
## 2. Modelling Prebiotic Chemistry: From Individual Reactions to a Network

While performing a bottom-up synthesis of a computational reaction network representation, going from a small set of reaction species to a network often requires making approximations due to time and computing resource limits. One must be wary of the sacrifice in the accuracy of model predictions when such approximations are made. The advantages of the different approaches to modelling, from quantum chemistry to graph theory, must be weighted based on the scale of the network and the desired accuracy. For example, rigorous quantum chemistry can provide a more precise estimation of the mechanism, kinetics or outcome of a single reaction, while graph theory-based modelling, with its less costly computations, can provide a convenient framework for the synthesis, visualization and analysis of large scale CRNRs from a network theory perspective. Here, we contrast some of the approaches that have gained traction in prebiotic chemical modelling.

The first of these sets of approaches use quantum chemistry-based computations, which can provide accurate predictions of reaction outcomes (e.g., [34]). Computational quantum chemical approaches have been used to understand prebiotic reaction pathways (e.g., [35–37]); however, these approaches often scale poorly due to the cost of computation involved, which currently limits their use in simulating complex prebiotic networks [38]. Besides computational resource issues, such approaches may rely on prior knowledge of intermediate transition states or efficient searches for these intermediates on the reaction's potential energy surface (PES) (e.g., [39]). More recently, computational efficiency of the global optimization problem algorithm has been explored by combining quantum chemistry techniques with network analysis methods such as exploration of optimal thermodynamic and stoichiometric pathways [40] or random sampling processes [34].

Chemical graph theory approaches, on the other hand, offer a way to streamline handling of large CRNRs [41,42]. Graph theory is commonly used in computational chemistry since molecules can be precisely represented as graphs [43], with nodes representing atoms and edges representing chemical bonds between them. These molecular “graphs” are then transformed by applying user-defined reaction templates that guide the synthesis of products using “seed” molecules. The reaction templates are rules that search for a particular pattern in a molecule, then apply transformations by modifying the edges of the graphs. The accuracy of reactions performed this way thus has a clear dependence on the selected reaction mechanism, and can be tuned to be appropriately restrictive or permissive. For more reliable reaction predictions using graph theory, one can use estimates of thermodynamic parameters to determine the feasibility of a reaction. A schematic describing how multiple rounds of reaction generation leads to the synthesis of a full network is shown in Figure 3.

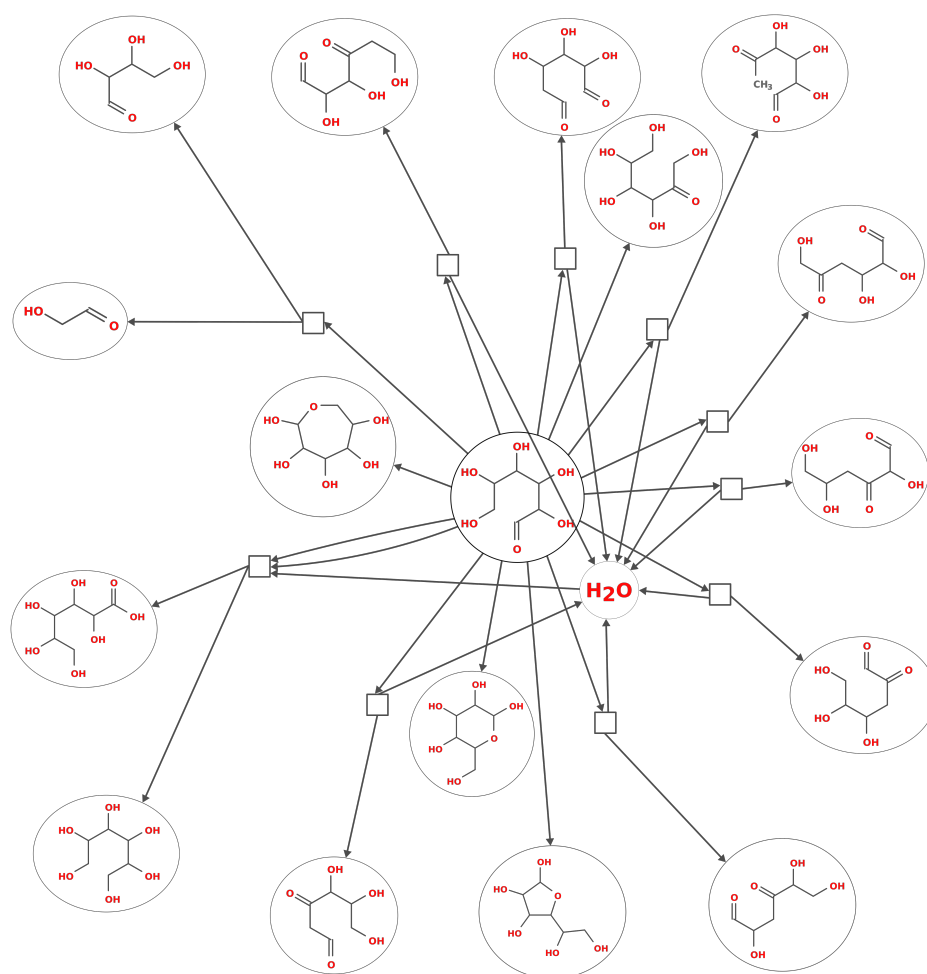




**Figure 3.** Schematic for the forward synthesis of a network using graph theory-based tools. Typically, a set of “reaction rules” are loaded, which specify how certain substructures are to be altered during a reaction. All the rules are then applied combinatorially to a set of initial reactants, which gives a set of products as the graph transformations happen that give product molecules, connected by “edges” representing reactions. Constraints that certain molecules (or “graphs”) with certain substructures should not form can be used to filter out unstable species. Iterative application of these rules to the product suite at each step gives a complete CRNR.

Graph theory-based tools have been developed to use intuitive ways of representing molecules (e.g., simplified molecular input line entry system, SMILES) and codifying reactions, for instance, the human-readable Graph Modeling Language (GML) format used in tools such as MØD, a software package developed for graph-based cheminformatics [41]. Graph theory also allows for various generalizations of sophisticated chemical phenomena, for example, allowing concepts such as autocatalysis and tautomerism, among others, to be formalized. Graph grammars can encode generalizable reaction mechanisms (referred to in Figure 3 as “reaction rules”). An early stage CRNR generated for the glucose degradation reaction using the graph-grammar approaches is shown in Figure 4.

Various tools for constraining the feasibility of reactions using thermodynamic calculations based on quantum chemistry methods exist. Quantum chemical calculations have been employed to provide more accurate determinations of the course of reaction pathways in prebiotic reaction networks [44]. Density functional theory (DFT) methods have been used to characterize thousands of molecules and chemical reactions [45] but are expensive in terms of required computing resources. However, there are classical group contribution methods (GCMs) that are less computationally demanding [46]. Joback and Reid’s classic GCMs can be used on the chemical reaction spaces with the help of tools such as JRgui [47], eEquilibrator [48] and the Benson group’s additivity methods [49]. The eEquilibrator program makes use of component contribution methods, which are a modification of GCM methods [50]. Semi-empirical approaches to studying prebiotic chemical thermodynamics have been made using software packages such as MOPAC [51]. Automated approaches like AutoMeKin combine molecular dynamics (MD), graph theory algorithms and Monte Carlo simulations to discover likely relevant reaction mechanisms [52]. Quantum mechanics/molecular mechanics (QM/MM) and MD simulations have been combined to study prebiotic nucleic acid analogues [53] and lipids [54]. Kua et al. [55], benchmarked thermochemical estimations for compounds thought to be important in protometabolism derived from eEquilibrator to those derived from more accurate yet computationally demanding DFT quantum chemical methods, and found them to be remarkably similar.



**Figure 4.** A primitive CRNR for glucose degradation reaction, generated using the graph grammar-based program MØD. The in silico synthesis of this CRNR was done as per the general methods outlined in Figure 3. The reaction rules used were selected based on prior knowledge of mechanisms known to dominate in this chemical system under pH and temperature conditions of interest. A single cycle of reaction rule application (i.e., one “generation”) is shown here. In the visual representation used here, molecules are shown in ovals, while reaction nodes are shown as squares.

Interactive frameworks for exploring chemical reaction space have been developed, for example Molpher software [56]. Bespoke automated computational approaches for generating CRNRs can be constructed modularly, for example by integrating graph grammar operations such as those used in MØD [41], reactive molecular dynamics tools such as ReacNetGenerator [57] and Python (programming language frameworks) such as Reaction Mechanism Generator (RMG) [58,59], CGRtools (Condensed Graph of Reaction) [60], and Rule Input Network Generator (RING) for generating CRNRs from complex reactive systems [61], among others. Other tools potentially useful for constructing such pipelines include pReSt [62] to discover novel chemistries in automated CRNRs, and CERENA (ChEmical REaction Network Analyzer) [63] to model stochastic chemical kinetics in chemical reaction networks.

For the sake of simplicity, many computational studies relying on graph theory use flattened molecular representations that lack stereochemical information, but stereochemical information can be encoded in such frameworks [64]. Such information can be used to gain insight into the kinetic mechanisms of stereochemical symmetry breaking which evidently occurred during the emergence of homochiral biological systems [65]. Efficient stereochemical handling in chemical reaction networks is so far limited to relatively small systems [66], partially due to complexities in representing Cahn—Ingold—Prelog rules to

unequivocally label stereoisomers. The ability to more agilely model stereochemical transformations would be a major advance for this field. Open source stereoisomer generation methods include RDKit's EnumerateStereoisomers module [67] and MAYGEN [68], among many others.

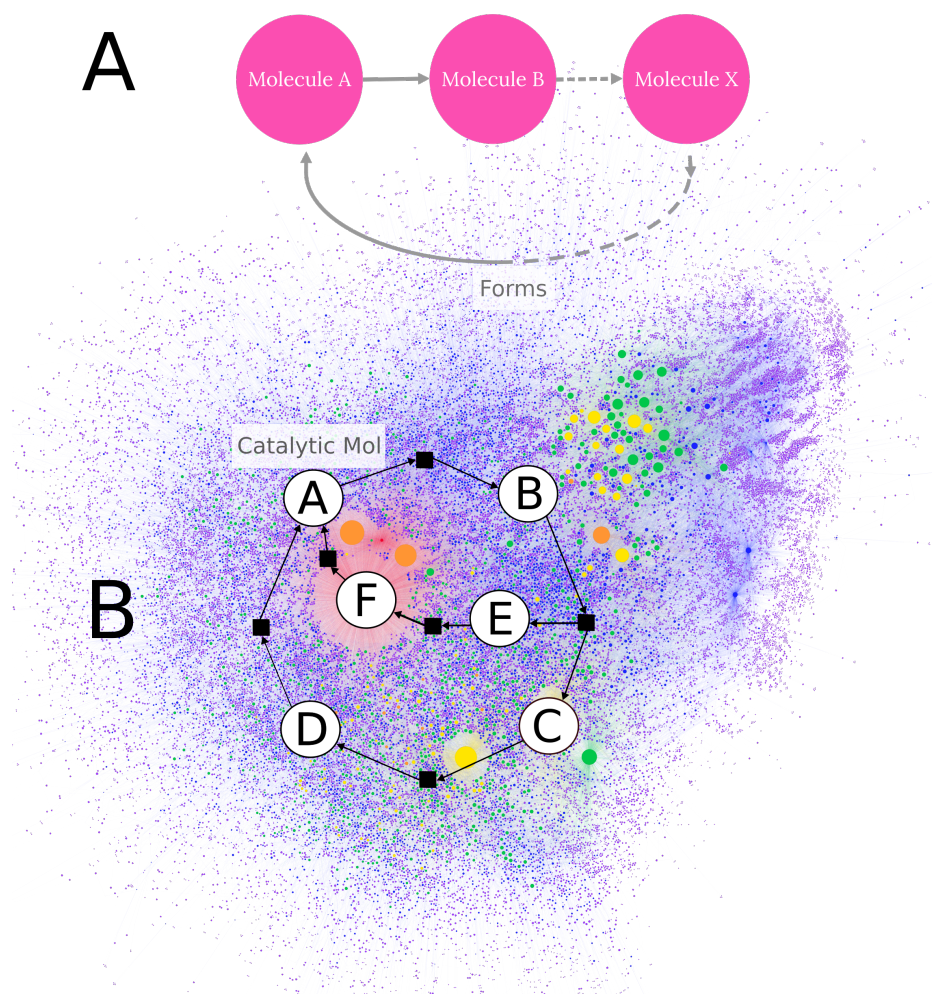
### 3. Detection of Autocatalytic Motifs in Computed Chemical Networks

A common feature of origins of life models is the involvement of self-replicating molecules or systems [69–74], which can be quantified as autocatalytic sets within CRNs [75]. Autocatalysis represents a range of phenomena of variable complexity potentially responsible for many processes of prebiotic interest [76], for example the formose reaction e.g., [77] and formaldehyde-catalyzed HCN oligomerization [78]. The detection of the emergence of autocatalysis is a major goal of prebiotic chemistry [79], and autocatalytic reactions may represent an important link between prebiotic chemistry, primitive metabolism and modern biochemistry [80–82]. It remains to be determined how common autocatalytic reaction systems are; they may be rather common but simply hard to detect, depending on their kinetics, complexity and context, or truly relatively rare.

There are two fundamentally different types of autocatalytic networks: ones which produce autocatalysis through topological network effects and those which generate feedback catalysis [83], in which some products of the reaction networks serve as true catalysts (as opposed to network catalysts) for existing or novel reactions. Both types can potentially fundamentally change the evolution of CRNs [79,84,85].

The former type depends most heavily on thermodynamic considerations: in order for a large number of reactions to proceed favorably and at a comparable pace, they must have low activation energy barriers. Indeed a frequent aspect of many proposed prebiotic chemical reactions is their seeding by high free-energy species generated by the action of environmental energy sources (e.g., HCHO or HCN generated by electric discharges, photochemically, etc.). CRNs can be driven by various energy inputs (e.g., [9]), including radioactive decay [86–89], though it might be expected that solar radiation may be more important due to its larger flux [90], and the dissipation of potential chemical energy likely plays a role in the development of CRNs. Dissipative chemistry may partially explain the growth of complex CRNs [91]. Semenov et al. [92] studied a nonenzymatic autocatalytic reaction network using continuous flow reactors and found them to display oscillatory behavior. CRNs may offer methods for designing such chemical systems rapidly from first principles. An autocatalytic loop motif in a sample CRNR is illustrated in Figure 5.

Andersen and colleagues [93] have proposed the use of integer hyperflows, which explore how varying stoichiometric relationships among CRN pathways may affect the overall flux of material through them as tools for the universal definition of autocatalysis in chemical reaction networks. Chemical cycles such as the reverse tricarboxylic acid cycle have been explored computationally [94,95], and graph grammar methods can also be applied to study alternative prebiotic pathways such as Eschenmoser's glyoxylate pathway [96].



**Figure 5.** General features of autocatalysis and a specific example of an autocatalytic cycle detectable within CRNRs. **(A)** The basic idea of autocatalysis is that a sequence (or network) of reactions begins with a specific molecule A and produces more than one copy of A, assuring that the cycle produces more A than it consumes. **(B)** A concrete example of such a cycle within a larger CRNR. Here, two paths,  $A \rightarrow B \rightarrow E \rightarrow F \rightarrow A$  and  $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ , contribute to produce stoichiometrically larger quantities of A. The CRNR illustrated here was produced using five rounds of reaction generation in the glucose degradation chemistry discussed above. The layout of the graph was executed using Gephi. The size of the nodes corresponds to each node in-degree. Each color represents a new generation.

#### 4. Use of Machine Learning (ML) for Understanding CRNs

ML has transformed computational chemistry and also holds immense potential for exploring chemical systems [97,98]. ML and neural networks (NNs) have been used to explore molecular structures, reactions and reaction mechanisms (e.g., [99]), and these methods can be applied for exploration of the generated CRNRs. The number of chemical reactions or transformations a molecule can participate in can be examined using automated reaction template analysis [100]. Much like graph theory-based implementations, deep learning has been applied to generating molecular structures and predicting their properties [101]. The use of ML in cheminformatics to predict plausible reaction mechanisms (e.g., [102,103]) has allowed approaches such as NN-based methods to predict chemical reaction space [104]. Combined graph theory-NN methods have also been used to predict the activation energies of organic reactions [105].

ML methods have been coupled with cheminformatics molecular descriptors [106] and structure-based reactivity estimation approaches to predict reaction outcomes [107,108]. Deep

learning, which is a subset of ML, has been used in chemical reaction prediction [109] and to predict reaction yields [110] using interfaces like IBM RXN for Chemistry [111,112], which can be further modified to predict enzymatic reactions [113], and being open-source, these approaches can be readily modified to meet user requirements. Meuwly [114] reviewed the utility of ML methods for chemical reactions. To date, we are not aware of careful comprehensive comparisons of these methods which would suggest one approach is better than another, merely that applying such approaches culls CRNR outputs.

## 5. Problem-Specific Cheminformatic Tools and Approaches

### 5.1. Computing Molecular Descriptors

It is often useful to characterize chemical species quantitatively using molecular structure based metrics. “Molecular descriptors” estimate properties such as octanol/water partition coefficients (LogP), aqueous solubility [115], and drug-likeness [116], among various other properties including topological and geometric ones. Computed descriptors provide a way to compare a wide range of species with varied structural and chemical properties, and to identify particular molecules with certain desirable properties or collections of properties among very large datasets. Descriptor-based analyses are often used for chemical space exploration, e.g., [117]. Tools for descriptor computation include PaDEL [118], RDKit [119], ChemDes [120,121], Mordred [122], CDK-GUI [123] and Py-BioMed [124]. There is some overlap with respect to the descriptors each of these packages computes, in addition to unique functionalities of each package.

### 5.2. Broad Functionality Chemoinformatics Tools

RDKit [119] is a widely used general purpose cheminformatics package with functionalities for molecular manipulation, curation, library building, and molecular analysis that also computes numerous molecular descriptors supporting various cheminformatics input formats such as SDF, SMILES, MOL and reaction templates such as SMARTS and SMIRKS [125]. Similar toolkits exist which have at least some degree of overlap in functionality, such as CDK [123], Indigo Toolkit [126], and OpenBabel [127], and are available as a bundle within a single package called Cinfony [128], which provides a simplified programmable application programming interface (API) for cheminformatic operations, and Chembench [129], which is a publicly accessible cheminformatics web portal to mine and model chemical data.

### 5.3. Handling Isomerism

Representations of molecular structures are approximations of real chemical bonding, and chemists have developed several concepts, such as resonance and tautomerism, to deal with nuances glossed over by these shorthand representations. These phenomena have collectively been termed delocalization-induced molecular equality [130], and the multiplicity of equivalent representations of the same compound can create complications in the computational generation of reaction networks, as they may introduce meaningless redundancies. Tautomerism is a challenging problem in computational chemistry [131] and tautomers may represent unique formalisms for understanding how chemicals engage in reactions, thus making their representation meaningful. There are many open-source software packages and libraries available for treatment of computationally generated tautomers and isomers. Tautomer generation can be accomplished using open-source tools such as AMBIT [132], TautGen [133], and RDKit’s MolVS Wrapper [134], which has been used to enumerate all possible tautomers in small molecule libraries [135]. Each of these has its own unique approaches to enumerate and prioritize tautomeric structures. Commercial software packages, including OpenEye [136], Chemaxon [137] and CACTVS [138], are also able to enumerate and rank tautomeric structures in terms of their relative importance. Databases like TautoBase [139] may also help in the comparison and evaluation of tautomers.

It is often important to find the lowest energy conformers of chemical species to predict energetically plausible reaction mechanisms. Several methods are available to explore PES for this purpose, each with its own nuances and utility. Applications like Molassembler [140] combine molecule generation and conformer exploration methods in



a single package. Conformer generation tools using semi-empirical approaches include DataWarrior [141], OMEGA [142], Balloon [143], Confab [144], ConfGen [145], Frog2 [146] and RDKit, which use various force-field estimations and algorithms [147,148]. Packages employing molecular mechanics such as Tinker 8 [149] have also been used for conformer generation, which allows thorough searches for low-energy conformers [150].

Graph-based conformer clustering methods, such as AutoGraph, are helpful for generating ensembles of lowest-energies conformers after conformers have been processed with semi-empirical methods [151].

#### 5.4. Miscellaneous Tools

Libraries such as ChemPy [152] and Catalyst.jl [153] were built to explore the dynamics of chemical reaction networks by solving systems of coupled continuous or stochastic differential equations and handling of chemical kinetics processes. Apart from Python, several cheminformatics tools and libraries have been developed in other languages as well, including, in the R programming language, such resources as ChemMine tools [154], ChemmineR [155] and rcdk [156]. The MolecularGraph.jl package provides cheminformatic capabilities using the Julia language [157].

Platforms like Dask [158] help bring the power of parallel computing to cheminformatics for large-scale, rapid analysis and manipulation of cheminformatic data. Calculated molecular fingerprints are used for rapid substructure matching and similarity searching [159] and can be calculated with tools like chemfp [160]. Molecular Set Comparator [161] uses cheminformatic approaches to compare two sets of datasets and properties like Tanimoto distances [162] are used to compare similarities in molecular representations as reduced to a two-dimensional mapping using t-distributed stochastic neighbor embedding (t-SNE, [163]). Statistical tools like principal component analysis (PCA) can also help in similarity clustering, which can be useful for identifying where similar types of species are generated among CRNs. Various visualization tools, including chemical scaffold networks and trees, can be generated and analyzed using automated scaffold graphs [164] or deep learning methods [165]. These may assist in the interpretation of the organization of these features as they are generated in CRNs.

Generating pKa data for organic molecules can be done by using proprietary software such as ChemAxon, and there has been considerable interest in computationally predicting pKa values, either by using ML methods coupled to QSAR (Quantitative Structure-Activity Relationship) [166,167] or by using Graph NN methods [168], developing interactive applications such as MolGpka [169].

Stoichiometric network analysis and flux balance analysis for such networks [170,171] can help get kinetic information off the reactions participating in the network.

## 6. Experimental Vetting of the Computational Methods

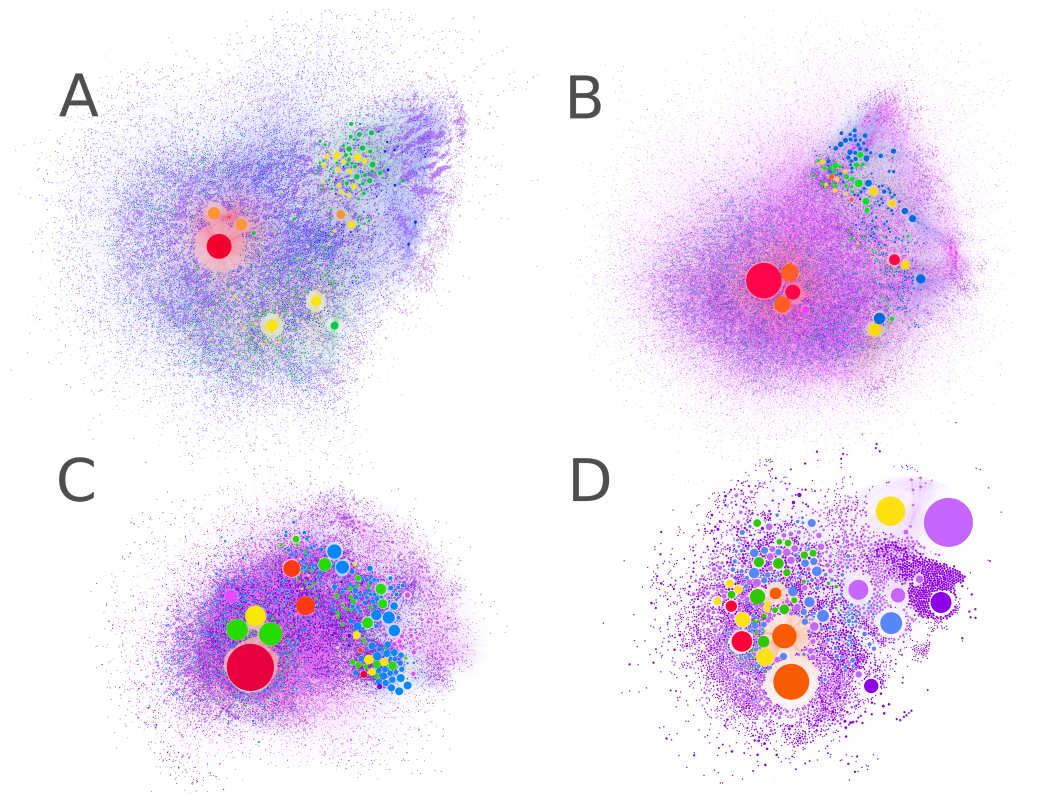
Experimental validation of computational models can be accomplished through comparison with products detected using chemical analysis [172,173], for example by using integrated NMR and mass spectral approaches [174–176], which provide complementary information regarding chemical diversity and bonding. Van Krevelen diagrams are used by geochemists to characterize large sets of chemical species by plotting the atomic ratios of certain elements. Automated R and Python libraries exist that can generate Van Krevelen diagrams from MS data [177,178]. Kendrick mass defect (KMD) analysis has been used to study large chemical networks using FT-ICR-MS data [32,179], allowing easy identification of homologous series of molecules. R programming language libraries to rapidly analyze such data to generate Kendrick mass defect (KMD) diagrams are commonly used [180]. To see if a CRNR is able to synthesize known metabolites, it may be useful to refer to databases like REAXYS [181], the Human Metabolites Database (HMDB, [182]), and the Kyoto Encyclopedia of Genes and Genomes (KEGG, [183]). The R package biodb [184] provides an interface for querying such chemical databases, while open source packages like Webchem enable scraping of several databases of interest [185]. Similar software packages are available to query these databases, such as PubChemPy for PubChem [186]. Many of the compounds and reactions submitted to the databases discussed in this section

are biologically, medically, or industrially relevant. There is a need unified for libraries and databases of prebiotic relevance that accurately predict compounds that should be expected to occur in abiotic systems. Creating databases specific to prebiotic considerations would help train the data used in Machine Learning (ML) models.

### 7. Visualization of Chemically Relevant Datasets

Visualizing dense datasets is a challenge in computational chemistry [187] and there are various interactive approaches for displaying “big data”, such as the use of minimum spanning trees in TMAP [188] and WebMolCS [189] for interactive visualization of chemical space. CRNRs can also be visualized using Gephi [190] (see Figure 6). To visualize networks using Gephi, firstly, a “source-target” table is required that lists how the nodes are connected, indicating which reactive node or product of a certain reaction is relevant. Subsequently, a layout based on the Gephi Force-Atlas 2 algorithm is made to show how to group the clusters according to its connectivity. Some additional tools that are used to visualize CRNRs are CytoScape [191], NetworkX [192], Graph Tool [193] and ReNView [194].

Graph databases such as Neo4J [195] can aid in developing pattern searching algorithms, along with enhanced visualizations of the reaction sequences generated. The associated SMILES format for cheminformatics can be visualized and embedded using tools such as SmilesDrawer [196] and Leruli [197], which help in the development of web-based representations of datasets for easier community access.



**Figure 6.** Gephi representations of CRNRs for (A) glucose degradation, (B) formose reaction, (C) pyruvic acid degradation, and (D) the reaction of  $\text{HCN} + \text{NH}_3$ . All CRNRs were generated using the graph-grammar techniques previously discussed, and the figures shown here were produced using the Gephi software. Such visualization tools are helpful for broad visual classification of CRNRs. The size of the nodes in these graphs is proportional to the node’s in-degree as a function of how many edges reach it. Node color indicates the generation in which each compound was first generated. The code for generating this figure is described in Supplementary Materials.



## 8. Conclusions

There has been considerable progress in computational methods to predict the outcomes of organic reactions. Detailed and accurate modelling procedures exist for the prediction of most single-step reactions and for reactions with ambiguous mechanisms. However, applying such techniques to concatenated reactions is often resource-expensive and difficult to scale to the degree that it would be of benefit to prebiotic chemists. Less sophisticated techniques involving lesser computational costs, such as those based on graph theory, exist, but they come at the expense of reliability of the predictions, as they are not based on estimates of some physical or chemical parameters that drive reactions.

To accommodate the needs of prebiotic chemists, it is likely preferable to employ a blend of chemical graph theory that is guided by thermodynamic energy estimates to determine reaction feasibility using semi-empirical methods (e.g., the eEquilibrator API). Such semi-empirical methods provide a good approximation consistent with low-level quantum mechanical treatments. Having generated a network, one can study the essential features of the chemistry that is involved. In this context, being able to identify self-replicating features such as autocatalytic cycles could be useful for prebiotic chemists. Although an agreed-upon definition of autocatalysis remains to be defined, it is possible to detect topological features in networks that one considers to resemble autocatalysis.

Studying large networks and analyzing their product suites manually can be difficult. One can make use of statistical metrics to quantify broad, important features of the network as a whole, or find clusters within the network. Additionally, one can make use of molecular descriptors to quantitatively assess structural properties of the products of the chemistry and see how they vary within the network. As for any scientific model, the ultimate test for these computer models is an agreement with the experiment. One ought to test if the CRNR predictions can match with observations, which could be gathered using mass spectrometry or analytical detections of experimental syntheses.

There are many potential applications of these computational tools to astrobiology, including understanding prebiotic chemistry and the origins of life. In addition to concerted studies to increase the accuracy of prediction methods, the development of user-friendly open-source pipelines will allow much greater community development of methods for rapidly exploring prebiotic chemistry *in silico* to understand real-world phenomena.

**Supplementary Materials:** The code for generating Figures 2 and 6 is available online on GitHub: <https://github.com/ssiddhantsharma/sharmaaryacruzcleaves2021>.

**Author Contributions:** Writing—original draft preparation, S.S., A.A., R.C., H.J.C.II; writing—review and editing, S.S., A.A., R.C., H.J.C.II; visualization, S.S., A.A., R.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the Blue Marble Space Institute of Science (BMSIS) for organizing the YSP2020 program (S.S., A.A., R.C., H.J.C.II). H.J.C.II would like to thank the Earth-Life Science Institute (ELSI) and the ELSI Origins Network (EON) for financial support during the initial development of this work. EON was supported by a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. S.S. would like to acknowledge the SETI Forward Award from the SETI Institute. R. C. wishes to acknowledge FONDECYT (Convenio n°208-2015-FONDECYT) for his Master scholarship. The authors would also like to acknowledge helpful discussions on the subject and proofreading of the draft possible through the insights of Johanna Huhtassari, Cole Mathis, Kjell Jorner and Harrison Brodsky Smith. The authors acknowledge Huan Chen for the support in collecting experimental FT-ICR-MS data for Figure 2. The authors would also like to thank the editor for the invitation to submit to this issue.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Robinson, W.; Daines, E.; van Duppen, P.; de Jong, T.; Huck, W. *Environmental Conditions Drive Self-Organisation of Reaction Pathways in a Prebiotic Reaction Network*; 2021. Available online: <https://www.researchsquare.com/article/rs-775456/v1> (accessed on 20 October 2021).
2. Cleaves, H.J. Prebiotic chemistry: What we know, what we don't. *Evol. Edu. Outreach* **2012**, *5*, 342–360. [[CrossRef](#)]
3. Cleaves, H.J., II. Prebiotic chemistry: Geochemical context and reaction screening. *Life* **2013**, *3*, 331. [[CrossRef](#)] [[PubMed](#)]
4. Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **2014**, *2014*, 1. [[CrossRef](#)]
5. Islam, S.; Powner, M.W. Prebiotic systems chemistry: Complexity overcoming clutter. *Chem* **2017**, *2*, 470–501. [[CrossRef](#)]
6. Pérez-Villa, A.; Pietrucci, F.; Saitta, A.M. Prebiotic chemistry and origins of life research with atomistic computer simulations. *Phys. Life Rev.* **2020**, *34–35*, 105–135. [[CrossRef](#)]
7. Cheng, G.J.; Zhang, X.; Chung, L.W.; Xu, L.; Wu, Y.D. Computational organic chemistry: Bridging theory and experiment in establishing the mechanisms of chemical reactions. *J. Am. Chem. Soc.* **2015**, *137*, 1706–1725. [[CrossRef](#)]
8. Andersen, J.L.; Andersen, T.; Flamm, C.; Hanczyc, M.M.; Merkle, D.; Stadler, P.F. Navigating the chemical space of HCN polymerization and hydrolysis: Guiding graph grammars by mass spectrometry data. *Entropy* **2013**, *15*, 4066–4083. [[CrossRef](#)]
9. Tran, Q.P.; Adam, Z.R.; Fahrenbach, A.C. Prebiotic reaction networks in water. *Life* **2020**, *10*, 352. [[CrossRef](#)]
10. Yi, R.; Tran, Q.P.; Ali, S.; Yoda, I.; Adam, Z.R.; Cleaves, H.J.; Fahrenbach, A.C. A continuous reaction network that produces RNA precursors. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 13267–13274. [[CrossRef](#)] [[PubMed](#)]
11. Vasas, V.; Szathmáry, E.; Santos, M. Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1470–1475. [[CrossRef](#)] [[PubMed](#)]
12. Butch, C.J.; Meringer, M.; Gagnon, J.S.; Cleaves, H.J. Open questions in understanding life's origins. *Commun. Chem.* **2021**, *4*, 1–4. [[CrossRef](#)]
13. Shapiro, R. Small molecule interactions were central to the origin of life. *Q. Rev. Biol.* **2006**, *81*, 105–125. [[CrossRef](#)]
14. Cronin, L.; Walker, S.I. Origin of life. Beyond prebiotic chemistry. *Science* **2016**, *352*, 1174–1175. [[CrossRef](#)]
15. Meringer, M.; Cleaves, H.J. Exploring astrobiology using *in silico* molecular structure generation. *Philos. Trans. R. Soc. A* **2017**, *375*, 20160344. [[CrossRef](#)] [[PubMed](#)]
16. Walton, C.; Rimmer, P.B.; Williams, H.; Shorttle, O. Prebiotic chemistry in the wild: How geology interferes with the origins of life. *ChemRxiv* **2020**. [[CrossRef](#)]
17. Surman, A.J.; Rodriguez-Garcia, M.; Abul-Haija, Y.M.; Cooper, G.J.T.; Gromski, P.S.; Turk-MacLeod, R.; Mullin, M.; Mathis, C.; Walker, S.I.; Cronin, L. Environmental control programs the emergence of distinct functional ensembles from unconstrained chemical reactions. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5387–5392. [[CrossRef](#)] [[PubMed](#)]
18. Gromski, P.S.; Henson, A.B.; Granda, J.M.; Cronin, L. How to explore chemical space using algorithms and automation. *Nat. Rev. Chem.* **2019**, *3*, 119–128. [[CrossRef](#)]
19. Wołos, A.; Roszak, R.; Żądło Dobrowolska, A.; Beker, W.; Mikulak-Klucznik, B.; Spólnik, G.; Dygas, M.; Szymkuć, S.; Grzybowski, B.A. Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. *Science* **2020**, *369*, eaaw1955. [[CrossRef](#)]
20. Dewyer, A.L.; Argüelles, A.J.; Zimmerman, P.M. Methods for exploring reaction space in molecular systems. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1354. [[CrossRef](#)]
21. Coley, C.W. Defining and exploring chemical spaces. *Trends Chem.* **2021**, *3*, 133–145. [[CrossRef](#)]
22. Walker, S.I.; Mathis, C. Network theory in prebiotic evolution. In *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids*; Springer: Cham, Switzerland, 2018; pp. 263–291. [[CrossRef](#)]
23. Smith, H.B.; Kim, H.; Walker, S.I. Scarcity of scale-free topology is universal across biochemical networks. *Sci. Rep.* **2021**, *11*, 6542. [[CrossRef](#)]
24. Das, T.; Ghule, S.; Vanka, K. Insights into the origin of life: Did it begin from HCN and H<sub>2</sub>O? *ACS Cent. Sci.* **2019**, *5*, 1532–1540. [[CrossRef](#)]
25. Magrino, T.; Pietrucci, F.; Saitta, A.M. Step by step strecker amino acid synthesis from ab initio prebiotic chemistry. *J. Phys. Chem. Lett.* **2021**, *12*, 2630–2637. [[CrossRef](#)]
26. Marshall, S.M.; Mathis, C.; Carrick, E.; Keenan, G.; Cooper, G.J.T.; Graham, H.; Craven, M.; Gromski, P.S.; Moore, D.G.; Walker, S.I.; et al. Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nat. Commun.* **2021**, *12*, 3033. [[CrossRef](#)]
27. Liu, Y.; Mathis, C.; Bajczyk, M.D.; Marshall, S.M.; Wilbraham, L.; Cronin, L. Exploring and mapping chemical space with molecular assembly trees. *Sci. Adv.* **2021**, *7*, 39, doi:10.1126/sciadv.abj2465. [[CrossRef](#)] [[PubMed](#)]
28. López-López, E.; Bajorath, J.; Medina-Franco, J.L. Informatics for chemistry, biology, and biomedical sciences. *J. Chem. Inf. Model.* **2021**, *61*, 26–35. [[CrossRef](#)]
29. Pirhadi, S.; Sunseri, J.; Koes, D.R. Open source molecular modeling. *J. Mol. Graph. Model.* **2016**, *69*, 127–143. [[CrossRef](#)]
30. González-Medina, M.; Naveja, J.J.; Sánchez-Cruz, N.; Medina-Franco, J.L. Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* **2017**, *7*, 54153–54163. [[CrossRef](#)]

31. Medina-Franco, J.L.; Sánchez-Cruz, N.; López-López, E.; Díaz-Eufracio, B.I. Progress on open chemoinformatic tools for expanding and exploring the chemical space. *J. Comput.-Aided Mol. Des.* **2021**. [[CrossRef](#)] [[PubMed](#)]
32. Guttenberg, N.; Chen, H.; Mochizuki, T.; Cleaves, H.J. Classification of the biogenicity of complex organic mixtures for the detection of extraterrestrial life. *Life* **2021**, *11*, 234. [[CrossRef](#)] [[PubMed](#)]
33. Rimmer, P.B.; Helling, C. A chemical kinetics network for lightning and life in planetary atmospheres. *Astrophys. J. Suppl. Ser.* **2016**, *224*, 9. [[CrossRef](#)]
34. Lee, K.; Kim, J.W.; Kim, W.Y. Efficient construction of a chemical reaction network guided by a Monte Carlo tree search. *ChemSystemsChem* **2020**, *2*, e1900057. [[CrossRef](#)]
35. Barone, V.; Biczysko, M.; Puzzarini, C. Quantum chemistry meets spectroscopy for astrochemistry: Increasing complexity toward prebiotic molecules. *Acc. Chem. Res.* **2015**, *48*, 1413–1422. [[CrossRef](#)] [[PubMed](#)]
36. Simm, G.N.; Reiher, M. Context-driven exploration of complex chemical reaction networks. *J. Chem. Theory Comput.* **2017**, *13*, 6108–6119. [[CrossRef](#)]
37. Simm, G.; Vaucher, A.C.; Reiher, M. Exploration of reaction pathways and chemical transformation networks. *J. Phys. Chem. A* **2019**, *123*, 385–399. [[CrossRef](#)]
38. Nghe, P.; Hordijk, W.; Kauffman, S.A.; Walker, S.I.; Schmidt, F.J.; Kemble, H.; Yeates, J.A.M.; Lehman, N. Prebiotic network evolution: six key parameters. *Mol. Biosyst.* **2015**, *11*, 3206–3217. [[CrossRef](#)]
39. Kim, Y.; Kim, J.W.; Kim, Z.; Kim, W.Y. Efficient prediction of reaction paths through molecular graph and reaction network analysis. *Chem. Sci.* **2018**, *9*, 825–835. [[CrossRef](#)]
40. Blau, S.M.; Patel, H.D.; Spotte-Smith, E.W.C.; Xie, X.; Dwaraknath, S.; Persson, K.A. A chemically consistent graph architecture for massive reaction networks applied to solid-electrolyte interphase formation. *Chem. Sci.* **2021**, *12*, 4931–4939. [[CrossRef](#)]
41. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. A software package for chemically inspired graph transformation. *arXiv* **2016**, arXiv:1603.02481.
42. Ratkiewicz, A.; Truong, T.N. Application of chemical graph theory for automated mechanism generation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 36–44. [[CrossRef](#)] [[PubMed](#)]
43. Temkin, O.N.; Zeigarnik, A.V.; Bonchev, D.G. *Chemical Reaction Networks*; CRC Press: Boca Raton, FL, USA, 2020.
44. Pearce, B.K.D.; Ayers, P.W.; Pudritz, R.E. A consistent reduced network for HCN chemistry in early earth and Titan atmospheres: Quantum calculations of reaction rate coefficients. *J. Phys. Chem. A* **2019**. [[CrossRef](#)]
45. St. John, P.C.; Guan, Y.; Kim, Y.; Etz, B.D.; Kim, S.; Paton, R.S. Quantum chemical calculations for over 200,000 organic radical species and 40,000 associated closed-shell molecules. *Sci. Data* **2020**, *7*, 244. [[CrossRef](#)] [[PubMed](#)]
46. Kolska, Z.; Za'bransky, M.; Randova, A. Group contribution methods for estimation of selected physico-chemical properties of organic compounds. In *Thermodynamics—Fundamentals and Its Application in Science*; IntechOpen: Rijeka, Croatia, 2012. [[CrossRef](#)]
47. Shi, C.; Borchardt, T.B. JRgui: A Python program of Joback and Reid method. *ACS Omega* **2017**, *2*, 8682–8688. 7b01464. [[CrossRef](#)]
48. Beber, M.E.; Gollub, M.G.; Mozaffari, D.; Shebek, K.M.; Noor, E. eQuilibrator 3.0—A Platform for the Estimation of Thermodynamic Constants. 2021. Available online: <http://xxx.lanl.gov/abs/2103.00621> (accessed on 20 October 2021).
49. Python Group Additivity (pgradd) Documentation. Available online: <https://vlachosgroup.github.io/PythonGroupAdditivity/> (accessed on 20 October 2021).
50. Noor, E.; Haraldsdóttir, H.S.; Milo, R.; Fleming, R.M.T. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* **2013**, *9*, e1003098. [[CrossRef](#)] [[PubMed](#)]
51. Stewart, J.J.P. MOPAC: A semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1–103. [[CrossRef](#)] [[PubMed](#)]
52. Rodríguez, A.; Rodríguez-Fernández, R.; Vázquez, S.A.; Barnes, G.L.; Stewart, J.J.P.; Martínez-Núñez, E. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *J. Comput. Chem.* **2018**, *39*, 1922–1930. [[CrossRef](#)]
53. Kaur, S.; Sharma, P.; Wetmore, S.D. Can cyanuric acid and 2,4,6-Triaminopyrimidine containing ribonucleosides be components of prebiotic RNA? Insights from QM calculations and MD simulations. *ChemPhysChem* **2019**, *20*, 1425–1436. [[CrossRef](#)]
54. Kahana, A.; Lancet, D. Protobiotic systems chemistry analyzed by molecular dynamics. *Life* **2019**, *9*, 38. [[CrossRef](#)]
55. Kua, J.; Hernandez, A.L.; Velasquez, D.N. Thermodynamics of potential CHO metabolites in a reducing environment. *Life* **2021**, *11*, 1025. [[CrossRef](#)]
56. Hoksza, D.; Škoda, P.; Voršilák, M.; Svozil, D. Molpher: A software framework for systematic chemical space exploration. *J. Cheminf.* **2014**, *6*, 7. [[CrossRef](#)] [[PubMed](#)]
57. Zeng, J.; Cao, L.; Chin, C.H.; Ren, H.; Zhang, J.Z.H.; Zhu, T. ReacNetGenerator: An automatic reaction network generator for reactive molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2020**, *22*, 683–691. [[CrossRef](#)]
58. Gao, C.W.; Allen, J.W.; Green, W.H.; West, R.H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **2016**, *203*, 212–225. [[CrossRef](#)]
59. Liu, M.; Grinberg Dana, A.; Johnson, M.S.; Goldman, M.J.; Jocher, A.; Payne, A.M.; Grambow, C.A.; Han, K.; Yee, N.W.; Mazeau, E.J.; et al. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *J. Chem. Inf. Model.* **2021**, *61*, 2686–2696. [[CrossRef](#)]
60. Nugmanov, R.I.; Mukhametgaleev, R.N.; Akhmetshin, T.; Gimadiev, T.R.; Afonina, V.A.; Madzhidov, T.I.; Varnek, A. CGRtools: Python library for molecule, reaction, and condensed graph of reaction processing. *J. Chem. Inf. Model.* **2019**, *59*, 2516–2521. [[CrossRef](#)] [[PubMed](#)]

61. Gupta, U.; Le, T.; Hu, W.S.; Bhan, A.; Daoutidis, P. Automated network generation and analysis of biochemical reaction pathways using RING. *Metab. Eng.* **2018**, *49*, 84–93. [CrossRef]
62. Gupta, U.; Vlachos, D.G. Learning chemistry of complex reaction systems via a python first-principles Reaction rule Stencil (pReSt) generator. *J. Chem. Inf. Model.* **2021**, *61*, 3431–3441. [CrossRef]
63. Kazeroonian, A.; Fröhlich, F.; Raue, A.; Theis, F.J.; Hasenauer, J. CERENA: ChEmical REaction Network Analyzer—A toolbox for the simulation and analysis of stochastic chemical kinetics. *PLoS ONE* **2016**, *11*, e0146732. [CrossRef] [PubMed]
64. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. Chemical graph transformation with stereo-information. In *Graph Transformation*; Springer: Cham, Switzerland, 2017; pp. 54–69. [CrossRef]
65. Laurent, G.; Lacoste, D.; Gaspard, P. Emergence of homochirality in large molecular systems. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012741118. [CrossRef]
66. Coley, C.; Green, W.H.; Jensen, K.F. RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application. *J. Chem. Inf. Model.* **2019**, *59*, 2529–2537. [CrossRef] [PubMed]
67. rdkit.Chem.EnumerateStereoisomers Module—The RDKit 2021.03.1 Documentation. Available online: <https://www.rdkit.org/docs/source/rdkit.Chem.EnumerateStereoisomers.html> (accessed on 20 October 2021).
68. Yirik, M.A.; Sorokina, M.; Steinbeck, C. MAYGEN: An open-source chemical structure generator for constitutional isomers based on the orderly generation principle. *J. Cheminf.* **2021**, *13*, 48. [CrossRef] [PubMed]
69. Gánti, T. Organization of chemical reactions into dividing and metabolizing units: The chemotons. *Biosystems* **1975**, *7*, 15–21. [CrossRef]
70. Kauffman, S.A. Autocatalytic sets of proteins. *J. Theor. Biol.* **1986**, *119*, 1–24. [CrossRef]
71. Wächtershäuser, G. Before enzymes and templates: Theory of surface metabolism. *Microbiol. Rev.* **1988**, *52*, 452. [CrossRef]
72. Kauffman, S.A. *The Origins of Order*; Oxford University Press: Oxford, UK, 1993.
73. Vaidya, N.; Manapat, M.L.; Chen, I.A.; Xulvi-Brunet, R.; Hayden, E.J.; Lehman, N. Spontaneous network formation among cooperative RNA replicators. *Nature* **2012**, *491*, 72–77. [CrossRef]
74. Tjhung, K.F.; Shokhirev, M.N.; Horning, D.P.; Joyce, G.F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2906–2913. [CrossRef]
75. Kauffman, S.; Steel, M. The expected number of viable autocatalytic sets in chemical reaction systems. *Artif. Life* **2021**, *27*, 1–14. [CrossRef]
76. Virgo, N.; Ikegami, T.; McGregor, S. Complex autocatalysis in simple chemistries. *Artif. Life* **2016**, *22*, 138–152.10.1162/ARTL\_a\_00195. [CrossRef] [PubMed]
77. Jeilani, Y.A.; Nguyen, M.T. Autocatalysis in formose reaction and formation of RNA nucleosides. *J. Phys. Chem. B* **2020**, *124*, 11324–11336 [CrossRef]
78. Schwartz, A.W.; Goverde, M. Acceleration of HCN oligomerization by formaldehyde and related compounds: Implications for prebiotic syntheses. *J. Mol. Evol.* **1982**, *18*, 351–353. [CrossRef] [PubMed]
79. Hordijk, W.; Steel, M. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* **2004**, *227*, 451–461. [CrossRef]
80. Kun, Á.; Papp, B.; Szathmáry, E. Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. *Genome Biol.* **2008**, *9*, R51. [CrossRef] [PubMed]
81. Preiner, M.; Xavier, J.C.; do Nascimento Vieira, A.; Kleineremanns, K.; Allen, J.F.; Martin, W.F. Catalysts, autocatalysis and the origin of metabolism. *Interface Focus* **2019**, *9*, 20190072. [CrossRef]
82. Steel, M.; Hordijk, W.; Xavier, J.C. Autocatalytic networks in biology: Structural theory and algorithms. *J. R. Soc. Interface* **2019**, *16*, 20180808. [CrossRef]
83. Peng, Z.; Linderth, J.; Baum, D. A mechanism of abiogenesis based on complex reaction networks organized by seed-dependent autocatalytic systems. *ChemRxiv* **2021**. [CrossRef]
84. Luo, Y.; Epstein, I.R. Feedback analysis of mechanisms for chemical oscillators. In *Advances in Chemical Physics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; pp. 269–299. [CrossRef]
85. Xavier, J.C.; Hordijk, W.; Kauffman, S.; Steel, M.; Martin, W.F. Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. B* **2020**, *287*, 20192377. [CrossRef]
86. Adam, Z.R.; Fahrenbach, A.C.; Kacar, B.; Aono, M. Prebiotic geochemical automata at the intersection of radiolytic chemistry, physical complexity, and systems biology. *Complexity* **2018**, *2018*, 9376183. [CrossRef]
87. Adam, Z.R.; Fahrenbach, A.C.; Jacobson, S.M.; Kacar, B.; Zubarev, D.Y. Radiolysis generates a complex organosynthetic chemical network. *Sci. Rep.* **2021**, *11*, 1743. [CrossRef] [PubMed]
88. Paredes-Arriaga, A.; Meléndez-López, A.; Heredia, A.; Cruz-Castañeda, J.; Negrón-Mendoza, A.; Ramos-Bernal, S. Role of Na<sup>+</sup>-montmorillonite in the stability of guanine exposed to high-radiation energy in primitive environments: Heterogeneous models. *Radiat. Phys. Chem.* **2021**, *186*, 109509. [CrossRef]
89. Pastorek, A.; Fergus, M.; Čuba, V.; Šrámek, O.; Ivanek, O.; Civiš, S. Primordial radioactivity and prebiotic chemical evolution: Effect of  $\gamma$  radiation on formamide-based synthesis. *J. Phys. Chem. B* **2020**, *124*, 8951–8959. [CrossRef] [PubMed]
90. Miller, S.L.; Orgel, L.E. *The Origins of Life on the Earth*; Prentice-Hall: Upper Saddle River, NJ, USA, 1974.



91. Cafferty, B.; Wong, A.S.Y.; Semenov, S.N.; Belding, L.; Gmür, S.; Huck, W.T.S.; Whitesides, G.M. Robustness, entrainment, and hybridization in dissipative molecular networks, and the origin of life. *J. Am. Chem. Soc.* **2019**, *141*, 8289–8295. [[CrossRef](#)] [[PubMed](#)]
92. Semenov, S.N.; Kraft, L.J.; Ainla, A.; Zhao, M.; Baghbanzadeh, M.; Campbell, V.E.; Kang, K.; Fox, J.M.; Whitesides, G.M. Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* **2016**, *537*, 656–660. [[CrossRef](#)] [[PubMed](#)]
93. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. Defining autocatalysis in chemical reaction networks. *arXiv* **2021**, arXiv:2107.03086.
94. Zubarev, D.Y.; Rappoport, D.; Aspuru-Guzik, A. Uncertainty of prebiotic scenarios: The case of the non-enzymatic reverse tricarboxylic acid cycle. *Sci. Rep.* **2015**, *5*, 8009. [[CrossRef](#)]
95. Meringer, M.; Cleaves, H.J. Computational exploration of the chemical structure space of possible reverse tricarboxylic acid cycle constituents. *Sci. Rep.* **2017**, *7*, 17540. [[CrossRef](#)] [[PubMed](#)]
96. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. In silico support for Eschenmoser's glyoxylate scenario. *Isr. J. Chem.* **2015**, *55*, 919–933. [[CrossRef](#)]
97. Strieth-Kalthoff, F.; Sandfort, F.; Segler, M.H.S.; Glorius, F. Machine learning the ropes: Principles, applications and directions in synthetic chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154–6168. [[CrossRef](#)] [[PubMed](#)]
98. Keith, J.A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.R.; Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **2021**, *121*, 9816–9872. [[CrossRef](#)]
99. Stocker, S.; Csányi, G.; Reuter, K.; Margraf, J.T. Machine learning in chemical reaction space. *Nat. Commun.* **2020**, *11*, 1–11. [[CrossRef](#)]
100. Plehiers, P.P.; Marin, G.B.; Stevens, C.V.; Van Geem, K.M. Automated reaction database and reaction network analysis: Extraction of reaction templates using cheminformatics. *J. Cheminf.* **2018**, *10*, 11. [[CrossRef](#)] [[PubMed](#)]
101. Walters, W.P.; Barzilay, R. Applications of deep learning in molecule generation and molecular property prediction. *Acc. Chem. Res.* **2021**, *54*, 263–270. [[CrossRef](#)]
102. Kayala, M.A.; Baldi, P. ReactionPredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning. *J. Chem. Inf. Model.* **2012**, *52*, 2526–2540. [[CrossRef](#)] [[PubMed](#)]
103. Coley, C.W.; Barzilay, R.; Jaakkola, T.S.; Green, W.H.; Jensen, K.F. Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* **2017**, *3*, 434–443. [[CrossRef](#)]
104. Wei, J.N.; Duvenaud, D.; Aspuru-Guzik, A. Neural networks for the prediction of organic chemistry reactions. *ACS Cent. Sci.* **2016**, *2*, 725–732. [[CrossRef](#)]
105. Pathak, Y.; Mehta, S.; Priyakumar, U.D. Learning atomic interactions through solvation free energy prediction using graph neural networks. *J. Chem. Inf. Model.* **2021**, *61*, 689–698. [[CrossRef](#)] [[PubMed](#)]
106. Skoraczynski, G.; Dittwald, P.; Miasojedow, B.; Szymkuć, S.; Gajewska, E.P.; Grzybowski, B.A.; Gambin, A. Predicting the outcomes of organic reactions via machine learning: Are current descriptors sufficient? *Sci. Rep.* **2017**, *7*, 3582. [[CrossRef](#)] [[PubMed](#)]
107. Gao, H.; Struble, T.J.; Coley, C.W.; Wang, Y.; Green, W.H.; Jensen, K.F. Using machine learning to predict suitable conditions for organic reactions. *ACS Cent. Sci.* **2018**, *4*, 1465–1476. [[CrossRef](#)]
108. Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A structure-based platform for predicting chemical reactivity. *Chem* **2020**, *6*, 1379–1390. [[CrossRef](#)]
109. Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep learning for chemical reaction prediction. *Mol. Syst. Des. Eng.* **2018**, *3*, 442–452. [[CrossRef](#)]
110. Schwaller, P.; Vaucher, A.C.; Laino, T.; Reymond, J.L. Prediction of chemical reaction yields using deep learning. *Mach. Learn. Sci. Technol.* **2021**, *2*, 015016. [[CrossRef](#)]
111. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C.A.; Bekas, C.; Lee, A.A. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Cent. Sci.* **2019**, *5*, 1572–1583. [[CrossRef](#)]
112. Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V.H.; Haeuselmann, R.A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **2020**, *11*, 3316–3325. [[CrossRef](#)]
113. Kreutter, D.; Schwaller, P.; Reymond, J.L. Predicting enzymatic reactions with a molecular transformer. *Chem. Sci.* **2021**, *12*, 8648–8659. [[CrossRef](#)]
114. Meuwly, M. Machine learning for chemical reactions. *Chem. Rev.* **2021**, *121*, 10218–10239. [[CrossRef](#)] [[PubMed](#)]
115. Delaney, J.S. ESOL: Estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005. [[CrossRef](#)]
116. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90–98. [[CrossRef](#)]
117. Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.O. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.* **2021**, *5*, 240–255. [[CrossRef](#)]
118. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [[CrossRef](#)]

119. RDKit: Open-Source Cheminformatics. Available online: <https://rdkit.org/> (accessed on 20 October 2021).
120. Dong, J.; Cao, D.S.; Miao, H.Y.; Liu, S.; Deng, B.C.; Yun, Y.H.; Wang, N.N.; Lu, A.P.; Zeng, W.B.; Chen, A.F. ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminf.* **2015**, *7*, 60. [CrossRef] [PubMed]
121. Cao, D.S.; Xu, Q.S.; Hu, Q.N.; Liang, Y.Z. ChemoPy: Freely available python package for computational biology and chemoinformatics. *Bioinformatics* **2013**, *29*, 1092–1094. [CrossRef]
122. Moriwaki, H.; Tian, Y.S.; Kawashita, N.; Takagi, T. Mordred: A molecular descriptor calculator. *J. Cheminf.* **2018**, *10*, 4. [CrossRef] [PubMed]
123. Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliaskova, N.; Kuhn, S.; Pluskal, T.; Rojas-Chertó, M.; Spjuth, O.; et al. The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminf.* **2017**, *9*, 33. [CrossRef] [PubMed]
124. Dong, J.; Yao, Z.J.; Zhang, L.; Luo, F.; Lin, Q.; Lu, A.P.; Chen, A.F.; Cao, D.S. PyBioMed: A python library for various molecular representations of chemicals, proteins and DNAs and their interactions. *J. Cheminf.* **2018**, *10*, 16. [CrossRef]
125. Bento, A.P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open source chemical structure curation pipeline using RDKit. *J. Cheminf.* **2020**, *12*, 51. [CrossRef] [PubMed]
126. Indigo Toolkit. Available online: <https://lifescience.opensource.epam.com/indigo/> (accessed on 20 October 2021).
127. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33. [CrossRef] [PubMed]
128. O’Boyle, N.M.; Hutchison, G.R. Cinfony – combining open source cheminformatics toolkits behind a common interface. *Chem. Cent. J.* **2008**, *2*, 24. [CrossRef]
129. Capuzzi, S.J.; Kim, I.S.J.; Lam, W.I.; Thornton, T.E.; Muratov, E.N.; Pozefsky, D.; Tropsha, A. Chembench: A publicly accessible, integrated cheminformatics portal. *J. Chem. Inf. Model.* **2017**, *57*, 105–108. [CrossRef] [PubMed]
130. Delocalization-Induced Molecular Equality. Available online: <https://depth-first.com/articles/2021/06/17/delocalization-induced-molecular-equality/> (accessed on 20 October 2021).
131. Dhaked, D.K.; Ihlenfeldt, W.D.; Patel, H.; Delannée, V.; Nicklaus, M.C. Toward a comprehensive treatment of tautomerism in cheminformatics including in InChI v2. *J. Chem. Inf. Model.* **2020**, *60*, 1253–1275. [CrossRef] [PubMed]
132. Kochev, N.T.; Paskaleva, V.H.; Jeliaskova, N. Ambit-Tautomer: An open source tool for tautomer generation. *Mol. Inf.* **2013**, *32*, 481–504. [CrossRef] [PubMed]
133. Harańczyk, M.; Gutowski, M. Quantum mechanical energy-based screening of combinatorially generated library of tautomers. TauTGen: A tautomer generator program. *J. Chem. Inf. Model.* **2006**, *47*, 686–694. [CrossRef]
134. MolVS: Molecule Validation and Standardization—MolVS 0.1.1 Documentation. Available online: <https://molvs.readthedocs.io/en/latest/> (accessed on 20 October 2021).
135. Ropp, P.J.; Spiegel, J.O.; Walker, J.L.; Green, H.; Morales, G.A.; Milliken, K.A.; Ringe, J.J.; Durrant, J.D. Gypsum-DL: An open-source program for preparing small-molecule libraries for structure-based virtual screening. *J. Cheminf.* **2019**, *11*, 34. [CrossRef] [PubMed]
136. OpenEye Scientific. Available online: <https://www.eyesopen.com/> (accessed on 20 October 2021).
137. ChemAxon—Software Solutions and Services for Chemistry & Biology. Available online: <https://chemaxon.com/> (accessed on 20 October 2021).
138. Sitzmann, M.; Ihlenfeldt, W.D.; Nicklaus, M.C. Tautomerism in large databases. *J. Comput. Aided Mol. Des.* **2010**, *24*, 521–551. [CrossRef] [PubMed]
139. Wahl, O.; Sander, T. Tautobase: An open tautomer database. *J. Chem. Inf. Model.* **2020**, *60*, 1085–1089. [CrossRef]
140. Sobez, J.G.; Reiher, M. Molassembler: Molecular graph construction, modification, and conformer generation for inorganic and organic molecules. *J. Chem. Inf. Model.* **2020**, *60*, 3884–3900. [CrossRef]
141. Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. [CrossRef]
142. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584. [CrossRef]
143. Vainio, M.J.; Johnson, M.S. Generating conformer ensembles using a multiobjective genetic algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474. [CrossRef] [PubMed]
144. O’Boyle, N.M.; Vandermeersch, T.; Flynn, C.J.; Maguire, A.R.; Hutchison, G.R. Confab—Systematic generation of diverse low-energy conformers. *J. Cheminf.* **2011**, *3*, 8. [CrossRef] [PubMed]
145. Watts, K.S.; Dalal, P.; Murphy, R.B.; Sherman, W.; Friesner, R.A.; Shelley, J.C. ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534–546. [CrossRef] [PubMed]
146. Miteva, M.A.; Guyon, F.; Tuffery, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622–W627. [CrossRef]
147. Ebejer, J.P.; Morris, G.M.; Deane, C.M. Freely available conformer generation methods: How good are they? *J. Chem. Inf. Model.* **2012**, *52*, 1146–1158. [CrossRef]
148. Lewis-Atwell, T.; Townsend, P.A.; Grayson, M.N. Comparisons of different force fields in conformational analysis and searching of organic molecules: A review. *Tetrahedron* **2021**, *79*, 131865. [CrossRef]

149. Rackers, J.A.; Wang, Z.; Lu, C.; Laury, M.L.; Lagardère, L.; Schnieders, M.J.; Piquemal, J.P.; Ren, P.; Ponder, J.W. Tinker 8: Software tools for molecular design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289. [CrossRef]
150. Folmsbee, D.; Hutchison, G. Assessing conformer energies using electronic structure and machine learning methods. *Int. J. Quantum Chem.* **2020**, *121*, e26381 [CrossRef]
151. Tanemura, K.A.; Das, S.; Merz, K.M. AutoGraph: Autonomous graph-based clustering of small-molecule conformations. *J. Chem. Inf. Model.* **2021**, *61*, 1647–1656. [CrossRef] [PubMed]
152. Dahlgren, B. ChemPy: A package useful for chemistry written in Python. *Open Source Softw.* **2018**, *3*, 565. [CrossRef]
153. Rackaukas, C.; Nie, Q. DifferentialEquations.jl—A performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Software* **2017**, *5*, 15. [CrossRef]
154. Backman, T.W.H.; Cao, Y.; Girke, T. ChemMine tools: An online service for analyzing and clustering small molecules. *Nucleic Acids Res.* **2011**, *39*, W486–W491. [CrossRef]
155. Cao, Y.; Charisi, A.; Cheng, L.C.; Jiang, T.; Girke, T. ChemmineR: A compound mining framework for R. *Bioinformatics* **2008**, *24*, 1733–1734. [CrossRef] [PubMed]
156. Voicu, A.; Duteanu, N.; Voicu, M.; Vlad, D.; Dumitrascu, V. The rcdk and cluster R packages applied to drug candidate selection. *J. Cheminf.* **2020**, *12*, 3. [CrossRef] [PubMed]
157. Matsuoka, S.; Holy, T.; TagBot, J.; Richard. mojaie/MolecularGraph.jl: v0.9.0. *Zenodo* **2021**. [CrossRef]
158. Dask: Scalable Analytics in Python. Available online: <https://dask.org/> (accessed on 20 October 2021).
159. Cereto-Massagué, A.; Ojeda, M.J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63. [CrossRef]
160. Dalke, A. The chemfp project. *J. Cheminf.* **2019**, *11*, 76. [CrossRef]
161. Rajan, K.; Hein, J.M.; Steinbeck, C.; Zielesny, A. Molecule set comparator (MSC): A CDK-based open rich-client tool for molecule set similarity evaluations. *J. Cheminf.* **2021**, *13*, 5. [CrossRef]
162. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminf.* **2015**, *7*, 20. [CrossRef] [PubMed]
163. van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
164. Scott, O.B.; Chan, A.W.E. ScaffoldGraph: An open-source library for the generation and analysis of molecular scaffold networks and scaffold trees. *Bioinformatics* **2020**, *36*, 3930–3931. [CrossRef] [PubMed]
165. Lai, J.; Li, X.; Wang, Y.; Yin, S.; Zhou, J.; Liu, Z. AIScaffold: A web-based tool for scaffold diversification using deep learning. *J. Chem. Inf. Model.* **2021**, *61*, 1–6. [CrossRef]
166. Yang, Q.; Li, Y.; Yang, J.D.; Liu, Y.; Zhang, L.; Luo, S.; Cheng, J.P. Holistic prediction of the pKa in diverse solvents based on a machine-learning approach. *Angew. Chem. Int. Ed.* **2020**, *59*, 19282–19291. [CrossRef]
167. Mansouri, K.; Cariello, N.F.; Korotcov, A.; Tkachenko, V.; Grulke, C.M.; Sprankle, C.S.; Allen, D.; Casey, W.M.; Kleinstreuer, N.C.; Williams, A.J. Open-source QSAR models for pKa prediction using multiple machine learning approaches. *J. Cheminf.* **2019**, *11*, 60. [CrossRef] [PubMed]
168. Roszak, R.; Beker, W.; Molga, K.; Grzybowski, B.A. Rapid and accurate prediction of pKa Values of C–H acids using graph convolutional neural networks. *J. Am. Chem. Soc.* **2019**, *141*, 17142–17149. [CrossRef]
169. Pan, X.; Wang, H.; Li, C.; Zhang, J.Z.H.; Ji, C. MolGpka: A web server for small molecule pKa prediction using a graph-convolutional neural network. *J. Chem. Inf. Model.* **2021**, *61*, 3159–3165. [CrossRef]
170. Tomar, N.; De, R.K. Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene* **2013**, *521*, 1–14. [CrossRef]
171. Radojković, V.; Schreiber, I. Constrained stoichiometric network analysis. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9910–9921. [CrossRef]
172. Ruf, A.; d’Hendecourt, L.; Schmitt-Kopplin, P. Data-driven astrochemistry: One step further within the origin of life puzzle. *Life* **2018**, *8*, 18. [CrossRef] [PubMed]
173. Geisberger, T.; Diederich, P.; Steiner, T.; Eisenreich, W.; Schmitt-Kopplin, P.; Huber, C. Evolutionary steps in the analytics of primordial metabolic evolution. *Life* **2019**, *9*, 50. [CrossRef] [PubMed]
174. Scheubert, K.; Hufsky, F.; Böcker, S. Computational mass spectrometry for small molecules. *J. Cheminf.* **2013**, *5*, 12. [CrossRef]
175. Kuhn, S.; Colreavy-Donnelly, S.; de Souza, J.S.; Borges, R.M. An integrated approach for mixture analysis using MS and NMR techniques. *Faraday Discuss.* **2019**, *218*, 339–353. [CrossRef]
176. Howarth, A.; Ermanis, K.; Goodman, J.M. DP4-AI automated NMR data analysis: Straight from spectrometer to structure. *Chem. Sci.* **2020**, *11*, 4351–4359. [CrossRef] [PubMed]
177. Kew, W.; Blackburn, J.W.; Clarke, D.J.; Uhrin, D. Interactive van Krevelen diagrams—Advanced visualisation of mass spectrometry data of complex mixtures. *Rapid Commun. Mass Spectrom.* **2017**, *31*, 658–662. [CrossRef] [PubMed]
178. Brockman, S.A.; Roden, E.V.; Hegeman, A.D. van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics* **2018**, *14*, 48. [CrossRef] [PubMed]
179. Hughey, C.A.; Hendrickson, C.L.; Rodgers, R.P.; Marshall, A.G.; Qian, K. Kendrick Mass Defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73*, 4676–4681. [CrossRef] [PubMed]
180. Bramer, L.M.; White, A.M.; Stratton, K.G.; Thompson, A.M.; Claborne, D.; Hofmockel, K.; McCue, L.A. ftmsRanalysis: An R package for exploratory data analysis and interactive visualization of FT-MS data. *PLoS Comput. Biol.* **2020**, *16*, e1007654. [CrossRef]



181. Reaxys. Available online: <https://www.reaxys.com/> (accessed on 20 October 2021).
182. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2017**, *46*, D608–D617. [[CrossRef](#)]
183. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
184. biodb: An R Package for Accessing Biological and Chemical Databases and Developing or Extending New Connectors. Available online: <https://github.com/pkrog/biodb> (accessed on 20 October 2021).
185. Szöcs, E.; Stirling, T.; Scott, E.R.; Scharmüller, A.; Schäfer, R.B. webchem: An R Package to retrieve chemical information from the web. *J. Stat. Softw.* **2020**, *93*, 13 [[CrossRef](#)]
186. PubChemPy. Available online: <https://pubchempy.readthedocs.io/en/latest/> (accessed on 20 October 2021).
187. Awale, M.; Visini, R.; Probst, D.; Arús-Pous, J.; Reymond, J.L. Chemical space: Big data challenge for molecular diversity. *Chim. Int. J. Chem.* **2017**, *71*, 661–666. [[CrossRef](#)]
188. Probst, D.; Reymond, J.L. Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminf.* **2020**, *12*, 12. [[CrossRef](#)] [[PubMed](#)]
189. Awale, M.; Probst, D.; Reymond, J.L. WebMolCS: A web-based interface for visualizing molecules in three-dimensional chemical spaces. *J. Chem. Inf. Model.* **2017**, *57*, 643–649. [[CrossRef](#)]
190. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Oftware for Exploring and Manipulating Networks. In Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, CA, USA, 17–20 May 2009; [[CrossRef](#)]
191. Shannon, P. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
192. Hagberg, A.A.; Schult, D.A.; Swart, P.J. Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 19–24 August 2008; pp. 11–15.
193. Peixoto, T.P. *The Graph-Tool Python Ibrary*; 2014. Available online: [https://figshare.com/articles/dataset/graph\\_tool/1164194/14](https://figshare.com/articles/dataset/graph_tool/1164194/14) (accessed on 20 October 2021).
194. Gupta, U.; Vlachos, D.G. Reaction Network Viewer (ReNView): An open-source framework for reaction path visualization of chemical reaction systems. *SoftwareX* **2020**, *11*, 100442. [[CrossRef](#)]
195. Neo4j Graph platform—The Leader in Graph Databases. Available online: <https://neo4j.com/> (accessed on 20 October 2021).
196. Probst, D.; Reymond, J.L. SmilesDrawer: Parsing and Drawing SMILES-Encoded molecular structures using client-side JavaScript. *J. Chem. Inf. Model.* **2018**, *58*, 1–7. [[CrossRef](#)] [[PubMed](#)]
197. Leruli. Available online: <https://www.leruli.com/> (accessed on 20 October 2021).