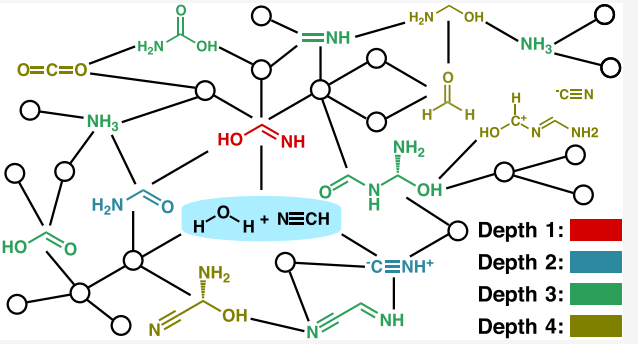
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Thermally Accessible Prebiotic Pathways for Forming Ribonucleic Acid and Protein Precursors from Aqueous Hydrogen Cyanide

Qiyuan Zhao, Sanjay S. Garimella, and Brett M. Savoie*

Cite This: https://doi.org/10.1021/jacs.2c11857
Read Online

ABSTRACT: The search for prebiotic chemical pathways to biologically relevant molecules is a long-standing puzzle that has generated a menagerie of competing hypotheses with limited experimental prospects for falsification. However, the advent of computational network exploration methodologies has created the opportunity to compare the kinetic plausibility of various channels and even propose new pathways. Here, the space of organic molecules that can be formed within four polar or pericyclic reactions from water and hydrogen cyanide (HCN), two established prebiotic candidates for generating biological precursors, was comprehensively explored with a state-of-the-art exploration algorithm. A surprisingly diverse reactivity landscape was revealed within just a few steps of these simple molecules. Reaction pathways to several biologically relevant molecules were discovered involving lower activation energies and fewer reaction steps compared with recently proposed alternatives. Accounting for water-catalyzed reactions qualitatively affects the interpretation of the network kinetics. The case-study also highlights omissions of simpler and lower barrier reaction pathways to certain products by other algorithms that qualitatively affect the interpretation of HCN reactivity.



INTRODUCTION

Reaction pathways that are capable of generating biologically relevant molecules under prebiotic constraints have been the subject of intense speculation.[1−10] The "RNA World" is a hypothesis that self-replicating RNA molecules first emerged that were capable of a range of catalytic behaviors.[11−16] This hypothesis has been partly spurred by the experimental discovery that hydrogen cyanide (HCN), a well-known prebiotic molecule, can participate in several pathways that form precursors to RNA, proteins, and lipids (Figure 1a).[17−26] Recently, the enumeration of potential reaction pathways involving HCN using known reaction templates has revealed many plausible conversions, even when limited to the realm of known reactions.[9] However, establishing the kinetic relevance of these pathways requires step-by-step reaction mechanisms that are difficult to experimentally obtain or computationally validate without a transition state search. While molecular reaction modeling has assisted in generating plausible transition states for some of the involved conversions, predictive explorations of prebiotic reactions are still in their infancy due to a combination of the computational cost and limited accuracy of contemporary reaction network exploration algorithms.[10,27]

Over the past few decades, molecular modeling has provided details on plausible intermediates and transition states associated with producing a variety of RNA and protein precursors (Figure 1a). Nevertheless, the history of computational modeling leading to focusing on HCN as an important prebiotic reactant is illustrative of the slow process of *ad hoc* reaction pathway proposals and characterizations. In 2007, Debjani et al. reported a pathway for synthesizing an HCN pentamer, adenine, from the HCN tetramer, AICN.[28] This observation triggered a search for even simpler reactants. Among the important milestones along the way, Sponer et al. derived the reaction pathway for adenine formation from a simple formamide dimer, *N*-(aminohydroxymethyl)-formamide,[29] and later Wang et al. used formamide as an even simpler starting point for water-catalyzed and autocatalyzed reaction mechanisms (orange box in Figure 1a).[30,31] Later in 2016, Kua et al. reported reaction pathways starting from HCN, NH_3, and H_2O to synthesize a formamide trimer, 1,3,5-triazinane-2,4,6-triol (blue box in Figure 1a).[32] This abbreviated history, spanning 10 years and many research groups, illustrates the important role played by reaction modeling in resolving specific pathways, but also the limitations and retrospective nature of typical explorations.

Received: November 7, 2022

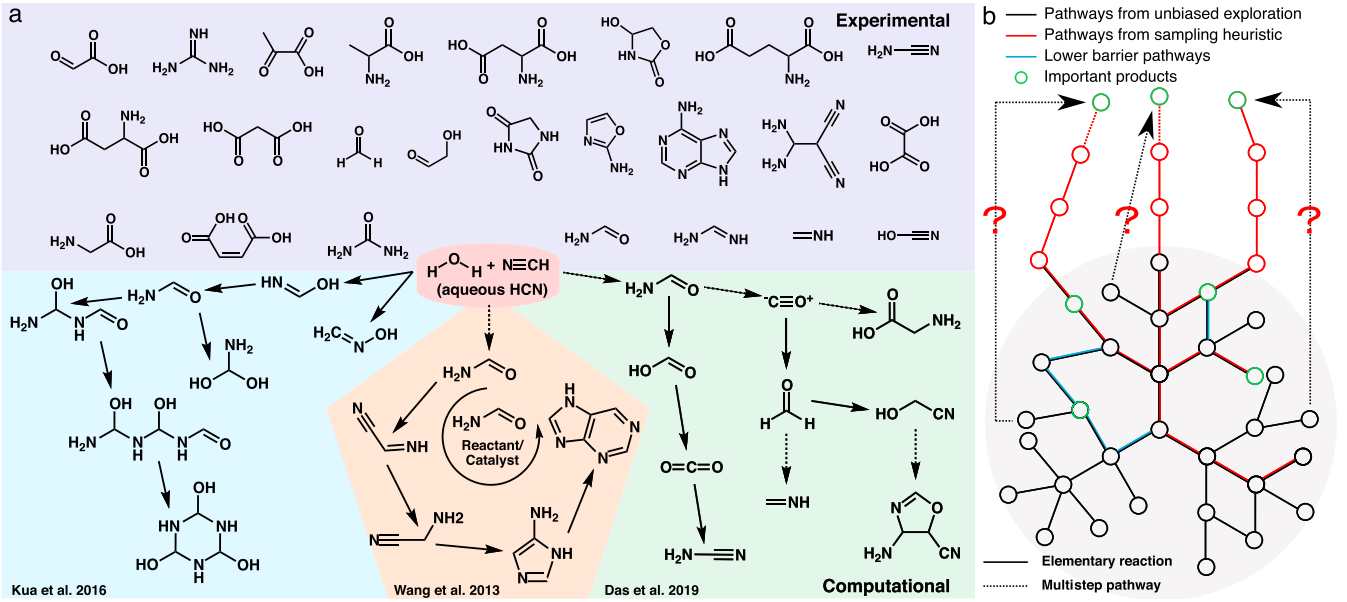


Figure 1. Diverse pathways and intermediates have been hypothesized for guiding the transition from prebiotic chemistry. (a) Previously synthesized RNA and protein precursors discovered in experiments (purple) and selected reaction mechanisms involving aqueous HCN discovered by computational approaches. (b) Comparison between wide (black) and deep (red) network exploration. Deep explorations may find pathways to interesting products but nevertheless miss kinetically relevant pathways.

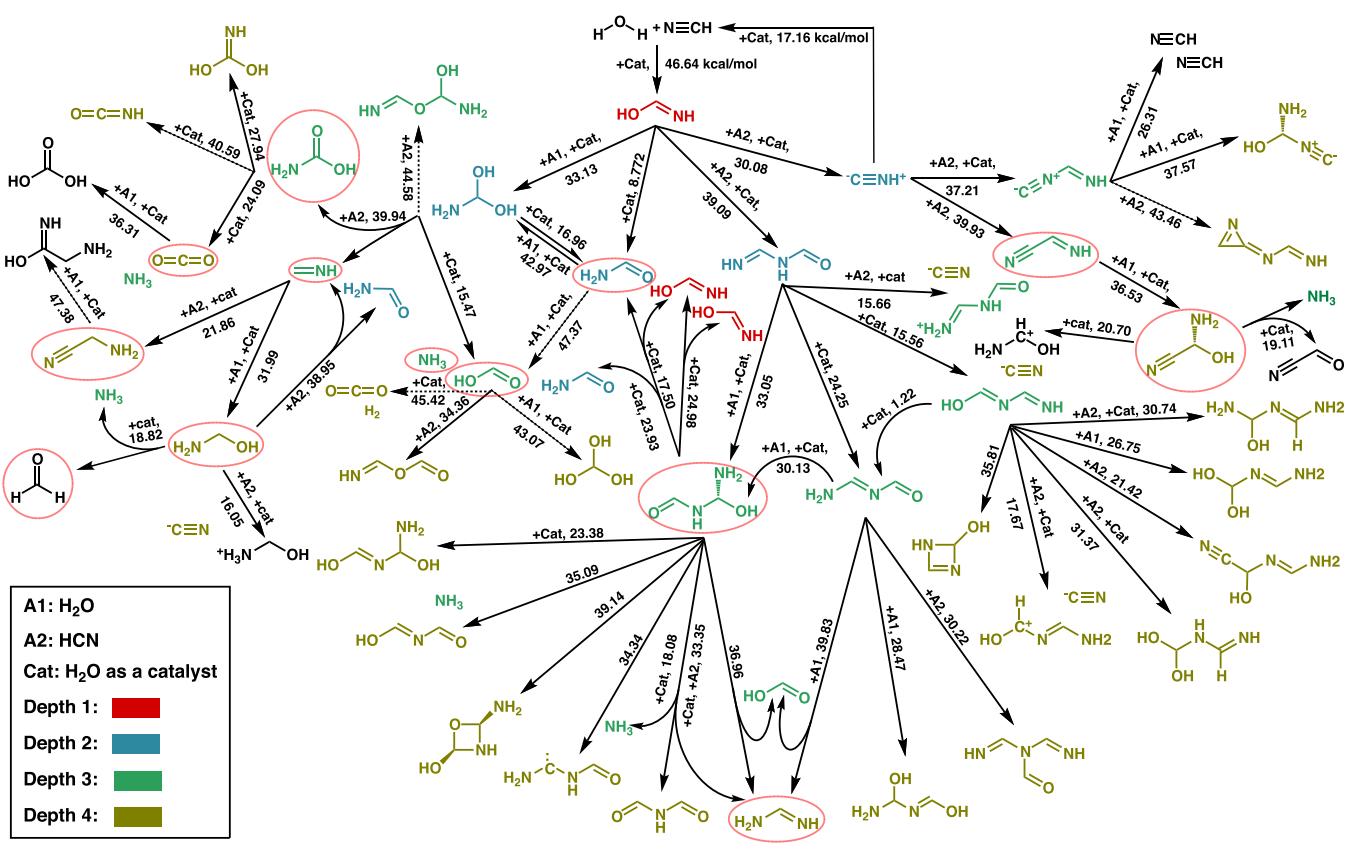


Figure 2. Summary of the reaction networks generated by comprehensive b2f2 exploration and kinetic modeling. The reaction network starting from water and HCN showing kinetically favorable reactions ($\Delta G^{\dagger} > 50$) with a maximum reaction depth of five. The arrows follow the direction of reaction exploration, (i.e., they point from the inputted reactants to the discovered products). The text "+A1/+A2" refers to a bimolecular reaction between the shown reactant and water or HCN, respectively, while the text "+cat" refers to a water-catalyzed reaction. The dotted arrows indicate reactions with a free energy of activation between 40 and 50 kcal/mol, which are less kinetically relevant in the whole reaction network. The circled products are putative RNA and sugar precursors.

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Because *ad hoc* reaction exploration only addresses a small number of hypothesis-driven pathways, many unexpected but important reactions might be left out of consideration (Figure 1 b).

Reaction exploration algorithms that do not rely on existing reaction hypotheses have the potential to move beyond rationalizing existing pathway proposals to discover authentically new prebiotic pathways.[33,34] The paradigmatic example for prebiotic chemistries is the use of the *ab initio* nanoreactor (AINR) algorithm by Wang et al. to generate reaction pathways to amino acids starting from simple molecules (CO, H_2, H_2O, and NH_3).[35] Das et al. further applied the AINR to study the biological precursors that can be formed from water and HCN under noncatalyzed conditions (green box in Figure 1a).[36] More recently, Stan et al. developed a fully automated postsimulation evaluation procedure for AINR and applied it to study the formose reaction network.[37] A major advantage of the AINR is its ability to generate long reaction sequences and correspondingly deep reaction networks. For example, Das et al. were able to identify a remarkable 11-step sequence for forming oxazole derivatives. Despite these compelling demonstrations, AINR uses violent reaction conditions to expedite reaction discovery (e.g., using boundary pistoning to induce reactions), which may bias the pathways that are discovered. Search biases pose essential challenges for interpreting computationally derived reaction networks, since missing a kinetically relevant reaction early in the network propagates into all of the explorations that are performed downstream of that point (Figure 1b). For this reason, it is worthwhile to revisit systems like aqueous HCN using exploration algorithms that possess some guarantee on which reactions are being considered.

Here, the potential reactions involving HCN and water out to four steps have been comprehensively explored to assess whether alternative routes to biologically relevant precursors like formaldehyde and formic acid exist that have been missed by algorithms focused on exploring deeper multistep reaction pathways. This exploration has been performed with Yet Another Reaction Program (YARP), an ultralow cost exploration method that uses generic reaction rules for exploration.[38−40] Similar to other contemporary exploration methods,[41−50] YARP couples a reaction exploration algorithm with quantum chemistry based transition state (TS) searches. A major advantage of YARP is that it can be run in a comprehensive mode such that all reactions of a generic class can be attempted, thus providing guarantees on the scope of the search. In the present case, TS searches were performed using YARP for all reactions involving breaking two bonds and forming two bonds (b2f2) with water and HCN as reactants. In organic chemistry terms, this exploration includes pericyclic and polar reactions, but excludes radical reactions since all reactants are closed-shell. The products for which valid TSs were discovered were then used as reactants, along with HCN and water, for another set of b2f2 searches. This process was repeated until all b2f2 channels out to four reactions from aqueous HCN had been characterized (Figure 2). For each reaction, an additional single water-catalyzed b3f3 reaction was attempted to explore the potential impact of water catalysis on the free energies of activation ($\Delta G^{\dagger}$). For detailed YARP settings and parameters and the workflow of water-catalyzed reaction mechanism generation, we direct readers to the SI. This "wide" exploration of aqueous HCN reactivity discovered several novel transformations and much shorter pathways to biological precursors than have previously been reported from experiments or in earlier AINR-based studies of the same system.

RESULTS AND DISCUSSION

Reaction Network Starting from Water and HCN. The overall reaction network that was generated by the comprehensive b2f2 exploration is shown in Figure 2. Only reactions exhibiting $\Delta G^{\dagger} < 40$ kcal/mol are considered kinetically accessible based on the environmental temperature of ∼100 °C under which these reactions are thought to occur.[36] Reactions with $\Delta G^{\dagger} > 50$ are not shown in the figure, but all reactions and products can be found in the SI. While the focus here is on thermally-accessible reaction channels, a smaller network exploration with hydroxyl radical available as a reactant is provided at the same level of theory for comparison with this network (Figure S3). Such radicals have been hypothesized to be potentially available from ionizing radiation, but a detailed exploration is beyond the scope of the current study.[8,51]

The only kinetically accessible product that is formed in the first step of aqueous HCN reaction exploration is formamidic acid. After the second-step reaction enumeration and TS characterization, formamide ($\Delta G^{\dagger}$ = 9.77 kcal/mol), hydrogen isocyanide (30.08), aminomethanediol (33.13), and *N*-(iminomethyl)formamide (39.09) were identified as kinetically accessible products of reactions involving formamidic acid, water, and HCN. The low barrier height signals a strong kinetic preference for formamide formation, which agrees with previous studies.[32,36] After three reaction steps, pathways are discovered to carbamic acid and methanimine (from aminomethanediol, $\Delta G^{\dagger}$ = 39.94 kcal/mol), formic acid and ammonia (from both formamide and aminomethanediol, with $\Delta G^{\dagger}$ = 47.37/15.47 kcal/mol, respectively), *N*-(aminohydroxymethyl)formamide (from *N*-(iminomethyl)-formamide, $\Delta G^{\dagger}$ = 33.05 kcal/mol), and iminoacetonitrile (from hydrogen isocyanide, $\Delta G^{\dagger}$ = 39.93 kcal/mol), among other products. Of the two pathways that produce formic acid, only the formamide pathway has been reported before. However, the current network predicts that the newly discovered pathway through aminomethanediol is kinetically preferred. The fourth step of exploration discovered kinetically accessible pathways to 20 additional products, including carbon dioxide, aminoacetonitrile, and aminomethanol, products that have been highlighted in several computational and experimental studies.[24,24,36] As with formic acid, a lower barrier reaction pathway for carbon dioxide formation via carbamic acid ($\Delta G^{\dagger}$ = 24.09 kcal/mol) is newly reported here. At the fourth step of exploration there are also several products that have been previously invoked as important intermediates, including aminomethanol, aminoacetonitrile, 2-amino-2-hydroxyacetonitrile, and carbon dioxide. To make this connection explicit, an additional step of exploration was performed on these species, which yielded additional channels for forming formaldehyde ($\Delta G^{\dagger}$ = 18.82 kcal/mol), formyl cyanide ($\Delta G^{\dagger}$ = 19.11 kcal/mol), and carbonic acid ($\Delta G^{\dagger}$ = 36.31 kcal/mol).

The network also includes several examples of backtracking, where a nominally lower depth product forms preferentially from a reactant that is deeper in the network. For instance, formamidic acid and formamide can be formed from the unimolecular decomposition of *N*-(aminohydroxymethyl)-formamide with $\Delta G^{\dagger}$ = 17.50 kcal/mol. Considering that

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

formamidic acid is the only kinetically accessible product in the first reaction step and formamide is the major product of the second reaction step, the bimolecular reaction of these two compounds to form *N*-(aminohydroxymethyl)formamide ($\Delta G^{\dagger}$ = 27.95 kcal/mol) creates a kinetically important reaction cycle. Similar backtracking reactions, e.g., the reaction of formamide with methylimine to form aminomethanol and the dimerization of HCN, also spontaneously appear in the reaction network. Thus, an indirect product of the comprehensive exploration performed here is the discovery of low barrier transition states for a wide range of bimolecular reactions, not just those involving water and HCN.

To test the kinetic relevance of competing *N*-(aminohydroxymethyl)formamide reactions, kinetic simulations of the reaction network were performed using Cantera,[52] and the cumulative net reaction fluxes were used to identify the kinetically important channels and intermediates (Figure 3).

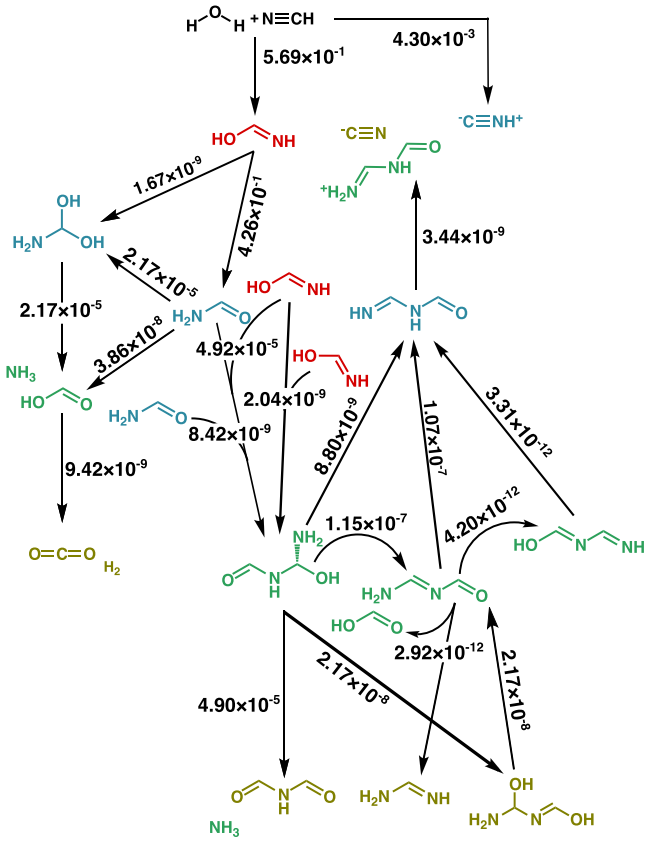


Figure 3. A subnetwork composed of the 20 reactions with the highest cumulative net flux based on the kinetic simulations. The arrows represent the direction of the net flux, and the numbers represent the fraction of cumulative net flux for each reaction. The color of each product matches that in Figure 2, and represents the number of reaction steps in the shortest path to form that product.

Based on the kinetic modeling, the bimolecular reaction of formamide and formamidic acid is the largest flux pathway yielding *N*-(aminohydroxymethyl)formamide. Two other pathways also show appreciable but lower flux (Figure 3). In a mass spectrometry study of HCN reactivity, Andersen et al. hypothesized that *N*-(aminohydroxymethyl)formamide, formamide, and *N*-methanimidoylformamide form an autocatalytic loop, where formamide reacts with *N*-methanimidoylformamide to form *N*-(aminohydroxymethyl)formamide, followed by the decomposition of *N*-(aminohydroxymethyl)formamide to two formamide molecules.[27] Our reaction network reproduces this hypothesized cycle, but also identifies other intermediates (e.g., two isomers of *N*-methanimidoylformamide, NC═NC═O and N═CN═CO, which cannot be deduced by mass spectrometry) and seven decomposition pathways of *N*-(aminohydroxymethyl)formamide that siphon off concentration flux from this loop. The kinetic modeling result indicates that the major source of *N*-(aminohydroxymethyl)formamide comes from formamide and its isomer, and *N*-(aminohydroxymethyl)formamide mainly converts into formamidine, which is an initial reactant for the prebiotic synthesis of various products.[53−55]

A total of 13 putative RNA and sugar precursors were discovered in the reaction network (circled products in Figure 2). These species have either been proposed by computational studies as important intermediates or identified by experimental work as precursors for biotic molecules (see Table S1 for the detailed summary and Figure S2 for the concentration distribution). In addition to the neutral closed-shell species, some charged species and diradical species were also discovered during the exploration. For instance, cyanide, a common product in HCN polymerization,[56] is connected by three kinetically favorable pathways in the reaction network (Figure 2).

Finally, we note that many additional reactions with higher barriers were discovered but not displayed in Figure 2. Some of these reactions might indeed be more competitive when simulated in bulk water rather than with implicit solvent as has been done here. For example, formic acid decomposition to carbon monoxide is expected to be relatively facile, but this reaction is excluded from Figure 2 due to its relatively high barrier under the simulated conditions (54 kcal/mol). The detailed geometries of all the products shown in this network and corresponding transition state structures can be found in the SI.

Pathway Comparisons for Producing Small Molecules. The present comprehensive exploration discovered new reaction pathways to formic acid, carbon dioxide, and formaldehyde (Figure 4). To establish the kinetic relevance of these new pathways, they have been compared with the energy diagrams for the pathways discovered by Das et al. using the AINR and the pathways predicted by the kinetic analysis to have the largest concentration flux. For the energy diagram, the pathway with the lowest free energy of activation for the rate-limiting step was considered (i.e., overall barrier, $\Delta G^{\dagger}_{max}$), while for the analysis of reaction net flux, the fraction of the cumulative reaction net flux of each reaction was used to determine the most important channels. Since all of the discovered multistep reaction pathways share the same first step, the overall barrier reported here excludes the barrier of the first step (i.e., water + HCN to formamidic acid). For formic acid formation (Figure 4a), the pathway indicated by red ($G^{\dagger}_{max}$ = 47.37 kcal/mol) was discovered here by YARP and also reported by Das et al.,[36] while the black (33.13), green (36.96), and blue (42.97) pathways are exclusive to the YARP exploration and more kinetically competitive than the red pathway. The red pathway is partly shared with the newly reported green pathway. The much lower overall barrier of the new pathway reflects the kinetic preference for formamide to first react with formamidic acid to form *N*-(aminohydroxymethyl)formamide. The decomposition of *N*-

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

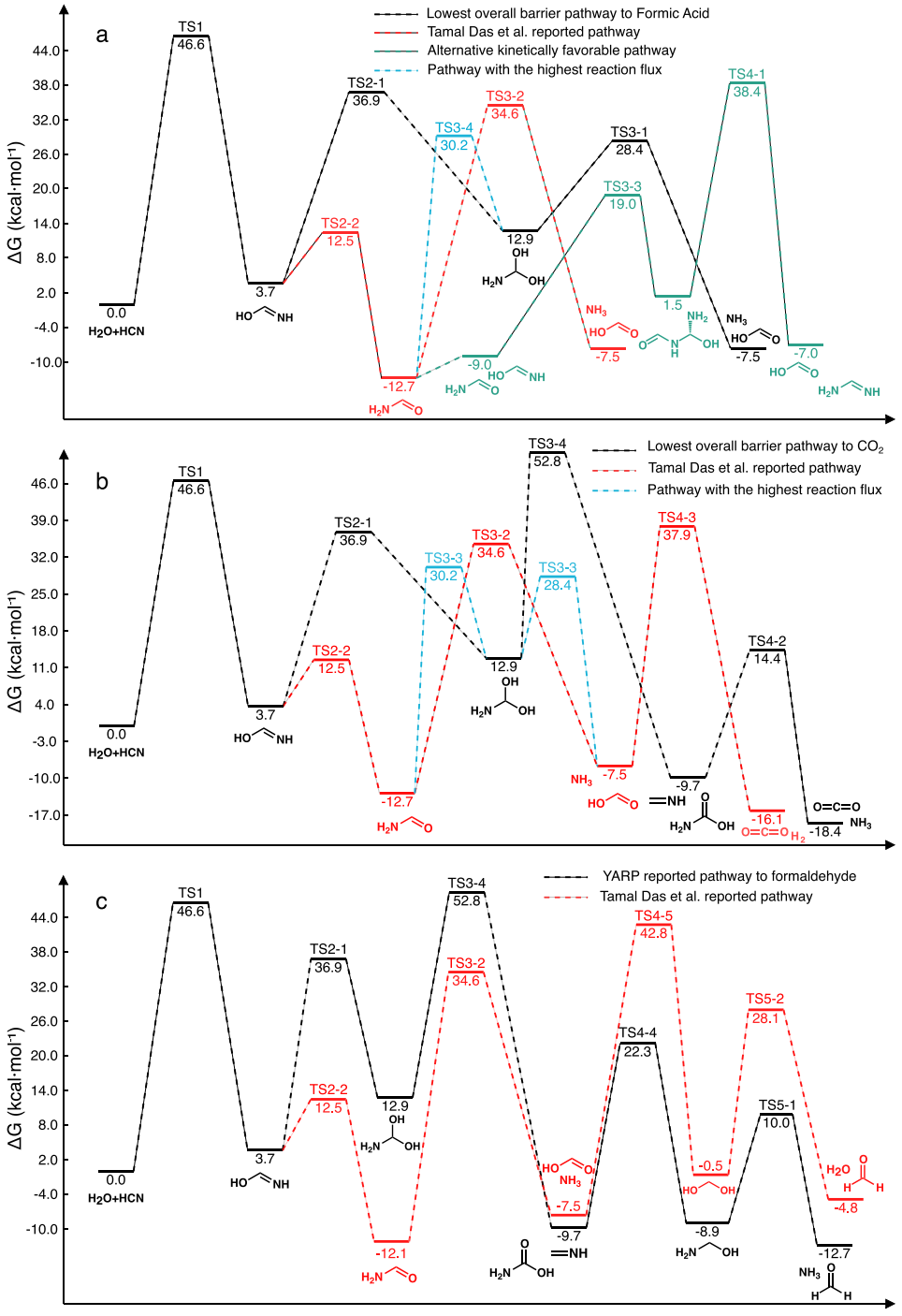


Figure 4. Energy diagrams of kinetically favorable pathways leading to the formation of (a) formic acid, (b) carbon dioxide, and (c) formaldehyde. Black lines refer to the lowest overall barrier pathways identified by the comprehensive exploration, while red lines represent the pathways reported by Das et al. Green lines in panel a refer to an alternative kinetically favorable pathway, while blue lines in panels (a) and (b) refer to the pathway with the highest reaction net flux through the kinetic modeling.

(aminohydroxymethyl)formamide to formic acid is the rate-limiting step in the green pathway but with a much lower activation energy of 36.96 kcal/mol. In addition, a fundamentally new black pathway involving aminomethanediol as an intermediate rather than formamide results in the lowest overall barrier. Surprisingly, the blue pathway, where aminomethanediol also serves as an intermediate but is generated through formamide, shows the highest net flux during kinetic simulations (Figure 3). This is due to a much higher concentration of formamide than aminomethanediol (12 orders of magnitude) that compensates for the somewhat higher barrier of the rate-limiting step for the formamide channel. The kinetic analysis reveals the potential errors involved in only considering the lowest overall barrier when judging the relevance of pathways, since fluxes can be qualitatively impacted by competing channels that are not typically shown on such diagrams.

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

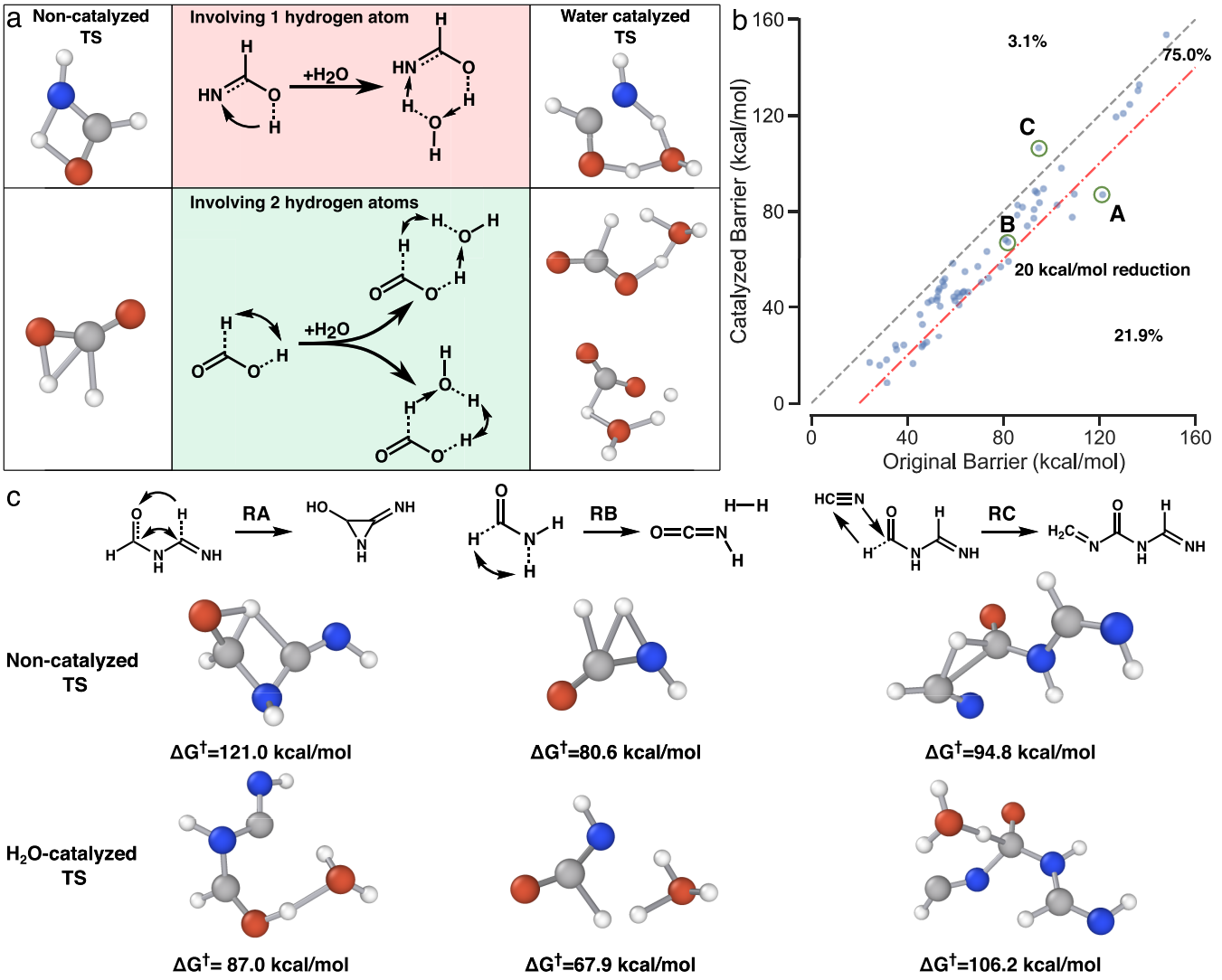


Figure 5. Overview of the study of water-catalyzed reaction mechanisms. (a) The generation of water-catalyzed reaction mechanisms. For reactions involving one hydrogen atom, the potentially catalytic pathway is singular. For H_2 elimination reactions, two distinct catalytic pathways are possible. (b) Correlation plots comparing the activation energies of 64 pairs of water-catalyzed and noncatalyzed reactions. (c) Three water-catalyzed reaction examples representing the maximum decrease (RA), the average effect (RB), and the maximum increase (RC) on the activation energy when including water.

The reaction pathway for carbon dioxide formation reported by Das et al. follows the sequence of formic acid formation (from Figure 4a) with a final step of H_2 elimination (red line in Figure 4b). The comprehensive search also discovered this pathway with an overall barrier of 47.37 kcal/mol (the second rate-limiting step happens on the H_2 elimination step with $\Delta G^{\dagger}$ = 45.42 kcal/mol, which is also fairly high). An alternative pathway with lower overall barrier was also discovered (black), where carbon dioxide is produced by carbamic acid decomposition with the formation of carbamic acid from aminomethanediol being the rate-limiting step ($\Delta G^{\dagger}$ = 39.94). However, due to a similarly large concentration difference between formic acid and carbamic acid (8 orders of magnitude), the reaction pathway with the highest net flux (blue lines in Figure 4b) occurs with aminomethanediol and formic acid as intermediates.

Only one pathway to producing formaldehyde was discovered within five reactions of aqueous HCN (black line in Figure 4c). The first three steps of this pathway are shared with the carbon dioxide forming pathway, but once methanimine is formed, it reacts with water to form aminomethanol, which then decomposes into formaldehyde and ammonia. The rate-limiting step of this pathway is methanimine formation ($G^{\dagger}_{max}$ = 39.94 kcal/mol). The most kinetically favorable formaldehyde formation pathway reported by Das et al. is completely different because the conditions of the nanoreactor produce H_2, which then reacts with formic acid. To compare the overall barrier of these pathways, we reoptimized the transition states of the last two steps provided by Das et al. at the same level of theory as used in this study (red line in Figure 4c). The overall barrier of this pathway is 50.3 kcal/mol, coming from the barrier of H_2 addition to formic acid and is >10 kcal/mol larger than the formaldehyde formation pathway discovered by the comprehensive search.

Water as an Important Catalyst. Another source of error in network exploration is failing to localize the lowest barrier TS for a given reaction. For example, the aqueous HCN network contains many proton transfers that could potentially be catalyzed by water (Figure 5a).[36,57] To interrogate the

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

importance of including water as a catalyst, water-catalyzed TS searches were performed. For this comparison, the $\Delta G^{\dagger}$ threshold on selecting reactions was removed to also assess the effect of water catalysis on previously discarded high activation energy reactions.

Intended TSs (i.e., TSs validated to correspond to the putative reaction by an intrinsic reaction coordinate calculation) were found for the catalyzed and noncatalyzed pathways of 64 reactions (Figure 5b). On average, a 12.54 kcal/mol reduction of activation energy occurred from using water as a proton shuttle, and only two reactions exhibited a higher catalyzed barrier. The largest reduction was 34 kcal/mol, while 14 (21.9%) reactions exhibited a >20 kcal/mol reduction. The large magnitude of these reductions illustrates the peril of overinterpreting reaction networks without a consistent treatment of potentially catalyzed alternative pathways.

The water-catalyzed reactions representing the maximum decrease (Figure 5c, RA), the average change (Figure 5c, RB), and the maximum increase (Figure 5c, RC) of the activation energy were analyzed in detail. Reaction RA involves the formation of a strained three-membered ring. The TS of the uncatalyzed pathway contains a fused ring-like structure and exhibits a $\Delta G^{\dagger}$ of 121.0 kcal/mol. However, with water acting as a proton shuttle, a more stable seven-membered ring-like TS structure reduces the $\Delta G^{\dagger}$ by 34 kcal/mol. Similarly, in reaction RB, a six-membered ring-like water-catalyzed TS has a lower barrier than a four-membered ring-like TS in the uncatalyzed case. Forming more stable five-, six-, and seven-membered ring-like TS structures explains most of the cases where the water-catalyzed mechanism results in a reduced $\Delta G^{\dagger}$. Only in reaction RC does the involvement of water increase the reaction barrier. Moreover, no intended TS can be identified for some of the water-catalyzed reaction mechanisms (e.g., the reactions without "+cat" in Figure 2), indicating the necessity of considering and characterizing both water-catalyzed and noncatalyzed mechanisms to build the reaction network.

CONCLUSIONS

This study revisits the question of whether aqueous HCN forms a plausible prebiotic bridge to several biologically relevant precursors using a state-of-the-art reaction network exploration algorithm. A key element of this exploration was performing a breadth-first type exploration that provided guarantees on exhaustive exploration of nonradical pathways to a depth of four reactions. Even with only four-step exploration, the comprehensive search (re)discovered more than 10 important RNA and sugar precursors that can be produced by a series of reactions with an overall barrier lower than 40 kcal/mol. More importantly, this exploration identified new pathways for the formation of formic acid, carbon dioxide, and formaldehyde that have significantly lower overall barriers and higher reaction net flux compared with previous reports. These findings illustrate the power of contemporary reaction exploration algorithms, like YARP, and suggest that they could find broader applicability in this field interpreting complex reaction data (e.g., from high-resolution mass spectrometry) or for hypothesis testing. Nevertheless, the comparisons with other algorithms also reveal the ongoing challenges of missing kinetically relevant reactions during network exploration. While it is relatively easy to construct a narrative around a partially explored network, the later discovery of missing reactions can fundamentally alter whether a given pathway is kinetically relevant. Nevertheless, the comprehensive approach adopted here is not a general prescription for this problem since it is too costly for deep explorations. For example, the reported network is comprehensive but shallow and of necessity excluded potential bimolecular reactions (e.g., involving H_2 and NH_3) and potential radical reactants (e.g., hydroxyl radicals generated from ionizing radiation). Thus, developing exploration policies that are kinetically comprehensive and avoid exponential growth of the reaction search space is critical for deep explorations.

There are several practical possibilities for striking this balance with the current approach. First, the current exploration only allowed water and hydrogen cyanide to participate in bimolecular reactions, which limits the scope of exploration. This restriction could be lifted by using the results of kinetic modeling (e.g., as shown in Figures 3 and S2) to identify species with sufficient abundance to participate in bimolecular reactions. Second, although the five-layer reaction network illustrates the reaction mechanisms of forming some important RNA precursors, some other products, like glycine and glycolaldehyde, are left out due to the insufficient exploration depth. Here, too, kinetic modeling could be used to guide deeper exploration of kinetically relevant branches but shallow exploration of kinetic dead-ends. Third, the current study has only considered water-catalyzed reactions. However, more complex catalytic pathways are expected, including double-water-catalyzed[32] and autocatalyzed reaction mechanisms,[31,58,59] that might further reduce the effective activation energies of some of these pathways.

ASSOCIATED CONTENT

Data Availability Statement

The authors declare that the data supporting the findings of this study are available within the paper and its Supporting Information files. The version of YARP used in this study is available through GitHub under the GNU GPL-3.0 License [https://github.com/zhaoqy1996/YARP/tree/main/version2.0].

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c11857.

Full list of reactions involved in the reaction network, including atom-mapped smiles of reactant and product and the corresponding activation energy and free energy of activation; geometries of reactant, product, and transition state of reactions in the reaction network (ZIP)

Computational costs of exploration; computational method for reaction network exploration, water-catalyzed mechanism generation, kinetic modeling; summary of previously reported products that are discovered in the reaction network; a reaction network starting from hydrogen cyanide and hydroxyl radical (PDF)

AUTHOR INFORMATION

Corresponding Author

Brett M. Savoie − *Davidson School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47906, United*

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

States; orcid.org/0000-0002-7039-4039;
Email: bsavoie@purdue.edu

Authors

Qiyuan Zhao − *Davidson School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47906, United States*

Sanjay S. Garimella − *Davidson School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47906, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c11857

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work performed by Q.Z., S.S.G., and B.M.S was made possible by the Office of Naval Research (ONR) through support provided by the Energetic Materials Program (MURI grant number: N00014-21-1-2476, Program Manager: Dr. Chad Stoltz). B.M.S. also acknowledges partial support for this work from the Dreyfus Program for Machine Learning in the Chemical Sciences and Engineering and the Purdue Process Safety and Assurance Center.

■ REFERENCES

(1) Dyson, F. *Origins of Life*, 2nd ed.; Cambridge University Press, 1999.
(2) Gesteland, R. F.; Cech, T.; Atkins, J. F. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World*, 2nd ed.; Cold Spring Harbor, NY, 1999.
(3) *Life's Origin: The Beginnings of Biological Evolution*, 1st ed.; University of California Press, 2002.
(4) Bada, J. L. How life began on Earth: a status report. *Earth Planet. Sci. Lett.* **2004**, *226*, 1−15.
(5) Steel, M.; Penny, D. Common ancestry put to the test. *Nature* **2010**, *465*, 168−169.
(6) Sutherland, J. D. The origin of life—out of the blue. *Angew. Chem., Int. Ed.* **2016**, *55*, 104−121.
(7) Sutherland, J. D. Opinion: studies on the origin of life—the end of the beginning. *Nat. Rev. Chem.* **2017**, *1*, 1−7.
(8) 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. U. S. A.* **2020**, *117*, 13267−13274.
(9) 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*, No. eaaw1955.
(10) Tran, Q. P.; Adam, Z. R.; Fahrenbach, A. C. Prebiotic reaction networks in water. *Life* **2020**, *10*, 352.
(11) Rich, A. On the problems of evolution and biochemical information transfer. *Horiz. Biochem.* **1962**, 103−126.
(12) Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367−379.
(13) Gilbert, W. Origin of life: The RNA world. *Nature* **1986**, *319*, 618−618.
(14) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99−123.
(15) Muchowska, K. B.; Varma, S. J.; Moran, J. Nonenzymatic metabolic reactions and life's origins. *Chem. Rev.* **2020**, *120*, 7708−7744.
(16) Benner, S. A.; Bell, E. A.; Biondi, E.; Brasser, R.; Carell, T.; Kim, H.-J.; Mojzsis, S. J.; Omran, A.; Pasek, M. A.; Trail, D. When did life likely emerge on Earth in an RNA-first process? *ChemSystemsChem.* **2020**, *2*, No. e1900035.
(17) Bernal, J. D. The physical basis of life. *Proc. Phys. Soc. A* **1949**, *62*, 597.
(18) Urey, H. C. On the early chemical history of the earth and the origin of life. *Proc. Natl. Acad. Sci. U. S. A.* **1952**, *38*, 351−363.
(19) Miller, S. L. Production of some organic compounds under possible primitive earth conditions1. *J. Am. Chem. Soc.* **1955**, *77*, 2351−2361.
(20) Miller, S. L.; Urey, H. C. Organic compound synthesis on the primitive Earth: Several questions about the origin of life have been answered, but much remains to be studied. *Science* **1959**, *130*, 245−251.
(21) Oró, J. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions. *Nature* **1961**, *191*, 1193−1194.
(22) Steinman, G.; Lemmon, R. M.; Calvin, M. Cyanamide: a possible key compound in chemical evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *52*, 27−30.
(23) Ritson, D.; Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **2012**, *4*, 895−899.
(24) Ruiz-Bermejo, M.; Zorzano, M.-P.; Osuna-Esteban, S. Simple organics and biomonomers identified in HCN polymers: An overview. *Life* **2013**, *3*, 421−448.
(25) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **2015**, *7*, 301−307.
(26) Ruiz-Bermejo, M.; de la Fuente, J. L.; Pérez-Fernández, C.; Mateo-Martí, E. A Comprehensive Review of HCN-Derived Polymers. *Processes* **2021**, *9*, 597.
(27) 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.
(28) Roy, D.; Najafian, K.; von Ragué Schleyer, P. Chemical evolution: The mechanism of the formation of adenine under prebiotic conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17272−17277.
(29) Šponer, J. E.; Mládek, A.; Šponer, J.; Fuentes-Cabrera, M. Formamide-based prebiotic synthesis of nucleobases: a kinetically accessible reaction route. *J. Phys. Chem. A* **2012**, *116*, 720−726.
(30) Wang, J.; Gu, J.; Nguyen, M. T.; Springsteen, G.; Leszczynski, J. From formamide to purine: An energetically viable mechanistic reaction pathway. *J. Phys. Chem. B* **2013**, *117*, 2314−2320.
(31) Wang, J.; Gu, J.; Nguyen, M. T.; Springsteen, G.; Leszczynski, J. From formamide to purine: A self-catalyzed reaction pathway provides a feasible mechanism for the entire process. *J. Phys. Chem. B* **2013**, *117*, 9333−9342.
(32) Kua, J.; Thrush, K. L. HCN, formamidic acid, and formamide in aqueous solution: a free-energy map. *J. Phys. Chem. B* **2016**, *120*, 8175−8185.
(33) Sharma, S.; Arya, A.; Cruz, R.; Cleaves, H. J., II Automated Exploration of Prebiotic Chemical Reaction Space: Progress and Perspectives. *Life* **2021**, *11*, 1140.
(34) Butch, C. J.; Meringer, M.; Gagnon, J.-S.; Cleaves, H. J. Open questions in understanding life's origins. *Commun. Chem.* **2021**, *4*, 1−4.
(35) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. chem.* **2014**, *6*, 1044−1048.
(36) Das, T.; Ghule, S.; Vanka, K. Insights into the origin of life: Did it begin from HCN and H2O? *ACS Cent. Sci.* **2019**, *5*, 1532−1540.
(37) Stan, A.; Esch, B. v. d.; Ochsenfeld, C. Fully Automated Generation of Prebiotically Relevant Reaction Networks from Optimized Nanoreactor Simulations. *J. Chem. Theory Comput.* **2022**, *18*, 6700−6712.

(38) Zhao, Q.; Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nat. Comput. Sci.* **2021**, *1*, 479−490.
(39) Zhao, Q.; Hsu, H.-H.; Savoie, B. M. Conformational Sampling for Transition State Searches on a Computational Budget. *J. Chem. Theory Comput.* **2022**, *18*, 3006−3016.
(40) Zhao, Q.; Savoie, B. M. Algorithmic Explorations of Unimolecular and Bimolecular Reaction Spaces. *Angew. Chem., Int. Ed.* **2022**, *61* (46), No. e202210693.
(41) Shang, C.; Liu, Z. P. Stochastic surface walking method for structure prediction and pathway searching. *J. Chem. Theory Comput.* **2013**, *9*, 1838−1845.
(42) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385−1392.
(43) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Comput. Chem.* **2014**, *35*, 166−173.
(44) Suleimanov, Y. V.; Green, W. H. Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods. *J. Chem. Theory Comput.* **2015**, *11*, 4248−4259.
(45) Zimmerman, P. M. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol. Simul.* **2015**, *41*, 43−54.
(46) Simm, G. N.; Reiher, M. Context-driven exploration of complex chemical reaction networks. *J. Chem. Theory Comput.* **2017**, *13*, 6108−6119.
(47) Grambow, C. A.; Jamal, A.; Li, Y. P.; Green, W. H.; Zador, J.; Suleimanov, Y. V. Unimolecular reaction pathways of a γ-ketohydroperoxide from combined application of automated reaction discovery methods. *J. Am. Chem. Soc.* **2018**, *140*, 1035−1048.
(48) Ismail, I.; Stuttaford-Fowler, H. B.; Ochan Ashok, C.; Robertson, C.; Habershon, S. Automatic proposal of multistep reaction mechanisms using a graph-driven search. *J. Phys. Chem. A* **2019**, *123*, 3407−3417.
(49) Van de Vijver, R.; Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Comput. Phys. Commun.* **2020**, *248*, 106947.
(50) Unsleber, J. P.; Grimmel, S. A.; Reiher, M. Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks. *J. Chem. Theory Comput.* **2022**, *18*, 5393−5409.
(51) Yi, R.; Hongo, Y.; Yoda, I.; Adam, Z. R.; Fahrenbach, A. C. Radiolytic synthesis of cyanogen chloride, cyanamide and simple sugar precursors. *ChemistrySelect* **2018**, *3*, 10169−10174.
(52) Goodwin, D. G.; Moffat, H. K.; Schoegl, I.; Speth, R. L.; Weber, B. W. *Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes*, version 2.6.0; https://www.cantera.org (accessed 02/115/2023).
(53) Shen, C.; Yang, L.; Miller, S.; Oró, J. Prebiotic synthesis of histidine. *J. Mol. Evol.* **1990**, *31*, 167−174.
(54) Voet, A.; Schwartz, A. Prebiotic adenine synthesis from HCN—evidence for a newly discovered major pathway. *Bioorg. Chem.* **1983**, *12*, 8−17.
(55) Ferus, M.; Knížek, A.; Petera, L.; Pastorek, A.; Hrnčířová, J.; Jankovič, L.; Ivanek, O.; Šponer, J.; Křivková, A.; Saeidfirozeh, H.; et al. Formamide-Based Post-impact Thermal Prebiotic Synthesis in Simulated Craters: Intermediates, Products and Mechanism. *Front. Astron. Space Sci.* **2022**, *9*, 103.
(56) He, C.; Lin, G.; Upton, K. T.; Imanaka, H.; Smith, M. A. Structural investigation of HCN polymer isotopomers by solution-state multidimensional NMR. *J. Phys. Chem. A* **2012**, *116*, 4751−4759.
(57) Assary, R. S.; Redfern, P. C.; Greeley, J.; Curtiss, L. A. Mechanistic insights into the decomposition of fructose to hydroxy methyl furfural in neutral and acidic environments using high-level quantum chemical methods. *J. Phys. Chem. B* **2011**, *115*, 4341−4349.
(58) Slavova, S.; Enchev, V. Self-catalytic mechanism of prebiotic reactions: From formamide to purine bases. *Int. J. Quantum Chem.* **2020**, *120*, No. e26362.
(59) Enchev, V.; Slavova, S. Self-catalytic mechanism of prebiotic reactions: from formamide to pterins and guanine. *Phys. Chem. Chem. Phys.* **2021**, *23*, 19043−19053.

https://doi.org/10.1021/jacs.2c11857
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX