

Environmental conditions drive self-organisation of reaction pathways in a prebiotic reaction network

William Robinson

Radboud University Nijmegen

Elena Daines

Radboud University Nijmegen

Peer van Duppen

Radboud University Nijmegen

Thijs de Jong

Radboud University Nijmegen

Wilhelm Huck (✉ w.huck@science.ru.nl)

Radboud University Nijmegen <https://orcid.org/0000-0003-4222-5411>

Article

Keywords: chemical reactions, self-organisation, environmental conditions, reaction pathways, prebiotic reaction network

Posted Date: August 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-775456/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Chemistry on June 6th, 2022. See the published version at <https://doi.org/10.1038/s41557-022-00956-7>.

Abstract

The evolution of life from the prebiotic environment required a gradual process of chemical evolution towards greater molecular complexity. Elaborate prebiotically-relevant synthetic routes to the building blocks of life have been established. However, it is still unclear how functional chemical systems evolved with direction using only the interaction between inherent molecular chemical reactivity and the abiotic environment. Here, we demonstrate how complex systems of chemical reactions exhibit well-defined self-organisation in response to varying environmental conditions. This self-organisation allows the compositional complexity of the reaction products to be controlled as a function of factors such as feedstock and catalyst availability. We observe how Breslow's cycle contributes to the reaction composition by feeding C₂ building blocks into the network, alongside reaction pathways dominated by formaldehyde-driven chain growth. The emergence of organised systems of chemical reactions in response to changes in the environment offers a potential mechanism for a chemical evolution process that bridges the gap between prebiotic chemical building blocks and the origin of life.

Introduction

The origin of life from the prebiotic environment required extensive development of increasingly sophisticated chemical systems, with only environmental factors and inherent chemical activity as the driving forces.¹ Conditions on early Earth allowed for a wide range of chemical reactions which could have given rise to a diverse range of structurally complex organic molecules.²⁻⁷ Recent work has shown how these simple starting materials may produce many of the components found in extant metabolic networks, hinting at the possibility of a prebiotic origin of a core metabolism.^{4,8,9} Yet, for a functional genetic or metabolic system to emerge from a mixture of feedstock molecules, increasingly organised and interconnected reaction pathways must be forged between them.¹⁰⁻¹³ Chemical reactivity alone is not sufficient to dictate the formation of one pathway over another,¹⁴ and it is therefore difficult to conceive how on prebiotic Earth, chemical systems became organised to avoid the formation of intractable mixtures. Environmental conditions could provide a directing force for the emergence of (pre)biotic systems.^{1,15} There is a dearth of understanding and experimental studies of how reactive and environmental information translates into the self-organisation of chemical reaction networks. Therefore, creating a conceptual framework in which chemical reactivity and reaction conditions combine to organise dynamic, out-of-equilibrium reaction networks is key to elucidating how inanimate matter evolved towards life.

Here, we demonstrate how simple sets of reactions between chemical feedstocks present in a model prebiotic reaction system form well-defined compositional patterns via the interaction between environmental conditions and chemical reactivity rules. Employing the formose reaction as a model system in a series of flow reactions, we studied the responsiveness of the system to a broad range of environmental factors. By characterising the propagation of periodic input modulations through chemical reaction networks,¹⁶⁻¹⁹ we were able to infer the underlying reaction connectivity between formose

products and unravel the self-organisational response of the network to environmental conditions. Our results demonstrate how patterns of chemical reactivity and environmental conditions may give rise to organised systems of chemical reactions, offering the first glimpse of a possible mechanism for chemical evolution.

Results And Discussion

The formose reaction is a prebiotically-plausible model system of sugar-forming reactions using formaldehyde as a C₁ building block (Fig. 1a).²⁰⁻²² It broadly consists of five reactions: enolate formation/protonation, aldol addition, retro-aldol and Cannizzaro reactions (Scheme S1). Much of its core reactivity is catalysed by hydroxide and divalent metal ions such as Ca²⁺.²³ Conceptually, any given monosaccharide or enol(ate) compound in the formose reaction may be converted into another via application of the aforementioned reaction types. Therefore, a range of feedstock monosaccharides may be used to initiate the reaction.

Unconstrained, recursive application of the limited set of reaction classes operative in the formose reaction produces a so-called combinatorial explosion of compounds (Fig. 1a). A number of studies have explored means to contain the potential generation of intractable mixtures of compounds formed in the formose reaction using thermodynamic constraints.²⁴⁻²⁶ However, relatively little data has been collected for comprehensively rationalising the formose reaction under out-of-equilibrium conditions.^{22,26-30} Such conditions are a key characteristic of living systems and of high relevance on prebiotic Earth,¹² upon which the conditions were dynamic and modulated on a variety of timescales. In out-of-equilibrium chemical reactions, kinetic, rather than thermodynamic, properties govern the reaction behavior and product distribution.¹¹ Therefore, molecular reactivity is a prime controller of such systems.

Here, out-of-equilibrium conditions were induced in the formose reaction using flow conditions in a continuous stirred-tank reactor (CSTR, Fig. 1B). The compositional and reaction connectivity responses of this model system were studied in response to variations in an overarching environment. The environment, a collection of 17 input variables, included concentration variations of formaldehyde, CaCl₂ and NaOH as well as temperature and the nature of initiating sugar (glycolaldehyde, dihydroxyacetone, erythrose or ribose, Extended Data 1-4 and Supplementary Information 2).

Experiments were performed to measure steady-state equilibrium compositions of the formose reaction. In addition, investigations were also performed in which the input concentrations of initiating sugars were modulated sinusoidally. Measuring the transfer of input modulation to product compounds (Extended Data 4) provided a handle on which estimations of the underlying reaction pathways of the formose reaction could be based.¹⁶⁻¹⁸

The composition of the CSTR was continually sampled from its outflow. Following appropriate derivatisation,³¹⁻³³ samples were analysed by GC-MS and HPLC (Fig. 1c; see Materials and Methods for full details). Analysis of the chromatographic peaks and mass spectra provided a compositional pattern

for each sample (Extended Data **3**, five examples are shown in Fig. 2a), comprising of varying amounts of the 52 compounds detected within the data set.

To visualise the relationships between the average compositions and kinetic signatures generated for each condition, hierarchical clustering was performed using a correlation-based pairwise dissimilarity metric (Materials and Methods). The resulting dendrogram (depicted qualitatively in Fig. 2b, see also Figure S2) represents the relative relationships between reaction outcomes. Pie plots placed on the 'leaf' positions represent normalised average product distributions. Longer paths between leaves represent lower similarity. The inset panels indicate how the values of key environmental parameters vary across the branches of the dendrogram.

Each branch of the dendrogram arises as a result of a dominant environmental factor. Fine-tuning of compositions within branches result from the mixing of additional condition variables. For instance, branch I is characterised by relatively low concentrations of formaldehyde (**1**, ≤ 50 mM) and inputs combining glycolaldehyde (**2**) dihydroxyacetone (**3**) and erythrulose (**9**). Its constituent compositions are distinguished by relatively high amounts of C₆ compounds (green sector hues). Following branch I from its tip towards the center of the tree, more diverse sets of compounds are produced, including both branched and linear C₄ and C₅ compounds (denoted by hues of red). Interestingly, varying the concentration of **1** (with fixed inputs **3**, CaCl₂ and NaOH of 50, 15 and 30 mM, respectively) results in a series of compositional transitions (Fig. 2c), manifesting as a series of 'jumps' of varying magnitude across the dendrogram (Fig. 2d). Beginning in branch I ($[1] \leq 50$ mM), the compositions consist of mostly of an α -hydroxymethyl-aldohexose (**32**) and an α , β -(hydroxymethyl)-aldotetrose (**37**). Within the series of experiments with $[1] \leq 50$ mM, the composition remains in branch I but the molecular diversity increases as the $[1]$ increases. The contributions of **32** and **37** in the reaction mixture decrease in favor of the generation of α -hydroxymethyl-glyceraldehyde (**7**), **9** and ribulose (**20**). Increasing $[1]$ to 50 mM results in a jump towards the center of the tree, suggesting the beginning of a more significant compositional transition. Compounds **7**, **20** and an α -hydroxymethyl-aldotetrose (**14**) become prominent. Further increasing the $[1]$ to above 50 mM results in a significant compositional transition from branch I to branch VII, highlighting a transition in the molecular complexity of the system. Threose (**10**), lyxose (**18**), two 3-ketohexoses (**29**, **30**) and a new α -hydroxymethyl-aldopentose (**34**) are added to the composition. Compounds **32** and **37** are almost completely depleted. Thus, the concentration of feedstock molecule **1** controls a thresholded compositional transition whose dynamic range is in the region $[1] = 0$ -100 mM.

Ca²⁺ and hydroxide are involved in several reaction types of the formose reaction. Ca(OH)₂ has been noted to have greater activity in catalysing the formose reaction than Sr²⁺ or Ba²⁺ hydroxides.³⁴ Fe(OH)₃ has also been reported to be active in catalysing formose reactions.³⁵ Though definitive characterisation is not yet available, it is generally understood that Ca²⁺ binds divalently to enolates,^{34,36} and may participate in organising intermediates in aldol addition reactions. The principal role of hydroxide is in α -proton abstraction and enolate formation from monosaccharides, which has been noted to be a key rate-

limiting step in the reactions of this class of compounds.^{37–39} Hydroxide is also involved as a reactant in Cannizzaro reactions.

A range of $[\text{Ca}^{2+}]:[\text{NaOH}]$ input ratios (remaining below the solubility limit of $\text{Ca}(\text{OH})_2$) were explored in the data set maintaining fixed concentrations of **1** and **3**. A demonstrative subset of the data (Fig. 2e) crosses three branches of the dendrogram (II, III and VII, Fig. 2f). Beginning at low $[\text{Ca}^{2+}]:[\text{NaOH}]$ (in branch VII), compositions similar to those previously described in the high [1] regime are found. Increasing the $[\text{Ca}^{2+}]:[\text{NaOH}]$ ratio induces a jump from branch VII to II due to increases in the relative proportions of **9** and **18** in comparison to **7**. Further increasing the ratio lowers the population of **7** and **18**, eventually creating a composition dominated by **9**. Notably, compositions recorded for $[\text{NaOH}] = 2.5$ mM and $[\text{Ca}^{2+}] = 20$ to 52 mM are compositionally very similar, mainly dominated by **9**. Furthermore, at concentration of $[\text{CaCl}_2] = 52$ mM and $[\text{NaOH}] = 20$ mM, erythrose (**8**) and **20** are particularly prominent.

Other environmental conditions such as varying the initiator sugar identity lead to distinctive compositions. Branches IV, V and VI result from using **9**, ribose (**19**) and the dimer of **2**, respectively as initiators. When the temperature in the reactor from 10 to 40°C (branch VII), the reaction composition remained remarkably constant. At 10°C there is a relatively lower concentration of **10**. There is also a slight divergence of the concentrations of **18** and **20** with increasing temperature. Therefore, the influence of temperature on the steady state composition of the formose reaction is modest in the range investigated.

The observed compositional variations are a direct result of the translation of the input conditions through the underlying formose reactions. To elucidate the structure of the self-organised reaction networks, we exploited the principles of the transfer of input modulation to different compounds in the network,^{16–19} rule-based pathway reaction network generation^{40–42} and graph searching.⁴³ The framework provides a direct translation of the experimental data into a descriptive set of reactions responsible for the compositions observed (see Materials and Methods for details). Briefly, a reaction network was generated by recursively applying a set of reaction rules to an initial set of compounds (**1**, the dimer of **2** and NaOH). This network was used as a basis for a set of pathway searches. It has been shown that periodic perturbations applied to reactants in reactions systems are transferred to products and co-reactants.^{16–19} The amplitudes (here, modulations in compound concentration) measured in response to the periodic input decrease as a function of connectivity with respect to the input.

Figure 3a shows a series of representative timecourse measurements that follow compound concentration changes in response to a modulated input of **3**. The search procedure was used to estimate the self-organisational response of the formose reaction pathways in response to increasing formaldehyde concentration (initiated by **3**, Fig. 3b). When the [1] is 5 mM, the operative reaction pathways are mainly accounted for by a small set of reactions between C_3 species to form C_6 compounds (green pathways). Increasing [1] to 10 mM, triggers expansion of the repertoire of reactions. The number of possible pathways for formaldehyde addition (red) increases, with a corresponding

increase in the number of proton transfer pathways (black). However, 1-based chain growth pathways do not appear to completely account for the observed behavior. Although compound **14** (a branched aldopentose) has a lower amplitude than **9**, consistent with chain growth via formaldehyde addition, **20** (ribulose) and **18** (lyxose) have stronger couplings to the input modulation than expected (Fig. 2a). There is a very low concentration of **12** and **13** (3-ketopentoses), which would be required in the formaldehyde growth pathway towards 2-ketopentoses and aldopentoses. For these reasons, we attribute the production of **20** to the reaction between **2** and the enolate of **3** and **4** (blue pathways). Further increasing [1] (> 50 mM) builds C₁ growth pathways from the previously described set of pathways, producing **29** and **30** as a product of the formaldehyde addition to enolates derived from **18** and **20**.

The pathway searching protocol was applied to experimental data for which input modulation was applied, affording lists of reactions across the set of conditions. These reactions describe how each product distribution was created from the carbon-containing inputs. Each reaction was assigned a class. Re-casting the lists of reactions as counts of each reaction class provides a condensed view of how formose reactivity adapts to environmental conditions (Fig. 3c).

Following the branches of the dendrogram traversed in Fig. 2d (variation in [1]), reveals key reaction characteristics which govern the various reaction outcomes. A significant feature of branch I is the relatively low proportion of formaldehyde aldol addition reactions. The majority of the reactivity is accounted for by monosaccharide-enolate reactions between C₃ compounds, which are responsible for creating products **32** and **37** (Fig. 4a, branch I).^{38,44} Moving to branch VII (higher [1]), the repertoire of reactions is expanded, and aldol addition reactions involving **1** are added to the network. In particular, reactions in which the α-carbon is bound to a hydrogen or glycol group are promoted. A range of protonation/ deprotonation reactions are promoted in branch VII. Deprotonation is favored at less sterically hindered positions (where the α-carbon is bound to a hydrogen or methoxy group, e.g. follow the sequence **3**, **9**, **12**, **20**, **29** in Fig. 4a, branch VII). Protonation is favored at α-carbons bound to methoxy groups. Interestingly, the amount of monosaccharide-enolate reactions also increases, suggesting that some monosaccharide products interact with other members of the network as reactants.

The formose reaction reorganises in a different manner when the [Ca²⁺] and [NaOH] are varied. At high [Ca²⁺] (52 mM) and low [NaOH] (2.5 mM) a limited set of pathways is formed, the majority of which may be accounted for via formaldehyde addition and proton transfer reactions terminating at **14** via **9** (the pathway connecting **3**, **9** and **14** in Fig. 4b). Interestingly, the linear C₅ **12** and **13** are not formed in significant quantities, so the reaction hits the 'dead end' branched **14**. However, the system unexpectedly avoids the formation the branched C₄ compound **7** under these conditions. Decreasing the [Ca²⁺]:[NaOH] ratio (35 mM:10 mM) leads to a more significant contribution of formaldehyde-controlled pathways. The branched C₄ **7** is created, whilst the branched C₅ **14** is demoted and at the same time the population of C₅ species and instances of Cannizzaro reactions are increased. As the ratio is further decreased, the conditions and reaction pathways begin to resemble those found for the high [1] region, as described above.

In contrast to the prevailing views of the formose reaction, our data indicate that formaldehyde-based chain growth pathways do not completely account for the observed behavior. Rather, reactions between C₂ and C₃ compounds are key chain-building reactions.^{20,46,47} Surprisingly, we observe the emergence of a self-organised cyclic set of reactions that explain how the C₂ monosaccharide **2** must be created from retroaldol reactions, as described in Breslow's proposed mechanism for autocatalysis in the formose reaction (Fig. 4c).⁴⁵ Though usually seen as an autocatalytic mechanism our results show how it can contribute to formose reactivity as a generator of **2**, **3** and **4** (and their enolates) as building blocks embedded in a set of pathways in which chain growth occurs via formaldehyde addition. As such, the Breslow cycle can be envisaged as a source of new reaction pathways through which monosaccharides may be built. These reactions between the formose reaction products are an excellent example of how underlying patterns in chemical reactivity define reaction outcomes. Thus, we propose that reinforcement of molecular diversity in the formose reaction does not necessarily occur via promotion of an initiating species (glycolaldehyde). Rather, diversity may be promoted by the activation of a class of reactions in which longer carbon chains are synthesised from building blocks with units of greater than one carbon.

Conclusion

Here, we have demonstrated how an environment directs the self-organisation of reaction pathways in the model prebiotic formose reaction network to create well-defined product compositions. Our analysis is based on detailed HPLC and GC-MS characterisation of the formose reaction under out-of-equilibrium conditions, combined with graph pathway analysis informed by chemical knowledge.

In the absence of a directing force, the recursive application of the small number of chemical reactions operative in the formose reaction potentially leads to a wide range of possible reaction pathways and compositions. However, we have found that in the diverse environments studied, the formose reaction is induced into using only a subset of these pathways, depending on the local conditions. This refinement of reaction pathways results in well-defined product compositions, in contrast to the intractable mixtures expected of a combinatorially detonated reaction route. We were able to estimate the organisation of the underlying reaction pathways via analysis of the time-resolved propagation of periodically changing inputs. This analysis revealed that sets of reactions can respond collectively to environmental conditions.

The self-organisation we observe is reminiscent of theoretically predicted fine-tuning of reaction network behaviour to environments.^{48,49} Our work represents an important breakthrough in understanding how molecular systems adapt abiotically to the environment. Environmental conditions control reaction types to varying degrees in complex abiotic reaction networks, leading to well-defined reaction pathways and product compositions. Such translation of information from the environment into its embedded chemical reaction networks hints at how reaction networks of biological importance may be the result of the abiotic self-organisation of systems of reactions in the absence of specific catalysis or genetic inheritance. Applying our methodology to other prebiotically relevant reaction networks that include

cyanosulfidic⁵⁰ or α -ketoacid^{8,9} reactivity could shed new insight into an environment-driven formation of a primitive core metabolic networks furnishing the building blocks of life.

Declarations

Acknowledgements

This work was supported by funding from the Simons Collaboration on the Origins of Life (SCOL; award 477123) and the Dutch Ministry of Education, Culture and Science (Functional Molecular Systems Gravity program 024.001.035). We thank Peter van Galen, Rob de Graaf, Jesús del Pozo Mellado and Guilherme Tripodi for support with GC-MS analysis.

Author Contributions

- Data curation: WER
- Formal analysis: WER
- Funding acquisition: WTSH, WER
- Investigation: WER, ED, PD, TJ
- Methodology: WER, ED
- Project Administration: WTSH
- Resources: WTSH
- Software: WER
- Supervision: WTSH
- Visualisation: WER
- Writing – original draft: WTSH, WER
- Writing – review & editing: all co-authors

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

All data supporting the findings of this study are available within the paper and Extended Data and Supplementary Information.

Code Availability

References

1. Surman, A. J. *et al.* Environmental control programs the emergence of distinct functional ensembles from unconstrained chemical reactions. *Proc. Natl. Acad. Sci. USA* **116**, 5387–5392 (2019).
2. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
3. Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).
4. Springsteen, G., Yerabolu, J. R., Nelson, J., Rhea, C. J. & Krishnamurthy, R. Linked cycles of oxidative decarboxylation of glyoxylate as protometabolic analogs of the citric acid cycle. *Nat. Commun.* **9**, 91 (2018).
5. Becker, S. *et al.* Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
6. Ritson, D. J., Mojzsis, S. J. & Sutherland, John. D. Supply of phosphate to early Earth by photogeochemistry after meteoritic weathering. *Nat. Geosci.* **13**, 344–348 (2020).
7. Wołos, A. *et al.* Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. *Science* **369**, eaaw1955 (2020).
8. Muchowska, K. B., Varma, S. J. & Moran, J. Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **569**, 104–107 (2019).
9. Stubbs, R. T., Yadav, M., Krishnamurthy, R. & Springsteen, G. A plausible metal-free ancestral analogue of the Krebs cycle composed entirely of α -ketoacids. *Nat. Chem.* **12**, 1016–1022 (2020).
10. Wu, L.-F. & Sutherland, J. D. Provisioning the origin and early evolution of life. *Emerging Topics in Life Sciences* **3**, 459–468 (2019).
11. Pascal, R., Pross, A. & Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **3**, 130156 (2013).
12. Semenov, S. N. *et al.* Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* **537**, 656–660 (2016).
13. Semenov, S. N. *et al.* Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.* **7**, 160–165 (2015).
14. Jinich, A. *et al.* A thermodynamic atlas of carbon redox chemical space. *Proc. Natl. Acad. Sci. USA* **117**, 32910–32918 (2020).
15. Sasselov, D. D., Grotzinger, J. P. & Sutherland, J. D. The origin of life as a planetary phenomenon. *Sci. Adv.* **6**, eaax3419 (2020).
16. Samoilov, M., Arkin, A. & Ross, J. Signal Processing by Simple Chemical Systems. *J. Phys. Chem. A* **106**, 10205–10221 (2002).

17. Roszak, R., Bajczyk, M. D., Gajewska, E. P., Hołyst, R. & Grzybowski, B. A. Propagation of Oscillating Chemical Signals through Reaction Networks. *Angew. Chem. Int. Ed.* **58**, 4520–4525 (2019).
18. Urmès, C. *et al.* Periodic reactor operation for parameter estimation in catalytic heterogeneous kinetics. Case study for ethylene adsorption on Ni/Al₂O₃. *Chemical Engineering Science* **214**, 114544 (2020).
19. Mettetal, J. T., Muzzey, D., Gomez-Urbe, C. & van Oudenaarden, A. The Frequency Dependence of Osmo-Adaptation in *Saccharomyces cerevisiae*. *Science* **319**, 482–484 (2008).
20. Kim, H.-J. *et al.* Synthesis of Carbohydrates in Mineral-Guided Prebiotic Cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).
21. Delidovich, I. V., Simonov, A. N., Taran, O. P. & Parmon, V. N. Catalytic Formation of Monosaccharides: From the Formose Reaction towards Selective Synthesis. *ChemSusChem* **7**, 1833–1846 (2014).
22. Simonov, A. N. *et al.* Selective synthesis of erythrulose and 3-pentulose from formaldehyde and dihydroxyacetone catalyzed by phosphates in a neutral aqueous medium. *Kinet. Catal.* **48**, 550–555 (2007).
23. Zafar Iqbal & Senad Novalin. The Formose Reaction: A Tool to Produce Synthetic Carbohydrates Within a Regenerative Life Support System. *Curr. Org. Chem.* **16**, 769–788 (2012).
24. Lambert, J. B., Gurusamy-Thangavelu, S. A. & Ma, K. The Silicate-Mediated Formose Reaction: Bottom-Up Synthesis of Sugar Silicates. *Science* **327**, 984–986 (2010).
25. Ricardo, A. Borate Minerals Stabilize Ribose. *Science* **303**, 196–196 (2004).
26. Colón-Santos, S., Cooper, G. J. T. & Cronin, L. Taming the Combinatorial Explosion of the Formose Reaction via Recursion within Mineral Environments. *ChemSystemsChem* **1**, e1900014 (2019).
27. Huskey, W. P. & Epstein, I. R. Autocatalysis and apparent bistability in the formose reaction. *J. Am. Chem. Soc.* **111**, 3157–3163 (1989).
28. Weiss, A. Homogeneously catalyzed formaldehyde condensation to carbohydrates. *J. Catal.* **16**, 332–347 (1970).
29. Weiss, A. H., Socha, R. F., Likholobov, V. A. & Sakharov, M. M. Formose sugars from formaldehyde. *Appl. Catal.* **1**, 237–246 (1981).
30. Kopetzki, D. & Antonietti, M. Hydrothermal formose reaction. *New J. Chem.* **35**, 1787 (2011).
31. Haas, M., Lamour, S. & Trapp, O. Development of an advanced derivatization protocol for the unambiguous identification of monosaccharides in complex mixtures by gas and liquid chromatography. *J. Chromatogr. A* **1568**, 160–167 (2018).
32. Becker, M., Liebner, F., Rosenau, T. & Potthast, A. Ethoximation-silylation approach for mono- and disaccharide analysis and characterization of their identification parameters by GC/MS. *Talanta* **115**, 642–651 (2013).
33. Becker, M. *et al.* Evaluation of different derivatisation approaches for gas chromatographic–mass spectrometric analysis of carbohydrates in complex matrices of biological and synthetic origin. *J. Chromatogr. A* **1281**, 115–126 (2013).

34. Ziemecki, S. Homogeneously catalyzed condensation of formaldehyde to carbohydrates VI. Preparation and spectroscopic investigation of complexes active in formaldehyde condensation. *J. Catal.* **50**, 455–463 (1977).
35. Weber, A. L. Prebiotic sugar synthesis: Hexose and hydroxy acid synthesis from glyceraldehyde catalyzed by iron(III) hydroxide oxide. *J. Mol. Evol.* **35**, 1–6 (1992).
36. Khomenko, T. Homogeneously catalyzed formaldehyde condensation to carbohydrates IV. Alkaline earth hydroxide catalysts used with glycolaldehyde co-catalyst. *J. Catal.* **45**, 356–366 (1976).
37. Kooyman, C., Vellenga, K. & De Wilt, H. G. J. The isomerization of d-glucose into d-fructose in aqueous alkaline solutions. *Carbohydr. Res.* **54**, 33–44 (1977).
38. Gutsche, C. David. *et al.* Base-Catalyzed Triose Condensations. *J. Am. Chem. Soc.* **89**, 1235–1245 (1967).
39. Nagorski, R. W. & Richard, J. P. Mechanistic Imperatives for Aldose – Ketose Isomerization in Water: Specific, General Base- and Metal Ion-Catalyzed Isomerization of Glyceraldehyde with Proton and Hydride Transfer. *J. Am. Chem. Soc.* **123**, 794–802 (2001).
40. Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. F. Inferring chemical reaction patterns using rule composition in graph grammars. *J. Syst. Chem.* **4**, 4 (2013).
41. Andersen, J. L., Flamm, C., Merkle, D. & Stadler, P. F. Rule Composition in Graph Transformation Models of Chemical Reactions. *MATCH Commun. Math. Comput. Chem* **80**, 661–704 (2018).
42. Simm, G. N. & Reiher, M. Context-Driven Exploration of Complex Chemical Reaction Networks. *J. Chem. Theory Comput.* **13**, 6108–6119 (2017).
43. Bajczyk, M. D., Dittwald, P., Wołos, A., Szymkuć, S. & Grzybowski, B. A. Discovery and Enumeration of Organic-Chemical and Biomimetic Reaction Cycles within the Network of Chemistry. *Angew. Chem. Int. Ed.* **57**, 2367–2371 (2018).
44. Berl, W. G. & Feazel, C. E. The Kinetics of Hexose Formation from Trioses in Alkaline Solution. *J. Am. Chem. Soc.* **73**, 2054–2057 (1951).
45. Breslow, R. On the Mechanism of the Formose Reaction. *Tetrahedron Lett.* 22–26 (1959).
46. Simonov, A. N., Pestunova, O. P., Matvienko, L. G. & Parmon, V. N. The nature of autocatalysis in the Butlerov reaction. *Kinet. Catal.* **48**, 245–254 (2007).
47. Delidovich, I. V., Simonov, A. N., Pestunova, O. P. & Parmon, V. N. Catalytic condensation of glycolaldehyde and glyceraldehyde with formaldehyde in neutral and weakly alkaline aqueous media: Kinetics and mechanism. *Kinet. Catal.* **50**, 297–303 (2009).
48. Pross, A. & Pascal, R. How and why kinetics, thermodynamics, and chemistry induce the logic of biological evolution. *Beilstein J. Org. Chem.* **13**, 665–674 (2017).
49. Horowitz, J. M. & England, J. L. Spontaneous fine-tuning to environment in many-species chemical reaction networks. *Proc. Natl. Acad. Sci. USA* **114**, 7565–7570 (2017).
50. 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.* **7**, 301–307 (2015).

List Of Supplementary Materials

Materials and Methods

Details of materials, experimental and analysis methods are available in the Materials and Methods section.

Supplementary Materials

For information on the contents of the supplementary materials, please see `ExtendedDataAndSI_Guide.md`.

Schemes, Text and Figures supporting the main text and given in Supplementary Information

Extended Data

All conditions, data and reaction SMARTS templates are included in Extended Data.

Analysis Programs

All programs used to analyse and plot the data are available on GitHub (<https://github.com/huckgroup/formose-2021.git>).

Figures

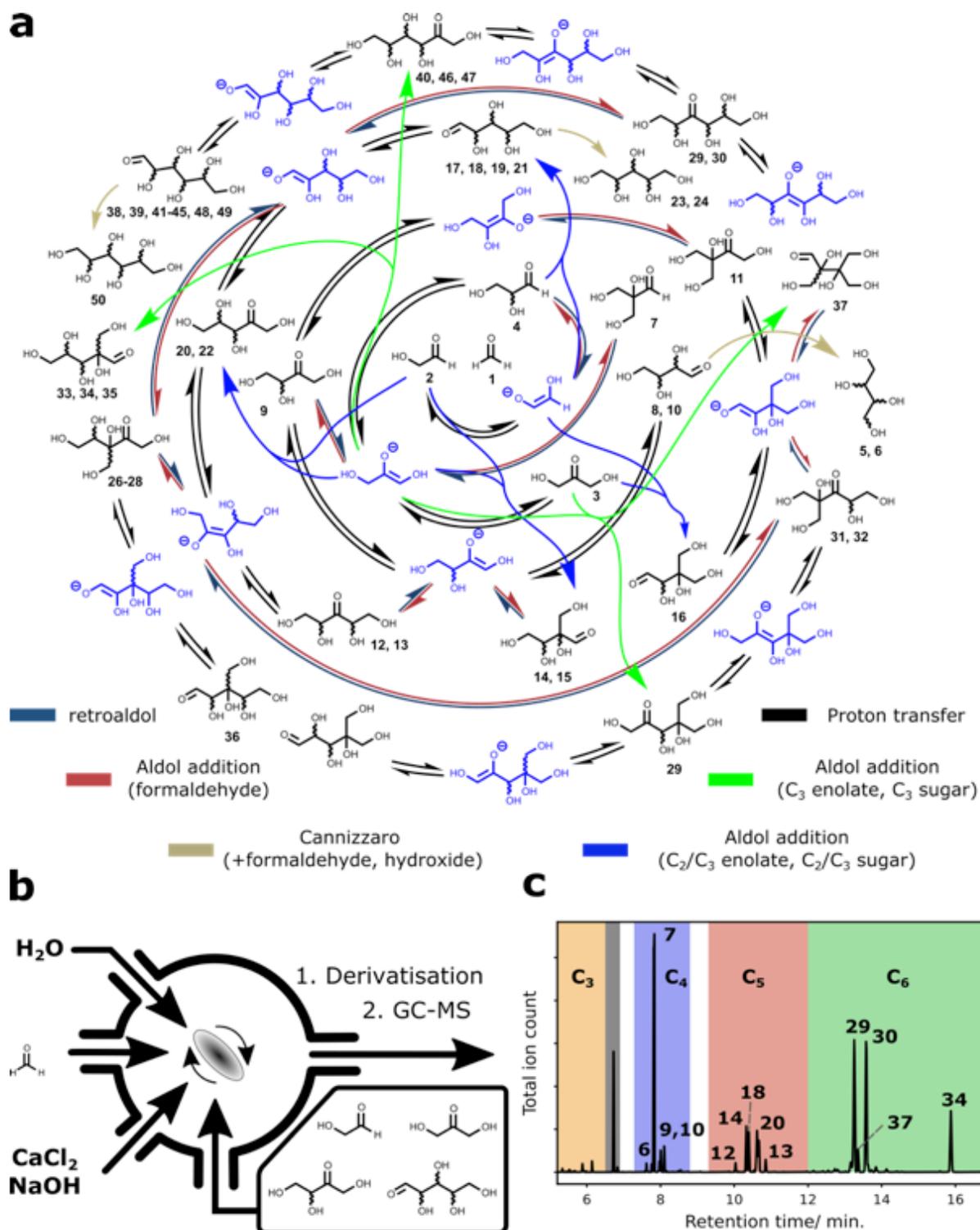


Figure 1

Background to this work. a: Schematic overview of the formose reaction. Key reaction pathways are coloured according to the reaction types (given below the scheme, wavy lines indicate both stereoisomers possible. Details of reaction types are given in Scheme S1). b: Schematic drawing of the experimental setup. Syringe pumps containing formaldehyde, an initiator sugar, CaCl₂, NaOH and water are connected to the inlets of a continuous stirred-tank reactor (CSTR). The outlet of the reactor was sampled

continually. c: An example GC-MS chromatogram indicating reactor composition. Regions of the chromatogram are coloured according to carbon chain length. Peaks are labelled according to the compound number (panel (a) Figure S1).

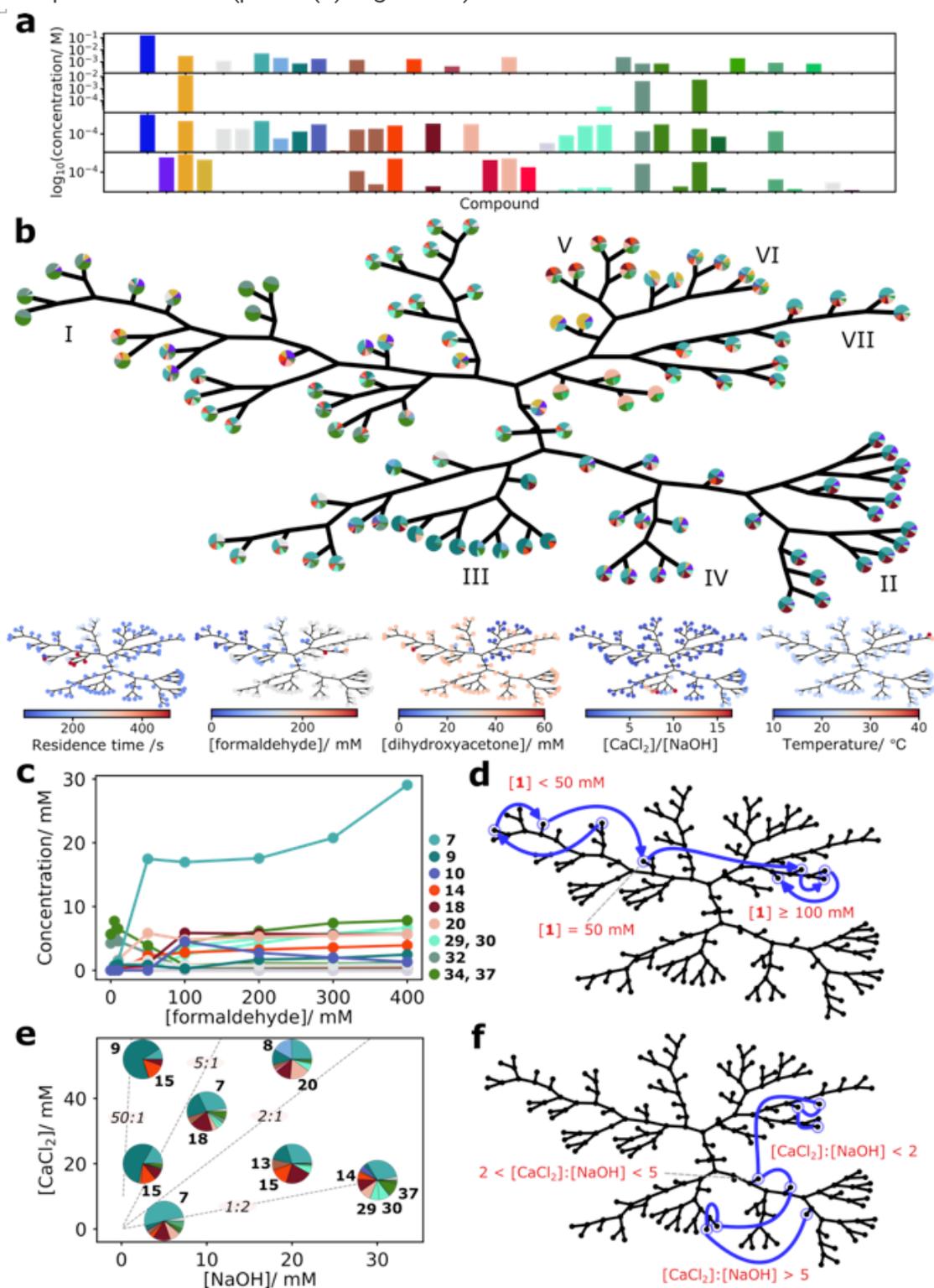


Figure 2

Description of the reaction data set collected in this work. a: Example product distributions demonstrating the compositional diversity covered by the data sets. Bars are coloured according to compound, with C4

alcohols in grey, C4 sugars in blue, C5 compounds in red and C6 compounds in green. b: The relationships between product compositions and compound response profiles in the data set when clustered using $1 - \text{Pearson Correlation Coefficient}$ as a dissimilarity measure and an average linkage algorithm (Materials and Methods). Pie plots represent experimental product composition, coloured similarly to panel a. Roman numerals are annotated to aid discussion in the main text. Underneath, key conditional features are mapped onto the dendrogram layout to illustrate how their values affect the compositional similarity between samples. c: The variation in composition of the formose reaction in response to increasing formaldehyde concentration. d: The compositional path taken through the dendrogram as the formaldehyde concentration. f: The variation in composition of the formose reaction in response to varying CaCl_2 and NaOH concentrations. Pie charts are presented similarly to those in panel b and $[\text{CaCl}_2]:[\text{NaOH}]$ ratios are indicated as annotated dashed lines. g: The compositional path taken through the dendrogram as $[\text{CaCl}_2]$ and $[\text{NaOH}]$ are varied.

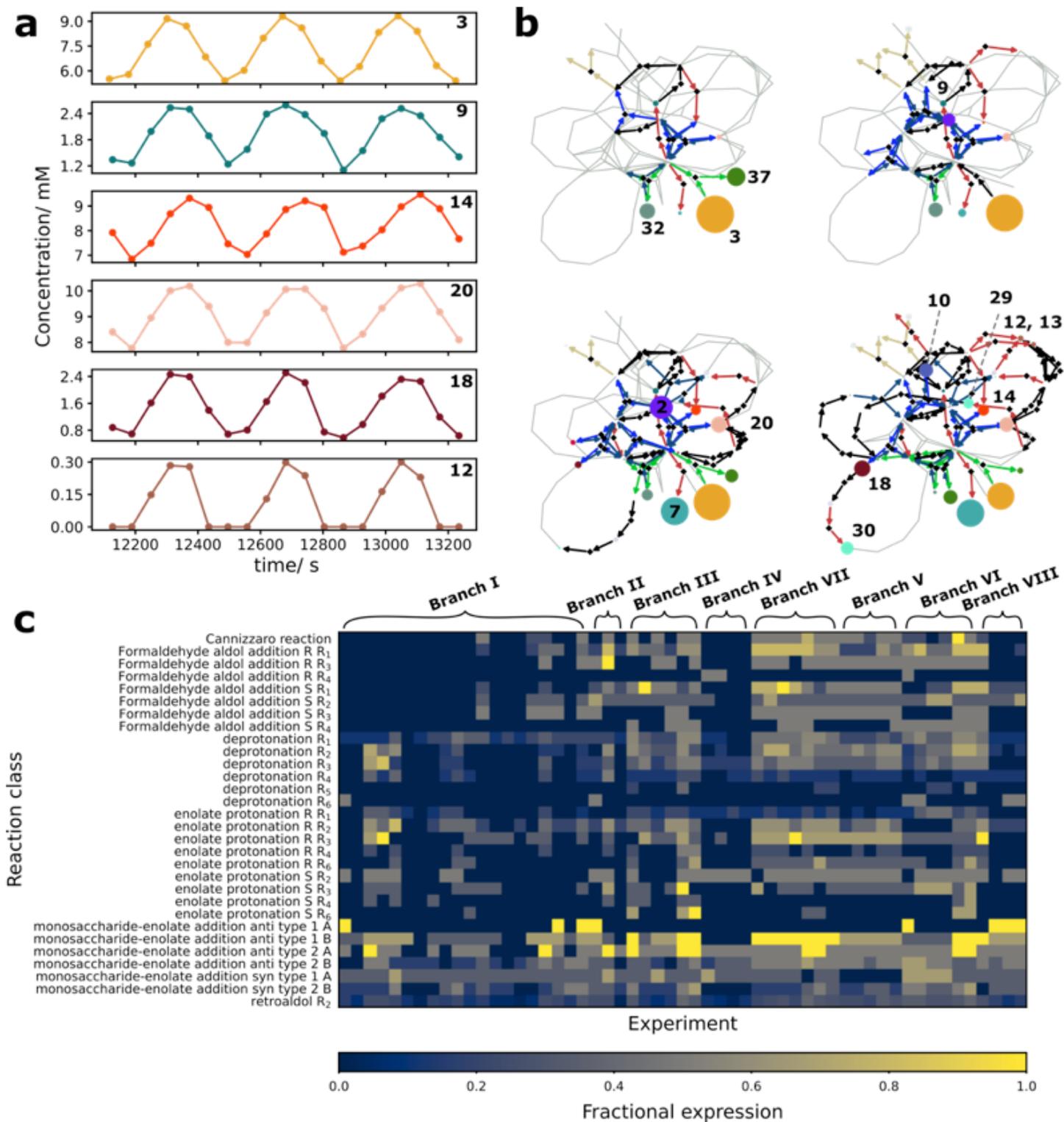


Figure 3

a: Selected time courses from a single experiment. Traces are numbered according to the structures give in Figure 1a. b: Bipartite graph representations of how the reaction pathways of the formose reaction vary with increasing formaldehyde concentration. Reactions (black diamonds) connect by directed edges to reactants and products (dots coloured according to compound), with key compound numbers annotated. The grey lines indicate the network formed from the union of all reaction pathways in this series of data.

c: A heatmap showing how the determined reaction types vary between branches of the dendrogram in Figure 2b. Fractional expression refers to the counts of the reaction class observed compared to the number of reactions in the class in a reaction network representing the union of the reactions found for all datasets.

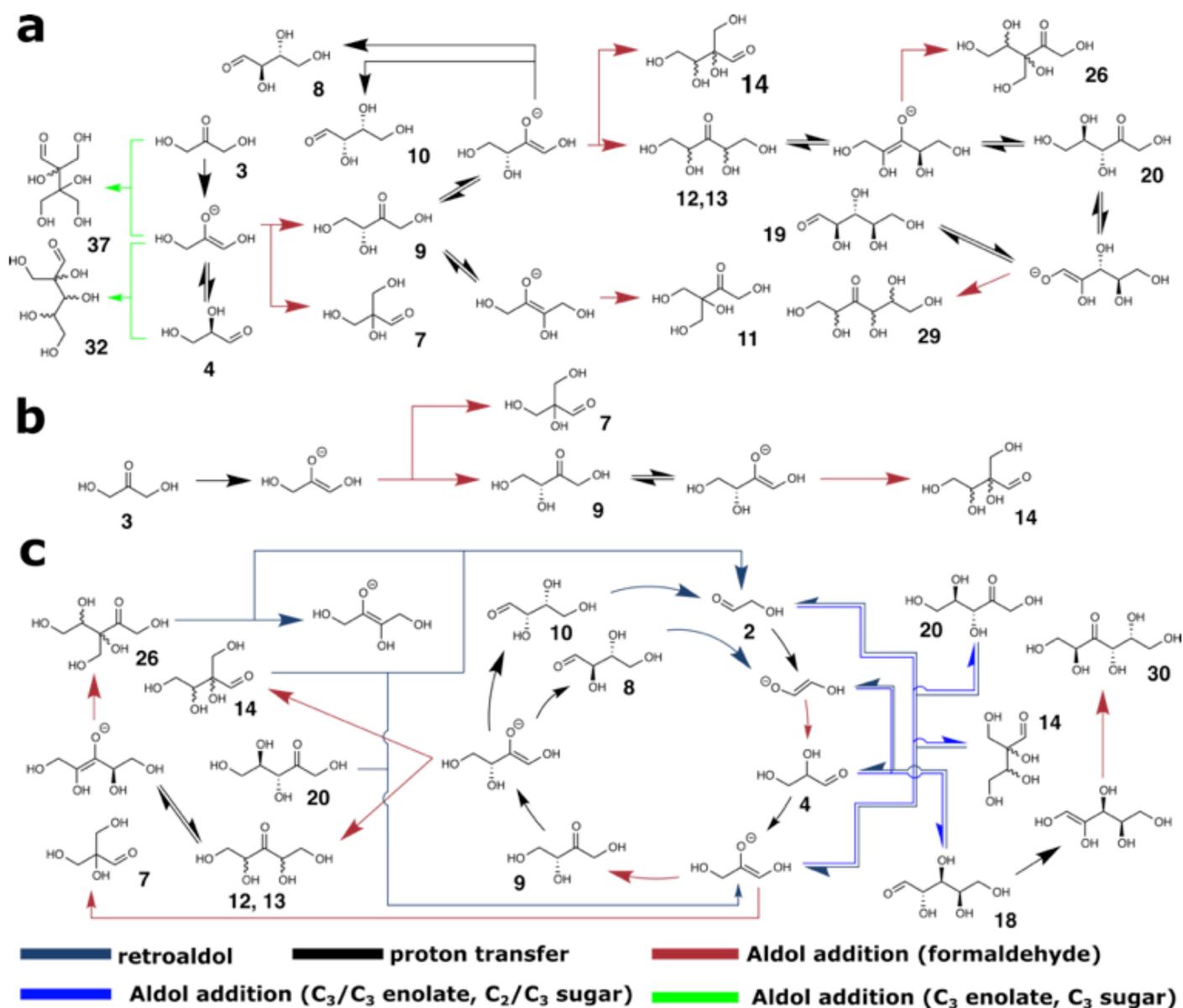


Figure 4

Key reaction pathways in the discussion a: The shift in reaction pathways from branch I to branch VII controlled by formaldehyde concentration. b: A subset of reaction pathways from panel a which account for the majority of reaction behavior in branches II and III. c: Relation of the apparent reaction pathways to the Breslow cycle,⁴⁵ indicating how the cycle acts as a source of C₂ building blocks for the formose reaction, whilst its constituents may be involved in off-cycle formaldehyde chain-growth reaction pathways. The reaction arrows are coloured according to the legend below the figure.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [RobinsonMaterialsAndMethods.pdf](#)
- [RobinsonSupplementaryInformation.pdf](#)
- [ExtendedData.zip](#)