Supplementary Information

Trajectory-dependent Compositional Outcomes in a Prebiotic Reaction Network

Peer van Duppen[1], William E. Robinson[1], Elena Daines[1], and Wilhelm T. S. Huck[1,*]

[1] Institute for Molecules and Materials, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

* Corresponding author. Email: w.huck@science.ru.nl

Table of Contents

Discussion of parameters chosen in this study

Selection of reaction conditions

Previous work[1] has shown that $[H_2C=O]_{in}$ controls a compositional transition between 0 – 100 mM (where $[DHA]_{in}$ = 50 mM, $[NaOH]_{in}$ = 30 mM, $[CaCl_2]_{in}$ = 15 mM). We thus selected average input concentrations of formaldehyde, dihydroxyacetone, NaOH and $CaCl_2$ from within this range which would provide an access to an appropriate range of reaction outcome compositions.

Dynamic inputs of $[Ca(OH)_2]$ were generated by sampling from normal distributions (μ = 15 mM, σ = 2.89 mM or 5.75 mM) generated using Numpy[2] (1.22.2.) with function numpy.random.normal().

Selection of $Ca(OH)_2$ amplitude series

For the amplitude series $[NaOH]_{in}$ and $[CaCl_2]_{in}$ were cycled around the steady state average with irregular amplitude sampled from a normal distribution (μ_{in} = 15 mM, σ_{in} = 2.89 mM or 5.75 mM) of flow rates. The flow rates where switched every 45 seconds, thus not allowing the catalyst to reach a new steady state concentration in the reactor and maintain a gradient change. We did not allow the NaOH and $CaCl_2$ syringes to reach a negative flow rate. By sampling around the steady state of 1541.25 µL h^{-1} for both syringes, we attained a maximum standard deviation of σ = 5.75 mM into the reactor. In the amplitude series the other amplitude was chosen to have half the maximum standard deviation of the inflow rate (σ = 2.89 mM).

Selection of multiple frequency input signal

We wanted to construct an input signal with different embedded frequencies to study the sensitivity of key reactions to different time scales of $Ca(OH)_2$ fluctuations. For the input

frequencies we wanted the time scales to be significantly different. However, we wanted to avoid going to steady state (rate > 120 s) or filtering out the amplitude as a consequence of the residence time in the reactor (rate → 0 s). From a theoretical sinusoidal input oscillation at a 120 second period (comparable to rate = 120 s) the reactor amplitude decays 75 % with respect to a 60 second period (comparable to rate = 30 s). Therefore, we chose not to modulate with a rate higher than 30 seconds (60 second period / 2). The 60 second rate of change was chosen since it was double the short 30 s rate of change and half the 120 s rate of change.

Supplementary figures

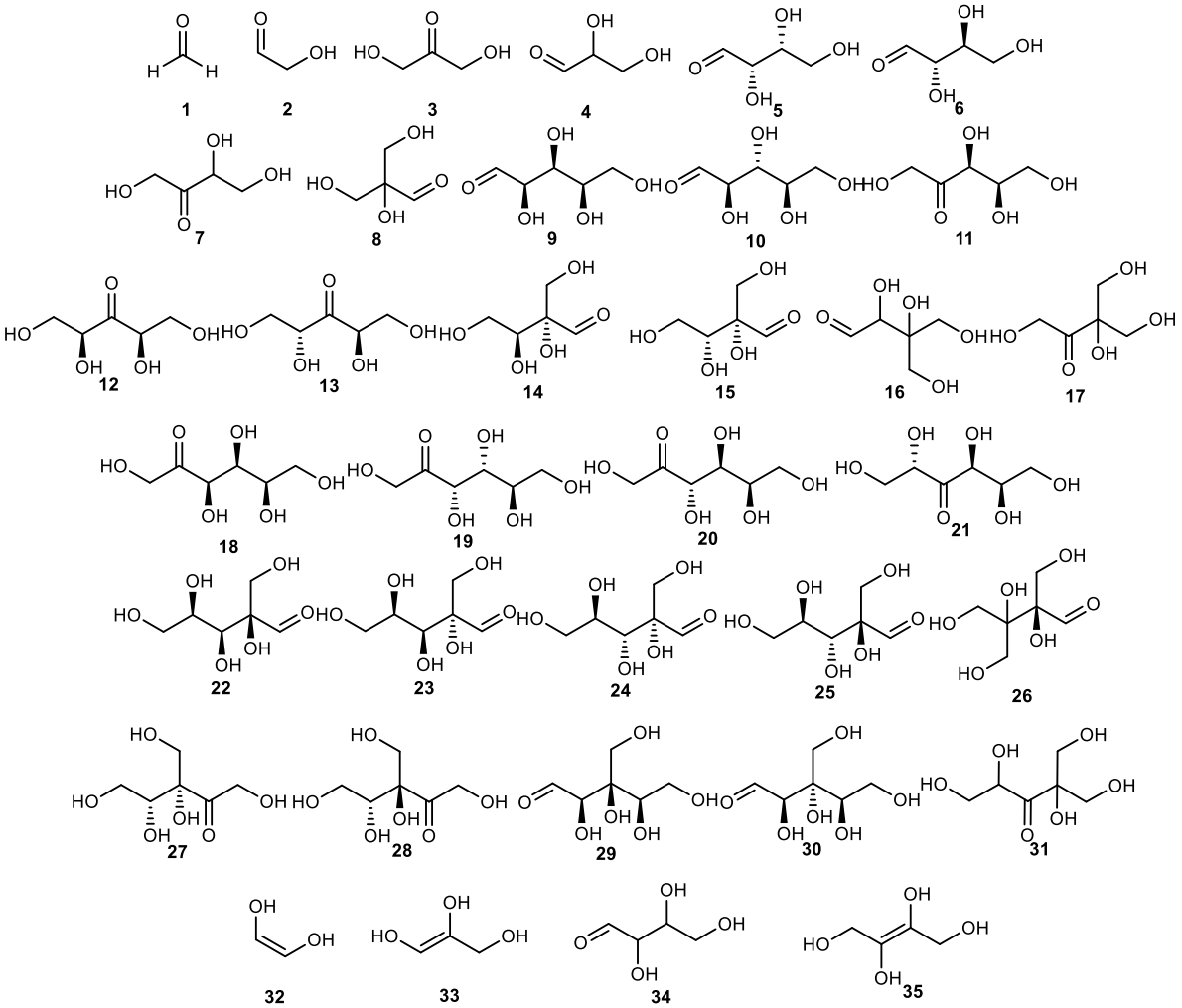

Figure S1. Structures of the compounds discussed in this work, with the corresponding numbering scheme (**1** – **35**).

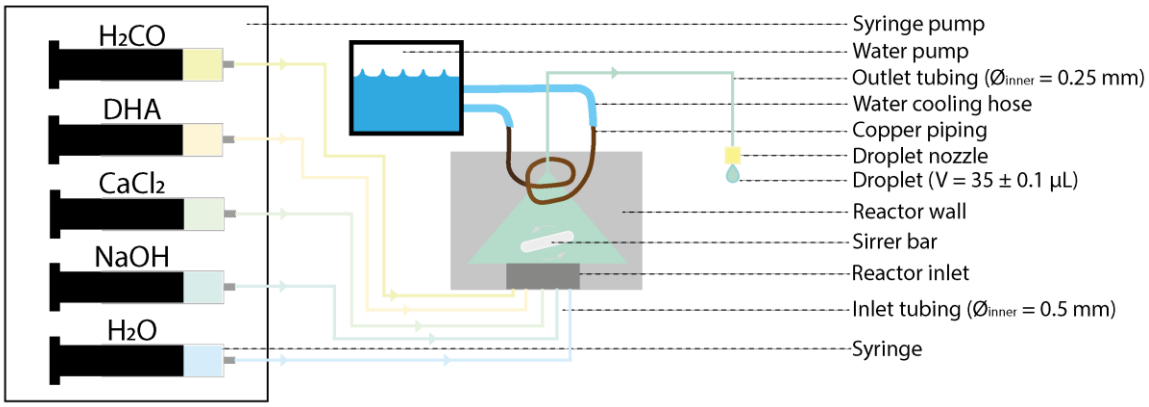

Figure S2. A detailed schematic representation of the flow reactor setup. The scheme indicates how five inlet syringes were connected to the bottom of a reactor constructured from polydimethylsiloxane (side view). The cone shaped reactor was temperature controlled and had the outlet at the top, which was connected to a droplet nozzle.

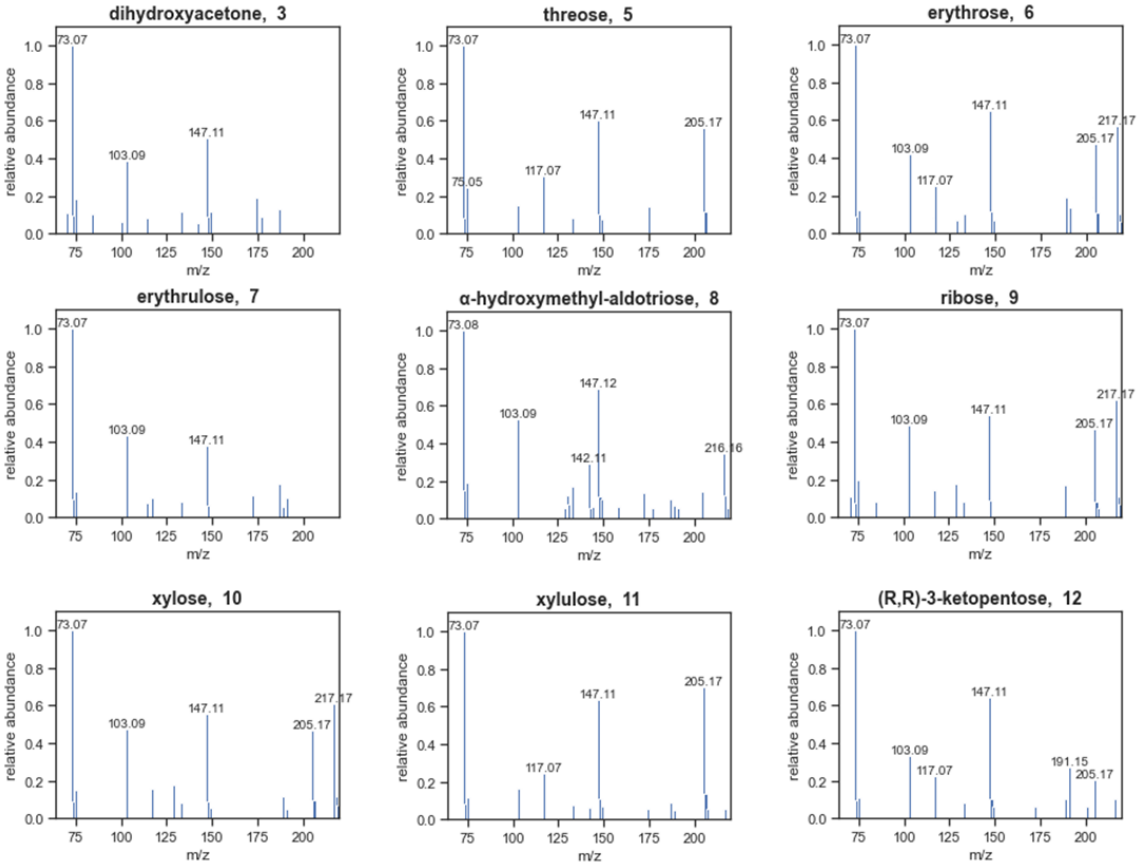

Figure S3. Mass spectra from GC-MS analysis of DHA (**3**), threose (**5**), erythrose (**6**), erythrulose, (**7**), α-hydroxymethyl-aldotriose (**8**), ribose (**9**), xylose (**10**), xylulose (**11**) and (R,R)-3-ketopentose (**12**).

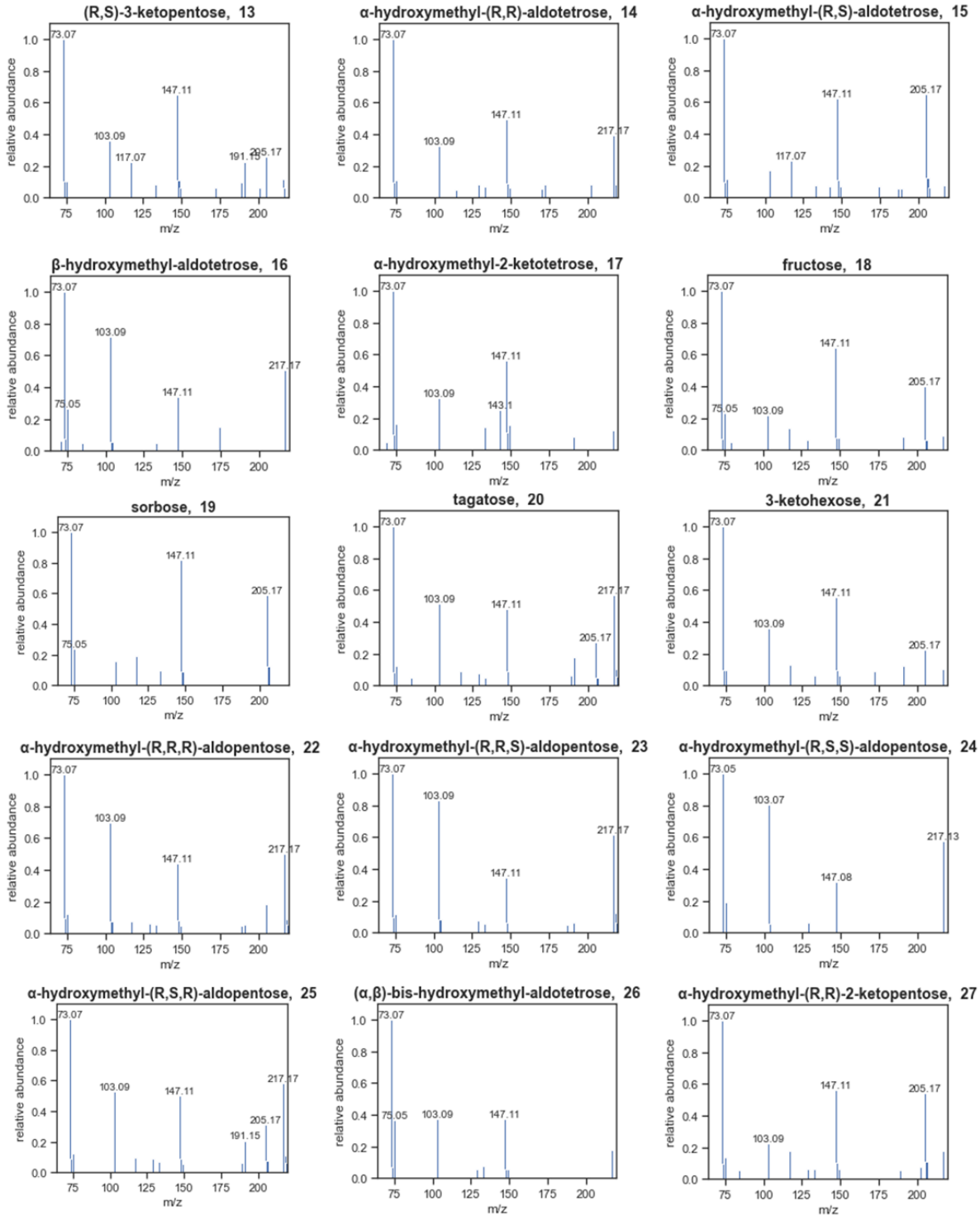

Figure S4. Mass spectra from GC-MS analysis for (R,S)-3-ketopentose (**13**), α-hydroxymethyl-(R,R)-aldotetrose (**14**), α-hydroxymethyl-(R,S)-aldotetrose (**15**), β-hydroxymethyl-aldotetrose (**16**), α-hydroxymethyl-2-ketotetrose (**17**), fructose (**18**), sorbose (**19**), tagatose (**20**), 3-ketohexose (**21**), α-hydroxymethyl-(R,R,R)-aldopentose (**22**), α-hydroxymethyl-(R,R,S)-aldopentose (**23**), α-hydroxymethyl-(R,S,S)-aldopentose (**24**), α-hydroxymethyl-(R,S,R)-aldopentose (**25**), (α,β)-bis-hydroxymethyl-aldotetrose (**26**) and α-hydroxymethyl-(R,R)-2-ketopentose (**27**).

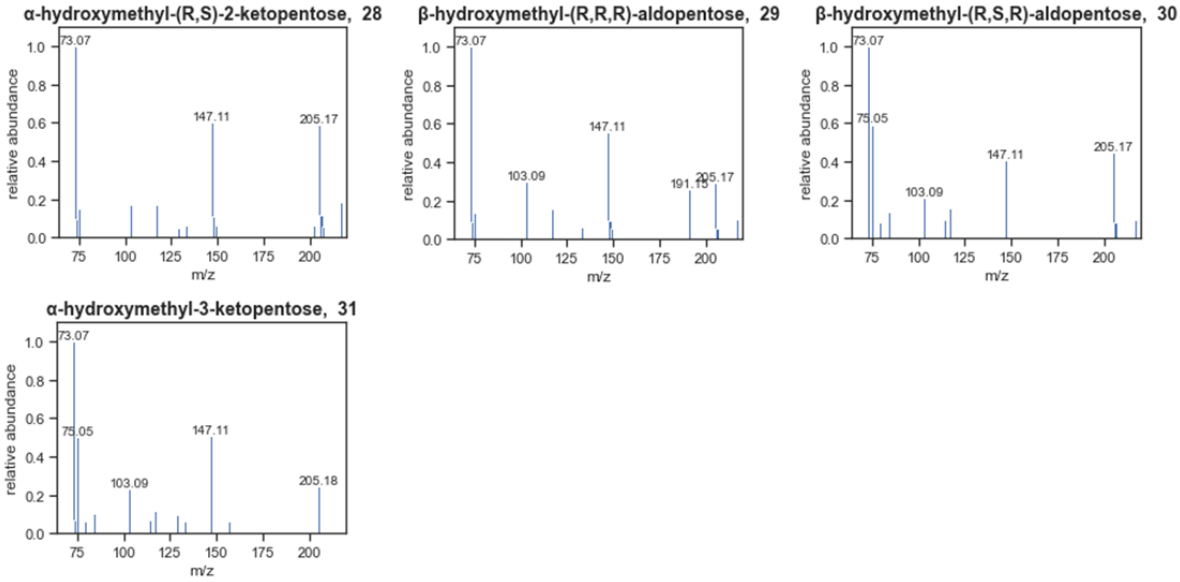

Figure S5. Mass spectra from GC-MS analysis for hydroxymethyl-(R,S)-2-ketopentose (**28**), β-hydroxymethyl-(R,R,R)-aldopentose (**29**), β-hydroxymethyl-(R,S,R)-aldopentose (**30**) and α-hydroxymethyl-3-ketopentose (**31**).

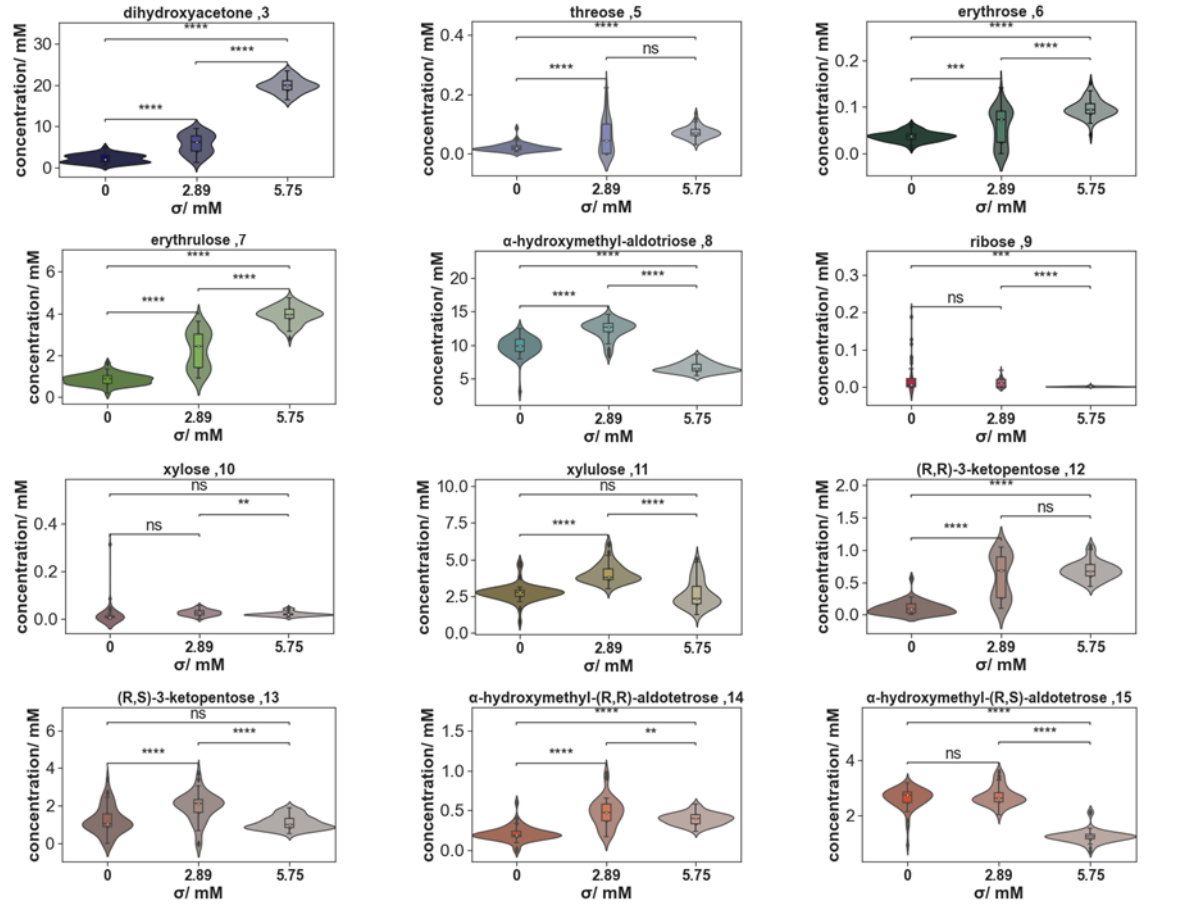

Figure S6. Violin plots of concentration distributions for identified compounds (**3**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **12**, **13**, **14**, **15**) at steady state, low amplitude and high amplitude ($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 20 mM H_2C=O, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

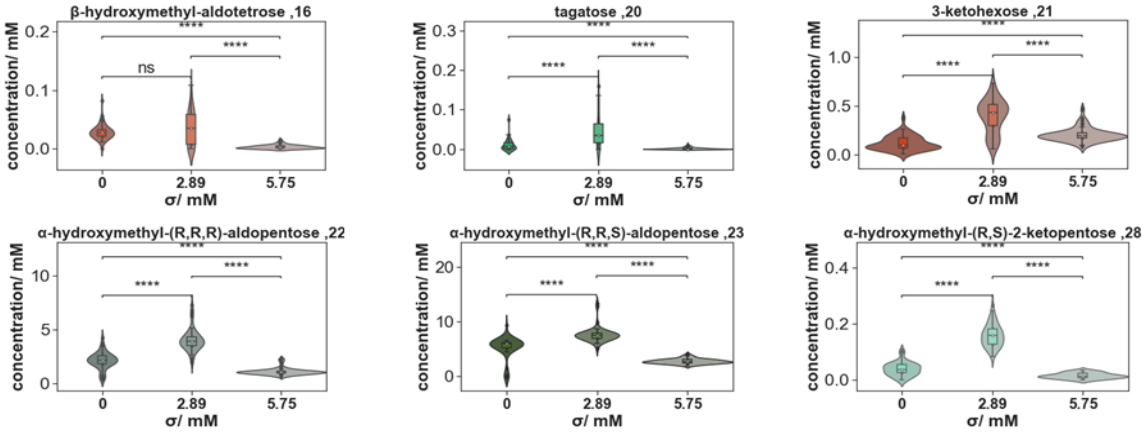

Figure S7. Violin plots for concentration profiles for the identified compounds (**16**, **20**, **21**, **22**, **23**, **28**) at steady state, low amplitude and high amplitude($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 20 mM H_2C=O, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

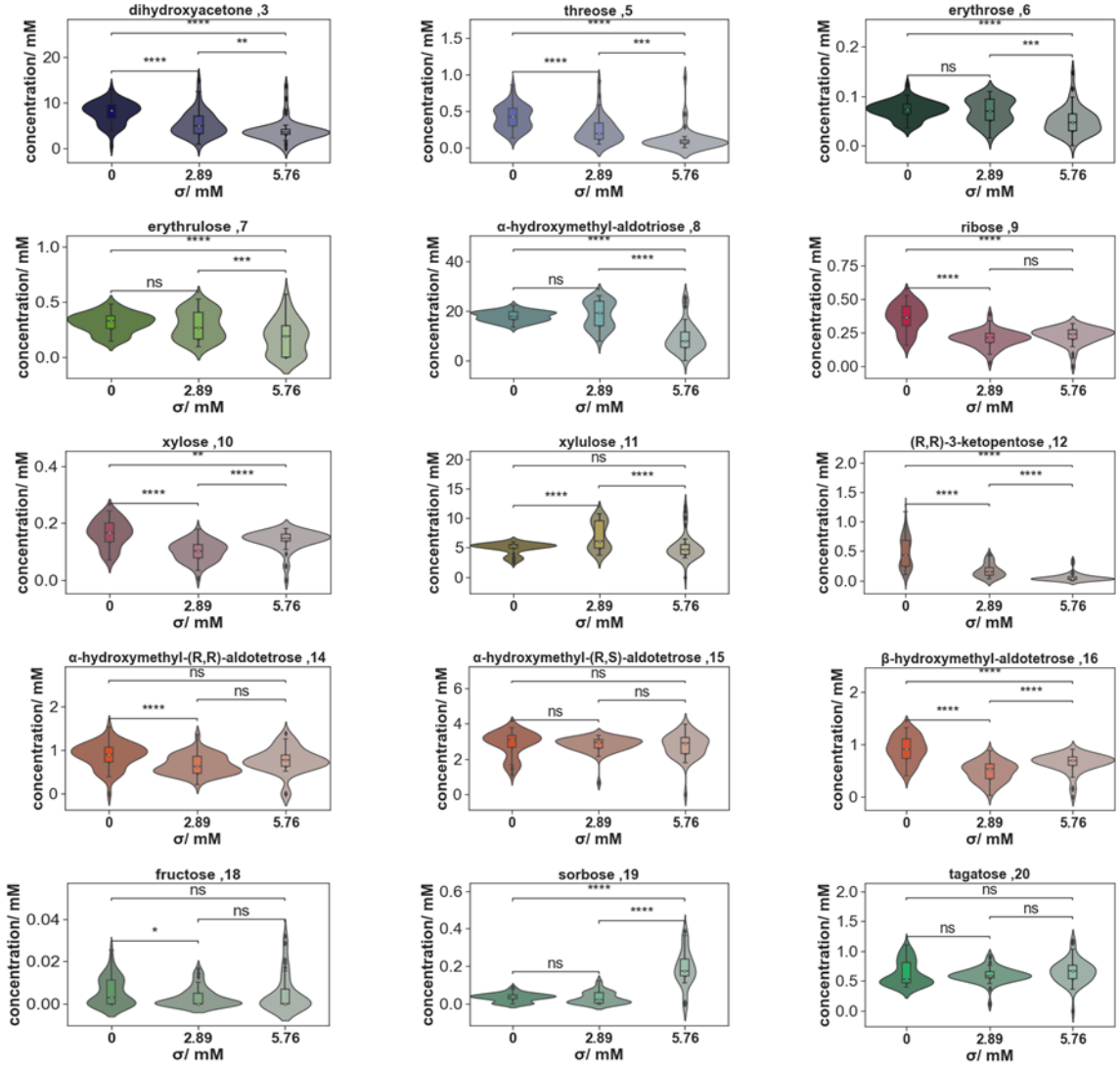

Figure S8. Violin plots of the concentration distributions for identified compounds (**3**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **12**, **14**, **15**, **16**, **18**, **19**, **20**) at steady state, low amplitude and high amplitude ($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 50 mM H_2C=O, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

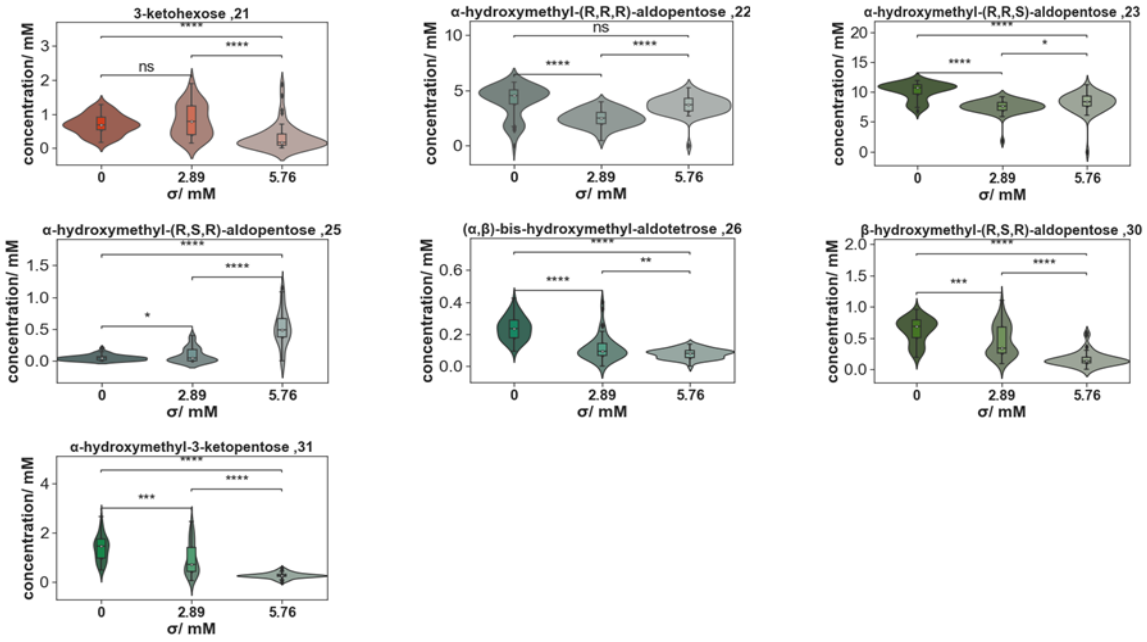

Figure S9. Violin plots for concentration distributions of the identified compounds (**21**, **22**, **23**, **25**, **26**, **30**, **31**) at steady state, low amplitude and high amplitude($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 50 mM $H_2C=O$, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

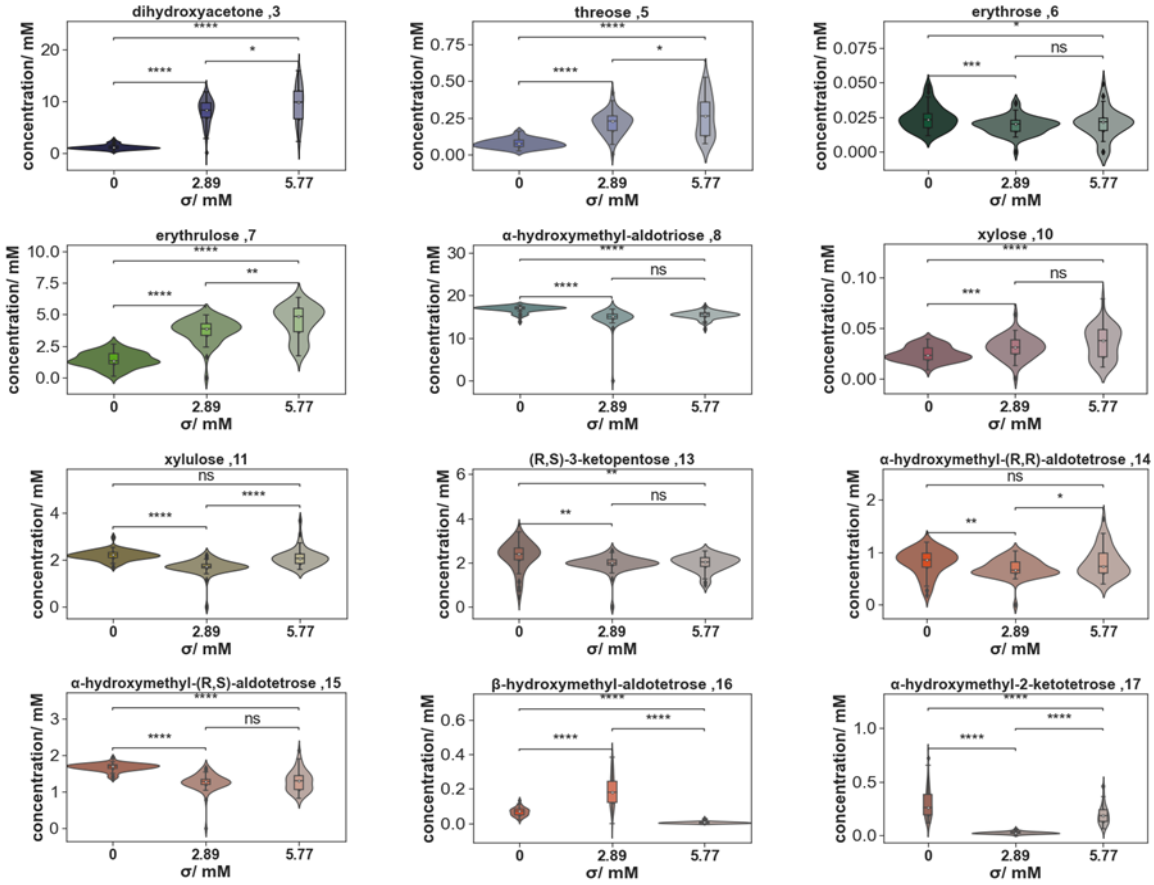

Figure S10. Violin plots for concentration distributions for the identified compounds (**3**, **5**, **6**, **7**, **8**, **10**, **11**, **13**, **14**, **15**, **16**, **17**) at steady state, low amplitude and high amplitude ($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 100 mM $H_2C=O$, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

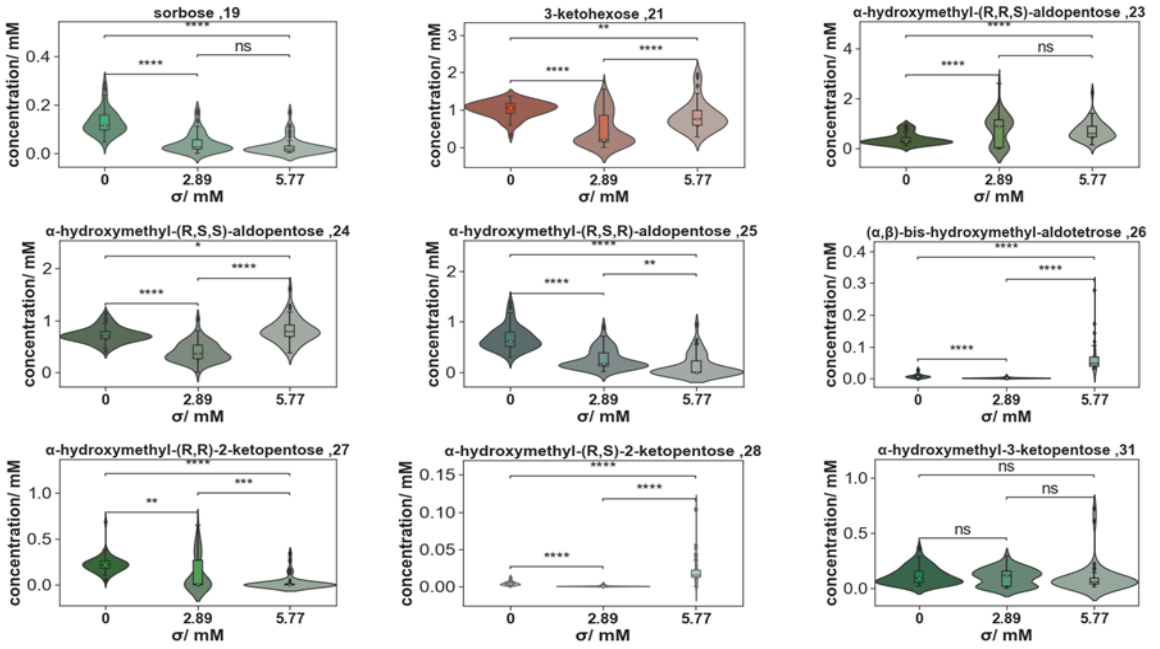

Figure S11. Violin plots for concentration distributions for the identified compounds (**19**, **21**, **23**, **24**, **25**, **26**, **27**, **28**, **31**) at steady state, low amplitude and high amplitude ($\sigma_{[Ca(OH)2]in}$ = 0, 2.89 and 5.75 mM). Conditions: 100 mM $H_2C{=}O$, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

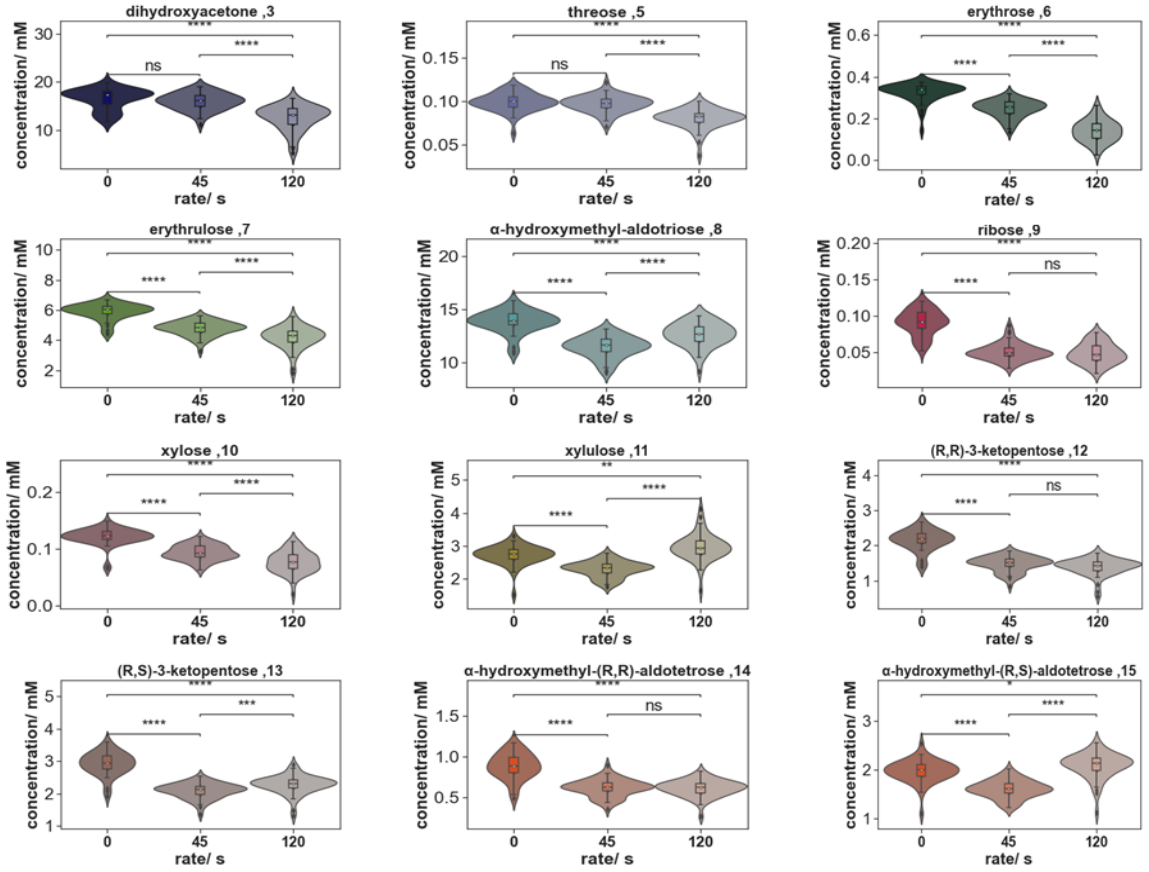

Figure S12. Violin plots for concentration distributions for the identified compounds (**3**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **12**, **13**, **14**, **15**) at steady state, 45 s and 120 s rate of change ($\sigma_{[Ca(OH)2]in}$ = 0, 5.75 and 5.75 mM). Conditions: 50 mM $H_2C{=}O$, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

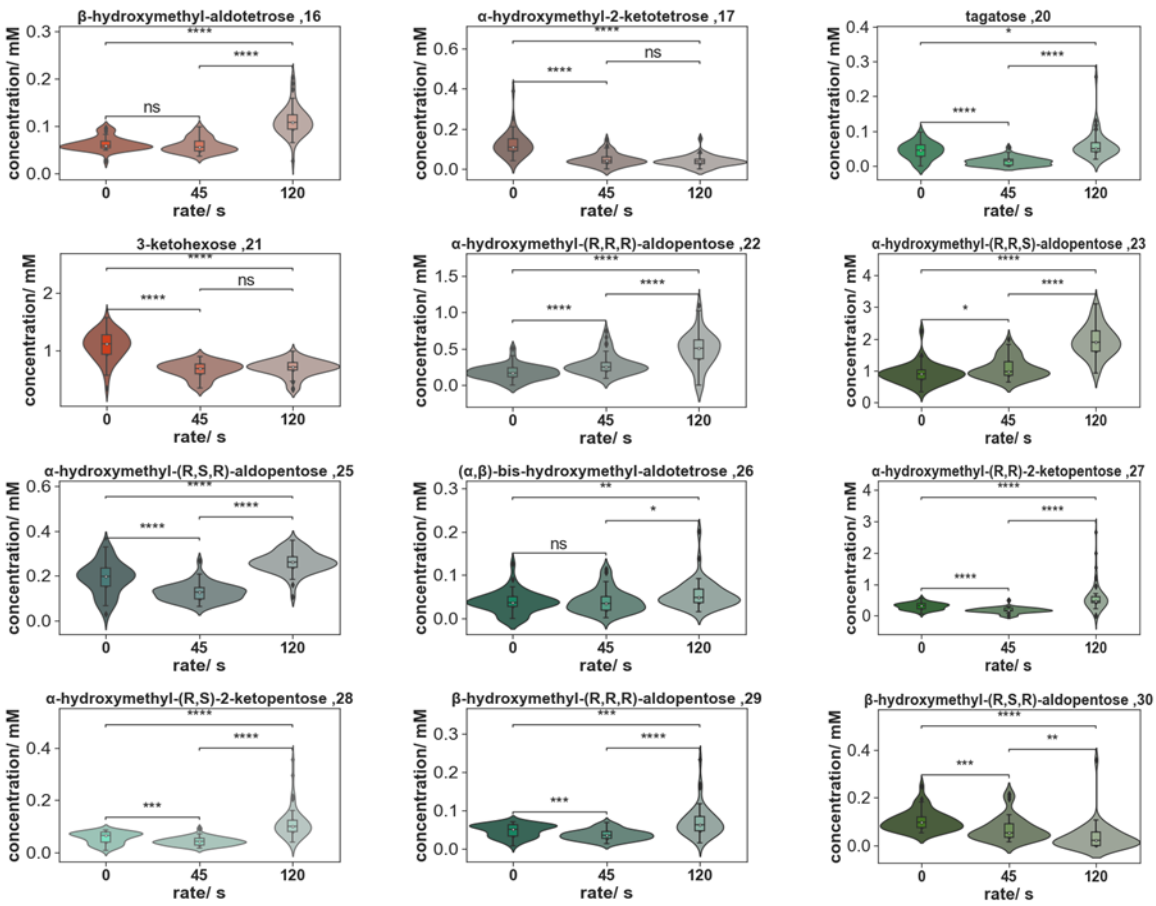

Figure S13. Violin plots for concentration distributions for the identified compounds (**16**, **17**, **20, 21**, **22**, **23**, **25**, **26**, **27**, **28**, **29**, **30**) at steady state, 45 s and 120 s rate of change ($\sigma_{[Ca(OH)2]in}$ = 0, 5.75 and 5.75 mM). Conditions: 50 mM H_2C=O, 50 mM DHA, 30 mM NaOH, 15 mM $CaCl_2$.

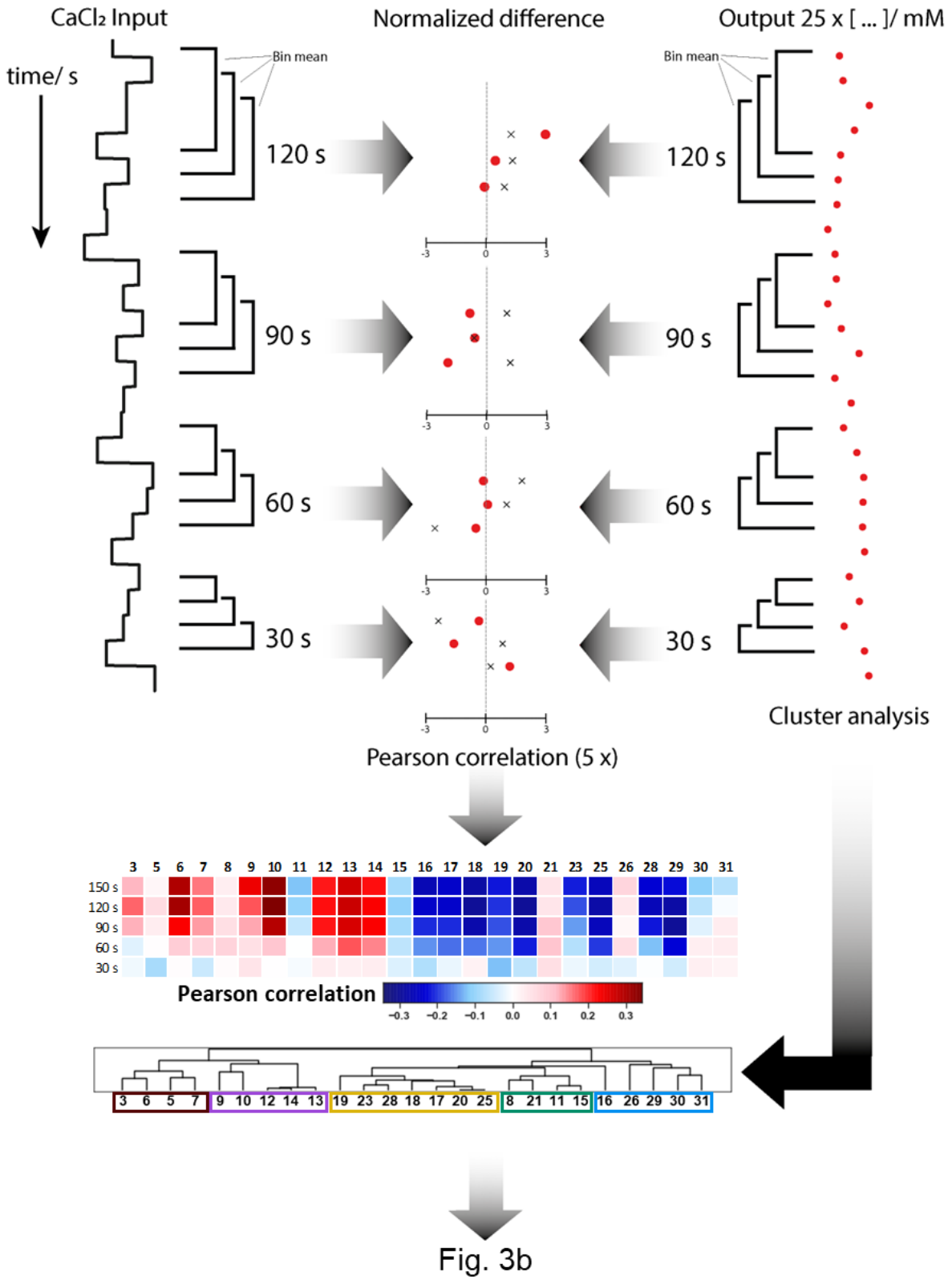

Figure S14. Scheme summarizing the correlation analysis on the 50 mM $H_2C{=}O$ network, perturbed with a complex input signal with three embedded rates of change (120 s, 60 s, 30 s). An example input concentration profile for $Ca(OH)_2$ and an output concentration profile for an example compound are shown on respectively the left and right hand side. From these, at different time-intervals – 150s, 120s, 90s, 60s and 30s – samples were binned and a mean concentration was calculated. Subsequently, for each time-interval, the concentration

difference was calculated between consecutive bin means to create a 'differential'. Both the input and output differential were normalized for each time step. A Pearson correlation was calculated between corresponding input and output differentials. The calculated correlations for each compound were superimposed on the identified cluster to obtain figure **3b**.

Supplementary tables

Table S1. Constants for fitted quadratic GC calibration curves for C_4, C_5 and C_6 compounds.

Compound	A	B
DHA, 3	0.028749	0.118714
threose, 5	0.143626	0.789532
erythrose, 6	-0.38468	2.13775
erythrulose, 7	0.143519	0.222204
ribose, 9	0.339712	1.487932
xylose, 10	0.048971	2.23855
xylulose, 11	0.695714	-0.17935
fructose, 17	0.589499	2.049516
sorbose, 18	0.08	2.354559
tagatose, 19	-0.10622	1.017795
8	0.071759	0.111102
12 to 16	0.229793	0.618952
21 to 31	0.250408	1.677321

The same calibration curves have been used as previously reported[1], based on: $[compound] = A \times (\frac{peak\ integral}{internal\ standard\ integral})^2 + B \times \frac{peak\ integral}{internal\ standard\ integral}$. We estimated a calibration curve for non-calibrated compounds (8, 12 – 16, 21 – 31) by taking the average calibration curve for the calibrated sugars of similar length.

References

1. Robinson, W. E., Daines, E., van Duppen, P., de Jong, T. & Huck, W. T. S. Environmental conditions drive self-organization of reaction pathways in a prebiotic reaction network. *Nat. Chem.* **14**, 623–631 (2022).
2. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).