Supplementary Material – Cavitation-induced synthesis of biogenic molecules on primordial Earth

Natan-Haim Kalson\textsuperscript{1,4}, David Furman\textsuperscript{2,3} and Yehuda Zeiri*\textsuperscript{1,3}

\textsuperscript{1}Biomedical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
\textsuperscript{2}Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, Hebrew University of Jerusalem, Jerusalem 91904, Israel
\textsuperscript{3}Division of Chemistry, NRCN, P.O. Box 9001, Beer-Sheva 84190, Israel
\textsuperscript{4}The Zuckerberg Institute for Water Research, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede-Boqer Campus, 8499000 Midreshet Ben-Gurion, Israel

\textsuperscript{*}Equal Contribution

Description of computational method

All simulations were performed using the ReaxFF\textsuperscript{1-3} implementation in the LAMMPS\textsuperscript{4-5} package, employing the force field given by Monti et al\textsuperscript{6}. The simulation stages are as follows: (1) energy minimization followed with (2) annealing by assigning velocities to all atoms using a Gaussian distribution for a goal temperature of 50K and lowering the temperature back to 0 Kelvin for a total of 4000 simulation steps. (3) An NVT at T=300K, using a Berendsen thermostat, with a damping coefficient of 25fs for a total of 25ps. (4) NPT stage using a damping coefficient of 25fs and 250fs for the temperature and pressure, respectively, for a total of 100ps. During this run, a restart file is saved every 10ps and used for the next stage (5) system shock stage is generated by compressing the simulation box from both sides while keeping full periodic boundary conditions. Practically, this is done by moving the simulation box faces perpendicular to the x axis towards each other. Box boundaries along the shock direction were given constant velocities of 3.5km/s and 4.5km/s. when the kinetic energy of the system reached a maximum, this signified that the shock fronts met and the system is now fully shocked. See Figure S1 below.
Figure S1. Kinetic energy of system 3, during shock stage with $V_{\text{impact}} = \frac{4.5 \text{km}}{s}$. The maximum value of the kinetic energy marks the point in which the two shock fronts meet. This configuration of the system is used as the conditions corresponding to the bubble collapse stage. At this stage, the time evolution of the system continues for 50ps in the micro-canonical ensemble. Initial box length, $dx=33.48\text{Å}$, compressed box length, $dx=17.28\text{Å}$. The maximum in kinetic energy was at $t = 180\text{fs}$.

6 From this point in time, the simulation box boundaries were fixed and dynamics were allowed to continue in the micro-canonical ensemble for 50ps. System cooling was reached by three consecutive steps: 7 box expansion back to its size with the original density (1g/cm$^3$). This process was accompanied by some thermal cooling but the system did not reach room temperature due to lack of a large enough thermal bath reservoir (i.e. cold ocean) of a thermal bath as in the sea. This stage was followed by 8 a temperature ramp at a rate of $-158.4K/ps$ (example given for system 9). Finally, 9 an equilibration stage at constant temperature ($T=300^6K$) to reach an isothermal system. This is shown schematically in Figure S2 below.
Temperature during cooling stages of system 9. First, we compare the thermal cooling during volume expansion over different time lengths from 1ps to 100ps. Then, we take the system expanded over 10ps and continue cooling it using a temperature ramp that follows the relative cooling rate from the expansion stage. The simulation is finished by 100ps of simulation at 300K.
Description of reactants used

The different systems simulated in the present study are shown in Table S1. All have an initial density of 1 g/cm$^3$ and all include 1000 H$_2$O molecules.

Table S1. Description of simulated systems: composition and size

| system | Initial volume [nm$^3$] | Carbon source | Nitrogen source | HCN |
|--------|-------------------------|---------------|-----------------|-----|
| 1      | 36.8921599              | 100 CO        | 100 N$_2$       | -   |
| 2      | 39.5490222              | 100 CO$_2$    | 100 N$_2$       | -   |
| 3      | 34.9049033              | 100 CH$_4$    | 100 N$_2$       | -   |
| 4      | 37.3943069              | 100 CO        | 100 NH$_3$      | -   |
| 5      | 40.0511692              | 100 CO$_2$    | 100 NH$_3$      | -   |
| 6      | 35.4070503              | 100 CH$_4$    | 100 NH$_3$      | -   |
| 7      | 41.3798162              | 100 CO        | 100 N$_2$       | 100 HCN |
| 8      | 44.0366784              | 100 CO$_2$    | 100 N$_2$       | 100 HCN |
| 9      | 39.3925596              | 100 CH$_4$    | 100 N$_2$       | 100 HCN |
| 10     | 41.8819631              | 100 CO        | 100 NH$_3$      | 100 HCN |
| 11     | 44.5388254              | 100 CO$_2$    | 100 NH$_3$      | 100 HCN |
| 12     | 39.8947066              | 100 CH$_4$    | 100 NH$_3$      | 100 HCN |
**Short description of analyses tools**

Results were analyzed using a in-house code in which the LAMMPS dump file and bonds table output were used as input, while the output contained information regarding the chemical identity of species in the system.

Identification of the various species is done by comparing the state of the systems between consecutive time steps and identifying reactants and products. Uniqueness of species and reaction routes are determined by comparison of chemical properties of species (constituent atoms and bonds) and reactions (reactants, products and bonding changes throughout the reaction).

Using sqlite3 databases allows to query species and/or reactions according to specific criteria. An example case is the search for species of specific structure and composition, as was done in the case of prebiotic related species.
Influence of impact velocity

The influence of the shock wave generation velocity is manifested mainly in the steady state pressure and temperature reached in the compressed phase. Typical differences between shock wave generation at $v_{\text{impact}}$ values of 3.5 and 4.5 km/sec, applied for example to system number 3 are $T = 4000K$ and $P = 33GPa$ for the low velocity and $T = 5700K$ and $P = 47GPa$ for the high velocity. These differences lead to a marked increase in the rate of reaction rates in the system. A manifestation of this behavior is presented in Figure S3 below that summarizes first order rates fitted to the reactants decomposition in systems 1, 2, 3, 6 and 9 for $v_{\text{impact}}=3.5$ km/sec and $v_{\text{impact}}=4.5$ km/sec. Inspection of these results clearly shows that the relative magnitude of the decomposition rates at the two shock generation velocities are very similar and the main difference is that the rates for $v_{\text{impact}}=4.5$ km/sec are higher.

Figure S3. Fitted first order initial reactants decomposition rates for five systems as obtained for $v_{\text{impact}}=3.5$ and $v_{\text{impact}}=4.5$ km/sec.
Intermediate reaction products during the bubble collapse phase

**Table S2.** List of the most abundant stable intermediate species formed during the bubble collapse phase. Note that a large amount of species with low formation probability as well as inorganic species are not included.

| System            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |
|-------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| **Acetylene**     | 1  | 3  | 456| 144| 14 | 1312| 29 | 51 | 439| 264| 185| 1409|
| **Cyanamide**     | 885| 606| 745| 220| 262| 3994|3651|3636|1386|1557| 80 |
| **Cyanomethanimine** | 2  | 4  | 30 | 15 | 33 | 8 | 7 |
| **Formaldehyde**  | 198| 155| 615| 415| 300| 110 | 435| 547| 572| 488| 543| 323|
| **Formaldimine**  | 60 | 69 | 691| 270| 116| 85 | 394| 372|1414| 597| 487| 825|
| **Formamadine**   | 48 | 35 | 192| 47 | 18 | 2 | 382| 252| 882|227 |204 |97 |
| **Formamide**     | 296| 261| 708| 342| 300| 19 | 1179|1000|1268| 836| 860| 226|
| **Formic acid**   | 397| 24 | 338| 387| 524| 1 | 702| 826|326 | 518 |751 |38 |
| **Hydroxylamine** | 21646| 23433| 8362| 3815| 7982| 70 | 24990|27281|13842|7589|12304|1172|
| **Isocyanic acid** | 12217| 10945| 6777| 3986| 6329| 21 | 33083|30893|13788|11828|18406|682|
| **Methane**       | 2  | 208| 26 | 4 | 2223| 4 | 6 | 171 |52 | 22 |1000|
| **Methanediol**   | 6  | 1  | 31 | 18 | 16 | 13 | 15 | 12 | 37 | 32 | 17 | 33 |
| **Methanol**      | 2  | 2  | 265| 60 | 20 | 819| 17 | 8 | 184 |89 | 51 | 585|
| **Methanolamine** | 7  | 8  | 1  | 34 | 12 | 38 | 42 | 30 | 172| 89 | 63 | 201|
| **Methylamine**   | 84 | 22 | 1  | 170| 10 | 10 | 144| 55 | 30 | 514|
| **Urea**          | 95 | 22 | 97 | 43 | 15 | 301| 206| 533| 182|139 | 30 |
**Calculation of the mass distributions**

The identity of the different species in the system are revealed once the bonding scheme among the atoms is calculated. This is done for every predetermined integration step (10fsec in the present case). The mass spectra are obtained by calculating the intensity of each mass by the average number of occurrences during 10 psec. Two typical mass spectra, obtained from calculation during 10psec starting at $t=5$ and at $t=40$psec are shown for systems 1, 3 and 9 in Figures S4, S5 and S6. It is easy to see that the mass spectra obtained for the two systems do not change much during the time evolution of the system. This suggests that the intermediate species distribution obtained after about 10psec is nearly constant.

![Mass distributions](image)

*Figure S4.* Mass distributions for system number 1 calculated from data obtained during 10psec starting at $t=5$psec (top) and at 40psec (bottom).
Figure S5. Mass distributions for system number 3 calculated from data obtained during 10psec starting at t=5psec (top) and at 40psec (bottom).
Figure S6. Mass distributions for system number 9 calculated from data obtained during 10psec starting at t=5psec (top) and at 40psec (bottom).
**Table S3.** List of the prebiotic chemistry related species that remained after bubble collapse and cooling stages. Note that X represents the identification of at least one molecule in the cold expended system. Many more species are formed however they do not play any significant biologically related role and hence were left out.

|                | System |
|----------------|--------|
|                | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |
| Cyanamide      | X  |    |    |    |    |    |    | X  |    |    |    |    |
| Isocyanic acid | X  | X  | X  |    | X  | X  | X  | X  | X  |    |    |    |
| Urea           | X  |    |    |    |    |    |    |    |    |    |    |    |
| Hydroxylamine  | X  | X  | X  |    | X  | X  | X  | X  |    |    |    |    |
| Acetylene      |    | X  |    |    |    |    |    |    |    | X  |    |    |
| Formamide      |    |    |    |    |    |    |    |    |    |    |    | X  |
| Formaldehyde   |    |    |    |    |    |    |    |    |    |    | X  |    |
System size convergence

By varying the total number of molecules, and keeping an initial density of 1 gr/cm$^3$, comparison of the effect of simulation size was examined. The 5x systems had 5 times as many molecules compared to the systems examined in this work, having 500 carbon source molecules, 500 nitrogen source molecules and 5000 water molecules.

**Figure S7.** Comparison of Temperature as function of simulation box size. (a) system 1 – CO + N$_2$ (b) system 3 – CH$_4$ + N$_2$. 
Figure S8. Comparison of pressure as function of simulation box size (a) system 1 – CO + N₂ (b) system 3 – CH₄ + N₂.

Figure S9. Comparison of relative number of H₂O molecules as function of simulation box size. (a) system 1 – CO + N₂ (b) system 3 – CH₄ + N₂.
Figure S10. Comparison of relative number of NH$_3$ molecules divided by the number of nitrogen atoms in the system as function of simulation box size. (a) system 1 – CO + N$_2$ (b) system 3 – CH$_4$ + N$_2$.

Figure S11. Comparison of relative number of CO molecules divided by the number of carbon atoms in the system as function of simulation box size. (a) system 1 – CO + N$_2$ (b) system 3 – CH$_4$ + N$_2$. 
Figure S12. Comparison of average molecules weight of molecules containing at least 1 carbon atom as function of simulation box size. (a) system 1 – CO + N₂ (b) system 3 – CH₄ + N₂.
System dilution

By varying the total number of water molecules, and keeping an initial density of $1\,gr/cm^3$, the effect of solute concentration was examined. The 10x diluted system had 10 times as many water molecules compared to the systems examined in this work, having 100 carbon source molecules, 100 nitrogen source molecules and 10000 water molecules.

Figure S13. Comparison between results obtained for two different concentrations of carbon and nitrogen sources. The "base simulation box" contains 1000 water molecules together with 100 molecules of CH$_4$ and 100 molecules of N$_2$. The "diluted" simulation contains the same amount of CH$_4$ and N$_2$ molecules but amount of water molecules was increased to 10,000. The
variation of the following quantities were monitored during the simulation: (a) temperature (b) pressure (c) normalized variation of water molecules (d) normalized formation of NH₃ molecules (e) normalized formation of CO molecules (f) average molecular weight of carbon containing species. The normalization in Figures d and e is with respect to initial number of N₂ and CH₄ respectively.

**Carbon cluster example**

For systems starting with CH₄ and NH₃ carbon clusters were formed. This is an example of a cluster formed in system 6 (no HCN as starting reactant)

![Carbon cluster example](image)

*Figure S14. An example of typical carbon cluster formed in the simulation of system 6.*

**References**

1. Bhoi, S.; Banerjee, T.; Mohanty, K., Molecular dynamic simulation of spontaneous combustion and pyrolysis of brown coal using ReaxFF. *Fuel* **2014**, *136*, 326-333.
2. Gamallo, P.; Prats, H.; Sayos, R., ReaxFF molecular dynamics simulations of CO collisions on an O-preadsorbed silica surface. *Journal of Molecular Modeling* **2014**, *20* (4), 2160.
3. Chen, B.; Diao, Z. J.; Lu, H. Y., Using the ReaxFF reactive force field for molecular dynamics simulations of the spontaneous combustion of lignite with the Hatcher lignite model. *Fuel* **2014**, *116*, 7-13.
4. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *Journal of Computational Physics* 1995, 117 (1), 1-19.

5. Aktulga, H. M.; Fogarty, J. C.; Pandit, S. A.; Grama, A. Y., Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques. *Parallel Computing* 2012, 38 (4-5), 245-259.

6. Monti, S.; Corozzi, A.; Fristrup, P.; Joshi, K. L.; Shin, Y. K.; Oelschlaeger, P.; van Duin, A. C. T.; Barone, V., Exploring the conformational and reactive dynamics of biomolecules in solution using an extended version of the glycine reactive force field. *Physical Chemistry Chemical Physics* 2013, 15 (36), 15062-15077.