Proposal funded by the VILLUM Experiment Programme Fall 2018

Searching for the origin of life using a computational search engine

Jan H. Jensen, Department of Chemistry, University of Copenhagen, Denmark
Email: jhjensen@chem.ku.dk, Twitter: @janhjensen

In a nutshell

Life is essentially an organized network of chemical reactions (metabolic pathways) that can create copies of itself given a source of energy. How was this complex reaction network formed from the simple molecules that were present on the early Earth? I will answer this question by simulating how simple reaction networks evolve starting from different combinations of building blocks and reaction conditions. Computer simulations will allow me to search many more combinations than is possible experimentally, thereby increasing the chances of finding reaction networks that resemble those found in modern cells. Finding a plausible explanation for how life originated on Earth will not only have profound implications for how we view ourselves and other species, but also give us a much better idea of how likely life is to have evolved on other planets.

Research idea and context

Starting with the famous Urey-Miller experiment numerous studies[1] have shown that carbohydrates, lipids, amino acids, and nucleotides can be made from simple molecules such as methane, carbon dioxide, hydrogen sulphide, phosphate, molecular nitrogen under plausible early Earth conditions without enzyme catalysts. However, examples of complex and sustainable reaction networks that resemble metabolism have only appeared relatively recently.[2] The main problem with such experimental studies is the very large number of combinations of building blocks and reaction conditions that must be tried. Reaction conditions such as high pressure that mimic extreme environments such as hydrothermal vents pose additional practical obstacles to experimental work. A thorough and systematic *experimental* investigation of prebiotic metabolism is thus not practical in the foreseeable future.

My idea is to replace the experiments with simulations. However, this will require a fundamental change in how such simulations are performed. The thermodynamics and kinetics parameters of a chemical reaction can now be routinely predicted using quantum chemical calculations and be used to model the behaviour of reaction networks under any conditions using kinetic models.[3] However, the setup and analysis of these calculations can be quite complicated and are usually done manually. In addition, the calculations can be quite time-consuming - even on supercomputers. These obstacles must be removed before these methods can be used for high throughput studies of prebiotic chemistry.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27233v1 | CC BY 4.0 Open Access | rec: 21 Sep 2018, publ: 21 Sep 2018

Proposed method of solution

I am developing an automated way to computationally explore reaction networks by combining recent work by Zimmerman[4] and Woo Youn Kim.[5] Furthermore, I have shown that the connectivity-based hierarchy approach developed by Segupta and Raghavachari[6] can be used to considerably improve the accuracy of very fast approximate quantum mechanical methods,[7] so that thermodynamic parameters can be predicted accurately and efficiently. A similar study for kinetic parameters in being planned. Combining these advances will allow me to reliably and efficiently predict reaction networks starting from a set of reactants and I propose to use this method to perform a systematic study of prebiotic chemistry looking for plausible metabolic pathways and replication via autocatalysis.

In concrete terms, reactant molecules (and possible catalysts such as metal ions) are given as input to a program that then generates all possible elementary reactions and predicts the thermodynamic and kinetic stability of the products. This process is then repeated for the most stable products until a biologically relevant molecule, e.g. an amino acid, emerges, thereby establishing a viable metabolic pathway from simple building blocks. The efficiency of this pathway is then tested for various start concentrations, pH, temperature, and pressure. The entire process is repeated systematically for all combinations of likely simple chemical building blocks to map out the most likely network of metabolic pathways and their intersections.

Figure 1 shows an example of the kind of metabolic cycle I would like to discover using my reaction search engine, using only pyruvate, glyoxylate, hydrogen peroxide, and ammonia (to create α-hydroxyglycine from glyoxylate) as input.

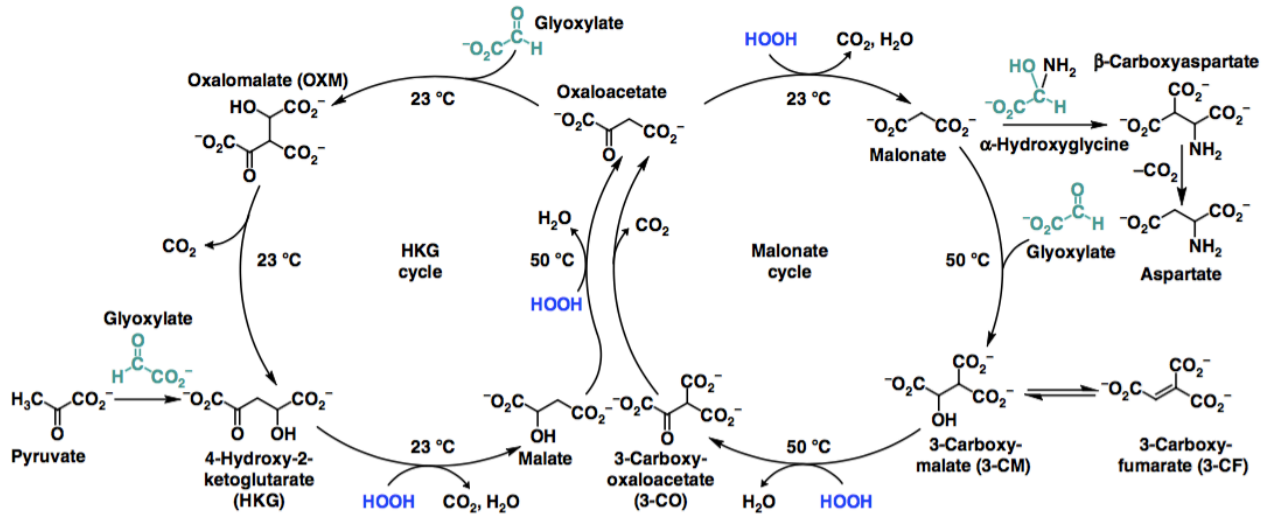

Figure 1. Protometabolic analogue of the citric acid (Krebs) cycle discovered by Springsteen et al.[2] Reproduced under the CC-BY 4.0 licence.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27233v1 | CC BY 4.0 Open Access | rec: 21 Sep 2018, publ: 21 Sep 2018

Major gains and obstacles

My goals for the first two years of this project are (1) to completely automate all aspects of the calculation, (2) to establish that the accuracy of the fast quantum methods is sufficient for the task at hand, (3) to make the method user friendly such that any chemist can use it, and (4) demonstrate proof-of-concept by finding protometabolic analogues such as the one shown in Figure 1. The systematic search through possible reaction will continue for many years after that, but will not require a lot of human intervention. The code will be freely distributed so that other scientists can join the search. I envision something similar to the current search for large prime numbers where interested parties download the search software that generates and evaluates possible candidates using spare CPU cycles. It is only fitting that the search for our origins is a global effort.

The initial obstacle will be to develop a robust and automated computational methodology that requires little human intervention. Several software packages have to be interfaced and any errors that occur at any step has to be detected and corrected automatically. The next major obstacle to overcome is the accuracy of the computed kinetic and thermodynamic parameters have to be quite high to yield even reasonable results. While it is possible to achieve such accuracy using current methods the computational cost is too high to be used for high throughput screening. Instead I have to learn when and where the fast quantum chemical methods fail and then improve their accuracy. Given the large amount of data that will be generated a part of this study, machine learning will be the best way to achieve this.

Appropriateness

Most chemists would call the idea of finding the origin of life by automated computational search for chemical reaction networks wildly optimistic at best. They would argue that the necessary tools do not exist and even if they did, the chances of an automated search finding something as complex as a prebiotic metabolic network are zero. However, I am convinced that all components that are necessary to make the search engine already exist, and we will not know what such a search will find until we perform it. While conventional funding agencies would reject this argument, it is precisely what the VILLUM Experiment Programme was designed to support.

Probable objections

Automated computational simulations of chemical reactions is in its infancy[8,9] and has never been applied to something as complex as metabolic reaction pathways. Furthermore, the combinations of building chemical blocks that must be tried in order to have some reasonable chance of finding something important is very large and such a large scale computational investigation of chemical reactivity has never been attempted. To this I would say that the tools necessary to undertake such a problem only have become available very recently[3-9] and therefore never have been combined into a single unified methodology. But once that is done it makes this ambitious undertaking possible.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27233v1 | CC BY 4.0 Open Access | rec: 21 Sep 2018, publ: 21 Sep 2018

References

[1] Ruiz-Mirazo, Kepa, Carlos Briones, and Andres de la Escosura. "Prebiotic systems chemistry: new perspectives for the origins of life." *Chem. Rev* 114 (2014): 285-366.
[2] Springsteen, Greg, et al. "Linked cycles of oxidative decarboxylation of glyoxylate as protometabolic analogs of the citric acid cycle." *Nature communications* 9.1 (2018): 91.
[3] Kalek, Marcin, and Fahmi Himo. "Mechanism and Selectivity of Cooperatively Catalyzed Meyer–Schuster Rearrangement/Tsuji–Trost Allylic Substitution. Evaluation of Synergistic Catalysis by Means of Combined DFT and Kinetics Simulations." *Journal of the American Chemical Society* 139.30 (2017): 10250-10266.
[4] Zimmerman, Paul M. "Automated discovery of chemically reasonable elementary reaction steps." *Journal of computational chemistry* 34.16 (2013): 1385-1392.
[5] Kim, Yeonjoon, et al. "Efficient prediction of reaction paths through molecular graph and reaction network analysis." *Chemical Science* (2018).
[6] Sengupta, Arkajyoti, and Krishnan Raghavachari. "Solving the Density Functional Conundrum: Elimination of Systematic Errors To Derive Accurate Reaction Enthalpies of Complex Organic Reactions." *Organic letters* 19.10 (2017): 2576-2579.
[7] Grimme, Stefan, Christoph Bannwarth, and Philip Shushkov. "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z= 1–86)." *Journal of chemical theory and computation* 13.5 (2017): 1989-2009.
[8] Rappoport, Dmitrij, et al. "Complex chemical reaction networks from heuristics-aided quantum chemistry." *Journal of chemical theory and computation* 10.3 (2014): 897-907.
[9] Simm, Gregor N., and Markus Reiher. "Context-Driven Exploration of Complex Chemical Reaction Networks." *Journal of chemical theory and computation* 13.12 (2017): 6108-6119.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27233v1 | CC BY 4.0 Open Access | rec: 21 Sep 2018, publ: 21 Sep 2018