In-silico Simulation of the Origin

N Sowmya Manojna * **Sahana Gangadharan** **

** Indian Institute of Technology, Madras,
(e-mail: sowmyamanojna@smail.iitm.ac.in)
** Indian Institute of Technology, Madras,
(e-mail: be17b038@smail.iitm.ac.in)*

Abstract: About seventy years ago, Stanley L. Miller and Harold C. Urey demonstrated the first evidence for the prebiotic synthesis of life. Their path-breaking experiment simulated conditions that are very similar to primitive Earth - atmosphere consisting of highly reducing gas mixture such as methane, ammonia and carbon dioxide, water as the solvent and high electric current discharges accounting for lightning. Upon testing the prebiotic "soup" for the presence of organic compounds, several amino acids were found. The main objective of our study is to perform in-silico simulations, similar to the Urey-Miller experiment to study the formation of amino acids.

A network biology approach, where in nodes represent elements, edges represent the covalent bonds connecting these elements and connected components represent compounds, is adopted for this study. The number of edges that each node can have, is constrained by the valency of the element. The edge weights depict the Gibbs free energy variation between the initial compound to the final compound. The simulated annealing approach would be used to scan the reaction space in our approach.

Keywords: Network biology, Amino acid synthesis, Simulated annealing, Prebiotic earth, Miller's experiment, Computational simulation, Gibbs free energy

1. INTRODUCTION

This project is an in-silico attempt to study the prebiotic synthesis of amino acids with a quest to answer several fundamental questions such as - "How did amino acids evolve from the preliminary set of compounds?", "What are the key intermediate steps involved in their formation?" and so on. More importantly, we aspire to find simple explanations for some of the most complex questions in Science.

2. WORK DONE

2.1 Thermodynamic data collection

Thermodynamic data used in this project is largely sourced from the webserver of RMG - Reaction Mechanism Generator. RMG is developed jointly by William H. Green's research group at MIT and Richard H. West's research group at Northeastern University. (Gao et al. (2016))

Thermodynamic data available in the RMG web server is segregated in the form of libraries and the data in most libraries are represented in one of the two formats - NASA format and Group additivity format.

The NASA format comprises of thermodynamic equations of heat capacity at constant pressure, standard enthalpy change and standard entropy change as a function of temperature. The Group additivity format comprises of highcharts of heat capacity at constant pressure, standard enthalpy change and standard entropy change as a function of temperature.

As no index of compounds was available in the database and parsing of highcharts using Python was tough, all thermodynamic data in the NASA format alone was parsed using Python and stored in a CSV format. The data parsed was segregated according to libraries and their corresponding labels.

Functions were written to access these coefficients (compound wise - given library, index number and across libraries) and calculate the standard enthalpy change and standard entropy change at any given temperature. Further, the standard Gibbs free energy change of the compound at a given temperature was calculated, using the above data.

2.2 Network Construction

Initial Network: Graph theory based approach has been used for this study, where the nodes represent elements and edges represent covalent bonds that connect nodes. Hence, a connected component of the graph depicts a compound in our model.

The initial nodes of our network will be carbon, hydrogen, nitrogen and oxygen as recognised as key elements that constitute the chemical composition of prebiotic earth, namely methane, water, ammonia and hydrogen gas.

Based on the concentration calculation from Miller's original paper (Miller (1953)) and other in-silico models (Saitta and Saija (2014)), we began the simulation with H:C:O:N ratio of 4:1:1:1, with a scaling factor of 8. Nodes indexed by numbers are introduced in the network. The first 32 nodes represent 'H', the next 8 represent 'C' and so on.

Based on the possible reaction mechanisms proposed by Miller and J. Bada (Miller (1953), Bada (2013)) for amino acid synthesis, initial edges resulting in the formation of 2 methane and 2 ammonia are considered and their edge weights correspond to the differences in Gibbs free energy.

Once the initial nodes and edges are defined, valencies and formal charges are assigned to respective nodes. For each of the connected components in the graph, free energy of formation, degree of unsaturation and the formal charge is calculated.

Identification of potential edges: A list of all independent nodes is maintained. Similarly, a list of all connected components is maintained. From the available information and specified attributes, a list of all valid edges is drawn by:

Table 1. Potential edge identification

Looping through	Comparing with
list of nodes	list of nodes
list of nodes	list of connected components
list of connected components	list of connected components

The validity of edges between nodes is checked based on valency, while the validity of edges between connected components is checked based on the degree of unsaturation. The edge weights depict the differences in standard Gibbs free energy between the final connected component (formed after the bond formation) and the initial connected compounds or nodes.

3. PROJECT WORK PLAN

3.1 Simulated Annealing

From the mechanisms proposed by Miller and other in-silico papers (Saitta and Saija (2014), Miller and Van Trump (1981)), it has been known that the energy of the intermediate compounds is higher than the reactant set and the product set. By calculating the Gibbs free energy of the reactants and products, we realised that the products are at a relatively lower energy state than the reactants. In order to scan our reaction space and to access the products which are present at a relatively lower energy state, we decided to use a Simulated Annealing approach. This would enable us to initially accept moves that lead to increase in the Gibbs free energy and later accept moves that reduce the Gibbs free energy.

3.2 Simulations with the addition of H_2S

So far we have only considered methane, water, ammonia and hydrogen as our initial compound set. In order to understand the synthesis of sulphur containing amino acids (Parker et al. (2011), Van Trump and Miller (1972)), we would be introducing nodes that represent sulphur and the reactions that follow would be studied.

4. PROBLEMS FACED

4.1 Data Collection

Databases with Gibbs free energy of several prebiotic organic compounds at very high temperatures were scarce.

Prior to accessing the RMG thermodynamic database, we had parsed thermodynamic data available in NIST-JANAF Thermochemical Tables webserver. Additionally, data for some of the key intermediates such as formaldehyde and formamide were obtained from vast literary surveys.

Quantum mechanical approaches were also studied and we had begun using Quantum Espresso (QE), a software that provided pseudo-potential energy information of different compounds as an alternative to corresponding Gibbs free energies of the compounds. QE enabled us to simulate events such as electro jumps and proton movement, which could occur at quantum mechanical timescale only. However, we changed our approach back to using Gibbs free energy once we got the required data from the RMG database.

4.2 Network Construction

As `networkx` doesn't support multiple nodes to have the same name, we weren't able to name the nodes based on elemental symbols (as generally required in molecular representation). In order to work around this problem, we named the nodes numerically and then used dictionaries to map them to their respective valencies and formal charges. Representation of multiple bonds between two nodes wasn't supported in `networkx` Graphs and hence, we used dictionaries as edge attributes to keep track of the number of covalent bonds between two nodes.

5. FUTURE CHALLENGES

Determination of key annealing parameters and hyper parameters. Other libraries that make visualisation of molecules more explicit would be explored.

REFERENCES

Bada, J.L. (2013). New insights into prebiotic chemistry from stanley miller's spark discharge experiments. *Chem. Soc. Rev.*, 42, 2186–2196. doi:10.1039/C3CS35433D.

Gao, C.W., Allen, J.W., Green, W.H., and West, R.H. (2016). Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications*, 203, 212 – 225. doi:https://doi.org/10.1016/j.cpc.2016.02.013.

Miller, S.L. (1953). A production of amino acids under possible primitive earth conditions. *Science*, 117(3046), 528–529. doi:10.1126/science.117.3046.528.

Miller, S.L. and Van Trump, J.E. (1981). The strecker synthesis in the primitive ocean. In Y. Wolman (ed.), *Origin of Life*, 135–141. Springer Netherlands, Dordrecht.

Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A., and Bada, J.L. (2011). Primordial synthesis of amines and amino acids in a 1958 miller H2S-rich spark discharge experiment. *Proceedings of the National Academy of Sciences*, 108(14), 5526–5531. doi:10.1073/pnas.1019191108.

Saitta, A.M. and Saija, F. (2014). Miller experiments in atomistic computer simulations. *Proceedings of the National Academy of Sciences*, 111(38), 13768–13773. doi:10.1073/pnas.1402894111.

Van Trump, J.E. and Miller, S.L. (1972). Prebiotic synthesis of methionine. *Science*, 178(4063), 859–860. doi:10.1126/science.178.4063.859.