Mapping the Chemical Universe of Biomolecules for Astrobiology

Markus Meringer (DLR), Melissa Ilardo (UCPH), Jay T. Goodwin (Emory University), H. James Cleaves (ELSI)

Understanding the origins of life is a central question of Astrobiology. Computer methods offer unique means to approach this challenge. There is strong evidence that biomolecules underwent a process of chemical evolution before life originated, and then a protracted period of biological evolution formed the universal genetic code shared by all extant terrestrial biology. The universal genetic code serves as an interface between informational biopolymers, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), and functional biopolymers, the proteins. The monomeric building blocks of proteins are amino acids; those of RNA are nucleotides, or deoxynucleotides in the case of DNA. In order to obtain a better understanding of the selection rules which guided chemical and early biochemical evolution, our approach is to computationally generate exhaustive sets of biomolecule analogues, to calculate their physico-chemical properties, and to simulate adaptive processes that might have led to the biochemical foundations of life as we know it.

Comprehensive libraries of amino acids had already been generated previously as part of a cooperative project with the University of Hawaii's NASA Astrobiology Institute (UHNAI) on "Creating a Reference Set of Amino Acids Structures for Use in Multiple Astrobiology Investigations". We continued these research activities during four weeks' on-site collaboration at Tokyo Tech's Earth Life Science Institute (ELSI). We focused on two main directions:
1. Extending the studies on the amino acid alphabet by analyzing the amino acid libraries prepared in [1].
2. Enumerating alternative nucleotide structures by extending the methods used in [1] and analyses of the obtained libraries.

It is believed that size, hydrophobicity and charge are the three physico-chemical properties of amino acids most likely responsible for their selection during the evolution of the genetic code. An analysis of the first two of these properties had already been conducted as part of a master's thesis [2] at UHNAI. During our collaboration at ELSI we acquired suitable software to calculate amino acid charge in terms of acid dissociation constants (pK_a values) of its functional groups. In this way we mapped a 1913-membered amino acid library into 3-dimensional space (Fig. 1). Subsequent statistical analysis confirmed that the set of the 20 genetically encoded amino acids covers this chemical space almost optimally in terms of range and evenness of these three physico-chemical properties. However, in contrast to previous results using much smaller comparison sets, now for the first time 20-membered sets being better than the encoded amino acids were found. This raises questions as to whether potential alien biochemistries using such alternative amino acid alphabets could outperform terrestrial life, *e.g.* in terms of protein folding capabilities. The results of these studies are published in *Nature Scientific Reports* [3].

Another focus of our collaboration was the definition and generation of nucleotide analogue libraries. RNA plays a central role as the intermediary carrier of genetic information in transcription and translation. Given fundamental structural constraints, such as the ability to form complementary base-pairs and to be linked into a linear covalent polymer, a variety of structural isomers of RNA could potentially function as genetic platforms. Using structure

generation software [4, 5], all of the potential structural isomers of the natural ribosides ($C_5H_9O_4B$, where B is a nucleobase) that can potentially serve as building blocks of nucleic acids were enumerated. Database queries showed that only very few of these computed isomers had been described previously. Together with extensive studies about which physico-chemical properties might have led to the selection of β-ribofuranosylnucleosides over a multitude of alternative structures during evolution, these results have been accepted by *Astrobiology* [6]. The structural space of nucleotides beyond $C_5H_9O_4B$ is the object of ongoing studies (Fig. 2).
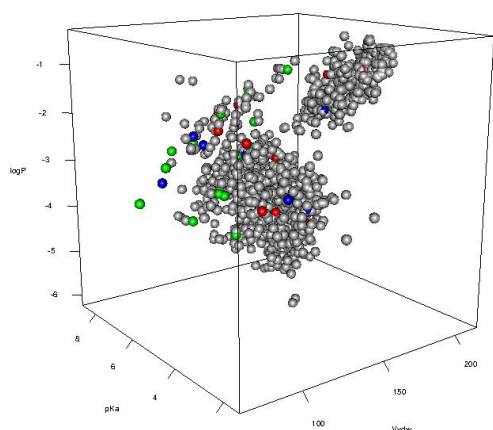
Figure 1. Chemical space of amino acids, represented by size, hydrophobicity and charge in terms of Van der Waals volume V_{vdw}, partition coefficient logP and logarithmic dissociation constant pK_a. Green spheres represent the 20 coded amino acids, blue and red spheres show two of the rare better sets. The top-right cluster is formed by somewhat larger-sized aromatic compounds.
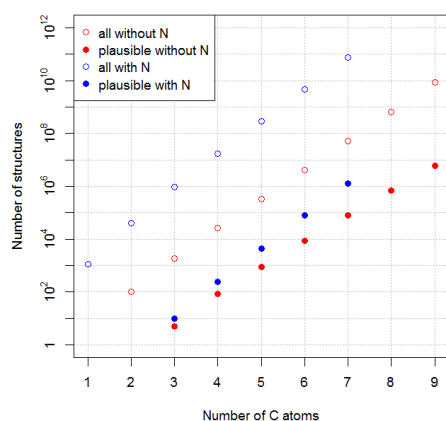
Figure 2. Numbers of nucleotide analogue structures as a function of the number of carbon atoms in the molecule. Enumeration is based on the formula spaces $C_nH_{5-(2n+1)}O_{2-4}B$ in absence of N and $C_nH_{5-(2n+3)}N_{1-2}O_{0-4}B$ with N. As was observed in the amino acid libraries, there is an exponential growth in the of number of structures with increasing number of C atoms, and a considerable discrepancy (by a factor of 10^3-10^5) in the corresponding numbers of mathematically possible and chemically plausible structures.

References:

[1] M. Meringer, H. J. Cleaves and S. J. Freeland. Beyond Terrestrial Biology: Charting the Chemical Universe of α-Amino Acid Structures. J. Chem. Inf. Model. 53(11): 2851-2862, 2013.
[2] M. Ilardo: Natural Selection and the Amino Acid Alphabet. Master's Thesis, University of Hawaii, 2013.
[3] M. Ilardo, M. Meringer, S. J. Freeland, B. Rasulev, H. J. Cleaves: Extraordinarily Adaptive Properties of the Genetically Encoded Amino Acids. Sci Rep. 5: 9414, 2015.
[4] M. Meringer: Structure enumeration and sampling. Handbook of Chemoinformatics Algorithms. Edited by J.-L. Faulon and A. Bender, Chapman & Hall: 233-267, 2010.
[5] A. Kerber, R Laue, M. Meringer, C. Rücker, E. Schymanski: Mathematical Chemistry and Chemoinformatics. Structure Generation, Elucidation and Quantitative Structure-Property Relationships. De Gruyter, 2013.
[6] H. J. Cleaves, M. Meringer, J. T. Goodwin: 227 Views of RNA: Is RNA Unique in Its Chemical Isomer Space? Accepted by Astrobiology.