Biomolecules in Astrobiology

Markus Meringer (DLR), H. James Cleaves (BMSIS), Stephen J. Freeland (UHNAI)

Astrobiology is the study of the origin, distribution and future of life in the universe. At the moment we know only one instance of life in the universe, that found on our planet. However, due to the accelerated discovery of extrasolar planets, questions regarding the origins and evolution of life are attracting increasing interest. Another driving force here is the discovery of extremophiles, organisms able to survive under extreme physicochemical conditions, e.g. with respect to temperature, pressure, acidity, etc. These new facts, mainly explored during the past two decades, have led to a re-examination of an age-old question: Is extraterrestrial life possible, does it exist, would it necessarily have to share the same biochemical framework as terrestrial life or could it be organized completely differently?

A key feature of all terrestrial life forms is the genetic code. In simple terms, the genetic code maps information to function. Information about life's composition is stored in DNA, a polymer of nucleotides. Function is realized by proteins. Proteins are polymers of another type of biomolecules, amino acids. Proteins give cells structure and perform multiple tasks within a living organism's metabolism, such as molecule transport and catalysis of chemical reactions. Interestingly, all known life forms with very few exceptions share the same genetic code, and proteins are built up from a unique but universal set of 20 genetically encoded amino acids.

Important questions arise from this fact: Why did terrestrial life select exactly these 20 amino acids out of a mathematically almost infinite number of possibilities? Is this only a random result of early evolution on Earth, or could there perhaps be universal rules behind this "choice"? Gayle Philip and Stephen Freeland came up with some original approaches to asking these questions [1]. They found that the genetically encoded amino acids exhibit a broad, even distribution of some key physicochemical properties, which sets them apart from any alternative set of amino acids drawn randomly from the superset of amino acids that was available for early evolution. This super set comprised about 60 amino acids which were likely available from abiological synthesis.

However, the "chemical space" of possible amino acids is much, much larger; see [2] and Figure 1 and 2. In order to either prove the hypothesis of Philip and Freeland, or to disprove and improve this approach, it is necessary to generate and examine much larger sets of amino acids. Dedicated computer programs, so-called structure generators use methods from graph-theory, combinatorics, group-theory and algebra to construct virtual chemical compound libraries with given constraints [3]. In order to define the structural constraints of the required α-amino acids, to code the input for the structure generation software, and to keep the sizes of constructed amino acid libraries and the required computational resources within acceptable limits, a four week on-site cooperation was initiated by the authors and carried out at the University of Hawaii's NASA Astrobiology Institute (NAI), financially supported by NAI's 2012 Director's Discretionary Fund.

The results include two virtual amino acid libraries that were generated using two different approaches. A large, "unique" library of 121,044 structures limited at an upper bound of six carbon atoms, which covers the space of molecular formulas as completely as possible, and a smaller "combined" library of 3,846 structures, which includes all coded amino acids. Figure 2 shows the composition of these libraries itemized by the number of carbon atoms. A detailed description of all methods and results will be published in [4].

References:

[1] Philip, G. K. and S. J. Freeland: Did evolution select a nonrandom "alphabet" of amino acids? Astrobiology 11(3): 235-240, 2011.
[2] Cleaves, H. J., 2nd: The origin of the biologically coded amino acids. J. Theor. Biol. 263(4): 490-498, 2010.
[3] Meringer, M.: Structure enumeration and sampling. Handbook of Chemoinformatics Algorithms. Edited by J.-L. Faulon and A. Bender, Chapman & Hall: 233-267, 2010.
[4] Meringer, M., H. J. Cleaves and S. J. Freeland. Beyond Terrestrial Biology: Charting the Chemical Universe of α-Amino Acid Structures. In preparation.

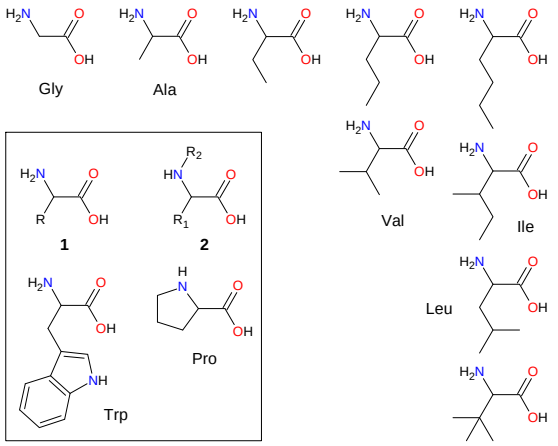

Figure 1: Small aliphatic α-amino acids with up to six carbon atoms. The coded amino acids among them are labelled by their abbreviation (Glycine, Alanine, Valine, Isoleucine, Leucine). Inset: 19 of the 20 coded amino acids are represented by generic structure **1**. R denotes the sidechain. Tryptophane, the largest coded amino acid, includes eleven carbon atoms. Coded amino acid Proline has another generic representation, depicted in structure **2**.

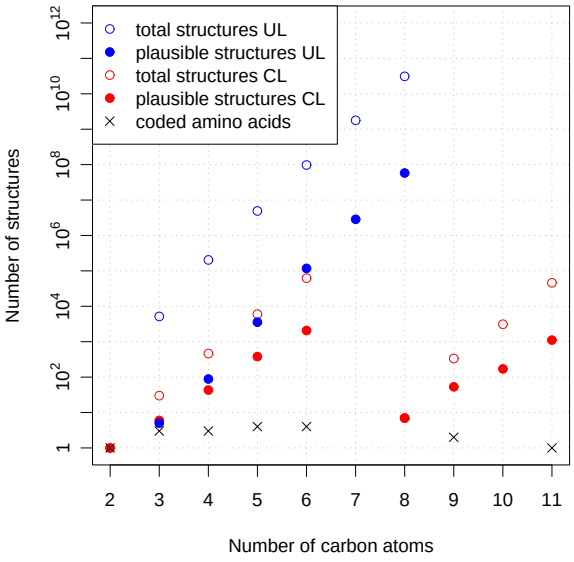

Figure 2: Sizes of α-amino acid libraries calculated during the study. In order to reduce the total set of mathematically possible structures to those structures which are chemically plausible, a list of 156 forbidden substructures was compiled. The plot shows the results of this effort. Compared to the unique library (UL), the combined library (CL) contains also structures of type **2** and structures with a larger number of carbon atoms.