Computational Exploration of the Chemical Space of Nucleic Acid-Like Compounds

M. Meringer[1], C. Butch[2,3], P. Burger[3], J. Goodwin[3], H. J. Cleaves II[2,4,5,6*]
[1]German Aerospace Center (DLR), Earth Observation Center (EOC), Münchner Straße 20, 82234 Oberpfaffenhofen-Wessling, Germany,[2]Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-IE-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan, [3]Chemistry, Emory University, Atlanta, Georgia, USA, [5]Institute for Advanced Study, Princeton, NJ 08540, [6]Blue Marble Space Institute of Science, 1515 Gallatin St. NW, Washington, DC 20011, [7]Center for Chemical Evolution, Georgia Institute of Technology, Atlanta, GA 30332

* cleaves@ias.edu

Introduction: Biological information storage for life as we know it is carried out by the nucleic acids DNA and RNA. However, these may be optimized end-states for life on Earth, or there may be other types of molecules which are similarly capable of carrying out these functions, perhaps used in alien biochemistries or in earlier biochemical states [1]. A number of these have already been synthesized in the lab [2-5], however the set of molecules compatible with this function may be far larger. Understanding this wider "chemical space" may give insights into what makes the biological isomers unique, as well as whether there are other isomers more easily accessed by abiotic chemistry.

Methods: Using graph theory-based structure generation [6], we have exhaustively computed the chemical isomer space of the natural ribosides (compounds of formula $C_5H_9O_4B$, where B is a nucleobase) [7] as well as a much wider range of formulas from C3 to C8, including those also containing N. We further culled these sets to include only likely chemically stable compounds, as well as only compounds containing at least two points of attachment for incorporation into a linear polymer via dehydration reactions or via the attachment of a suitable linker such as phosphate.

Having generated these sets, it is then possible to compute *in silico* some of the chemical properties of these compounds, including their three-dimensional conformations, van der Waals volumes, number of freely rotable bonds and solvent accessible surface areas, among many others. These computations then enable quantitative comparison of the generated structures.

Results: We find there are potentially millions of structural isomers, and many more stereoisomers, of nucleic acid-like compounds over this formula range. However, only a fraction of these are likely able to adopt conformations allowing for complementary strand-recognition. Further compatibility considerations may make the truly nucleic-acid-like set still smaller, though there may also be new strand-recognition motifs which have not been explored yet *in vitro* in these sets.

References:

[1] Joyce GF et al. (1987) Proceedings of the National Academy of Sciences USA, 1987. 84:4398-402. [2] Wilds CJ et al. (2002) Journal of the American Chemical Society 124:13716-21. [3] Egholm, M et al. (1992) Journal of the American Chemical Society 114:1895-1897. [4] Zhang L et al. (2005) Journal of the American Chemical Society 127:4174-4175. [5] Pinheiro VB and Holliger P (2012) Current Opinion in Chemical Biology 16:245-252. [6] Meringer M and Cleaves HJ (In Press) Philosophical Transactions of the Royal Society A. [7] Cleaves HJ et al. (2015) Astrobiology 15:538-558.