

Exploring the History of the Genetic Code – Making Proteins from Ancient Alphabets

Matilda S. Newton, Dana Morrone and Burckhard Seelig

All known life today is based on the standard set of twenty amino acids. However, the “universal” code encoding this set is a continually evolving “work in progress” which is predicted to have originated from a smaller number of primordial amino acids. Therefore, proteins employed by early forms of life relied on a significantly reduced chemical diversity of their amino acid building blocks. We aim to practically test this hypothesis through the synthesis of "reduced amino acid alphabet libraries" of random proteins. The libraries are comprised of 5 (considered most ancient), 9 and 16 amino acids and are compared to a control library of the standard 20. We aim to compare the libraries for the ability to form structured proteins and perform simple biological functions, using a selection method that allows the selection of functional proteins from a pool of trillions. By generating tangible empirical data in a field of research that has been largely dominated by theoretical approaches, this project has the potential to have a broad impact on our understanding of the history of the standard amino acid alphabet.