COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 9, Issue: 14, e201402004, http://dx.doi.org/10.5936/csbj.201402004

CSBJ

Open questions in origin of life: experimental studies on the origin of nucleic acids and proteins with specific and functional sequences by a chemical synthetic biology approach

Katarzyna Adamala [a,b,#], Fabrizio Anella [a,c,#], Rafal Wieczorek [a,d,#], Pasquale Stano [a], Cristiano Chiarabelli [a], Pier Luigi Luisi [a,*]

Abstract: In this mini-review we present some experimental approaches to the important issue in the origin of life, namely the origin of nucleic acids and proteins with specific and functional sequences. The formation of macromolecules on prebiotic Earth faces practical and conceptual difficulties. From the chemical viewpoint, macromolecules are formed by chemical pathways leading to the condensation of building blocks (amino acids, or nucleotides) in long-chain copolymers (proteins and nucleic acids, respectively). The second difficulty deals with a conceptual problem, namely with the emergence of specific sequences among a vast array of possible ones, the huge "sequence space", leading to the question "why these macromolecules, and not the others?"

We have recently addressed these questions by using a chemical synthetic biology approach. In particular, we have tested the catalytic activity of small peptides, like Ser-His, with respect to peptide- and nucleotides-condensation, as a realistic model of primitive organocatalysis. We have also set up a strategy for exploring the sequence space of random proteins and RNAs (the so-called "never born biopolymer" project) with respect to the production of folded structures. Being still far from solved, the main aspects of these "open questions" are discussed here, by commenting on recent results obtained in our groups and by providing a unifying view on the problem and possible solutions. In particular, we propose a general scenario for macromolecule formation via fragment-condensation, as a scheme for the emergence of specific sequences based on molecular growth and selection.

Mini Review Article

1. Introduction

One of the main open questions in the field of the origin of life is the biogenesis of proteins and nucleic acids as ordered sequences of monomeric residues, possibly in many identical copies. The first important consideration is that functional proteins and nucleic acids are – chemically speaking – *copolymers*, i.e., polymer formed by several different monomeric units, ordered in a very specific way. There are some indications of how homo-oligo/polypeptides [1-3] and homo-oligo/polynucleotides [4-6] can be obtained. Experimental results generally focus on exploring the conditions for achieving the desired condensation reactions and demonstrating, as proof-of-concept, that oligo- and polymerization can occur from the chemical viewpoint.

Attempts to obtain copolymers, for instance by a random polymerization of monomer mixtures, yield a difficult to characterize mixture of all different products. To the best of our knowledge, there is no clear approach to the question of the prebiotic synthesis of macromolecules with an ordered sequence of residues. Note that this problem has been beautifully solved by the invention of genetic code and the translation apparatus; however this does not solve the original question about how primitive functional molecules have arisen under prebiotic conditions.

The copolymeric nature of proteins and nucleic acid challenges our understanding of origin of life also from a theoretical viewpoint. The number of all possible combinations of the building blocks (20 amino acids, 4 nucleotides) forming copolymers of even moderate length is 'astronomically' high, and the total number of possible combinations it is often referred as the "sequence space". Simple numerical considerations suggest that the exhaustive exploration of the sequence spaces, both for proteins and nucleic acid, was physically not possible in the early Universe, both for lack of time and limited chemical material. Then, the question becomes: even taking for granted a prebiotic chemical route to copolymerization (i.e., a condensation reaction occurring in primitive conditions that does not produce only simple homopolymers), how could early, functional biological macromolecules have been selected? Do other functional biological macromolecules exist in still-unexplored regions of sequence space? What do they look like? Could living systems have been originated if chemical evolution was allowed to start from other regions of sequence space?

In order to answer these questions and shed some light in the mystery of origin of life, we started some years ago a multifaceted research program aimed at clarify some aspects of the problem, by

[a]*Sciences Department, University of Roma Tre; Viale Guglielmo Marconi 446, 00146 Rome, Italy*
[b]*MIT Media Lab, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[c]*Department of Bionanoscience, Kavli Institute of Nanoscience Delft University of Technology Lorentzweg 1, 2628 CJ, Delft, The Netherlands*
[d]*FLinT Center, Institut for Fysik, Kemi og Farmaci (IFKF), University of Southern Denmark; Campusvej 55, 5230 Odense, Denmark*

[#]These Authors are listed in alphabetic order and contributed equally to the work

* Corresponding author. Tel.: +39 0657336329
E-mail address: Pier Luigi Luisi (luisi@mat.ethz.ch)

following a chemical synthetic biology approach. As it will be discussed below, we have recently proposed that small organic molecules could have played a key role in favoring condensation reactions (organocatalysis), and that the stepwise condensation of short, easy-to-form oligomers can give rise to very specific long sequences, in high copy number, due to the interplay between reactivity, selection and solubility ("fragment condensation"). Finally, we have devised a strategy to explore the sequence space of proteins and nucleic acid based on random sequences that make use of modern molecular biology methods, at the aim of search for those sequences that are folded, because folding is the most important pre-requisite for any biochemical activity.

Although these approaches do not represent the solution, they may indicate a very useful path in the right direction.

2. Formation of peptide bonds and phosphodiester bonds by organocatalysis

Because of their place as the most fundamental biopolymers, prebiotic synthesis of peptides and nucleic acids has been an object of many investigations. The earliest experiments on the condensation of amino acids were conducted by simply heating the solutions of amino acids [7] and various thermal condensation and dehydration reactions between amino acids or amino acid precursors appeared in the literature [8,9]. Subsequent, more sophisticated approaches included hydrolysis of HCN polymers [10], condensation on clay environments [11] and in the presence of different mineral surfaces [12], or by atmospheric gases [13,14]. Modern methods, such as copper-catalyzed and salt-induced peptide synthesis [15,16], or condensation of *N*-carboxyanhydrides [17] allow for formation of peptides that are longer (ca. 10-mer) and can be formed from various proteinogenic amino acids.

Analogously, the first attempts at prebiotic condensation of nucleic acids constituted of simple heating experiments [18], or usage of metal ions [19]. The most effective methods that have eventually emerged are montmorillonite condensation [20,21] and water/ice eutectic approach [22,23].

Although there is number of interesting approaches for prebiotic condensation of both peptides and nucleic acids, they all have some shortcomings; mainly in the formation of specific sequences, and in chemoselectivity and regioselectivity. This problem might be successfully attacked if we take similar approach as life did – by exploiting organocatalysis. Enzymes are basically highly evolved organocatalysts. The investigation of prebiotic analogues possessing organocatalytic properties can shed light on the origin of molecular recognition and reactivity in the synthesis of proteins and nucleic acids and also provide novel perspectives (to be tested in the laboratory) about the "co-evolution" of peptides and nucleotides. This is perhaps the most intriguing aspect of organocatalysis in origin of life scenarios.

2.1 Peptide bonds

The simplest dipeptide Gly-Gly was reported to possess some proteolytic abilities [24-26] – but by far the best example of proteolytic activity is given by the dipeptide Ser-His [27]. Like Ser-His, also Ser-His-Gly is an active catalyst (see Figure 1A for structures), as well as similar di- and tri-peptides [27-29].

It is useful to remind that a catalyst does not change the free energy of reactants and products, but just enhance the reaction rate. If, by any means, the energetics of a reaction is inverted (i.e., by changing conditions so that the reaction goes in opposite direction), the same catalyst will still enhance the reverse reaction rate. Therefore, since Ser-His is endowed with proteolytic activity, under specific conditions it also acts as a catalyst for peptide *synthesis*.

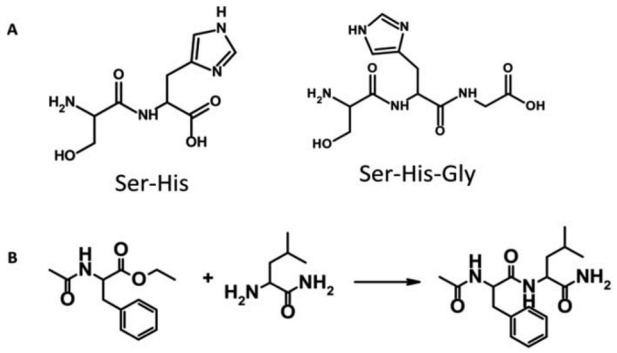

Figure 1. (A) The structure of catalytic dipeptides Seryl-Histidine (Ser-His) and Seryl-Histidine-Glycine (Ser-His-Gly). (B) The model reaction catalyzed by the dipeptides.

Ser-His (and Ser-His-Gly) catalyzes peptide bond formation between a *C*-terminus of an activated amino acid and the free *N*-terminus of another amino acid. The activating group on the *C*-terminus is a carboxylic acid ester, most commonly in the form of methyl or ethyl ester.

Together with the condensation product, the main compound accumulating during the reaction is the product of hydrolysis of the ester of the activated amino acid. The yield of the hydrolysis product (the free acid), under most typically explored conditions, pH 7-8, in water-organic solvent mixture, accounts for half or more of the observed total substrate conversion. Thus the accumulation of the desired peptide is hampered by the hydrolysis [28].

The most extensively studied reaction catalyzed by Ser-His is condensation between two natural amino acid substrates: a *N*-acetylated phenylalanine ethyl ester and leucine amide (see Figure 1B), to give the hydrophobic dipeptide *N*-acetylated phenylalanyl-leucinamide [28]. The reaction has almost no uncatalyzed background product formation, and the product accumulation in presence of small peptide catalysts (most notably Ser-His or Ser-His-Gly) is easily detectable thanks to the high UV absorbance of the product and substrates. HPLC analysis is usually employed to monitor the progress of the reaction.

To date, the only known peptide condensation reactions efficiently catalyzed by small peptides are the ones with a great change in hydrophobicity between product, the byproducts and the substrates. For the reaction to proceed to the desired direction, the difference in solubility between substrates, the side product (the hydrolyzed activated amino acid) and the main product must be enough to remove the main product from the reaction mixture. Otherwise the only compound accumulated with significant quantity will be the hydrolysis product from the ester activated amino acid substrate.

The removal of the produced peptide from the reaction mixture can happen through the precipitation or through the migration of the product to a different phase. In the first case, the substrates must be fully soluble in the reaction mixture, and the product must be nearly completely insoluble [28]. In the second case, the reaction must be carried out in a two-phase system, for example with a lipid bilayer membrane as a hydrophobic environment. The product of the reaction accumulates in the bilayer membrane, thus being effectively removed from the reaction environment, shifting the equilibrium toward the formation of the dipeptide product [30].