JACK W. SZOSTAK†

†*Dept. of Chemistry, University of Chicago, 5735 S Ellis Ave., Chicago, IL 60637, USA*

From Prebiotic Chemistry to the Beginnings of Biology

Current Status of Research on the Origin of Life

How, given an appropriate geophysical environment, inputs of energy and prebiotically synthesized chemicals, did life actually begin? Progress in understanding the conversion of simple starting materials such as cyanide into the chemical building blocks of biology has been summarized in the accompanying article by John Sutherland. Here I will focus on our efforts to understand how those prebiotically available chemicals spontaneously assembled into the first primitive cells, commonly referred to as protocells. Since the beginning of biology corresponds to the beginning of Darwinian evolution, those protocells must have contained a genetic polymer capable of encoding advantageous functions in its sequence. Moreover the chemistry of the cell and its environment must have been sufficient to drive replication of this material, which would have generated the variation that is the substrate for evolution. In addition, this genetic material must have been segregated into spatially localized units so that any advantage provided by a particular sequence could accrue to itself and not to other unrelated sequences.

It has become clear in recent years that the most likely primordial genetic material was in fact something very much like modern RNA. This conclusion was by no means obvious even a few years ago. Going back somewhat further, Jerry Joyce and the late Leslie Orgel concluded, on the basis of difficulties encountered in explaining the prebiotic synthesis of ribonucleotides and the nonenzymatic copying of RNA sequences, that RNA might have been preceded by some simpler ancestral genetic polymer. This outlook sparked a remarkable outburst of creative chemistry in the form of a search for alternative genetic polymers that might have been easier to synthesize and/or replicate than RNA. The highlight of this approach was the tour de force synthesis by Eschenmoser of diverse families of nucleic acids all capable of self-association into antiparallel Watson-Crick base-paired duplexes. Given the many seemingly plausible alternatives to RNA, the question arose as to why, in the end, did life settle on RNA and not something?

Our Recent Contributions

Much of our recent work has focused on obtaining a mechanistic understanding of the chemistry of nonenzymatic template-directed RNA copying. Somewhat surprisingly, our efforts in this regard have led not only to a framework for the chemical copying of arbitrary RNA sequences and a model for replication, but also to a potential explanation for why RNA won out over all potential competitors. We started by addressing some of the peculiar aspects of the kinetics of template-directed primer extension that had first

been noted by Orgel. The most puzzling of these observations was the very slow rate of primer extension when only a single activated monomer can bind next to the primer, and the dramatic acceleration of the reaction when a second activated monomer binds downstream of the first monomer [1]. This catalytic effect had been unexplained for some 25 years, until it was rediscovered by Noam Prywes in my lab [2] and then explained by Travis Walton [3] as the consequence of the generation of a highly reactive covalent intermediate which forms by the reaction of two activated monomers with each other. A series of structural studies led by Wen Zhang in my lab culminated in the observation by time resolved crystallography of all steps in the primer extension reaction, from monomer binding to intermediate formation to formation of a new phosphodiester bond [4]. Subsequent studies showed that primer extension via reaction with imidazolium-bridged dinucleotides was both faster and of higher fidelity than reaction with activated monomers [5]. Our most recent studies have shown that monomers bridged to short oligonucleotides lead to even faster reaction, and, importantly, allows all four canonical nucleotides to be copied and incorporated at similar rates. We now believe that the formation of imidazolium-bridged intermediates is the key to the efficient and accurate nonenzymatic copying of RNA templates.

While the above mechanistic insights provide a potential explanation for how primordial RNA sequences could have been copied prior to the evolution of polymerase ribozymes, we still have to ask why RNA and not some other genetic polymer? To address this question we have begun to evaluate the copying kinetics of alternative genetic polymers. The two that we have studied in most detail are arabino- and threo-nucleic acids, since the corresponding mononucleotides are likely to form during the synthesis of ribonucleotides. We see that threo-nucleotides exhibit slower formation of the bridged intermediate, while both threo- and arabino-nucleotide bridged intermediates exhibit slower primer extension. Furthermore once a threo- or arabino-nucleotide is added to a primer, subsequent extension is also slower. In contrast, non-canonical nucleotides in the template can be copied over by primer extension with ribonucleotides with only modestly reduced rates. Other aspects of template heterogeneity such as 2′-5′ linkages and pyrophosphate linkages can also be copied over to generate a canonical RNA product. Based on these experiments it appears that RNA is intrinsically better able to take part in nonenzymatic copying chemistry, and thus over repeated rounds of copying chemistry, would have won out over competing nucleic acids. At this time it appears that potential alternatives to ribonucleotides are either harder to make, more susceptible to degradation, or are less able to take part in copying chemistry, leading to potential alternatives to RNA being filtered out at one or more stages.

Returning to the subject of nonenzymatic RNA replication, there are several problems that make replication more difficult than simple template copying. In a prebiotic situation, there is no way to supply defined the defined primers needed for replication of

a linear genome, while the replication of a circular genome would encounter severe topological difficulties. To overcome these and other issues, we have proposed a model for primordial RNA replication that we call the virtual circular genome or VCG model [6]. In this model a circular genomic sequence is represented by a collection of oligonucleotides that map to a circular sequence, but no actual circular molecules need exist. Replication is then driven in a distributed manner by oligonucleotide elongation by primer extension, with longer oligos acting as templates and shorter oligos acting as primer, downstream helpers and invaders for catalysis of strand displacement. We are actively engaged in experimental tests of this model.

At this point we can see at least the outlines of an overall prebiotic, nonenzymatic process leading to the synthesis and replication of a primordial RNA genome. But a protocell is not just replicating RNA – that RNA must be encapsulated within a replicating compartment boundary. By analogy with modern cells that boundary is likely to be a bilayer lipid membrane. However, primordial cell membranes must have been quite different than modern cellular membranes. Critically, they must allow the passive transfer of nutrients such as nucleotides from the external environment to the cell interior; in addition, they must be able to grow and divide solely in response to the chemistry and physics of the environment.

For 20 years now we have studied the properties of vesicles assembled from simple fatty acids, ranging from oleic acid/oleate as a convenient model system to decanoic acid/decanoate as a more prebiotically plausible system. Such vesicles are permeable to activated nucleotides and even to di- and tri-nucleotides, so that in principle RNA replication could take place internally but be fed from an external source of material. Vesicle growth can be driven by the addition of alkaline micelles to vesicles in a solution at lower pH. Ting Zhu in my lab showed that fast growth transforms initially spherical multilamellar vesicles into filamentous vesicles that are easily broken apart by mild shear forces into smaller daughter vesicles [7]. Subsequent slow growth allows the daughter vesicles to increase in size, so that the cycle can repeat. Thus a fluctuating environmental supply of fatty acids can drive repeated cycles of growth and division. More recently Anna Wang found conditions that lead to the spontaneous assembly of fatty acids into giant unilamellar vesicles. Fast addition of alkaline micelles again leads to a rapid increase in surface area, after which shape fluctuations lead to spontaneous division [8].

An alternative pathway for vesicle growth involves competition between vesicles. Irene Chen initially showed that osmotically swollen vesicles grow following an influx of iso-osmotic vesicles, because the tense membrane of the swollen vesicles can relax by absorbing fatty acid molecules from the empty vesicles [9]. This pathway relies upon the dynamic behavior of single chain amphiphiles, which exchange rapidly between vesicles. Subsequently, Itay Budin showed that fatty acid vesicles that contained a fraction of two

chain lipids could also grow by absorbing fatty acids from pure fatty acid vesicles, with growth into filamentous forms again leading to facile division [10]. However in this case the daughter vesicles do not have the same composition as the parental vesicles, so continued cycles would require in situ synthesis of two-chain lipids. Another route to competitive growth was found by Kate Adamala, who showed that the internal synthesis of a hydrophobic peptide could drive the competitive growth of fatty acid vesicles [11]. In this case continued cycles of growth and division would require the replication of an internal catalyst of peptide synthesis, presumably a ribozyme. Overall, fatty acid vesicles seem like ideal models for protocell membranes.

There is, however, a big problem: the high concentrations of Mg^{++} that are required for RNA copying chemistry rapidly disrupt and destroy fatty acid membranes. Finding solutions to this mutual incompatibility is critical to the field and is increasingly the focus of our attention. The first partial solution to this problem was identified by Kate Adamala, who showed that chelation of Mg^{2+} by citrate allowed RNA copying to proceed while protecting membranes. This allowed RNA synthesis to occur within model protocells [12], but unfortunately the model of growth following the addition of alkaline micelles was no longer effective. We are currently investigating the potential for competitive growth in the presence of Mg^{2+}-citrate, which could provide a solution to the compatibility problem noted above. However, such a solution would not be prebiotically realistic, and so our search for other solutions to this problem continues.

Our work with mixed fatty acid – phospholipid vesicles points to a potential solution, with interesting consequences. Such vesicles are stable to the presence of moderate levels of Mg^{2+}, and in the presence of even higher levels of Mg^{2+}-citrate they exhibit enhanced permeability to activated nucleotides, which allows RNA template copying chemistry to proceed well internally. We know they can grow by absorbing fatty acids from fatty acid vesicles – but in order to maintain a constant membrane composition, we would need a nonenzymatic pathway by which fatty acids could be transformed into phospholipids. An interesting possibility is that the same chemistry that activates phosphates for RNA synthesis could also activate carboxylates for esterification and phospholipid synthesis [13]. Another attractive possibility involves vesicle membranes that contain single chain cyclophospholipids, which also confer tolerance to Mg^{2+}, and which would be easier to synthesize from fatty acids. An important direction for future research is therefore to explore synthetic pathways that would allow maintenance of a steady state membrane composition during multiple cycles of vesicle growth and division.

Outlook for the Future

In summary, the twin goals of nonenzymatic RNA replication and sustained vesicle growth and division are in sight, and I expect to see experimental demonstrations of these key processes in the near future. The integration of these two processes remains

challenging, but several approaches to solving the compatibility problem are under investigation, and I am optimistic that one or more solutions to this problem will become apparent in the coming years. At that stage a fully functional protocell system will open the doors to new modes of evolutionary exploration as we see how adaptation to a range of physical and chemical environments occurs in a reconstituted laboratory-scale version of the RNA World.

Acknowledgements

I thank the many brilliant students and postdoctoral fellows who have carried out the work described above over the past two decades. I also thank the Simons Collaboration on the Origins of Life, the National Science Foundation, and the Howard Hughes Medical Institute for supporting this work.

References

1. Wu T, Orgel LE. *J. Am. Chem. Soc.* **114**, 5496 (1992).
2. Prywes N, Blain JC, Del Frate F and Szostak JW. *eLife* **5**, e17756 (2016).
3. Walton T and Szostak JW. *J. Am. Chem. Soc*. **138**, 11996 (2016).
4. Zhang W, Walton T, Li L, Szostak JW. *eLife* **7**, e36422 (2018).
5. Duzdevich D, Carr CE, Ding D, Zhang SJ, Walton TS, Szostak JW. *Nuc. Acids Res.* **49**, 3681 (2021).
6. Zhou L, Ding D, Szostak JW. *RNA* **27**, 1 (2021).
7. Zhu TF and Szostak JW. *J. Am. Chem. Soc*. **131**, 5705 (2009).
8. Kindt J, Szostak JW, Wang A. *ACS Nano.* **14**, 14627 (2020).
9. Chen IA, Roberts RW and Szostak JW. *Science* **305**, 1474 (2004).
10. Budin I and Szostak JW. *Proc. Natl. Acad. Sci. USA.* **108**, 5249 (2011).
11. Adamala K and Szostak JW. *Nat. Chem.* **5**, 495 (2013).
12. Adamala K and Szostak JW. *Science* **342**, 1098 (2013).
13. Bonfio C, Russell DA, Green NJ, Mariani A, Sutherland JD. *Chem. Sci.* **11**(39), 10688 (2020).