MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL

January 15, 2015

Origins Project
Arizona State University
PO Box 871902
Tempe, AZ 85287-1902

Dear Chair of the Committee:

I am submitting this cover letter in conjunction with materials supporting my application for the Origins Project Postdoctoral Prize Lectureship, for which I was nominated by my postdoctoral advisor, Jack W. Szostak. You should also have received letters of recommendation from Prof. Nicholas Hud (my PhD advisor) and Prof. Loren Williams, both of the Georgia Institute of Technology. Below, I detail how my research interests in the emergence of life relate to the interests of the Origins Project. I also have been involved in origins of life-related outreach and science communication efforts, including an invited talk (hosted by Prof. Sara Walker) to the BEYOND Center at ASU about my origins research a year and a half ago. I describe my speaking engagements in further detail in my CV, which I have included with my application. Further information on my outreach efforts, as well as a summary of each of my publications written for the lay public, can be found at my professional website, [redacted] As an ASU alumnus and a researcher in the origins of life, I am especially interested in the efforts of the Origins Project. My research relates to the emergence of life at the molecular level, with a particular interest in how complex pools of molecules likely to have been produced *via* prebiotic chemical reactions may have influenced the behaviors of the earliest cells and enzymes. In this letter, I detail four projects from my postdoctoral studies informed by these interests.

By way of background: Great strides have been made in our understanding of biochemistry since the early days of molecular biology. For much of this time, the state of the art of analytical techniques required that researchers adopt a reductionist approach, limiting their studies to simplified systems involving one or a few purified biomolecules. This approach has yielded considerable dividends over the years, and our understanding of the chemical basis of biology has advanced greatly over time. However, life is messy and complex: even a single *E. Coli* cell contains on the order of 10^6 proteins and 10^5 tRNAs, and fully ⅓ of the mass of the cell is composed of solutes. Recent advances in DNA sequencing, mass spectrometry, and computing technology have enabled researchers to experimentally and computationally study a previously intractable range of complex "-omes," including the rich ensemble of RNAs, small molecules, and even molecular interactions within living cells.

If life is messy, "pre-life" was perhaps even messier. Without the sophisticated suite of well-defined enzyme catalysts present in modern life, prebiotic chemical reactions likely produced a wide array of molecules, including those found in life today, as well as countless others. For example, prebiotic copying reactions of informational polymers, like RNA and DNA (or related molecules), no doubt produced a different set of products than copying of RNA or DNA with contemporary protein enzymes. These products likely included the desired product of full length with the correct sequence, as well as a range of related molecules, including fragments of the desired product, products with mismatched nucleotides relative to the template, or products with slightly different chemical features. Such are the problems associated with prebiotic copying, as well as those raised by the possibility of informational polymers that were different (in nucleobases and/or backbone chemistry) than DNA and RNA, that in 1998, Leslie Orgel, a pioneering researcher in the origins of life, called the problem of one or more precursor polymers to RNA and DNA not extant in contemporary life a "gloomy prospect" (LE Orgel, *TIBS* (1998) 23:491-495).

While such problems with information transfer no doubt caused difficulties in the emergence of life, they were, of course, overcome; life emerged. Below, I discuss four projects from my recent research relating to different aspects of primitive (proto)-biochemistry that suggest ways by which life might have overcome such problems. In several cases, my colleagues and I have discovered ways in which the complexity found in prebiotic chemical mixtures could have afforded positive aspects that ultimately enabled the emergence of life.

Aaron E. Engelhart • Center for Computational and Integrative Biology
Massachusetts General Hospital/Harvard Medical School
Simches Research Center, CPZN7106 • 185 Cambridge Street • Boston, MA 02114
[redacted]

The first two projects I describe relate to a fundamental problem in the prebiotic copying of RNA. When the protein enzyme polymerases found in contemporary cells copy RNA, they produce a product with a homogeneous backbone, containing the well-known 3′-5′ phosphodiester linkage. All known prebiotic RNA copying reactions, however, produce a mixture of 3′-5′ and 2′-5′ linkages, as a result of the presence of both 2′- and 3′-hydroxyls on the ribose sugar. This was known for some time to have deleterious consequences for RNA duplexes; the presence of 2′-5′ linkages destabilizes the RNA double helix, depressing the T_M (Giannari, PA and MJ Damha, *Nucl. Acids. Res.* (1993) 21:4742-4749). Furthermore, helical 2′-5′ linkages are preorganized for transesterification; that is, complementary strands catalyze the degradation of strands containing these linkages (Usher, DA and AH McHale, *Proc. Nat. Acad. Sci. U.S.A.* (1976) 73:1149-1153). As a result, the investigators that made this discovery suggested the inherent instability of the 2′-5′ linkage may have afforded a proofreading mechanism in early life that provided selective pressure for the dominance of the 3′-5′ linkage in modern life.

This presents a problem in nonenzymatic RNA copying reactions, which generate, in general, at least 10-25% 2′-5′ linkages. We sought to answer two questions: 1) can functional RNAs, such as ribozymes and aptamers, tolerate this level of 2′-5′ substitution? and 2) what are the structural consequences of isolated 2′-5′ linkages within dsRNA? In the case of the first question, we found that 10-25% 2′-5′ linkages are well-tolerated within both an aptamer (which binds the biological cofactor FMN) and a ribozyme (the nuclease hammerhead ribozyme), with both these molecules maintaining function. Furthermore, we found that this level of 2′-5′ substitution within a ribozyme-length dsRNA (30 nt) enabled thermal strand separation, which was otherwise impossible in the same dsRNA with a fully 3′-5′-linked backbone. Thus, random mixtures of 2′-5′ linkages could not only have not been fatal in functional RNAs, but they could have afforded an capacity for multiple-turnover RNA replication that RNAs comprised of a homogeneous backbone would have lacked (Engelhart, AE; Powner, MW; and JW Szostak, *Nat. Chem.* (2013) 5:390-394). In another project, in which we sought to answer the second question, we obtained the first X-Ray crystal structures of RNA containing 2′-5′ linkages. We found that, while destabilizing, these linkages induce only local structural distortions. Furthermore, we showed direct evidence of the origins of 2′-5′ linkage-induced preorganization for strand degradation for the first time; this previously had been shown only by model building. Additionally, our results showed that, while helical 2′-5′ linkages are predisposed to degrade, they are stable for at least a week (the time required to grow a diffraction-quality crystal) under our conditions (Sheng, J et al, *Proc. Nat. Acad. Sci. U.S.A.* (2014) 111:3050-3055). Thus, in this work, we demonstrated that a likely product of prebiotic replication (RNA with randomly dispersed 2′-5′ linkages) maintained the capacity for catalysis and ligand binding, exhibited similar structural features to fully 3′-5′-linked RNA, and that, in some cases, these 2′-5′ linkages could have even afforded a positive trait.

My third and fourth projects both relate to another aspect of prebiotic copying of RNA. All known model prebiotic copying reactions for nucleic acids are relatively inefficient; given the inefficiency of stepwise synthesis, even fairly high coupling yields result in surprisingly low full-length product yields. Assuming a stepwise synthesis reaction, even at 99.5% coupling efficiency at each step, only 86% full-length product is produced in a 30-coupling reaction, while at a more prebiotically plausible 85% coupling efficiency, only 0.6% full-length product is produced. This presents a problem in the prebiotic synthesis of ribozyme catalysts, as most ribozymes with $k_{cat} > 1\ min^{-1}$ are ≥ 30 nt in length.

In the third project, we pursued methods to enable RNAs of as short a length as possible to still act as functional molecules. A partial solution had already been shown: most ribozymes can be split into two oligonucleotides, with each oligonucleotide comprising half of at least two structural base-paired stems. We sought to drive the length requirement (and, thus, the associated requirement for copying efficiency) down still further, by truncating the 3′-terminus of half-functional RNAs, generating shorter RNAs that could form marginally stable (but catalytically nonfunctional) structures. Each truncated strand had at its 5′-terminus a single-stranded poly(C) region that could act as a template for nonenzymatic primer extension when complexed with the other strand. By taking these unstable, transient complexes and subjecting them to nonenzymatic primer extension, the ends were polished and stems were regenerated, thus producing a stable, functionally active RNA.

Aaron E. Engelhart • Center for Computational and Integrative Biology
Massachusetts General Hospital/Harvard Medical School
Simches Research Center, CPZN7106 • 185 Cambridge Street • Boston, MA 02114
[redacted]

Through this strategy, we demonstrated that two ribozymes (the hammerhead and a *trans*-acting Diels-Alderase) and an aptamer (for malachite green) could be produced from half-ribozymes that were otherwise too short to be catalytically active. Thus, *via* this work, we showed that the requirements for copying reactions in prebiotic functional RNA systems could be substantially relaxed, and even incomplete ribozyme copies likely to be found in prebiotic RNA synthesis product mixtures could be made functional by subsequent nonenzymatic primer extension. This work was just published (Adamala, K; Engelhart, AE; and JW Szostak, *J. Am. Chem. Soc.* (2015) 137:483-489).

In the fourth project, we sought to examine the consequence of the presence of even shorter oligonucleotides on the function of primitive cells. This interest stemmed from previous work in the lab on lipid vesicles as models of primitive cells. Previously, colleagues in the lab had shown that fatty acid vesicles can grow competitively, as a result of their lipid composition (Budin, I and JW Szostak, *Proc. Nat. Acad. Sci. U.S.A.* (2011) 47:6141-6143), encapsulated catalysts producing membrane-binding components (Adamala, K and JW Szostak, *Nat. Chem.* (2013) 5:495-501), or osmotic stress due to vesicle contents (Chen, IA; Roberts, RW; and JW Szostak, *Science* (2004) 305:1474-1476). In each case, it was suggested that the propensity for membrane growth represented a primitive form of fitness for a lipid vesicle. While accumulating membrane components certainly is an important part of fitness for a cell, indiscriminate growth comes at the cost of diluting cellular contents. Towards solving this paradox, we sought and discovered a mechanism by which the growth of primitive cells could have recruited additional functional molecules (i.e., enzymes). Thus, as cells grew, catalysts would have come online and become active. This would have afforded a primitive mechanism for homeostasis in early life (Engelhart, AE; Adamala, KA; and JW Szostak, *Submitted*).

Most encouragingly, we have shown that such a mechanism would have arisen automatically, as a natural consequence of inefficient copying. As mentioned, low-efficiency stepwise synthesis likely produced, in addition to a small amount of full-length product, many copies of short oligonucleotides complementary to their template. We discovered that these short oligonucleotides act as concentration-dependent ribozyme inhibitors. At high concentration, we found that the presence of an excess of these short oligonucleotides reduced hammerhead ribozyme activity to undetectable levels. Upon 100-fold dilution (with enzyme:inhibitor stoichiometry constant), we observed ca. 40-fold (as expressed by apparent first-order rate constant) activation of this enzyme (in contrast to ca. 2-fold *inactivation* for the enzyme alone)! We also observed that the more modest (ca. 3-fold) concentration changes associated with competitive lipid vesicle growth can enable enzyme activation. An enzyme-inhibitor complex, when present in a lipid vesicle that grows at the expense of other vesicles, exhibited ca. 10-fold activation after growth (in contrast to ca. 2-fold inactivation for the enzyme alone). Furthermore, we showed that a pool of short oligonucleotides (5-6 nt) of random sequence were also capable of affording such a regulatory mechanism. Thus, we have shown the complex mixtures of RNA likely present in the earliest cells could have afforded one of the earliest homeostatic mechanisms (Engelhart, AE; Adamala, KA; and JW Szostak, *Submitted*).

My research throughout my postdoctoral studies has shown that, in many ways, the complex mixtures expected to have been produced in prebiotic reactions likely did not present an impediment to the emergence of life. Indeed, it appears that the chemical complexity associated with prebiotic reactions may have enabled phenomena and mechanisms for the emergence of life that would not have been possible if these reactions proceeded with the chemical homogeneity observed in modern biochemistry. I hope to have the opportunity to present these results and discuss them with members of the Origins Project and the broader ASU/Tempe community as the 2015 Origins Project Postdoctoral Prize Lecturer later this year. Please contact me if you have any questions. Thank you for your consideration.

Sincerely,

Aaron Engelhart

Aaron E. Engelhart • Center for Computational and Integrative Biology
Massachusetts General Hospital/Harvard Medical School
Simches Research Center, CPZN7106 • 185 Cambridge Street • Boston, MA 02114
[redacted] • [redacted] • [redacted]