Interpretation of Chemical Environments by RNA and the Implications to the Origins of Life

Brian Larson

Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Chemistry Commons

Let us know how access to this document benefits you.

Recommended Citation

10.15760/etd.1704

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Interpretation of Chemical Environments by RNA and the
Implications to the Origins of Life

by

Brian Larson

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Chemistry

Thesis Committee:
Niles Lehman, Chair
Albert S. Benight
John Perona

Portland State University
2014
Abstract

We describe the initial realization of behavior in the biosphere, which we term behavioral chemistry. If molecules are complex enough to attain a stochastic element to their structural conformation in such a way as to radically affect their function in a biological (evolvable) setting, then they have the capacity to behave. This circumstance is described here as behavioral chemistry, unique in its definition from the colloquial chemical behavior. This transition between chemical behavior and behavioral chemistry need be explicit when discussing the root cause of behavior, which itself lies squarely at the origins of life and is the foundation of choice. RNA polymers of sufficient length meet the criteria for behavioral chemistry and therefore are capable of making a choice.

We test these theoretical findings with an empirical system; since RNA simultaneously possesses evolvability and catalytic function we set to construct a model system. Firstly, during in vitro transcription of the Chlorella PBCV-1 pdg intron, we observed multiple RNAs of slower electrophoretic mobility along with the expected 98-nucleotide transcript. The preferences observed for TI or linear RNA for the pdg intron are manifestations of alternative phenotypic states and represent evidence of behavior at the chemical level. Secondly, we provide a basis for the marriage between empirical findings and the philosophical studies of biosemiotics. As well, provide a basis for the minimal criteria for an interpretative response and suggest that interpretation is evolvable. The studies hold relevance to the origins of life showing a minimal system capable of carrying out a choice based on the environment. We have shown this phenomenon is not an intrinsic characteristic to all RNAs and that the Tetrahymena ribozyme recognizes the presence of the object through the sensing of a sign.
Dedication

This thesis is dedicated to Aubree Ng, my biggest supporter and source of motivation throughout this process.
Acknowledgements

Dr. Niles Lehman

Albert S. Benight

John Perona

Lehman Lab members past and present

PSU chemistry faculty and staff

Christopher Southgate

Rob Jenson for his insights on free will.
# Table of Contents

Abstract................................................................................................................................i

Dedication........................................................................................................................... ii

Acknowledgments..............................................................................................................iii

List of Figures......................................................................................................................v

List of Tables.......................................................................................................................vi

Chapter 1: Introduction: The Chemical Origin of Behavior is Rooted in Abiogenesis.....1

Chapter 2: Kinetic traps in the Chlorella PBCV-1 pdg intron .............................................15

Chapter 3: Empirical demonstration of interpretive behavior in catalytic RNA: implications for the origins of life.................................................................30

References..........................................................................................................................46
List of Figures

Figure 1.1 Bounded chemical systems.................................................................12
Figure 1.2 RNA Folding pathways.......................................................................13
Figure 1.3 The evolution of behavior.................................................................14
Figure 2.1 Primary sequence and predicted secondary structure of the PDG intron...22
Figure 2.2 Gel electrophoresis of Chlorella virus PBCV-1 pdg intron ..................23
Figure 2.3 The pdg intron forms TIs during in vitro transcription......................24
Figure 2.4 Representation of the folding trajectory of the PDG intron...............25
Figure 2.5 Transcription reaction conditions have an effect on the percentage yield of TIs ........................................................................................................26
Figure 2.6 TI transcription yield as a percentage of total yield in reactions containing RNA oligonucleotides.................................................................27
Figure 2.7 TI transcription yield of mutated primary sequence..........................28
Figure 3.1 Example payoff matrix for interpretive behavior..............................41
Figure 3.2 5-error mutant Tetrahymena ribozyme activity assay......................42
Figure 3.3 Activity comparisons for wild-type and 5-error mutant Tetrahymena ......43
Figure 3.4 The DSNL ribozyme ........................................................................44
Figure 3.5 In vitro evolution of the DSNL ribozyme .........................................45
Figure 3.6 Results of In vitro evolution...............................................................46
List of Tables

Table 2.1 Changes to the chemical environment and the effect on available folding trajectories..........................................................................................................................29
Chapter 1

This chapter is adapted from: Larson, B.C.; Jensen, R.P.; Lehman, N. The Chemical Origin of Behavior is Rooted in Abiogenesis. *Life* 2012, 2, 313-322.

1.1 Introduction

Behavior is an integral feature of life. Behavior is manifest when a choice is possible, and a living entity responds to its environment in one of multiple possible ways. For relatively simple unicellular life, we tend to think of behavior as a deterministic response. Bacterial cells respond to a metabolite gradient in predictable ways, dictated by the biophysical processes of substrate uptake, second messenger activation, and flagella operation, for example. For relatively complex life such as sentient and conscious humans, we tend to think of behavior as a less predictable function—one that is enmeshed in the concept of free will [1].

Free will has been described as a modern *vitalism* [2], while Christian de Duve mentions, “we still know too little about the human mind to affirm categorically that it is a mere emanation of neural activity lacking the power to affect this activity.” [3] At this point, we are not going to indulge in the ageless philosophical debates of dualist and monist. Rather we consider behavior, defined as above and conceptualized in free will, the outcome of the interplay between genotype and phenotype based on evolution and stochasticity. If we consider the working definition of life as a self-sustaining chemical system able to evolve, then we can describe the manifestation and characteristics of this behavior as well as the progressive association with objects we call living opposed to a criterion to be alive.
As long ago as ~300 BCE the early atomist Epicurus hypothesized, in an attempt to refute the deterministic nature of the physical world and account for the behavior of choice, that all occurrences are due to small matter (atoms) colliding and interacting in voided space (kenos) and that the atoms are not restricted to straight lines; rather, they exhibit random swerves [4]. Epicurus introduced chance into the description of nature centuries before its being detailed by particle physics and quantum mechanics. Schrödinger expanded on the basis of choice saying, “The spatio - temporal process in the body of a living being which are in line with its intellectual activity and with its consciousness, or actions carried out in whatever way, are deterministic in a statistical sense, if not strictly deterministic in nature.” [5]

As conscious humans, our thoughts on choice and free will range between perceptions that a living entity functions as an absolute mechanism ultimately described by physical laws, and a constant personal awareness of making decisions of our own accord. Without anthropomorphizing molecules, we suggest that even the simplest of life must follow physical laws, but also be able to make a “choice” based on their environment and dependent on self-contained information.

Cashmore recently tackled this problem head on, and came to the conclusion that any atomic “swerve” was unable to lead to more than a perception of free will [2]. In his discussion of human behavior and free will he states, “in some ways, it might be more appropriate to replace ‘genetic and environmental history’ with ‘chemistry’—however, in this instance these terms are likely to be similar and the former is the one commonly used in such discussions.” Our premise is that this transition between genetic and environmental history and chemistry needs to be explicit when discussing the root cause
of behavior, which itself lies squarely at the origins of life. During abiogenesis on the Earth, enough information became available that molecules accumulated the required chemical repertoire to make “choices” based on the environment, i.e., substrate uptake or folding pathways, and transitioned from a system in which chemicals had behavior to one in which chemicals could behave.

1.2 Chemical Behavior vs. Behavioral Chemistry

In order to address the ultimate source of “choice” we must differentiate between chemical behavior and behavioral chemistry. Chemical behavior is the essentially deterministic outcome of a chemical reaction among one or a few relatively small molecules. This is a term commonly used by chemists to describe the predicted products of a reaction under study, with the implicit assumption that any deviation from this outcome is simply a failing on our part to understand all the energetic facets of the molecules involved. In other words, chemical behavior is a deterministic process, one that becomes sharper the more we know about a system. A good example is the Haber production of ammonia: \( \text{N}_2 + 3\text{H}_2 \rightarrow 2\text{NH}_3 \). Because of its historical and industrial importance, this is a very well studied reaction that has a deterministic outcome under typical reaction conditions. For example, at elevated temperatures and pressures, the negative enthalpy of this reaction (\( \Delta H = -92.2 \text{ kJ/mol} \)) will ensure that about 15% of nitrogen will be converted to ammonia, while the balance of the reactants will remain uncombined.

However, we would like to introduce the subtly different concept of behavioral chemistry. When the molecules involved are complex enough to attain a stochastic
element to their structural conformation in such a way as to radically affect their function in a biological (evolvable) setting, then they have the capacity to behave, and the analysis of such molecular-level behavior would be described as behavioral chemistry. Behavioral chemistry is derived from two processes, either or both of which can be in operation at any given point in time.

The first is the expression of quantum mechanical stochasticity “writ medium-sized” in the conformational energetics of macromolecules such as RNA and proteins. Here we mean that atomic swerves do not influence collections of small molecules in any meaningful way, but with the advent of polymeric macromolecules characteristic of life, the stochasticity begins to reassert an influence.

At first glance one might claim that the distinction here is arbitrary, because even small molecules can adopt conformational variations in time that can affect their subsequent reactivities; think of the sugar pucker of glucose for example. However, this property does not maintain sufficient complexity to warrant behavior, and does not impact evolution. That said, complexity is not unique to living systems and can be seen in many non-living physical systems as well as in simple mathematical models. Complexity cannot solely determine behavior. A pertinent example of this is the formose reaction, whereby the incubation of the two simple precursors formaldehyde and glycoaldehyde can lead to the spontaneous synthesis of many sugars, including ribose. The formose reaction is both notoriously complex and has the feature of autocatalysis, making it of prebiotic importance [6]. It is in fact the best known example of a self-organized chemical cycle that bears on the origins of life but is absent of genetic material. Although the formose reaction can generate a wide range of reaction products that vary with the
reaction conditions and phosphorylation states of the reactants, it fails to have the capacity to behave in the sense that we are defining. This is because the stochastic nature of the system, which originates in the quantum-mechanical nature of its component atoms, is mainly expressed in the set of side reactions that serve to convolute and disrupt the cycle, rather than to promote its continuation. The formose reaction, though autocatalytic, is essentially unbounded, meaning that it lacks a means to ensure survival of its “self.” The self is too poorly delineated (Figure 1a).

Once Nature had the capacity to synthesize information-bearing macromolecules though, the stochasticity of the system became embodied into the “behavior” of the molecules because now there was the possibility that a molecule was the system! In essence, a system requires both a genotype and a phenotype to be able to display behavior. The “self” is now clearly defined (Figure 1b); however, it can be a single self-replicating molecule or a network of related cooperators. Here we are using the example of RNA as the informational polymer, but the same conclusions were to apply if other polymers, or even inorganic lattices or compositional sets of macromolecules such as lipids, were the ancestral genotypes. Clearly, the advent of compartmentalization (protocellular life) would further enhance the establishment of a bounded genotype [7], thereby firmly entrenching behavior. The key to this first aspect of behavioral chemistry is that the information is physically bondable such that the possible states of the system can be quantized (discrete). Once available, a boundary such as a membrane prevents informational loss, allowing for heredity and evolutionary change. This characteristic of life is not possible in an unbounded system such as the formose reaction.
The second process that allows behavioral chemistry is the real possibility that macromolecules end up in kinetic traps during their folding [8]. When the lifetimes of these traps are of the same order of magnitude as the time in which these molecules need to function, then a second type of behavioral expression becomes possible. An example of behavioral chemistry in operation is the metal-ion-dependent kinetic traps observed during the folding of many self-splicing RNA introns, such as the group I intron from the rRNA gene of *Tetrahymena* [9,10]. Another example is the equilibrium between pseudoknotted and non-psuedoknotted structures observed in the yeast telomerase RNA [11]. In this case, depending on the folding pathway, the reaction coordinate of the RNA folding trajectory can pass through one of two possible routes to the lowest energy fold; each pathway has a intermediate fold of distinct energy representing kinetic traps (Figure 2). Which pathway is taken depends on where the RNA happens to be prior to initiating folding, a state that depends in turn on the composite atomic thermal fluctuations of the composite atoms in the RNA. Empirical studies with single catalytic RNA molecules have revealed that the subtleties of their conformational dynamics are so complex that partially unfolded molecules possess a type of “memory” in that they tend to return to the folded state that they occupied prior to denaturation [12,13]. The important point here is that the molecule can be physically observed to be in one of the intermediate states, and thus an immensely large number of microscopic atomic ensemble states can be momentarily binned into two discrete macroscopic structures. The translation of quantum mechanical variation into alternative phenotypic states is the first manifestation of behavior at the chemical level.
1.3 From Behavioral Chemistry to Biological Behavior

At some point during the origins of life on the Earth, molecular systems acquired one or both of the conditions needed to satisfy the variations of chemical behavior described above. For the best understood (but still debated) case of RNA during a hypothetical RNA World, it is entirely possible that these condition often applied simultaneously. Historical RNAs existed the maintained sufficient informational storage and followed folding pathways to kinetic traps. While maintaining these properties and RNA molecule in essence respond to environmental conditions, through changes of state. interpret them [14], and maintain specific conformations that represent primitive choice. This seemingly simple progression lead to what is referred to as organismal behavior and is outlined in Figure 3.

Firstly, Quantum mechanical and thermal fluctuations in single molecules get incorporated into informational polymers that can then display behavioral chemistry; the choices that arise from behavioral chemistry also lead to the manifestation of evolution. Natural selection operating on these molecules leads to chemical choice, then to heritable decision making, and ultimately to behaviors ascribed to intelligent organisms. This can be observed empirically in populations of RNA evolving in the test tube [15]. However it is important to realize that the characteristic of being able to adopt—and chose among—alternative phenotypes is also subject to natural selection. Genotypes that confer only single phenotypes may be advantageous today, but during biogenesis the ability to choose in direct response to environmental perturbations could actually have conveyed a selective advantage. This would have set up a feedback loop between genotype and behavior as depicted in Figure 3.
Complex collections of behaving molecules (i.e., organisms) with multiple genes and phenotypic products can collect heritable traits crafted by natural selection. The evolution of heterogeneous collections of molecules, through the advent of cellularity) further delineates ‘self’ from surrounding chemical environments. This second layer of separation, in addition to direct chemical bonding, adds richness to the spectrum of behavior by controlling which elements of the environment the self interacts with. Even single-celled organisms such as bacteria and protists could then express complex behavior as chemotaxis or phototaxis. Investigations on \textit{E. coli} have revealed that this bacteria expresses chemotaxis primarily through the actions of 15 proteins, such as specific chemoreceptors for amino acids. The functions of these proteins serve to bias an otherwise random walk powered by flagellar motion into a directional track toward certain ligands [16]. Directional and stabilizing selection operating on, ancestral alleles of these proteins could favor \textit{E. coli} genotypes that migrated up a, say, serine concentration gradient when cellular serine stores were low. If this migration were not 100% coincident with a concomitant need to migrate up a, say, galactose concentration gradient, then a molecular choice becomes apparent. The advent of behavior, which is both flexible by definition and robust through genetic underpinnings was key in the establishment of a new phenomenon, evolvability [17], which in turn ensured the evolutionary persistence of behavior itself.

This evolutionary persistence eventually to behavioral ecology. Now payoff matrices can describe various behaviors [18]. Molecules can express behavioral chemistry that can be quantified in terms of a payoff matrix; however, empirical demonstrations of this are lacking.
Importantly organisms evolved intricate and convoluted means of interacting with their environments and displaying classical behavior, as in a rat choosing paths within a maze (Figure 3). Having developed the ability to speculate about what the future may bring and how best to survive potential threats, natural selection may have presented multicellular life with the ability to transcend strict determinism and make decisions of their own free will [1]. Specific to the behavior of intelligent life manifested in the act of making a choice, there are a few requirements to fulfill before this higher level meaning of choice is relevant in an evolutionary context:

(1) There must be at least two options available to the chooser. From a molecular perspective, this would be described in terms of what is allowed energetically (e.g., Figure 2). Perhaps importantly, we can think of options in terms of different goals, or in terms of different routes to reach one goal; both fit nicely into the context of what is energetically allowed.

(2) There must be a determination of one option as being optimal for some reason. For intelligent life forms, this generally comes after a cost/benefit analysis of some kind (Figure 3), which relies upon some previous exposure to or knowledge of the system around which a choice is being made. In the absence of any knowledge or exposure, a decision relies on some kind of unrelated bias, or is arbitrarily based upon probability (as in the case of a coin toss). For a molecular system lacking intelligence, the only real option is to fall to bias or probability, both of which are essentially yielding to energetic factors again.
(3) Some action must be taken towards realizing the choice being made. From a physical perspective, the objective determination of a choice being made is reliant upon some observable difference relative to the pre-choice state. In some sense, this is an extension to the organismal level of the molecular memory discussed above for RNA. Unless it can be communicated between intelligent beings in some kind of language, the only evidence of a choice is some subsequent action, which can be observed. Specifically, the choice or the ability to make it must have a fitness consequence.

Even intelligent life follows the basic properties provided by behavioral chemistry. A singular genotype may express multiple phenotypes, genotypes must interpret and respond to their environments, and most importantly the phenotype and the ability to chose among them are subject to natural selection.

Dennett argues that determinism does not imply inevitability, and that furthermore even if the physical reality were deterministic then there indeed are real options to life, not just apparent ones. His premise is that the process of natural selection has operated on an extant physics to create organisms with free will and with responsibilities for their own actions [1]. Here we extend this idea to the origins of life and attempt to place the origins of behavior into a chemical context. Further experimental work is needed to test some of the ideas put forth in this paper.

1.4 Conclusions

Physical variations in molecules of a certain level of complexity can be trapped into quantum levels, and at this point, a transition from chemical behavior to behavioral
chemistry can happen. This is the origin of all forms of behavior, and this transition was more-or-less coincident with the origins of life. The boundaries that define the self allow choice to transcend quantum mechanics and become engrained in chemistry and described better by statistical thermodynamics and Newtonian mechanics. Empirical studies that test the hypothesis that molecules can exhibit measurable properties that would satisfy the constraints of a payoff matrix will shed much light on this issue. Whether behavioral chemistry is still an operative force in the expression and evolution of organismal behavior is uncertain. Moreover, behavioral chemistry may have been an ultimate determinant of free will. It has been argued that determinism, and therefore lack of free will, does not imply inevitability, and that furthermore even if the physical reality were deterministic then there indeed are real options to life, not just apparent ones. This premise is that the process of natural selection has operated on an extant physics to create organisms with free will and with responsibilities for their own actions [1]. Here we extend this idea to the origins of life and attempt to place the origins of behavior into a chemical context.
Figure 1.1 A) The formose reaction. Incubation of formaldehyde and glycoaldehyde leads to the autocatalytic synthesis of hexoses and pentoses B) Stick and line drawing representing RNA polymer secondary structure. In both panels grey circles represent the boundary of “self.” The boundary is ill defined for the formose reaction (a) but definite for the RNA (b).
Figure 1.2 RNA Folding pathways. A RNA (green) can achieve a lowest energy fold through two separate folded intermediates. In the first (left-hand) intermediate, a less stable helix is formed 3’ end of the molecule followed by the formation of a helix in the 5’ portion to complete the psuedoknot. In the second (right-hand) intermediate a stem-loop is formed in the 5’ end of the molecule followed by hydrogen bond breakage to allow a helix to form in the 3’ portion of the psuedoknot. The energy states of the folds are dependent on the reaction coordinate of the folding pathway (black line).
Figure 1.3 The evolution of behavior. Quantum mechanical and thermal fluctuations in single molecules get incorporated into informational polymers that can then display behavioral chemistry. Natural selection operating on these molecules leads to chemical choice, then to heritable decision making, and ultimately to behaviors ascribed to intelligent organisms.
Chapter 2

2.1 Introduction

Activity in RNA relies on specific ‘RNA folds.’ RNA structure, like protein structure, is described at the primary sequence structure level, the secondary level, the tertiary level, and the quaternary level. ‘RNA folds’ derive from a primary sequence responsible for the conserved base-pairing structures abundantly found in naturally occurring RNAs, as well as from the interaction of non Watson-Crick pairs to form specific phosphodiester geometries. But, they can best be described as a conserved 3D structure including all chemical properties such as bond geometries, Watson-Crick base pairing, base stacking, non Watson-Crick hydrogen bonding, and metal-ion and water binding (19). Specific folding motifs include tetraloops, hairpins, helices, etc., and can be predicted (20) as well as used to design function (21). These interactions allow RNA to bind small molecules with high selectivity moreover to catalyze a variety of chemical reactions.

As mentioned in Chapter 1, the RNA folding trajectory can sometime pass through multiple possible routes to the lowest energy fold; each pathway may have an intermediate fold of distinct energy (i.e., kinetic traps). The pathway taken depends on the RNA interaction with chemical environment prior to and during folding. RNA molecules can be physically observed to be in one of the intermediate states, and thus an immensely large number of microscopic atomic ensemble states can be momentarily binned into discrete macroscopic structures. The folding variation into alternative phenotypic states is the first manifestation of behavior at the chemical level.

Herein we report topological isomers (TIs) of distinct phenotypic states that form during in vitro transcription as 5-20% of the total transcription yield of ~100- nucleotide,
A+U-rich intron (83%), from the Chlorella PBCV-1 pdg (Figure 1)(22). The higher order topologies have previously shown to be converted to linear molecules following incubation in salt solutions for several hours (23).

These observed TIs are representative of long lasting kinetic traps only energetically available during transcription. The preference or choice given to a folding pathway (i.e., TI yield) is dependent on the chemical environment. The RNA products populating two specific energetic states under static conditions but derivable in ratio from changes to the static conditions are an example of behavioral chemistry as defined in Chapter 1.

2.2 Materials and Methods

Sources of oligonucleotides and DNA

DNA oligonucleotides were purchased from Integrated DNA Technologies. A dsDNA template of the native Chlorella virus PBCV-1 pdg intron with the T7 promoter sequence (5′-TAATACGACTCACTATAGTATGTAAATATTAATAATCACAACCTAATAATTCATAATCACAACCTAATTATATTATTTATTTATTTATTTAAAATTTCGATTTGTAATGTTTTTGCGAG-3′; the promoter sequence is underlined) was constructed by recursive gene synthesis (24,25). The E. coli promoter sequence (5′-CGAAAAACAGGTATTGACAAATGCAGTAAGATACAAATCG-3′) and various mutations were introduced during recursive gene synthesis by changing the appropriate oligonucleotide(s).
In vitro transcription

Transcription reactions generally contained 15 mM MgCl$_2$, 50 mM EPPS (pH 7.5), 5 mM dithiothreitol, 2 mM spermidine, 2 mM of each rNTP, and 50–200 U of T7 RNA polymerase (Fermentas or Ambion). Radiolabeled RNAs were generated by including 10–20 µCi of $\alpha$-32P•ATP (for internally labeled molecules), or 50–100 µCi of $\gamma$-32P•GTP in the transcription reaction. Following incubation, reactions were typically treated with 3 U DNase I (Fermentas) for 30–60 min at 37°C, quenched by adding EDTA to a final concentration of 20 mM, acid-phenol / chloroform extracted, and precipitated with ethanol. Products were separated on 5%, 8%, or 15% polyacrylamide / 8 M urea gels and visualized by phosphorimaging. Radiolabeled products were purified by phosphorimaging using Dip-N-Dot (26). To minimize conversion of the TIs to the linear form, gel-purified products were eluted from their gel slices in 10 mM EPPS (pH 7.5) / 1 mM Na$_2$EDTA at 4°C for 8 – 16 hrs, and desalted and concentrated using Nanosep MF and precipitation. Reaction products were detected by phosphorimaging on a Typhoon Trio+ imager and quantified using the ImageQuant software (GE Healthcare). *E. coli* RNA polymerase transcription reactions generally contained ~40–60 absorbance units of PCR-amplified DNA template (containing the T7 A2 promoter; (27,28) in 10 mM MgCl$_2$, 50 mM KCl, 40 mM Tris (pH 8.0), 5 mM dithiothreitol, 50 µg / mL BSA, 300 µM each NTP, and 4 U of *E. coli* RNA polymerase sigma-saturated holoenzyme (Epicentre) in a volume of 20 µL and were incubated for 90 minutes at 37°C. Reactions were terminated, visualized, and quantified as above.
Glyoxal denaturing of TIs

Roughly 1–5 pmol of 5′-radiolabeled TI or linear molecules (10 μL volume) were added to 10 μL of glyoxal loading dye (Ambion), and incubated for 30 min at 50°C. Products were separated on a non-denaturing 10% polyacrylamide next to their non-glyoxylated counterparts.

2.3 Results and Discussion

Topological isomer formation

During in vitro transcription of the Chlorella PBCV-1 pdg intron, we observed multiple RNAs of slower electrophoretic mobility along with the expected 98-nucleotide transcript. RNAs of slower electrophoretic mobility were not dependent on the RNA polymerase used and were observed in transcription products of both T7 and e.coli RNA polymerases (Figure 2a). Both polymerases tests allowed for access to a co-transcriptional folding trajectory. Although differences in polymerases produced TIs of varied conformation. Depending on the acrylamide percentage of the gel, the mobilities of these topological isomers varied relative to RNA size standards, indicating that they were not topologically linear (Figure 1). To verify that TIs were denatured, we electrophoresed the TIs and linear transcript both with and without glyoxal treatment (Figure 2b). The migration pattern of the glyoxylated TIs resembled their migration on urea-containing gels with glyoxal treatment, indicating that urea alone is sufficient to denature them (Figure 2b). Despite being denatured, however, the TIs were not converted to a linear topology and remain kinetically trapped. This combined variability electrophoritic mobility compared to the topologically linear molecules suggests that the
TIs truly are kinetically trapped topological isomers RNA and remain kinetically trapped after transcription. During transcription an energetic pathway is availed through the interaction of polymerase and RNA product; once the polymerase reaction is completed the TIs persist along a unique folding trajectory.

Environmental affects RNA Folding.
The likelihood of the RNA transcripts to become trapped as topological isomers is dependent on the reaction conditions found during transcription. When transcription reactions were carried out with lower than normal DNA input concentration, say 0.2X, regardless of the concentration of polymerase, either .5 units or 2 units in 30ul total reaction volume, the TIs are produced at the standard ~8% of total RNA yield. However, if the transcription reactions were carried out with higher than normal DNA input concentration, say 5X, then at low polymerase concentrations TI yield dropped to less than 1% of total RNA produce. In contrast, with the concentrations of DNA and polymerase both increased, the standard yield was again observed (Figure 3). The high DNA input low polymerase affords less interaction between the RNA products and polymerase enzyme, thus limiting the availability of co-transcriptional the folding trajectory that is responsible for producing TI products. In essence the TI yield can be ‘tuned’ based on the environmental conditions of the transcription reaction.

This tuning does not persist under all changes to the chemical environment. To show this we transcribed the RNA with 100pmols each of a single 12-mer antisense RNA oligonucleotide RT1-RT8 included in the reaction cocktail. The 12-mer oligos RT1-RT8 span the length on the transcripts after nucleotide 2 for the entire length. None of the
transcription reactions with interfering oligos disrupted the TI yield or changed the likelihood of RNA being found in the kinetic traps. The presence of antisense primers during transcription does not interrupt the interaction between polymerase and RNA. The TI products are yielded in ‘normal’ amounts.

Mutational effects on TI yield.

As previously show by A. Burton in 2010 (23) the predominance of the TI product can be effected through mutations in the primary sequence of the RNA transcript (Figure 2.7). With randomization of the last 17 nucleotides TI yield can be nearly abolished. Also notable is the effect of the single point G98U. This mutation relieves TI yield to less than 10 percent relative to the wild type and ~1% of total RNA. Sequence mutations can interrupt the polymerase/RNA interaction and block the TI folding trajectory.

2.4 Conclusion

In this case, depending on the co-transcriptional folding pathway, the reaction coordinate of the RNA folding trajectory can pass through one of two possible routes. The predominate route leads to linearly transcribed RNA products while the other route leads to the formation of a kinetically trapped topological isomer of the same sequence (Figure 2.4). The route preferred is directly influenced by the conditions that produces the RNA. We tested various environmental parameters and their effects on TI yield (Table 2.1). The pdg intron can decipher the environment and change the state preferred.
The preferences observed for TI or linear RNA for the pdg intron are manifestations of alternative phenotypic states and represent evidence of behavior at the chemical level. There exist a set of environmental conditions that the system, mainly the RNA, can ‘sense’ and choose a phenotypic response accordingly. However, all environmental conditions do not show the same perception or ability of the RNA to ‘sense.’ As expressed in Table 2.1 the RNA readily senses changes in polymerase concentration but only with high DNA input concentrations. On the other hand the RNA has a limited ability to sense these environmental changes with low DNA input. Also the RNAs ability to sense and make a correct choice based on the environment directly relies on the information contained within the sequence of polymers. The physical variations in the molecules, due to a level of complexity, are trapped into quantum levels and thus represents a transition from chemical behavior to behavioral chemistry. This system divulges some of the variation in sensing present in an RNA system. Biological polymers at the origin of life must have been able to interpret environmental stimuli and change a given course of action in response.

This system, however, has a very specific outcome related to highly evolved interactions between a protein enzyme and intron. The reason for the existence of TIs in the pdg intron has yet to be determined. While interesting in the context of behavioral chemistry, further empirical investigation into the physical confirmations of the TI products will shed light on the selective advantage of maintaining two discrete folding pathways. The data will also be reinforced as more variables (i.e., environmental conditions) are tested to tune the predominance or lack of TI products.
Figure 2.1 Primary sequence and predicted secondary structure of the PDG intron.
**Figure 2.2** Gel electrophoresis of Chlorella virus PBCV-1 pdg intron RNA and formation of topological isomers (TI) during in vitro transcription. Transcription product RNAs were separated by electrophoreses on denaturing gels, with either 5% or 15% polyacrylamide alongside linear RNA molecules of 100 and 197 nucleotides. The unlabeled band above the linear transcript contains linear overextension products (*i.e.*, n +1, n+2).
Figure 2.3 A) The pdg intron forms TIs during in vitro transcription with both T7 and E. coli RNA polymerases. B) Gel electrophoresis of glyoxal-treated linear and TIs alongside their non-glyoxylated counterparts.
Figure 2.4 Representation of the folding trajectory of the PDG intron. The transcription products (1) can proceed through two distinct folding pathways. One pathway (left) leads to a short lived intermediate (2) to a low energy fold (4) that is readily denatured and visualized as the linear product. Another pathway (right) continues to a long lived folding intermediate (3) that we call the topological isomer or TI.
Figure 2.5 Transcription reaction conditions have an effect on the percentage yield of TIs. A) Percentage yield of TI and Linear transcripts with low DNA input ~10 absorbance units containing high polymerase concentration (2 units per reaction) and low polymerase concentration (.5 units per reaction). B) Percentage yield of TI and Linear transcripts with low DNA input ~50 absorbance units containing high polymerase concentration (2 units per reaction) and low polymerase concentration (.5 units per reaction).
Figure 2.6 TI transcription yield as a percentage of total yield in reactions containing 100pmols each of a single 12-mer antisense RNA oligonucleotide RT1-RT8. These 12-mer oligos RT1-RT8 are designed each with a unique non-overlapping sequence that occur antisense to the transcripts after the second nucleotide. Each oligo corresponds to 12 nucleotides in the transcript sequence for the entire length (e.g., 12 nucleotides x 8 oligonucleotides + 2 nucleotides = 98 nucleotides = length of the transcript).
Figure 2.7 (Figure adapted from the dissertation of Aaron Burton reference 23). The yield of TI can be increased or minimized based on changes to the primary sequence. U94G, C96A, A97U, and G98U represent single nucleotide mutations while 96-98 and 81-98 refer to nucleotides mutated in random: CAG mutated to AUU and UUUGUAUGUUUUGCAG mutated to ACCAGUCGGACUCCCAUU respectively.
Denaturation of RNA product after completion of transcription reaction (Figure 2.3b)  
**Condition**  
**Effect on TI yield**  
**Conclusion**  
The TI are persistent to denaturation and are not readily converted to linear topologies  
RNA remain kinetically trapped after transcription. During transcription an energetic pathway is availed through the interaction of polymerase and RNA product. Once separated the TIs persist in an energetic well along a folding trajectory.
Chapter 3

3.2 Introduction

The origin of life relied on the catalytic transfer of information without the presence of previously evolved macromolecules. This event depended on non-enzymatic abiotic processes to build a catalytic information transfer system that can evolve new phenotypes. With the discoveries of catalytic RNAs such as RNase P (29) and the *Tetrahymena* ribozyme (30) and the “smoking gun” showing that the ribosome is a ribozyme (31), previous ideas of a historical period in which RNA maintained both phenotype and genotype of primitive life (32) have been reinforced. Relying on the RNA World Hypothesis to describe a polymer-based genetic/catalytic system requires the description of the synthesis of monomers, the condensation of the monomers to polymers, and the elongation of polymers, all resulting in a catalytic system that replicates the mature polymers. Insight into some of these processes has been shown in the work of Sutherland, prebiotically relevant synthesis of nucleotides (33), in the work of Monnard, condensation of activated nucleotides with metal catalysts in the eutectic phase of water (34), and in the work of Bartel, *in vitro* selection of a polymerase ribozyme (35). Somewhat unexplored in the current origins of life story is the historical foundation of the genotype-environment interaction. Understanding the intricacies of this interaction are of great interest.

The information contained within a given genotype, on the most fundamental level, must allow for environmental interactions and responses. At the origins of life, biological polymers must have been able to interpret environmental stimuli and change a
given course of action in response. In the context of this text this has been described as the transition between behavioral chemistry and chemical choice in Chapter 1. Primitive biological polymers (RNA) would not have been proficient at deciphering environments since evolution had only had the benefit of a short timescale. Interpretation of environments by biological processes has been thoroughly described in the field of biosemiotics (36). To further explore the basis of interpretation and evolution there in, we required a conceptually robust definition of interpretation. This definition shows that interpretation of the environment, intrinsic in all organisms, may be a consequential diagnostic in the search for proto-biotic entities and has been previously proposed as follows (14):

A response, \( R \), of an entity is a proper interpretation of \( X \) as a sign of \( O \) if and only if:

1. The entity has a property, \( Q \), of undergoing change of state \( \Delta S \) in response to some \( X \), where \( R \) is any actual instance of such a response;
2. (a) \( R \) tends to increase the probability of an effect of a certain general type, \( P \);
   (b) This tendency of \( R \) depends on a relation between \( X \) and \( O \), where the occurrence of \( X \) does not necessarily imply the occurrence of \( O \);
3. The property \( Q \) has been selected for the tendency of instances of \( R \) to actualize effects of general type \( P \).
Some biological properties such as structure exist for a function and have been evolved to produce a specific phenotype. However, structure does not solely determine fitness and as such some biological events or actions, define above as a change of state \( \Delta S \), also arise for function (the general type of outcome for which they have been selected). A special case of such functional changes of state are is that type of change of state which has been selected to occur in response to a particular state, or change in state, of the environment (14).

Interpretations are a further subset of these adaptive changes of state. In an interpretation, the trigger for the change of state (the ‘sign’) is related to but not obligatorily implied by some aspect of the environment (the ‘object’) that would make such a change in state beneficial. For example, an amoeba will crawl up a chemical gradient of attractant molecules towards the source. The chemotactic response is adaptive because it increases the chance of the amoeba obtaining nutrition (by ingesting the bacterium that is the source of the attractant). The response is interpretative because it is not the chemical gradient that is of benefit to the amoeba but the bacterium of which the attractant molecules are a sign. It is this separation between sign and object that distinguishes an interpretation from a direct adaptive response. This separation necessitates that interpretations are fallible and are thus not a deterministic response to a chemical environment. The amoeba may swim up a chemical gradient that has not, in this particular instance, been produced by a potential source of nutrition. Interpretative behavior may be adaptive in spite of its fallibility because on balance the benefits derived from obtaining information about the object without having to respond directly to the
object outweigh the costs of misinterpretations. The amoeba benefits from responding the
attractant gradient because it can thereby detect the presence of a bacterium at a distance.

The simplest conceivable scenario in which an interpretative response could occur
would involve a two-state entity in a two-state environment. Suppose that the
environment can be ‘favorable’ (F) or ‘unfavorable’ (U) and the entity has two possible
states, A and B. In environment F it is advantageous for the entity to be in state A and in
environment U it is advantageous for the entity to be in state B. The fitness of the entity
in a specific environment (e.g., environment F, entity state A, environment U, entity state
B, etc) may be communicated generally with a 2x2 payoff matrix (Figure 1). In this
simple payoff matrix we can define the conditions of each of the four quadrants: O1
(environment F, entity state A), O2 (environment U, entity state A), O3 (environment F,
entity state B), and O4 (environment U, entity state B). The ‘fitness’ (payoff) for the
entity is the sum of O1 to O4 and represents a correct interpretation. In the non-
interpretative (wild-type) configuration of the entity, its state (A or B) varies
independently of the state of the environment. An interpretative variant of the entity
might be capable of, say, responding to some sign that indicates (fallibly) that the state of
the environment is F, the response being a change from state B to state A. Such an entity
may have a selective advantage over the wild-type because it will increase the time that it
spends in the advantageous combination of environment F and state A. Such
interpretative responsiveness will only be adaptive, however, if it is not outweighed by
the costs of a misinterpretation. This cost will be a function of the degree of disadvantage
entailed by the entity being in state A in environment U, and the probability of being so
placed in such a relation to the environment by the fallible (\textit{i.e.}, less than perfect) correlation between the presence of the sign and environmental state F.

For responsiveness to the sign to have the potential to count as an interpretation it is thus necessary that there be both a potential advantage to the response in the appropriate environment, and a potential disadvantage if the response occurs in the inappropriate environment (a misinterpretation). This may be expressed with three inequalities that together provide the minimum criteria for interpretative behavior in a 2-environment 2-state-entity system. The relevant inequalities are: \( O_1 > O_2 \) (in state A the entity is better off in the favorable rather that the unfavorable environment); \( O_1 > O_3 \) (in the favorable environment the entity is better off in state A than state B); \( O_4 > O_2 \) (in the unfavorable environment the entity is better off in state B than state A, hence the \textit{cost} of a misinterpretation).

We can test these theoretical findings with an empirical system; since RNA simultaneously possesses evolvability and catalytic function we set to construct a model system (37). Specifically, we wished to test the hypothesis that a single biomolecule may exhibit properties consistent with the minimum criteria for interpretative responsiveness. In this chapter we report investigations of the properties of the group I self-splicing ribozyme from the ciliated protozoan \textit{Tetrahymena}. We show that the cation dependency of the catalytic properties of this ribozyme could, under appropriate conditions in a hypothetical two-state environment, provide the basis on which the inequalities required as minimal conditions for interpretative responsiveness could be satisfied. We demonstrate that such interpretative responsiveness is a potentially evolvable property in
variants of this ribozyme. Furthermore we provide evidence that interpretation is not intrinsic to catalytic RNAs and report on the *in vitro* evolution of an “un-interpretive” ligase ribozyme. The implications for origin of life research are discussed.

3.2 Materials and Methods

RNA preparation

The variants of the 393-nucleotide *Tetrahymena* ribozyme were transcribed *in vitro* from DNA templates as described previously (37). Two variants were chosen for study. The “wild-type” is the L–21 derivative of the naturally occurring sequence obtained from *T. thermopile* and used in one of the first published examples of a catalytic RNA (38). This ribozyme requires divalent cations for self-splicing or trans-splicing activity *in vitro*, and can utilize Mg$^{2+}$ and/or Mn$^{2+}$ but no other (39). The “mutant” is a five-error variant of the wild-type ribozyme that was discovered through *in vitro* selection for activity in 10 mM CaCl$_2$ (40). This variant has a catalytic efficiency in CaCl$_2$ that is five orders of magnitude less than that of the wild-type ribozyme in MgCl$_2$, but its activity Ca$^{2+}$ is nonetheless notable because the ionic radius of Ca$^{2+}$ is significantly larger than that of Mg$^{2+}$, and the five mutations are required to open up the active site enough to confer detectable catalytic activity (37). The two ribozymes were purified by electrophoresis through 8% polyacrylamide / 8M urea gels and resuspended in 0.1 mM EDTA at a concentration of 10 pmol/µl prior to use. The substrate for *in vitro* assays, the 25-mer RNA "S-1t" (41): 5’-GGCCCUCU•AAAUAUAAAUAUAUAUAUA-3’, was purchased commercially from IDT (Coralville, IA).
Ribozyme assays

The assays were performed in a "pick-up-the-tail" format (41,42). Successful reaction between the mutant *Tetrahymena* ribozyme and S-1t substrate results in a length change from 393 to 410 nt. The reaction products can be tracked through electrophoreses. The ribozyme was 5’ end labeled with $\gamma^\text{-32P}[\text{ATP}]$ by the the Optikinase enzyme (USB Biochemicals, Cleveland, OH). For assays, 0.5 µM ribozyme was incubated with 1.25 µM substrate for 280 min in four different reaction buffers, all of which were buffered to pH 8.3 with 30 mM EPPS. The reaction buffers (final concentrations) were: 25 mM MgCl$_2$ + 10 mM CaCl$_2$ (O1), 0.5 mM MgCl$_2$ + 10 mM CaCl$_2$ (O2), 25 mM MgCl$_2$ (O3), and 0.5 mM MgCl$_2$ (O4). Reactions were stopped by the addition of enough EDTA to chelate stoichiometrically all divalent cations in solution, and the products were electrophoresed through 15% polyacrylamide / 8 M urea gels and the resulting bands were detected by phosphorimaging on a Typhoon Trio$^+$ imager and quantified using the ImageQuant software (GE Healthcare).

Evolution of DSNL ribozyme

Hypermutagenic PCR produced products with nucleotides randomized at 6% degeneracy and used as templates for in vitro transcription. The transcribed RNA was purified for length homogeneity by denaturing PAGE. At the opening generation $\approx$6 copies of each $\approx$10$^{13}$ different RNA molecules (0.1 nmol) were allowed to ligate a chimeric DNA RNA substrate SDSLCE . For all selection generations, the ligation reaction was carried out in
the presence of 5 μM ribozyme, 25 μM substrate, 10 mM Mg$^{2+}$ or Co$^{2+}$, and 15 mM EPPS (pH 8.4) or Tris (pH 7.0) at 22°C.

### 3.3 Result/Discussion

The mutant ribozyme was initially challenged to react under conditions represented by each of the four quadrants of the payoff matrix: O1, O2, O3, and O4 (Figure 2A). Given the criteria previously discussed where O1>O2, O1>O3, and O4>O2 are required to fulfill the minimal qualifications of an interpretative entity, the reactions where allowed to continue to completion and the percentage of product yield was determined as representation for the payoff (Figure 2B). The mutant ribozyme satisfied the given criteria (Figure 2C).

To investigate further, the assays were replicated; this time, however, the wild-type ribozyme was challenged in place of the mutant. While the wild-type ribozyme also met the criteria for interpretation there existed a disparity in the inequalities of the payoff matrix when compared to the mutant (Figure 3.3A). Most importantly the relationship between O4 and O2 shows that the mutant has a lesser reactive increase not having Ca$^{2+}$ at low Mg$^{2+}$ concentration compared to the wild type. The reactive increase is seen for a continuum of Mg$^{2+}$ concentrations from 0.5mM to ~2mM when the two ribozymes are compared (Figure 3.3B,C).

The mutant is more dependent on the sign, while the wild-type has more “flexible” or “permissive” interpretation. Another way to elucidate the outcome is that the wild-type ribozyme is maximally interpretive and has less dependence on sign for
interpretation. The mutant has been evolved in the laboratory with greater metal ion
tolerance forcing a greater dependence on the sign. In greater evolutionary context
starting with a primitive ribozyme there existed selection for interpretation and through
many rounds or
generations of evolution (say billions of years) the wild-type ribozyme emerged. The
mutant ribozyme has been artificially selected for a more specific interpretation and thus
the mutant is selected for response to the sign \(X = Ca^{2+}\), not the object \(O = Mg^{2+}\).
The trait of interpretation is not connate to all catalytic RNAs as it is contingent on the
entity, \(Q\), undergoing a change of state, \(\Delta S\), specific to the sign, \(X\). To show this we
constructed a simple RNA ligase ribozyme (Figure 5) using the catalytic core and
structure of the previously reported DSL ribozyme (44) with adaptations to allow for in
vitro evolution experiments (45). The artificial ribozyme is active in \(Mg^{2+}\) and \(Co^{2+}\)
(although minimally in the latter) at the initial generation and not active in \(Ca^{2+}\). With
introduction of mutations at 3% per nucleotide position, the ribozyme was tasked to
catalyze the ligation of a short substrate. The selection was carried out in two parallel
selection lines carried out exclusively in \(Mg^{2+}\) and \(Co^{2+}\). The mutated populations was
catalytically enriched through eight selective generations (Figure 5). The resulting
populations, regardless of the metal ion present during selection, predominately
expressed a G4A mutation.

In this case, interpretation as described with the Tetrahymena system cannot be
achieved. The DSNL evolution does not provide separation from the sign \(X\) \((Co^{2+})\) and
the object $O$ ($\text{Mg}^{2+}$). After eight rounds of selection the population exhibits only an adaptive response to the presence of the substrate and independent of the sign $X$.

3.4 Conclusion

These studies provide a basis for the marriage between empirical findings and the philosophical studies of biosemiotics. As well, they provide a basis for the minimal criteria for an interpretative response and suggest that interpretation is evolvable. The studies hold relevance to the origins of life showing a minimal system capable of carrying out a choice based on the environment. We have shown this phenomenon is not an intrinsic characteristic to all RNAs. The *Tetrahymena* ribozyme recognizes the presence of the object through the sensing of a sign. The selective advantage of this interpretative responsiveness to the environment could be adjusted if there was selection for the benefit of a correct interpretation ($O_1$), reduced disadvantage if there is a misinterpretation ($O_2$), or the emergence of responsiveness to a sign with a higher correlation with the favorable environment. The data presented here could be used as a springboard to evolve a more interpretative ribozyme with maximum payoff in $O_1$ and $O_4$. Future investigations would benefit from in depth kinetic investigations of the system as well as investigations under varying environmental conditions, mainly ranging pH and temperature, in order to determine if the satisfaction of the inequities persists. These empirical finding shed light on the origin of all forms of behavior, and a historical transition that was more-or-less coincident with the origins of life. In this work a theoretical framework for the evolution of behavior, based in quantum mechanical variations and leading to biological traits, has been provided. Using this framework we discussed the ability of an RNA to use
molecular variations to behave specifically to and dependent on environmental interactions. While finally in this chapter we challenged a simple biological polymer to make a choice. Firmly grounded in a marriage of philosophy and empirical investigation, this work provides a starting point on a roadmap of future work approaching an ageless question from a unique perspective.
Figure 3.1 Example payoff matrix for interpretive behavior. Payoff values O1–O4 are evaluated for each pair-wise combination of environmental conditions and genotype traits as discussed in the text. The ion concentrations refer to those used in the assays of the *Tetrahymena* ribozyme, as described in Figures 2–4.
**Figure 3.2** 5-error mutant Tetrahymena ribozyme activity assay. **A**) Gel image of reaction products electrophoresed on a denaturing a 5% poly-acrylamide gel. Reaction conditions for O1-O4 are detailed in the methods. **B**) Payoff matrix with total percentage reacted for each condition. Percentage reacted was calculated by dividing the quantitated product divided by the quantitated ribozyme plus the product. **C**) The percentage yield ratios of O1:O2, O1:O3, and O4:O2 are all positive and follow the minimum criteria for interpretative behavior in a 2-environment 2-state-entity system.
Figure 3.3 Activity comparisons for wild-type and 5-error mutant Tetrahymena ribozymes. A) Graphical comparison of the payoff matrix inequalities for the wild-type and 5-error mutant ribozymes calculated from total percentage reacted. B) The fraction reacted with and without Ca\(^{2+}\) (at 10mM) for the wild-type ribozyme for a range of [Mg\(^{2+}\)] (at 0.5 mM to 25mM). C) The fraction reacted with and without Ca\(^{2+}\) (at 10mM) for the 5-error mutant ribozyme for a range of [Mg\(^{2+}\)] (at 0.5 mM to 25mM). The endpoints of the lines represent the assay conditions from Figure 2 and are labeled to match their respective payoff matrix quadrants.
Figure 3.4 The DSNL ribozyme. A. Sequence and secondary structure of the starting DSNL-0 ribozymes. Arrow represents 3’-5’ linkage with substrate at the 5’ end and ribozyme at the 3’ end. B. Metal and pH dependence of the initial DSNL construct. The ribozyme is minimally active in Co$^{2+}$. 
Figure 3.5 In vitro evolution of the DSNL ribozyme. The ligated products (1) were purified by PAGE and eluted to prevent amplification of selection artifacts (2). The eluted products were subjected to reverse transcription (3), PCR amplified (4), and transcribed to produce progeny RNAs (5). The combined processes of ligation, purification, DNA amplification, and transcription comprise a complete selection generation.
Figure 3.6 After eight generation of evolution the resulting populations predominately expressed a G4A of G3A mutation. The mutations (located in green box) occur need the site of reaction (shown by the arrow) and are thought to allow for flexibility in active site structure.
References


45. Voytek SB and Joyce GF. Emergence of a fast-reacting ribozyme that is capable of undergoing continuous evolution. *PNAS,* **2007**, *104*, 15288-15293