Spring 5-10-2016

The Foundations of Network Dynamics in an RNA Recombinase System

Jessica Anne Mellor Yeates
Portland State University

Recommended Citation

10.15760/etd.2915
The Foundations of Network Dynamics in an RNA Recombinase System

by

Jessica Anne Mellor Yeates

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

Doctor of Philosophy
in
Chemistry

Dissertation Committee:
Niles Lehman, Chair
John J. Perona
Albert S. Benight
Martin Zwick

Portland State University
2016
Abstract

How life originated from physical and chemical processes is one of the great questions still unanswered today. Studies towards this effort have transitioned from the notion of a single self-replicating entity to the idea that a network of interacting molecules made this initial biological leap. In order to understand the chemical kinetic and thermodynamic mechanisms that could engender pre-life type networks we present an empirical characterization of a network of RNA recombinase molecules. We begin with 1-, 2-, and 3-molecular ensembles and provide a game theoretic analysis to describe the frequency dependent dynamics of competing and cooperating RNA genotypes. This is then extended to 4- and 5-membered networks where varying topologies are compared and mechanisms that could lead to preferential growth and selection of genotypes are described. At the core of these network connections is ribozyme catalysis initiated through a 3-nucleotide base-pairing interface. With the development of a fluorescence anisotropy method, we are able to illustrate a correlation between these binding thermodynamics and network outcomes. Finally, we consider how the heterogeneity of the environment could impact network dynamics and develop a spectrum of spatial inducing methods in which our chemical populations can be probed. These experiments illustrate simple chemical dynamics of RNA interactions, yet these very processes are the foundation for building complexity and ultimately from where selection and evolvability derive.
Dedication

To God, my ultimate foundation in life.

My parents, Jim and Ruth Mellor, for providing me every opportunity with love, support, and guidance

My husband, Jeffrey Yeates, my eternal companion and best friend.
Acknowledgements

I would first and foremost like to thank my advisor, Dr. Niles Lehman for his incredibly supportive and wise mentorship. His guidance through this scientific endeavor has truly transformed my ability to consider and question the world around me. From the origins of life to grammatical correction he has always provided encouraging and thought provoking direction.

To my committee members, Drs. John J. Perona, Albert S. Benight, and Martin Zwick, a huge thank you for your time commitment, suggestions, corrections, collaborations, and conversations.

The work represented in this dissertation was certainly not done alone. I owe much gratitude to the many collaborators with whom I have had the privilege to work. Dr. Philippe Nghe, Dr. Shigeyoshi Matsumura, Harry Kemble, and Dr. Andrew Griffiths at ESPCI Paris Tech for hosting me in their laboratory to work with their microfluidics devices and for their collaboration with network experiments. Dr. Martin Nowak and Dr. Christian Hilbe at Harvard University for their mathematical expertise and collaboration on chemical game theory. Last, but not least a huge thank you to my undergraduate colleagues, Naomi Iverson for her work on folding heterogeneity in the recombinase ribozyme and Elizabeth Satterwhite for her work with fluorescence anisotropy.

A decade and a half in the laboratory has provided innumerous mentors and colleagues that have given more encouragement and support than I will ever
have time to fully thank. But to the past and present Lehman lab members, Portland State colleagues and friends, industrial co-workers, Missouri State fellow students, and my original stockroom lab partner Corrie Sifers, your pub outing conversations have been some of the most scientifically motivating times of my life.

Finally, to my family, this work would not have occurred without you. Mom and Dad you are the original mentors of my life and your example, wisdom, and love have made me who I am today. My sisters Christi Annin and Katie Shuck, you are the original scientific spark in my life and I will always thank you for being not just a sister but my partner in crime. To my husband and babies, you are my inspiration, my haven, and my oasis. Together.
Table of Contents

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

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

Acknowledgements ....................................................................................................... iii

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

List of Figures ............................................................................................................... vii

Chapter 1
Introduction .................................................................................................................... 1

Chapter 2
Chemical Game Theory ............................................................................................... 10

Chapter 3
Recombinase Networks ............................................................................................... 55

Chapter 4
Fluorescence Anisotropy for the Investigation of IGS and tag Binding Interactions in the Recombinase Ribozyme ......................................................................................... 79

Chapter 5
Spatial Heterogeneity in the Recombinase System .................................................... 103

References ..................................................................................................................... 136

Appendix
Mathematical Modeling ............................................................................................... 147
List of Tables

Table 1. Rate constants, $k_a$ (min$^{-1}$), for the 16 genotype variations of $\text{GMG}_\text{WX Y}_\text{CNU}$ .......................................................................................................................... 38

Table 2. Kruskal-Wallis test ............................................................................................................. 43

Table 3. Rate constants, predicted $\Delta G^\circ_{48}$, Dissociation rate constants ($K_D$), and, experimental $\Delta G^\circ_{48}$ binding for recombinase interactions................................................. 94
List of Figures

Figure 1. Recombinase ribozyme from *Azoarcus* group I intron. .......................... 9
Figure 2. Evolutionary game theory feedback loop. .................................................. 34
Figure 3. Payoff matrix and selection dynamics. ....................................................... 35
Figure 4. 2-piece recombinase reaction, $\text{GMG} \text{WXY}_{\text{CNU}} + Z$. .................... 36
Figure 5. IGS-tag sequence interaction. ................................................................. 37
Figure 6. Graph for the determination of autocatalytic rate constants. ................. 39
Figure 7. Use of IGS and tag nucleotides to predict 2-strategy outcomes. ....... 40
Figure 8. Schematic and example gel electrophoretic for payoff matrix quantification. .................................................................................................................. 41
Figure 9. Empirical 2x2 matrices for four broad categories of games. ............... 42
Figure 10. Schematic for serial dilution. ................................................................. 44
Figure 11. Serial dilution results for four categories of 2-strategy competitions. 45
Figure 12. 2-strategy interactions for the 3-strategy RPS scenario. .................... 46
Figure 13. Rock-scissors-paper 3-strategy game. .................................................. 47
Figure 14. Comparison of empirical game outcomes with ODE model outcomes. .................................................................................................................. 48
Figure 15. Excluded 2-strategy outcomes due to equalities. ......................... 50
Figure 16. 72 2-strategy results from predicted payoff matrices. .................... 51
Figure 17. Comparison of replicator dynamics with the chemical game dynamics. .................................................................................................................. 53
Figure 18. Altering genotype activity with environmental perturbations. ....... 54
Figure 19. Comparison of 5-membered overall network rates calculated from experimental rate constants vs. simplified Watson-Crick topology. ............... 73

Figure 20. Slow vs. Fast 5-membered networks. ........................................ 74

Figure 21. Determination of best giver/receiver in a 5-membered network. ...... 75

Figure 22. Addition of links and removal of nodes in a 5-membered network. .. 76

Figure 23. Competing cores - comparison of 4-membered networks. ............ 77

Figure 24. Native gels of the 16 genotypes of WXY. .................................. 93

Figure 25. FA ligand and receptor design. .................................................. 95

Figure 26. UA binding curve. ................................................................. 96

Figure 27. AC binding curve. ................................................................. 97

Figure 28. CA binding curve. ................................................................. 98

Figure 29. AA binding curve. ................................................................. 99

Figure 30. IGS-less binding curve. ......................................................... 100

Figure 31. GA and GC binding curves. .................................................... 101

Figure 32. “GGG” IGS aggregation evidence. ......................................... 102

Figure 33. 2-piece recombinase reaction $GUG_{WX_{CAU}} + YZ$. .................. 122

Figure 34. Cooperative cycle recombinase reaction. ................................. 123

Figure 35. 2-piece recombinase reaction $GUG_{WX_{CAU}} + YZ$ in aqueous, PEG, Ficoll, and Dextran environments. ........................................ 124

Figure 36. 4-piece recombinase reaction $GUG_{W_{CAU}} + X_{CAU} + Y_{CAU} + Z$ in aqueous, PEG, and Ficoll environments. .................................. 125

Figure 37. 4-piece recombinase reaction in varying concentrations of PEG. .. 126
Figure 38. Cooperative cycle recombinase reaction in varying concentrations of PEG. ................................................................. 127

Figure 39. Diffusion of radiolabeled RNA fragments on 0.2% agarose plates. 128

Figure 40. 2-piece recombinase reaction on a 2D agarose surface. .......... 129

Figure 41. 4-piece recombinase reaction on a 2D agarose surface. .......... 130

Figure 42. Cooperative cycle recombinase reaction on a 2D agarose surface. ........................................................................................................................................................................ 131

Figure 43. 2-piece recombinase reaction in artificial protocells. .............. 132

Figure 44. Cooperative cycle recombinase reaction in artificial protocells. ...... 133

Figure 45. Cooperative cycle recombinase reaction in artificial protocells with alternative encapsulation. ........................................................................................................ 134

Figure 46. Compartmentalization using droplet based microfluidics system. .. 135
Chapter 1  Introduction

The foundations for life, as we understand it on our planet, began some 4.5 billion years ago with the formation of the Earth (Plaxco & Gross, 2011) and by 4.1 Ga evidence for the presence of life existed (Bell et al., 2015). Within that time frame, physical and chemical processes would have produced increasingly more complex interactions, moving from simple inorganic molecules to biopolymers capable of replication and variation. In order to answer the question of how life originated and to even understand what life is, empirical proof-of-concept simple abiotic pathways demonstrating these transitions are needed.

The transition from simple inorganics to biological precursors (i.e., sugars, amino acids, nucleic acids, and lipids) has been recorded in the literature for over a century. The most famous of these experiments is the Miller-Urey experiment, in which amino acids were synthesized from ammonia, water, methane, and hydrogen in a “prebiotic” environment (Miller, 1953). Formaldehyde was demonstrated to form sugars in the autocatalytic formose reaction (Butlerow, 1861) and hydrogen cyanide can be used to synthesize a pyrimidine nucleobase (Oro, 1961). A route to abiotic purine nucleobase synthesis had eluded scientists for some time, yet once produced, a nucleotide in a semi-activated form with a 2’,3’-cyclic phosphate emerged (Powner et al., 2009). Lipids, being another
important category of biological molecules, can be synthesized through the reduction of sugars catalyzed by metal or mineral surfaces (McCollom et al., 1999; Rushdi & Simoneity, 2001), or through a primitive fatty acid synthesis (Weber, 1991), and can be delivered extraterrestrially (Sanford, 1996; Ehrenfreund & Chournley, 2000).

It should be noted that the above processes typically represent low yield, multi-component, racemic products, and are far from giving the highly specific building blocks of modern day biology. Yet, many of these biological building blocks have been detected on meteorites (Cronin & Chang, 1993) indicating they are more than plausible products. Additional studies have demonstrated pathways to increased specificity and stereoselectivity, for example through the use of borates (Ricardo et al., 2004), lipids (Sacerdote & Szostak, 2005), and phosphates (Mueller et al., 1990). Recently a flow chart of synthesis reactions demonstrated how all major classes of biological precursors could have been interconnected and produced (Patel et al., 2015). This diversity and cooperation of molecules and reactions truly invokes the imagination when considering how chemistry can spontaneously build complexity.

While abiotic synthesis of biological precursors and monomers represents one challenge for nascent life, condensation of these units into functional polymerized biomolecules represents a second challenge. Again, we see precursors
harnessing the functionality of the environment and/or other molecules. Abiotic oligomerization has been demonstrated using clays (Ferris et al., 1996), templates (nucleic acids, Kozlov & Orgel, 2000, proteins, Lee et al., 1996), and lipids (Rajamani et al., 2008). These processes produce in the lab, at most, the synthesis of an oligomer approximately 50 subunits long (i.e., a “50 mer”). For RNA, this length is actually sufficient for some small or minimal ribozymes (Ferre-D’Amare & Scott, 2010) and recently only 22 nucleotides (nt) have been proposed to lead to the evolution of the ribosome (Petrov et al., 2014). Yet, a path from functionality to an evolving, self-replicating chemical system is still very illusive.

There has been vast progress in making longer oligomers through classic polymerization. A ribozyme was in vitro evolved to elongate an RNA strand to over 200 nt (Attwater et al., 2013). While this length is impressive and exceeds the length of the ribozyme itself, the product does not represent a self-replicating sequence. An adjacent line of research uses the premise that collections of interdependent molecular species – or chemical networks – may have been a more robust pathway to life. This concept is not new; as early as the 1920’s Oparin suggested a bag of interacting organic chemicals (coacervates) could have lead to the first life forms. Even the aforementioned work on monomer formation and abiotic polymerization demonstrates the usefulness of multiple molecular species interacting (e.g., template interactions). Theoretical work on
chemical networks has been of great interest over the last several decades. Eigen and Schuster proposed the hypercycle and quasispecies (Eigen, 1971; Eigen & Schuster, 1977; 1978), Gánti introduced the chemotron (Gánti, 2003), and Kaufmann’s collectively autocatalytic set (CAS) (Kauffman, 1993) have been very influential ideas in the field. Yet an empirical approach along this pathway has been much more difficult to implement. This could be due to the traditional reductionist philosophy within chemistry or a lack of technology/methodology for multi component interactions (Ludlow & Otto, 2008). Regardless, basic cellular life consists of vast interconnected networks that produce emergent phenomena not easily predicted by the molecular species treated in isolation. An empirical understanding of how simple molecular networks develop and evolve is now needed to add critical insight to the non-life-to-life pathway.

It is within this niche that the following thesis work is produced. Here, the dynamics of multiple interacting sequences of short RNA oligomers are characterized. Through interactions with each other and/or the environment, outcomes are manifested that would not necessarily be predicted from reduction to chemical dynamics of single entities. Through a systematic study (of simple two-molecule competitions to increasing network interactions) this work can add to the tools and underlying concepts for the empirical study of chemical networks and their evolution.
The system that is used to study these chemical evolutionary dynamics is based on past work with the Azoarcus group I intron (Hayden & Lehman, 2006; Hayden et al., 2008; Vaidya et al., 2012). In vivo, this ribozyme splices itself out of a pre-tRNA\textsuperscript{Ile} sequence with catalysis relying on binding of a three-nucleotide internal guide sequence (IGS) with a tag sequence (tag) (Zaug et al., 1993). Previously, it was determined that this ribozyme can be fragmented into shorter oligomers, which can assemble non-covalently to restore catalytic activity (Fig. 1). By placing tag sequences at loop regions that are fragmented, these non-covalent complexes can then be catalyzed by other non-covalent complexes, through a recombination reaction, into fully functional covalent ribozymes in a reaction akin to the reverse in vivo function (Draper et al., 2008).

This self-assembling ribozyme system is particularly relevant for evolutionary dynamics studies for several reasons. First is the molecule itself, RNA. RNA has the unique attributes of being able to both carry information and have catalytic ability. The information carrying capacity is found at the global level of non-covalent interactions that form specific secondary and tertiary structures and at the local level of IGS/tag recognition, where the three nucleotides base pair and provide catalytic activity. The dual information carrying and catalytic nature of RNA is part of a strong body of research supporting an archaic RNA world that is thought to be a step between prebiotic chemistry and modern life (Joyce, 2002; Neveu et al., 2013). While there is evidence for metabolism (Branscomb &
Russell, 2013), membranes (Deamer & Weber, 2010), and peptides (Bowman et al., 2015) ushering in life, it seems most certain that RNA molecules were central to its origins.

A more subtle aspect of the self-assembling ribozyme that imbues this molecule with prebiotic-like qualities is the actual chemistry that it uses for catalysis. Unlike modern polymerization, which requires high energy activated nucleotides to elongate strands one monomer at a time, the Azoarcus ribozyme uses recombination. More specifically this is a net neutral free energy, phospho-ester transfer reaction, which links short oligomers together with the simultaneous loss of one or more nucleotides. Three catalytic events can take ca. 50-mer fragments to over 200 nt (Fig. 1). Recombination was one of the original chemical tools for building complexity in the universe (Gilbert, 1986) and in modern times is found across all phyla assisting in genotypic and phenotypic retention and variation (Lehman et al., 2010).

Lastly, because the product of this Azoarcus recombinase ribozyme is a second recombinase ribozyme, the reaction is autocatalytic (Hayden et al. 2008). This reaction does not represent self-replication in a Darwinian sense; this would necessitate exponential growth of replicators (Szathmáry & Gladkih, 1989). Autocatalytic systems, instead, represent a minimal self-replication (von Kiedrowski, 1986; Tjivikua et al., 1990). They are central to all origins-of-life
theories and represent a critical mechanism to move from prebiotic chemistry to self-replicating systems (Eigen, 1971; von Kiedrowski, 1986; Kauffman, 1993; Szathmáry, 2006). Within our system, the autocatalytic ability can be viewed as a pre-life selfish production of more of itself. Further though, this autocatalytic ability can be manipulated by varying the recombinase sequence to produce cross-catalytic or cooperative productions of other sequences. The sum of both the autocatalytic and cross-catalytic activity of the recombinase allows it to form large complex networks of interacting species with feedback capabilities (Vaidya et al., 2012, Vaidya et al., 2013).

To understand the development of molecular networks is to understand the kinetic and thermodynamic forces at play. Ultimately it is kinetic selection that leads to the surviving or “winning” molecules in a prebiotic world (Pross, 2009, 2013). This concept is exactly what I look to explore empirically in this dissertation. Chapter 2 begins with the introduction of evolutionary game theory (EGT) to chemical systems. EGT allows characterization of simple two or three molecular species interactions. Chapter 3 evaluates small networks of interacting recombinases and how the addition of network members (or nodes) and differing chemical connections (or links) affect an overall network’s “fitness”. In Chapter 4, I discuss the development of a fluorescence anisotropy (FA) method for quantitating the binding energy of the recombinase interactions. The binding interactions between the IGS and tag sequences represent a thermodynamic
mechanism for the molecular competitions and network formations in chapters 2 and 3. Finally in Chapter 5, I introduce the idea of spatial heterogeneity and encapsulation to our system. The ability to evaluate the recombinase ribozyme activity with increasing spatial deviation is developed and the potential influence these physical aspects have to affect the interactive landscape is considered. The bottom-up characterization of binding thermodynamics and 1- and 2-entity kinetics paired with the top-down characterizations of network topologies and environment influences lays a foundation for understanding how an RNA system is able to develop ever-increasing complexity through network formations.
Figure 1. Recombinase ribozyme from *Azoarcus* group I intron. (A) secondary structure illustrating the division of the ribozyme into four fragments, GUG\_W\_CAU, X\_CAU, Y\_CAU, and Z. The IGS (red) is located on the 5' end of the W fragment, tag sequences (orange) are located on the 3' end of fragments, and the junctions where catalytic closure occurs are outlined in orange. (B) Reaction scheme of the recombinase ribozyme. W, X, Y, and Z first associate via base pair interactions to form a non-covalent WXYZ molecule. The non-covalent WXYZ is then catalytically connected through three separate reactions at the W-X junction, X-Y junction, and Y-Z junction by other molecules of non-covalently or covalently linked WXYZ recombinases (color scheme is the same as A). (C) A zoomed-in view of the phosphor-ester transfer reaction connecting the Y-Z junction.
Introduction.

Game theory is a mathematical framework that helps predict outcomes between two or more entities choosing strategies within a competition. Developed in the 1940’s and 50’s, game theory has been widely used to analyze economic, military, and public health phenomena, just to name a few (von Neumann & Morgenstein, 1944; Nash 1950). In the 1970’s, John Maynard Smith and George Price applied game theoretic concepts to Darwinian competition (Maynard Smith & Price, 1973; Maynard Smith & Price 1982). Evolutionary game theory, as it was termed, removes the emphasis from rationally and intentionally chosen strategies and instead uses fixed strategies. The approach analyzes the subsequent frequency changes of those strategies within a population of competing “players”.

The basic idea of evolutionary game theory (EGT), as applied to biological systems, models the rules of natural selection and describes the population dynamics of two or more strategies (or phenotypes) competing (Page & Nowak, 2002). The outcome of a competition (i.e., who wins, loses, or coexists) is ultimately determined by the fitness of a phenotype, quantitated in a payoff matrix, and the frequency of a phenotype (Fig. 2). This mathematical analysis has provided explanations of biological phenomena that were not otherwise
accessible, for example the establishment of cooperative and altruistic traits within populations (Nowak et al., 2004; Nowak, 2006). Biochemical pathways in organisms (Pfeiffer et al., 2001), as well as collections of yeast (Dyken et al., 2013), bacteria (Kerr et al., 2002), and viruses (Turner & Chao, 1999) have been shown to operate in patterns predictable from game theoretic scenarios.

Recently, optimization principles from EGT are even being applied to population dynamics of cancer cells (Liao & Tlsty, 2014a; 2014b).

The fundamental mathematical equation of evolutionary game theory is the replicator equation:

$$\dot{x}_i = x_i [f_i (\bar{x}) - \phi(\bar{x})]$$  \hspace{1cm} [1]

Here, \(n\) phenotypes are present in a population, \(x_i\) is the frequency of the \(i^{th}\) type and \(\dot{x}_i\) is time derivative frequency of \(i\). The term \(f_i (\bar{x})\) is the fitness of \(i\), where \(f_i (\bar{x}) = \sum_{j=1}^{n} a_{ij} x_j\). The coefficients of this function are the entries in a payoff matrix \(A = [a_{ij}]\) and \(a_{ij}\) denotes the payoff for strategy \(i\) interacting with strategy \(j\). The function \(\phi(\bar{x})\) is the average fitness of the population. So, if \(f_i (\bar{x})\) is greater than \(\phi(\bar{x})\) then the frequency of \(i\) will undergo positive selection. Note, however, a central point to equation [1], the growth of type \(i\) is frequency dependent. The fitnesses in term \(f_i (\bar{x})\) are linked to the frequency of both \(i\) and each of the \(n\) phenotypes. Thus how one type grows over time is linked to how common that type is in the population, and the dynamics of each member of the
population is dependent on the network of interacting types, not just one type alone in a system (Nowak, 2006).

A common illustration of game theoretic models is the payoff matrix. This matrix, central to the replicator fitness function, can be written as an inequality and describe population outcomes or equilibriums (Fig. 3). Within a 2-player game, a 2 x 2 payoff matrix produces 24 possible strict ordinal rankings of the quadrant values (e.g., \(a > b > c > d\)). In biological scenarios this can be categorized into four broad categories that have evolutionary significance, strategy 1 dominates, strategy 2 dominates, coexistence, or bistability (Nowak, 2006). Note that a “strategy” in biological scenarios refers to a phenotype and payoff values produce a fitness term (Smith & Price, 1982; Sigmund, 2010). If \(a > c\) and \(b > d\), strategy 1 will dominate, indicating if the population is allowed to reach equilibrium strategy 1 will be the only remaining strategy. If \(c > a\) and \(d > b\), strategy 2 will dominate. A special case of this outcome is the strict ordinal ranking \(c > a > d > b\), representing the Prisoner’s Dilemma (PD) scenario. Here, while one strategy is expected to dominate in a single iteration of the game, the overall group payoff would be higher if the alternative strategy were iteratively invoked. This contrasts with a dominance ordinal ranking, \(a > c > b > d\), where the strategy that dominates will produce the highest payoff in a single iteration as well as for the group in an iterative game (also known as a Harmony scenario). A population will attain a coexistence of strategies if \(c > a\) and \(b > d\). Lastly, if \(a > c\) and \(d > b\) a
bistable equilibrium will result, which indicates no stable equilibrium exists for both strategies and either strategy 1 or 2 result.

While an empirical game theoretic demonstration in a purely chemical system is so far lacking (Hart, 2005), the concept has been alluded to in the literature (Bohl et al., 2014). With the finding of molecular cooperative and selfish behaviors (Vaidya et al., 2012) the idea was sparked that game theory analysis could be empirically relevant analysis for the study of competition within molecular populations. EGT offers not only a framework to characterize the population frequency dynamics in multiple species interactions but also a potential bridge between biological and chemical phenomena.

For our chemical game theoretic analysis, we are using the previously described self-assembling, autocatalytic, Azoarcus recombinase ribozyme (Hayden & Lehman, 2006; Hayden et al., 2008). This ca. 200 nucleotide (nt) catalytic RNA (found in nature as the Azoarcus pre-tRNA\textsuperscript{lle} group I intron (Reinhold-Hurek & Shub, 1992)) can be artificially fragmented into two, three, or four fragments, which can spontaneously self-assemble into a functional non-covalently bound recombinase. This recombinase complex can then catalyze the covalent assembly of a second non-covalent recombinase. Catalytic recombination is initiated through a 3-nt base-pairing interaction between an internal guide
sequence (IGS) on the 5’ end of one recombinase and the tag sequence (tag) on the 3’ end of a fragment on a second recombinase (Fig. 1). Importantly, changing the identity of the nucleotide triplets in the IGS/tag interaction will alter the specificity with which the RNAs react with one another (Draper et al., 2008).

For simplicity, we focused on the 2-piece assembly reaction, which can be symbolized as WXY + Z → WXYZ (Fig. 4). We allowed 4-fold variation in the middle nucleotides of the IGS and tag (denoted M and N, respectively) to confer a range of recognition interactions (Fig. 5). This design maintains the 3’ G-U wobble pair, canonical to group I introns (Damberger & Gutell, 1994) and a G-C Watson-Crick base pair on the 5’ side of the IGS/tag interaction. In order to apply a game theoretic analysis to our chemical system, we define three variables: what constitutes a player, what is a strategy or phenotype, and how is a molecular fitness quantitated (entries into the payoff matrix). The mathematical framework should also be considered, a chemical system, albeit a primitive replicator, does not produce classical Darwinian replicator dynamics such as those described in equation [1]. Chemical game theory instead needs to either consider a “prevolutionary” dynamics model, as in equation [2] below, or some hybrid of equations [1] and [2].

\[ \dot{x}_i = a_{ii}x_i - (d + a_{io} + a_{ii})x_i \quad [2] \]

A prevolutionary dynamics model, like in equation [2] incorporates the ideas of selection and mutation, but not replication (Nowak & Ohtsuki, 2008). Here, \( \dot{x}_i \) is
time derivative frequency of \(i\), and \(a_i\) is the rate constant of the chemical reaction from \(i'\) to \(i\). The term \((d + a_{i0} + a_{i1})x_i\) represents a decay or death term.

Using the above recombinase system and mathematical framework, we present a characterization of chemical players, strategies and payoff matrices. We then demonstrate the ability of these molecules to interact and compete in scenarios describable by a game theoretic analysis. Lastly, we show the ability to predict chemical population outcomes using this analysis.
Results.

Defining players, strategies, and payoffs. To characterize molecular players, strategies, and payoffs, we considered the biological definitions of these concepts as well as the mathematical requirements as defined in game theoretic models. In a biological setting, the player is an individual that uses a strategy defined as the phenotype (Smith & Price, 1982; Sigmund, 2010). Strategies in a game at the chemical level, or molecular reproduction strategies, would be the catalytic phenotypes displayed by a genotype, in our case the genotype is the primary nucleic acid sequence structure and the phenotype is its ability to covalently recombine WXYZ ribozymes.

Within the recombinase system, phenotypic variation is derived in the IGS and tag sequences of the WXY molecule (Fig. 5) (cf. Bohl et al., 2014). It is the recognition between these two triplets that initiates the recombination reaction and therefore what characterizes the catalytic strategy in our system. With a 4-fold variation in the middle nucleotide of the IGS and tag sequences, where the IGS is G<sub>M</sub>G and the tag is C<sub>N</sub>U, (denoted as “MN” in Table 1), 16 possible molecular strategies can be created to compete against and among each other to use the shared resource Z molecule to produce covalently-contiguous WXYZ ribozymes. These chemical phenotypes can further manifest as selfish autocatalytic strategies of like-genotypes and cooperative cross-catalysis of other genotypes (Vaidya et al. 2012)
It is thus this competition for covalently bound WXYZ that portends a payoff for the phenotype. The covalently bound WXYZ represents a successful genotype in an evolutionary scenario, contributing a roughly 2-fold higher catalytic activity than a non-covalent complex of WXY-Z, which is held together only by hydrogen bonds (Vaidya et al., 2012). Within replicator (Nowak, 2006) or prevolutionary (Nowak & Ohtsuki, 2008) mathematical models, payoff is in terms of a rate constant. For the Azoarcus recombinase we have opted to use the autocatalytic rate constant $k_a$ as the payoff, as this parameter best represents the differential dynamics of competing recombinase RNAs.

*Payoff matrix quantitation.* To quantitate payoff matrices for a simple 2-player interaction we first measured the autocatalytic rate constants of our 16 strategies in isolation (Hayden et al. 2008). As expected when the variable middle nucleotide of the IGS and tag are Watson-Crick pairs, much higher rates of catalysis occur, but all possible pairings allow some degree of catalysis (Table 1, Fig. 6). The characterization of the 16 individual catalytic abilities of our recombinases provides two important aspects to the game theoretic analysis. First, the data corresponds to the diagonal values in the payoff matrix, *i.e.* the payoff a player gets when interacting with its own genotype: $a$ or $d$. Second, the 16 IGS/tag individual catalytic capabilities offer a rudimentary prediction for the
off-diagonal terms in the payoff matrix, or how the IGS of one genotype cross-
catalyzes the tag of a second genotype: \( b \) and \( c \) (Fig. 7).

With the assumption that at the chemical level no two quadrants will be exactly
the same – and with diverse variation in our phenotype strategies – we should
have the chemical capability to experimentally realize any of the 24 possible
payoff matrix inequalities. Additionally we can assume that \( a > d \) without loss of
generality (otherwise one only needs to re-label the strategies), lowering the
number of possible outcomes to 12. Using our predicted matrices from the 16
individual kinetic tests, we initially choose four 2-player (or 2-genotype)
interactions to further measure the \( b \) and \( c \) quadrants of the payoff matrix. These
2-genotype interactions are predicted to result in each of the four broad
categories as illustrated in Figure 3, 1 dominates, 2 dominates (since we assume
\( a > d \), in this case we are using the predicted PD scenario, with the strict
inequality \( c > a > d > b \), as our second category), coexistence, and bistability.

In order to determine the rate constants for one genotype as acted upon by
another genotype, we used the same methodology as our individual kinetic
assays, except here the \(^{32}\text{P}\)-labeled low concentration \( \text{WXY}(< 0.003 \mu\text{M}) \) is one
genotype and the unlabeled high concentration \( \text{WXY}(1 \mu\text{M}) \) is of the other
genotype (Fig. 8). The labeling is then reversed in order to measure how the
other genotype acts on the first genotype. Having now all four experimental rate
constants within our matrix ($a$ and $d$ from measuring how each genotype acts on itself, and $b$ and $c$ quadrants containing the measurement of one genotype acting on the other) we indeed see that each of the four broad category inequalities are satisfied (Fig. 9). It should be noted here that the chemical payoff matrices are labeled with alternative game names from those illustrated in Figure 3. This was done to better describe chemical game dynamics as opposed to biological game dynamics and will be described further below. The Kruskal-Wallis non-parametric test was used to test the statistical significance of the rank order of the values in the 2 x 2 payoff matrices (Sokal & Rohlf, 1981; Meyer & Seaman, 2015). All rank order values for the games are statistically significant at the $P < 0.05$ value or better (Table 2).

Chemical game outcomes. From here we looked to understand what chemical outcomes would transpire in an evolutionary type setting. For a generation-like chemical setting we established a serial dilution framework in which two genotypes could compete in empirical game scenarios. Here, two WXY genotypes are mixed at some frequency, reacted for a brief period (corresponding to the kinetic linear phase of WXYZ production), and then a small fraction is carried to a new reaction vessel, which contains more raw materials (Mills et al., 1967; Vaidya et al., 2012; Wright & Joyce, 1997) (Fig. 10). We performed the serial dilution experiments using our pairs of WXY RNAs that demonstrated the four categories of payoff matrix inequalities (Fig. 9). Each pair
was run at an equimolar ratio and additionally our *selfish* (or bistability) scenario (AU vs. UC, \( a > c \) and \( d > b \)) was performed at 20:80 and 05:95 molar ratios, in an attempt to understand chemical bistability.

It should be noted that while we look to correlate our games with the biological scenarios, there is a key dynamic difference between the two. Here, “extinction” is not possible due to fresh material of both genotypes being added at each burst or generation. While this precludes the ability to see only one strategy take over, it is prebiotically relevant in that it simulates a cyclically replenished pool or some other form of abiotically produced fragments.

Among our four chemical games we observe a diverse set of outcomes that appear to correlate well with the biological scenarios (Fig. 11). When CG is pitted against GA, an example of a *dominance* strategy (or the biological version of 1 dominates), CG dominates because it can autocatalyze its own genotype far better than cross-catalyze GA (\( a > c \)) and GA cross-catalyzes CG better than it autocatalyzes its own genotype (\( b > d \)) (Fig. 11A). In contrast to this we observe the CA versus GG game (*counter-dominance*, an example of the Prisoner's *dilemma*), the weaker genotype, GG, dominates despite its relatively poor autocatalytic rate constant (Fig. 11B). This counter-intuitive result emerges due to the cross-catalytic capability of CA acting on GG (\( c > a \)), GG autocatalysis exceeds its cross-catalytic ability towards CA (\( d > b \)) while CA autocatalysis
exceeds that of GG autocatalysis \((a > d)\). At the chemical level, the interaction of CA with GG is very strong; the middle nucleotide of the IGS in CA forms a Watson-Crick interaction with the middle nucleotide of the tag of GG. Thus the distinction between the “cooperator” (CA) and the parasitic player, or “defector” (GG) becomes clear, and the counter-intuitive PD result emerges (thus termed *counter-dominance* here). To our knowledge this is the first example of this PD-like phenomenon at the molecular level.

When AC versus UU was tested (the coexistence inequality, termed *cooperation* here) both genotypes persisted at high frequency \((> 40\%)\) stably over time (Fig. 11 C). Here the cross-catalytic rates are expected to be higher than their corresponding autocatalytic rates, AC acting on UU is higher than AC acting on AC \((c > a)\) and UU acting on AC is better than UU acting on itself \((b > d)\). These high cross-catalytic rates lead to a cooperative stable frequency of both genotypes. Due to the idea that all our scenarios chemically produce “coexistence” due to the continual replenishment of genotypes, we termed this scenario *cooperation* to better describe the chemical dynamics.

Lastly, the AU versus UC payoff matrix predicts an inverse of this scenario where autocatalysis is more effective than cross-catalysis, \(a > c\) and \(d > b\) (termed bistability in the biological setting, and *selfish* scenario here). At the equimolar ratio we expected and do see AU dominate (albeit at a lower frequency than the
dominance games) (Fig. 11D). This dynamic is again confirmed when the ratio of AU is greatly reduced to 20:80, AU:UC (Fig. 11E). We begin to see strategy prevalence change when we further reduce our genotype ratio (Fig. 11F), however, at burst 6 the strategies shift back to AU. This data could indicate a bistable-like scenario in that we see a shift in the genotype outcomes. The switch back to AU could be due to the continual replenishment of fragments and the strong catalytic ability of AU (both auto-catalytically and cross-catalytically). However, additional game scenarios would need to be evaluated to confirm or re-define this dynamic.

3-strategy chemical game. With the characterization of 2-strategy games, we next pushed our system to a more complex 3-strategy game. The Rock-Paper-Scissors (RPS) scenario illustrates a thoroughly studied 3-strategy model where a cyclical arrangement of dominance relationships between pairs of strategies should lead to a stable co-existence of all three strategies. Because this scenario has been demonstrated in lizards (Sinervo & Lively, 1996) and bacteria (Kerr et al., 2002), we were interested to see if it could also occur among molecular genotypes. Based on our expectations from the results of other 2-strategy games, we chose three WXY genotypes, AA, UC, and GU, that we anticipated could generate a game similar to the RPS. In isolation, we would predict (from individual 16-player kinetics (as illustrated in Fig. 7) and from measured 2 x 2 payoff matrices (Fig.12A-C)) that GU beats AA, AA beats UC, and UC beats GU.
All rank order values for the games are statistically significant at the \( P < 0.05 \) value or better (Table 2) (Sokal & Rohlf, 1981; Meyer & Seaman). Note that here the AA versus UC game is a predicted *cooperation scenario* not dominance, however, each player contains a cross-catalytic Watson-Crick interaction with the next player that should provide the cyclical dynamics as in an RPS game, this would however be a weak RPS scenario. When we competed these genotypes against each other two-at-a-time in a serial-dilution format, we indeed saw that one genotype reaches at least 70% superiority in each game (Fig. 12 D-F). When all three genotypes were pitted against each other in the same reaction vessel their joint frequencies quickly attained values near equifrequency (ca. 30–40%) and remained there (Fig. 13A). The genotype UC was the furthest outlier from the 33% equality, which may be attributed to the AA genotype not fully dominating UC (as this is a *cooperation scenario*, not a *dominance scenario*). Notably, however, we do see the steady-state frequencies of each genotype in the RPS-like scenario converge on an internal point in a Simplex plot (Fig. 13B).

*Mathematical modeling.* In parallel with the experimental results, we collaborated with Christian Hilbe and Martin Nowak at Harvard University to create ordinary differential equation (ODE) models of this system to visualize more clearly the dynamics of chemical games (see Appendix for mathematical model properties). A simple model was first compiled in which the frequencies of two players were tracked in a flow reactor setting that is a continuous analog of the serial dilution
experiments. In this model, the frequency changes of the two strategies over time 
($\dot{x}$ and $\dot{y}$) are described by:

$$\dot{x} = ax + by - \phi x \quad \text{and} \quad \dot{y} = cx + dy - \phi y$$ [3]

Here $a$, $b$, $c$, and $d$ are the values in the payoff matrices from Fig. 9 and 12. The 
death (or dilution) term, $\phi = (a + c)x + (b + d)y$, guarantees that $\dot{x} + \dot{y} = 0$ and 
$x + y = 1$. This parameterization is appropriate because the reaction rate is a 
linear function of RNA abundances. This mirrors our serial dilution format in that 
reactions were only allowed to occur for five minutes before transfer to ensure they remained in their exponential growth phase (Fig. 6).

The unique equilibrium values, $\hat{x}$ and $\hat{y}$, for each game are given by:

$$\hat{x} = \frac{a - 2b - d + \sqrt{(a - d)^2 + 4bc}}{2(a + c - b - d)}$$ [4]

and $\hat{y} = 1 - \hat{x}$. This model closely predicted the qualitative outcomes of all seven 
example 2-strategy games described above (the four broad categories and three 
additional games from the RPS-like scenario). Although it was unfeasible 
empirically to perform all possible 2-strategy games, the seven games that we 
did empirically test spanned a broad range of input payoff matrix values and their 
steady-state player frequencies matched those from the model to within a few 
percent in all cases (Fig. 14A and B). Such strong concordance of model and
data underscores the utility of the game-theoretic framework to forecast complex
dynamics from lower-dimensional data.

Along with the 2-strategy games, we completed the 3 x 3 payoff matrix to
compare the ODE model with experimental data (Fig. 13C). Here, the diagonal
terms were derived from single genotype kinetic data (Table 1) and the off-
diagonal terms were derived from the results of the three 2-strategy payoff
matrices (Fig. 12 A-C). Once again we find the simple ODE model described
above was able to predict correctly the game outcome. From this matrix, we
calculated that there should be a stable internal equilibrium point (cf. Broom &
Rychtar, 2013) consisting of 39% UC, 35% AA, and 26% GU (Fig. 13D). This
result agreed qualitatively with the empirical data in the 3-strategy serial dilution
experiment where burst 8 contained an equilibrium frequency of 45% UC, 30%
AA, and 25% GU.

In addition, while we did not empirically measure all 120 possible 2-strategy
games, we were able to forecast results for many of these games by estimating
the four values of the payoff matrix (estimations were performed as described in
Figure 7. By eliminating cases with equal payoffs (e.g., AU vs. AC; see Fig. 15),
there are 72 possible 2-strategy games. The predicted outcomes suggest that
26% of scenarios should lead to the dominance scenario, 10% to the counter-
dominance or PD game, 32% to cooperation, and 32% to the selfish scenario
(Fig. 16). Figure 14C illustrates the correlation of this purely predicted data with the experimental results and ODE model results from experimental payoff matrices. While some variation does exist with the quantitative frequency outcomes, qualitatively the purely predicted results provide a good concordance with the empirical data.
Discussion.

In this chapter, we have developed a chemical game theoretic analysis based on autocatalytic RNA. By changing single nucleotides of RNA strands, we generate chemical players with strategies that compete for production of their genotype’s autocatalytic product from precursor molecules. The chemical kinetics of a serial dilution experiment can be accurately described by a system of ordinary differential equations. At the center of this system is a payoff matrix whose entries specify the catalytic rate constants of the different players generating their genotypes or other genotypes from the provided material.

Given the concordance between our experimental results and the ODE models that were constructed, the game-theoretic analysis appears to be a natural consequence of chemical kinetics. By representing molecular interactions in the 2x2 matrix with simple rate constants, we are able to accurately predict what would happen in various competition scenarios. This qualitatively extends to the three-genotype interaction level, and may extend further. Although, as the complexity of the system increases, the unpredictable interactions such as non-productive binding events (Vaidya et al., 2012) will begin to play more of an important role. Prior efforts have focused on either simulation models of similar RNA network dynamics (e.g., Eigen & Schuster, 1977; Nowak & Ohtsuki, 2008; Bohl et al., 2014), or on broad-scale experimental data (e.g. Vaidya et al., 2012;
Kim & Joyce, 2004; Lincoln & Joyce, 2009). Here, we demonstrate that experiment and modeling can agree in a chemical game-theoretic context.

Differences between chemical and biological systems leads to some subtle but important distinctions between what we describe here and classical evolutionary game theory. Biological systems reproduce via template-directed polymerization, which can be characterized by the replicator equation [1]. In our chemical system, there is no reproduction per se, as the process is more of a de novo synthesis rather than a replacement, and could be described as “pre-life” (Nowak & Ohtsuki, 2008) in which the replicator equation does not apply. Thus in pre-life game dynamics the specific outcomes of inter-genotypic competitions can be predicted in a fashion parallel to, but not identical to, those used to calculate Nash equilibria (Nash, 1950) or evolutionary stable strategies (Smith, 1982) in biological games (see also Appendix and Fig. 17). Nevertheless we address here a key facet of understanding chemical evolution and the dynamics of frequency changes of molecular genotypes. By characterizing the strategies employed by molecules in creating more molecules we can begin to understand how chemistry could be optimized though kinetic selective forces (Pross, 2011)

Additionally our game theoretic analysis allows us the ability to rapidly assess the evolutionary outcomes of contests among prebiotic genotypes. This illustrates for us what general outcome can arise (i.e. dominance vs. coexistence) and
importantly, which specific mechanism is operational in any given situation (cross- and/or autocatalysis). An ODE analysis alone can predict our genotype outcomes but the specific mechanism for this may be obscured. This deeper understanding of the specific mechanism at play potentially provides us the ability to engineer the system to create alternative scenarios. Preliminary work has been performed that demonstrates the ability to change the relative activity of genotypes through perturbation of thermodynamic forces (changing Mg$^{2+}$ and temperature variables, Fig. 18A) and through the alteration of resources (changing Z concentrations, Fig. 18B). Use of the payoff matrix and game theoretic analysis provides a quick assessment of how and why system and environmental changes can affect overall outcomes.

Lastly, chemical game theory provides a foundational understanding of how prebiotic network dynamics develop. While it has previously been shown that the Azoarcus RNA fragments spontaneously form complex interaction networks that change over time (Vaidya et al., 2012), the underlying step-by-step mechanisms of this process, or for any pre-organismal change in genotype frequencies, are not known. It is these simple one, two, and three entity interactions that define the local environment of a multi-species interaction system and thus the characterization of these connections provides a basis for an ability to describe the build-up of network dynamics (as discussed further in Chapter 3). Such interactions, driven mainly by non-covalent bonding strengths (as discussed in
Chapter 4), could then be used to understand the build-up of network complexity among prebiotic polymers that has been observed in other RNAs (Mutschler et al., 2015), proteins (Ashkenasy et al., 2004), and lipids (Segre et al., 2000). Thus a game-theoretic approach gives us an appreciation of how a chemical ecology could develop and evolve prior to biological replication and perhaps offer a bridge between simple chemical mechanisms and complex biological-like networks.
Materials and Methods.

RNA preparation

The WXY and WXYZ molecules, which are portions of the self-splicing group I intron from the isoleucine pre-tRNA in the purple bacterium *Azoarcus* (Reinhold-Hurek & Shub, 1992), were prepared by *in-vitro* transcription from DNA plasmid templates. The Z fragment was purchased from TriLink Biotechnologies (San Diego, CA) and was gel purified prior to use. All RNAs were resuspended in 1–10 μM solutions in 0.1 mM EDTA. For quantification, < 0.1 μM of the WXY-fragment was 5'-labeled with γ[^32]P•ATP using OptiKinase (USB, Cleveland, OH).

Self-assembly kinetics (Quadrants a and d of the payoff matrix)

See Figure X-5 and X-7 for a schematic of this process. Reaction mixtures containing WXY (1 μM), Z (1 μM), WXYZ (0 – 2 μM), and ^32^-labeled WXY (≤ 0.003 μM), all of the same IGS and tag genotype, were heated to 80 °C for 2 minutes then cooled to 48 °C. Time “zero” aliquots were drawn and quenched with equivolume quench solution (125 mM EDTA and 2X loading dye containing formamide and bromophenol blue). Reactions were initiated with the addition of reaction buffer (100 mM MgCl₂ and 30 mM EPPS, pH 7.5). Time point samples were drawn and immediately quenched with quench solution at 0.5, 1.0, 2.0, 5.0, 10, and 30 minutes. Samples were loaded on an 8% polyacrylamide/ 8M urea gel and WXY and WXYZ bands were separated. Visualization and quantification was possible via phosphorimaging on a Typhoon Trio+ variable mode.
phosphorimager (GE Healthcare) and accompanying ImageQuant software (GE Healthcare). A product ratio was calculated by comparing the RNA in the product WXYZ band to the unreacted WXY band (% reacted = [reacted / (reacted + unreacted)]*100%). Kinetic values were calculated as previously described (Hayden et al., 2008). Briefly, initial rates were calculated from the slope of the linear portion of the reaction curve from a plot of the product ratio versus time (total of \(n = 3\) trials for each concentration). For fast reactions this was \(\leq 5\) minutes and for slower reactions \(\leq 10\) minutes. The rate constant \((k_a)\) was calculated from the slope of the initial WXYZ concentration versus the initial rate of the reaction (Fig. X-5).

2-strategy kinetics (Quadrants b and c of the payoff matrix)

See Figure 8 for a schematic of this process. Reactions, visualization and quantitation were performed same as the self-assembly kinetics with the exception of the fragment genotypes. For the \(b\) quadrant, the initial mixture contained 1 μM of the player 2 genotype and a trace amount of \(^{32}\text{P}\)-labeled WXY (\(\leq 0.003\) μM) of the player 1 genotype. For the \(c\) quadrant, the initial mixture contained 1 μM of the player 1 genotype and a trace amount of \(^{32}\text{P}\)-labeled WXY (\(\leq 0.003\) μM) of the player 2 genotype.
Serial dilutions

A master mix reaction mixture was formed containing equimolar WXY genotype 1 (0.5 μM) and WXY genotype 2 (0.5 μM) and Z (1.0 μM). The mixture was then divided in two equal volumes. One part was doped with $^{32}$P-labeled WXY genotype 1 and the other part with $^{32}$P-labeled WXY genotype 2. The two reaction mixtures were then aliquoted into eight tubes each (one for each burst). (In the case of AU vs. UC, the original master mix made for the eight bursts was made at a 50:50, 20:80, or 05:95 ratio, and then divided into eight portions, and then this was used as above.) All tubes were heated up to 80 °C for 2 minutes and then cooled to 48 °C. The reaction in the first tube was initiated with the addition of reaction buffer (100 mM MgCl$_2$ and 30 mM EPPS, pH 7.5). At 5 minutes, 10% of the solution volume from tube #1 was transferred to tube #2, and tube #1 was placed on ice. Reaction buffer was immediately added to tube #2 while tube #1 was subsequently quenched with equal volume of quench solution. The transfer protocol was repeated through eight bursts. The two-part master mix containing $^{32}$P-labeled WXY was used as a negative control for the assay. Gel separation, visualization and quantitation was performed same as kinetic assays. Three-strategy serial dilutions (Figure X-13) were performed using the same protocol as above with the addition of a third genotype.
Figure 2. Evolutionary game theory feedback loop. The dynamics, and thus the outcomes, of evolutionary game theory competitions can be illustrated in a feedback loop. Here, players interact with a strategy (phenotype), which receives a payoff in terms of fitness. This fitness subsequently determines the frequency of the players/strategies, which then feeds back into the player/strategy interaction environment. The dynamic cycle then begins again.
**Figure 3. Payoff matrix and selection dynamics.** *(left)*The payoff matrix contains the payoffs $a$, $b$, $c$, and $d$ that each strategy receives for each interaction. The asterisk (*) denotes the strategy receiving the payoff. So, strategy 1 receives payoff $a$ when interacting with strategy 1 and payoff $b$ when interacting with strategy 2. Contrarily, strategy 2 receives payoff $c$ when interacting with strategy 1 and payoff $d$ when interacting with strategy 2. *(right)* The schematic describes how the payoff ordinal rankings relate to game outcomes as processed through a game theoretic model (adapted from Nowak, 2006).
Figure 4. 2-piece recombinase reaction, $\text{GMG}_\text{WXYZCNU} + Z$. Here, as in Figure 1, the WXY and Z fragments first associate via base pair interactions to form a non-covalent WXYZ molecule. In the payoff matrix this molecule is the row player. The non-covalent WXYZ is then catalytically connected with one recombination reaction at the Y-Z junction by another molecule of non-covalently or covalently linked WXYZ recombinase. In the payoff matrix the catalyzing molecule is the column player.
Figure 5. IGS-tag sequence interaction. The IGS-tag interaction can be manipulated to confer a range of recognition reactions and thus a range of catalytic abilities for the recombinase. The IGS-tag interaction is zoomed in where the IGS is shown in red with the middle nucleotide denoted “M” representing a 4-fold variation (A, C, G, or U nucleotide) and the tag is shown in orange with the middle nucleotide denoted “N” again representing 4-fold variation (A, C, G, or U). The gray structure illustrates a WXYZ molecule (either non-covalent or covalently linked) as it’s IGS (red) interacts with the tag (orange) on the WXY fragment (blue). The WXY fragment is associated with the Z fragment (green) via base pair interactions, additionally the Z fragment interacts with the WXYZ recombinase at the G-binding site (gray dotted lines).
<table>
<thead>
<tr>
<th>Genotype</th>
<th>$k_a$ (min$^{-1}$)</th>
<th>Std. error*</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>0.0415</td>
<td>0.0066</td>
<td>0.98</td>
</tr>
<tr>
<td>AU</td>
<td>0.0319</td>
<td>0.0011</td>
<td>1.00</td>
</tr>
<tr>
<td>UA</td>
<td>0.0197</td>
<td>0.0004</td>
<td>1.00</td>
</tr>
<tr>
<td>GC</td>
<td>0.0125</td>
<td>0.0021</td>
<td>0.97</td>
</tr>
<tr>
<td>GU</td>
<td>0.0091</td>
<td>0.0007</td>
<td>0.99</td>
</tr>
<tr>
<td>AC</td>
<td>0.0069</td>
<td>0.0002</td>
<td>1.00</td>
</tr>
<tr>
<td>UG</td>
<td>0.0049</td>
<td>0.0004</td>
<td>0.99</td>
</tr>
<tr>
<td>UC</td>
<td>0.0038</td>
<td>0.0002</td>
<td>1.00</td>
</tr>
<tr>
<td>UU</td>
<td>0.0022</td>
<td>0.0001</td>
<td>1.00</td>
</tr>
<tr>
<td>CA</td>
<td>0.0020</td>
<td>0.0000</td>
<td>1.00</td>
</tr>
<tr>
<td>CC</td>
<td>0.0016</td>
<td>0.0001</td>
<td>1.00</td>
</tr>
<tr>
<td>GG</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.99</td>
</tr>
<tr>
<td>GA</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.98</td>
</tr>
<tr>
<td>AA</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.92</td>
</tr>
<tr>
<td>CU</td>
<td>0.0004</td>
<td>0.0000</td>
<td>1.00</td>
</tr>
<tr>
<td>AG</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.99</td>
</tr>
</tbody>
</table>

*Standard error values based on three independent replicates.
Figure 6. Graph for the determination of autocatalytic rate constants. Initial rates were measured by doping in 0 μM, 0.5 μM, 1 μM, or 2 μM full-length WXYZ into reactions containing 1 μM GMG WXY CNU and 1 μM Z as described above and in Hayden et al., 2008; based on the method of von Kiedrowski, 1986. Each point represents the average of three independent trials. Main plot: data for the 11 fastest self-assembling genotypes, where rates were measured for reaction times of 5 minutes or less. Inset: data for the five slowest self-assembling genotypes, where rates were measured for reaction times of 10 minutes or less.
Figure 7. Use of IGS and tag nucleotides to predict 2-strategy outcomes.

(A) A depiction of how the MN notation translates into the a, b, c, and d values in the payoff matrix. (B) An example game of AU vs. UC is shown. The a value is the autocatalytic rate constant that results when \( G\text{AG}_{\text{WXY}}\text{CUU} \) is incubated with \( Z \). This reaction is determined by the strength of an A-U nucleotide pair within the interaction between an IGS triplet in one RNA fragment and the “tag” in another (Fig. 5). Likewise the d value is the autocatalytic rate constant that results when \( G\text{UG}_{\text{WXY}}\text{CCU} \) is incubated with \( Z \). The b and c values can be estimated by using the autocatalytic IGS/tag pairing in place of a measured IGS/tag interaction between two genotypes (as described in the gray bubbles). These estimated values of 0.0022 min\(^{-1}\) and 0.0069 min\(^{-1}\) compare with the measured values of 0.0028 min\(^{-1}\) and 0.0079 min\(^{-1}\), respectively (see Fig. 9).
Figure 8. Schematic and example gel electrophoretic for payoff matrix quantification. (A) Differential $^{32}$P labeling procedure. The schematic describes how the four values ($a$, $b$, $c$, and $d$) in the 2x2 matrices were obtained. Each value was obtained in a separate experiment. In the schematic, the $\gamma^{32}$P•[ATP]-labeled WXY molecule is a small ($<< 0.1\%$ by moles) dopant in a 100 mM MgCl$_2$ solution of 1 $\mu$M unlabeled RNA of the “same” genotype for quadrants $a$ and $d$ and the “other” genotype for quadrants $b$ and $c$, plus 1 $\mu$M of the common substrate resource ($Z$). The asterisk (*) in the matrix denotes the genotype receiving the payoff; this is read as the “column” genotype acting on the “row” genotype. The $a$ and $d$ values were obtained by the same experiments described in Table 1. The $b$ and $c$ values were obtained by doping genotype 1 into a solution containing 1 $\mu$M of genotype 2. Reaction rate constants were then determined by the initial slopes of percent product (WXYZ) formed as a function of time. Seven time points, at 0, 0.5, 1, 2, 5, 10, and 30 minutes were taken after incubation at 48$^\circ$C. (B) Example gel electrophoretic analysis used to obtain raw initial rate data for computing the above values.
Figure 9. Empirical 2x2 matrices for four broad categories of games. All values are averages of three independent trials, and each rank-order is significant at the $P < 0.05$ level or better (Table 2). In red, $a > c$ and $b > d$ indicating strategy 1 dominance, in blue $c > a > d > b$, representing the biological Prisoner’s Dilemma (PD) scenario but here termed counter-dominance. The game categories are re-named here as discussed in the text under “chemical game outcomes” to better reflect chemical dynamics as opposed to biological dynamics. In green $c > a$ and $b > d$ described as coexistence in the biological scenario but here termed cooperation. Lastly, in black, $a > c$ and $d > b$ indicating a biological bistable equilibrium but here termed selfish.
Table 2. Kruskal-Wallis test.

<table>
<thead>
<tr>
<th>Game</th>
<th>H</th>
<th>Critical value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG vs. GA dominance</td>
<td>7.821</td>
<td>6.897436</td>
<td>0.046</td>
</tr>
<tr>
<td>GU vs. AA dominance</td>
<td>9.359</td>
<td>6.897436</td>
<td>0.025</td>
</tr>
<tr>
<td>AC vs. UU cooperation</td>
<td>8.504</td>
<td>8.435897</td>
<td>0.009</td>
</tr>
<tr>
<td>UC vs. AA cooperation</td>
<td>10.385</td>
<td>6.897436</td>
<td>0.016</td>
</tr>
<tr>
<td>AU vs. UC selfish</td>
<td>10.385</td>
<td>8.435897</td>
<td>0.009</td>
</tr>
<tr>
<td>CA vs. GG counter-dominance</td>
<td>9.667</td>
<td>8.435897</td>
<td>0.009</td>
</tr>
<tr>
<td>GU vs. UC counter-dominance</td>
<td>8.197</td>
<td>6.897436</td>
<td>0.042</td>
</tr>
</tbody>
</table>
Figure 10. Schematic for serial dilution. After 5 min reaction times, correlating to the linear kinetic phase of the reaction, 10% of the reaction mixture is transferred to a new test tube containing fresh WXY and Z molecules (in the same amounts and ratios as in the initial tube), and fresh buffer is added to initiate the reaction.
Figure 11. Serial dilution results for four categories of 2-strategy competitions. Plots of relative frequencies of WXYZ genotypes as a function of time (bursts) in the serial-dilution format for the same four competitions described in Fig 9. (A) domination scenario, where $a > c$ and $b > d$, (B) counter-dominance, where $c > a > d > b$ (C) cooperation, where $c > a$ and $b > d$ (D) equimolar selfish scenario, where $a > c$ and $d > b$. (E) 20:80 AU:UC ratio for the selfish scenario (F) 05:95 AU:UC ratio for the selfish scenario. The three frequency differences were chosen to understand the chemical dynamics of a “bistable” biological game.
Figure 12. 2-strategy interactions for the 3-strategy RPS scenario. (A-C) The payoff matrices for each 2-strategy interaction present in the 3-strategy game. Values were determined the same as previous 2-strategy games in Fig. 9. (D-F) Plots of relative frequencies of WXYZ genotypes as a function of time (bursts) in the serial-dilution format for the color correlated payoff matrix. Serial dilution experiments were performed the same as the previous 2-strategy games (Fig. 10 and 11).
Figure 13. Rock-scissors-paper 3-strategy game. (A) Empirical data from a serial dilution experiment. The experiment was performed in the same fashion as in Fig. 10, 11, and 12, except here three genotypes (AA, UC, and GU) were separately tracked. Each pair of 2-genotype competitions is expected to give a clear winner (Fig. 12); but here near equal genotype frequencies result. (B) Simplex plot of the serial dilution empirical data showing the joint frequencies starting at the center and approaching a stable internal equilibrium point. (C) 3x3 matrix for the 3-strategy scenario. Values are compiled from the autocatalytic rate constants from Table 1 and from the appropriate measured 2-strategy payoff matrices (Fig. 12). (D) Predicted results from the ODE model with Simplex plot.
Figure 14. Comparison of empirical game outcomes with ODE model outcomes. (A) Experimental plots of relative frequencies of WXYZ genotypes as a function of time (bursts) in the serial-dilution format (same as Fig. 11A-C and E and Fig. 12D-F). (B) Predicted dynamics of the genotypes in these same competitions based on a simple ODE model in a flow reactor scenario using measured 2x2 interaction data (Fig. 9 and Fig. 12A-C) (C) Modeling results using estimated values in the 2x2 matrix, based on the method shown in Fig. 7.
Figure 15. Excluded 2-strategy outcomes due to equalities. Tabulation of the 48 2-genotype contests that generate estimated equal values in their 2x2 payoff matrices. There are 16(15)/2 = 120 distinct 2-strategy contests from 16 WXY genotypes, but 48 of these will generate values in the estimated payoff matrix that are equal, using the logic shown here. When two genotypes compete with the same M or N nucleotide, then the autocatalytic rate constant values (Table 1) will estimate two identical values in the matrix. For example, when AU competes vs. AC as shown here, the a and b values in the matrix both would derive from the AU autocatalytic rate constant value in Table 1. The a value represents an A-U pairing during through autocatalysis of GAGWXYCUU and Z, while the b value represents an A-U pairing during the catalytic cross-catalysis of GAGWXYCUU and Z by a GAGWXYCUU-containing ribozyme. Likewise in this particular contest, the c and d values would be estimated to be equal. Equal values in the payoff matrix do not lead to strict equilibria. Thus for the purposes of forecasting outcomes from all possible 2-strategy contests using only autocatalytic data, these 48 cases are excluded, leaving 72 possible contests, as shown in Fig. 16.
Figure 16. **72 2-strategy results from predicted payoff matrices.** Plots of outcomes of all 72 strict 2-strategy contests from predicted self-assembly matrices. Each panel shows the expected 2-strategy dynamics, for all 72 payoff matrices that can be derived from the autocatalytic data in Table 1. For details on how these matrices were obtained, see Fig.7. Based on the respective matrix inequalities, we have classified 32% of the games as Cooperation, 32% as
Selfish, 26% as Dominance, and 10% as Counter-dominance. These matrices were then used as our input to calculate the expected dynamics according to the kinetic equation [3] above. As this figure suggests, different 2-strategy contests can have remarkably different dynamics, even if the contests are taken from the same game class. The predicted equilibrium frequencies according to equation [4] above are depicted by the dotted black lines. Exact equilibrium frequencies are provided at the bottom of each panel.
Figure 17. Comparison of replicator dynamics with the chemical game dynamics. Based on the 3x3 experimental matrix displayed in Fig. 13C, the two graphs show the resulting replicator dynamics (left) and the dynamics according to the kinetic equation [4] above (right). Replicator dynamics predicts cyclical behavior: AA can be invaded by GU, GU can be invaded by UC, and UC in turn can be invaded by AA. The fixed point on the edge between UC and AA is unstable, and all orbits spiral towards the unique Nash equilibrium in the interior of the state space (with equilibrium proportions UC 35.8%, AA 50.4%, GU 13.8%). However, in this kinetic equilibrium, AA is considerably less abundant than predicted by the Nash equilibrium (UC 39.1%, AA 35.0%, GU 25.8%). Moreover, in the right graph, the edges of the Simplex are no longer invariant because absent strategies are introduced continually (as described in detail in the Mathematical Modeling section above).
Figure 18. Altering genotype activity with environmental perturbations. (A) The production of WXYZ after 4 hours at varying temperature and Magnesium concentrations for 7 individual genotypes. Note the relative activity changes between AU and GC; CA and GU; and GG, CU, AG. Product yield was determined using the same methods as in kinetic assays, except here the reaction was allowed to go for 4 hours and a yield was calculated by comparing the RNA in the product WXYZ band to the unreacted WXY band (% reacted = [reacted / (reacted + unreacted)]*100%). (B) Increasing Z concentration leads to a payoff matrix ordinal change causing the qualitative game outcome to shift from counter-dominance to a dominance game.
Chapter 3 - Recombinase networks

Introduction.

Molecular networks have recently become a focus of work towards understanding abiogenesis. Such a network is a collection of cooperative molecules acting as a unit of origin and evolution rather than a single selfish replicating entity. The very nature of a network's cooperative and distributed functionality imparts several advantages in the “harsh” prebiotic world. Distributing catalytic activity among members provides robustness against the loss of entities (Hinshelwood, 1952; Jain & Krishna 2001). Cooperative replication among a group of molecules may represent a more facile process than self-replication arising in a single entity (Kauffman, 1993, Shapiro 2007) and introduces a possible solution to the error threshold (Eigen, 1971; Eigen & Schuster, 1977; 1978a; 1978b; Higgs & Lehman, 2014). Lastly, cooperative molecular networks are not just possible and represented in the state of modern life but have demonstrated fitness benefits and selection preferences as compared to selfish entities (Smith & Szathmáry, 1995; Vaidya et al., 2012; Hordijk & Steel, 2013; Higgs & Lehman, 2014).

Theoretical work towards understanding a chemical network manifestation of life has looked extensively at autocatalytic sets (Hordijk et al., 2010; Vasas et al,
2012) or collectively autocatalytic sets (CAS) as originally presented by Kauffman (Kauffman, 1993). The basic concept of a CAS is that there is catalytic closure of all members of the CAS such that the set can self-replicate. Catalytic function need not reside in any one member but each member has at least one of the possible last steps of its formation catalyzed by some other member of the set. The CAS concept has been extended and the importance of raw materials (i.e., food) in the environment has been mathematically formalized as a reflexively autocatalytic and food-generated (RAF) set (Hordijk & Steel, 2004). CAS-like systems were further shown to be relevant in abiogenesis studies with the demonstration that they could arise spontaneously from a simulated pre-life chemical environment (Jain & Krishna, 2001). In Jain and Krishna (2001), the pre-life scenario was interconnected, undefined molecular species able to catalyze the ligation of reactant molecules to produce other catalytic species. The appearance of autocatalytic sets was not only shown to be inevitable but they were also able to recover from stochastic frequency fluctuations and extinctions.

While a CAS-like system does embody the self-sustaining aspect of life (as life defined as a self-sustaining chemical system capable of Darwinian evolution), an autocatalytic set itself has not been determined to be evolvable (Vasas et al., 2010). However, an autocatalytic set as part of a chemical network containing multiple autocatalytic sets could be selectable. If these sets are further
compartmentalized, competition could exist intra- and inter-compartmentally (Vasas et al., 2012).

Despite molecular networks potential role in ushering in biotic systems from abiotic systems, concrete examples of the chemical and physical dynamics of such networks are still lacking. Empirical work has demonstrated the formation of networks under thermodynamic and kinetic control. While thermodynamically controlled networks are not discussed here (see Corbett et al., 2006; Ludlow & Otto, 2008 for reviews), the experimental manipulations of network constituents at equilibrium using templating and molecular recognition are interesting and may have implications for molecular evolution (Cousins et al., 2000). Living systems, however, operate far from equilibrium and thus a kinetically controlled network, such as a CAS, is fundamental in understanding the chemical to biological shift.

Empirical manifestations of autocatalytic sets have been previously demonstrated in a simple two-ligase system (Lincoln & Joyce, 2009). More recently much larger networks consisting of the recombinase ribozyme have been experimentally demonstrated and mathematically modeled (Vaidya et al., 2012; Hordijk et al., 2013). While this work has shown the capability of interacting molecular species to act cooperatively and form networks, the underlying dynamics behind connections, topology, and how selective forces could act is largely unknown.
What is known is that network dynamics (from neural, metabolic, social and the world wide web) ultimately derive from local structures (i.e., connectivity kinetics between the entities and their links), the connections within the environment, and the long-range structure of the network or topology (Dorogovstev & Mendes, 2003). Thus it would seem that molecular networks could be characterized within the same framework. Recently, we presented a review of six key parameters that could influence and direct the evolvability of networks with the goal of experimental exploration in mind. These parameters include viable cores, connectivity kinetics, scalability, information control, resources, and compartmentalization (Nghe et al., 2015).

It is clear from previous work that autocatalytic networks are characterized by cores, where a core is a subset of entities (or nodes) in which each node is reachable from any other and a viable core is a core that is also self-sustaining (provided the food source) (Vasas et al., 2012; Nghe et al., 2015). However, it is not known how the connectivity of entities within a core affects growth or how the strengths of connections can impact the ability of a network to add or subtract members (i.e., the connectivity kinetics parameter). In Chapter 2 we performed an in-depth characterization of they dynamics between one, two, and three entities. Here, we explore how varying connection strengths and network genotypes influence the system dynamics of up to five entities. These characterizations illustrate how a network can be “stronger”, through faster
growth rates and robustness, or “weaker” with slower growth rates and/or more prone to network degradation. These concepts are thus indicative of how selection could act upon competing cores.

While the experiments here primarily discuss the parameters of connectivity kinetics and cores, the remaining parameters are discussed in detail in Nghe et al. (2015) and can additionally be found in previous literature or in subsequent chapters. More specifically, the parameter of scalability is briefly mentioned in the discussion here. Information control through negative feedback mechanisms was evaluated in the recombinase system in Vaidya et al. (2013). The effects of resource availability are briefly mentioned in Chapter 2 with changing the Z resource. Further, any environmental resource (i.e., Mg$^{2+}$), that alters the free energy ($\Delta G$) of the IGS/tag interaction (as measured in Chapter 4 and illustrated in Figure 18A), will alter the connectivity kinetics. Lastly, experimental methods and preliminary studies towards understanding the affects of compartmentalization will be discussed further in Chapter 5.
Results.

*The recombinase ribozyme and network nodes and links.* In order to empirically investigate chemical connectivity kinetics and cores of networks we used the recombinase ribozyme with varied auto- and cross-catalytic activity (Yeates et al., in press). Briefly, this system derives from the *Azoarcus* group I intron, which can be broken into roughly 50 nt fragments (Fig. 1). These fragments spontaneously re-assemble via base pair interactions to form a non-covalent, full length (ca. 200 nt), active, recombinase ribozyme. The non-covalent assembly then catalyzes the covalent linkage of other non-covalent RNA assemblies. The capability to form large networks derives from the ribozyme’s recognition interaction between the IGS and tag sequences, which instigate the catalytic reaction. In Yeates et al. (in press) it was demonstrated that the variation of IGS and tag sequences can lead to vastly diverse outcomes, among 2- and 3-entity interactions, depending on the competing genotypes. This work lays the initial foundation for understanding connections among interacting genotypes, and here we extend our networks to consider 4- to 5-membered networks.

The connectivity kinetics of a network ultimately describe which entities (also described as network members or *nodes*) are connected to which and how the strengths of such connections can impact the ability of a network to change with time. Network members are connected, or *linked*, through their ability to act catalytically on other members. The network growth here, as well as in Yeates et
al. (in press), is manifested as the rate of production of covalently linked WXYZ from WXY and Z fragments (Fig. 4). The covalently bound WXYZ represents a fitness advantage in an evolutionary scenario, contributing a roughly 2-fold higher catalytic activity than a non-covalent complex of WXY-Z, which is held together only by hydrogen bonds (Vaidya et al., 2012). Thus here we compare how varying connections (or links) impact overall network rates and at times the individual entity’s growth.

It should be noted that all IGS/tag combinations have some level of catalytic activity (Table 1). Thus our networks could be considered fully connected with the number of links equaling the number of nodes squared. However, as the network members increase, the strong catalytic connections of Watson-Crick (W-C) interactions become the main determinants of the network dynamics. We find this approximation even appropriate in our Rock-Paper-Scissors game in Yeates et al. (in press) where three W-C links (AA linked to GU linked to UC, which is in turn linked to AA) illustrate the cyclical dominance relationship (Fig. 13 and Fig. 23, 3-membered network). Therefore to simplify concepts moving forward the term “link” will represent W-C connections exclusively.

**Number of links and the “next best link”**: In collaboration with Philippe Nghe and Harry Kimble at ESPCI ParisTech, we first considered two aspects of connectivity that could impart network growth differences. The first is the effect of the number...
of links within the network and the second is differential link additions (in other words what is the “next best link” for a network to add). Previously, we demonstrated a rudimentary way to predict recombinase genotype outcomes using simple autocatalytic rate constants (see Fig. 7). Here we use this same concept of estimating cross-catalytic rate constants from autocatalytic rate constants to evaluate all combinations of 5-membered networks using our 16 genotypes (Table 1 and Fig. 19), where genotype as defined here is identical to Yeates et al. (in press) “MN” notation. Extraction of the slowest and fastest predicted 5-membered network illustrated a correlation between the number of links and the growth of the network. The slowest network, where “slow” refers to the initial rate of total WXYZ production by the network, contained 6 links. The number of links refers to the number of W-C interactions between the 5 members of the network (see schematic in Fig. 20A). The “fast” network in contrast, contains 11 W-C catalytic links between the 5 members of the network (Fig. 20B). This fast network is predicted to have an initial rate of total WXYZ production about 2.1 times faster than the slow network.

An experimental comparison of the two networks was performed where we tracked the fraction of all WXY genotypes converted to WXYZ over time (Fig. 20C). Our results confirmed the predicted outcome from theory: the fast network with the greater number of links does indeed produce a higher rate of total WXYZ production, (around 2.4 times faster) compared to the slow network. Additionally,
it appears the fast network, after 6 hours, produces a higher yield of WXYZ. Here we are only considering rates of WXYZ production, as this correlates with kinetically controlled systems such as a CAS or life itself. It may be fruitful in the future to consider how WXYZ yields and network equilibrium differences could contribute to network compositions and strengths.

The second aspect of network connectivity we considered was differential link additions within a network. In other words, how would the location of an added single link affect the overall network dynamics? Theoretical literature indicates that the best link addition connects the best “giver” node to the best “receiver” node (Jain and Krishna, 2001). To determine our best giver and receiver within the previously tested slow 5-membered network, we tracked each genotype individually within the network and compared the initial rate of reaction for each entity. Results indicated that GG is by far the best receiver of the network (Fig. 21), in that it receives the highest catalytic rate and production of WXYZ compared to the other members of the network. Given that CU is the W-C linked genotype to GG, thus producing a predicted rate constant almost an order of magnitude higher than any other network member provides to GG, we interpreted CU as the best giver of the network. In order to add a link to the best giver/receiver we doubled the concentration of CU and GG, thereby adding a link by increasing the “weight” (or catalytic rate) of the link between CU and GG (Fig. 21).
We then systematically evaluated adding a link between each node within the network and compared the overall network rates (Fig. 22B).

Our overall network rate comparisons between each of the link additions indicate the CU to GG added link produces the best overall network rate, providing evidence that the next best link is added between the best giver/receiver. The network concept of “the next best link addition is between the best giver/receiver” correlates logically with underlying chemical rate law concepts. In chemical terms, adding a catalytic link by increasing the nodes that create the link, is simply increasing the concentration of two of the genotypes of the network. If this concentration increase occurs where the rate constant is the highest, then this will create the greatest network rate gain.

As an alternative way to consider the next best link, we systematically removed each genotype, one at a time, and evaluated overall network growth (Fig. 22C). Here we see removal of either the CU or GG nodes produces the largest detrimental effect to overall network growth (Fig. 22D). Again, this provides evidence that the placement of a link between the best receiver/giver produces the strongest overall network, when compared to link additions elsewhere in the network. This data may further indicate a chemical mechanism for node/link preferences.
Of interest on a further note, the systematic removal of nodes allows us to consider the slow 5-membered network from the perspective of competing cores, where viable cores represent a set of entities that are fully connected and self-sustaining (provided the food source) (Vasas et al., 2012; Nghe et al., 2015). Cores are what characterize autocatalytic networks and are essentially the units of heredity in an autocatalytic network (Nghe et al., 2015). In Figure 22C we give an example of how the 5-membered slow network can be reduced to a smaller 4-membered core. This data additionally demonstrates that the outer network of AA, CU, GG, and UC represents the strongest network compared to any other 4-membered networks contained within this 5-membered network. Evidence of this is demonstrated in the removal of genotype GU, which provides the least detrimental effect on the network (Fig. 22D) and doubling the GU to UC link as well as the AA to GU link provides the least benefit to the network (Fig. 22B).

**Competing cores.** To further investigate the concept of competing cores, we evaluate how the connectivity and node differences affect a 3-membered network transition to a 4-membered network. Here we use the “inner” 3-membered network from the slow network (AA, GU, and UC; note this is also the RPS game from chapter 2) and evaluate how adding the 4th genotype with differing connectivities affects network growth. We performed five individual node additions, where each addition represents slightly different link characteristics (Fig. 23).
In general, we initially see a trend that the greater number of links leads to overall faster network growth. As we move from networks (A) and (B) with 1-link additions to networks (C) and (D) with 2- and 3-link additions, respectively, the rate of both the overall network and the 3-membered network increase. However, consider network (E); this could be thought of as a 1-link addition (albeit a selfish 1-link that is autocatalytic). Here we see an overall network rate that exceeds even the 3-link network in (D) This network comparison, between (D) and (E), brings to light the importance of the weight of the links. While all of these node additions are Watson-Crick links, there still exist a difference in catalytic capabilities. Genotype CG has the strongest autocatalytic rate constant and GC the weakest, among the W-C genotypes (Table 1). So, while network (E) produces a 3-membered growth rate similar to other 1-link networks (A) and (B), the overall network growth benefits from the very strong 4th node.

Network (E) also brings to light a case for network cooperativity or altruism. While the overall network rate is higher than networks (A) – (D), note the 3-membered rate in network (E) \( V_o = 0.0219 \, \mu M/min. \), this rate is one of the lowest among the compared networks and is even lower than the original 3-membered network rate \( V_o = 0.0234 \, \mu M/min. \) (Fig. 23). This indicates a disadvantage to the cooperative portion of the network (the 3-membered portion), yet despite this cost
of sharing resources with a selfish node, the overall network growth “wins” compared to the networks (A) – (D).

This concept is again demonstrated in networks (A) and (B). When we compare the 3-membered rates within networks (A) and (B), we find the (A) 3-membered rate is depressed, here (A) 3-membered $V_o = 0.0208 \, \mu M/min$ and (B) 3-membered $V_o = 0.0240 \, \mu M/min$. However, when we compare the overall network rates between networks (A) and (B), we find (A) more than compensates for and “wins” versus network (B) (overall network rates: (A) $V_o = 0.0261 \, \mu M$ and (B) $V_o = 0.0251 \, \mu M$) (Fig. X-5). Here, network (A) is sharing resources with a “selfish” node CU, it is additionally providing a W-C link to the “selfish” node. A cost is imposed on the cooperative portion of the network (a depressed 3-membered rate) yet the benefit of catalyzing covalent production of WXYZ for a fourth node is enough to benefit the network as a whole. When we compare this to the network (B) topology we see an inverted scenario. Here, the 3-membered portion of the network is receiving a W-C link from a fourth node. Yet, despite this “selfish” act by the network, there is little rate gain received.

The chemical mechanism underlying this seems fairly straight-forward. In network (A) four nodes are being catalytically converted, through W-C links, to covalently bound WXYZ. In network (B) only three nodes are being catalytically converted to W-C links, thus the overall network rate of WXYZ production is less
in (B) than (A). (Note that network rates in (A) and (B) are also measured in Figure 22D. Removing the GG node produces network (A) and removing the CU node produces network (B). In Figure 22D as well as in Figure 23 the overall growth rate of (B) is less than (A)). While additional scenarios will need to be evaluated, the two cases of networks (E) and (A) indicate cooperation and altruism by a network provides greater benefits than selfishness.

Lastly, we explore the advantage of network cooperativity from a slightly different angle with network (F). Here, the tag sequences from network (E) are swapped. Nodes UC and GU remain the same, however, CG and AA from network (E) become CA and AG (Fig. 23). This produces the fully Watson-Crick linked 4-membered network in (F). Chemically speaking all collective IGS and tags remain constant between (E) and (F) (both networks contain an A, C, G, and U IGS and both contain an A, C, G, and U tag). So here, chemically speaking, only the connectivity of the nodes is different, all links are weighted the same. From this angle, we again see cooperativity producing a greater overall network growth. Here, the fully cooperative network (F) produces WXYZ with an initial rate of 0.0446 µM/min and network (E) has an initial rate equal to 0.0419 µM/min. This increase in overall growth of network (F) vs. network (E) is thought to be due to network (F) incorporating the strong activity of CG into the network topology.
Among competing cores, we find network cooperativity produces advantages to the overall strength of the core (or rate of catalytic activity). This seemingly benevolent behavior resides in simple chemical dynamics where “cooperative” links produce more covalently bound WXYZ, which is able to overcome any “altruistic costs”. Specifically we found that:

- In network (E) vs. (D), the rate decrease of the 3-membered network from (E) due to harboring a selfish 4th entity in the system, is more than made up for in that member’s high WXYZ catalytic activity.
- In network (A) vs. (B), the rate decrease of the 3-membered network in (A) due to providing a W-C link to a selfish 4th entity in the system, is compensated for through the increase in node WXYZ production. Network (A) contains 4 nodes that are produced with W-C links versus network (B) where only 3 nodes are produced with W-C links.
- In network (E) vs. (F), the incorporation of the strong CG weighted link into the network allows the network to incorporate the weighting into more production of its species compared to network (E) where that link is isolated.
Discussion.

Here we have demonstrated concrete chemical trends that characterize how connectivity kinetics can lead to stronger or weaker networks and produce a foundation for selection among cores. The trends that impact network growth include the number or weight of links within a network, the preferential adding or weighting of links between specific nodes (i.e., best giver/receiver), and the cooperativity within the network. While these dynamics are specifically illustrated in the recombinase system, the overall outcomes reduce down to simple chemical dynamics of activity rates of entities within and between each other (nodes and links). Thus it would seem that regardless of the chemical system these trends would be transferable to other systems.

These trends, which demonstrate network growth differences, further hint at the ability of simple chemical mechanisms to drive positive and negative selection and thus potentially core evolvability. Furthermore, the observation of differential node activity indicates a mechanism for growth with node preferences, a common basis for scale-free networks to arise (Barbási & Albert, 1999). While these statements may be far-reaching due to the nature of the simple/minimal prebiotic networks empirically evaluated here, these systems do demonstrate that network behaviors can be generalized and empirically explained using simple chemical dynamics.
Materials and Methods.

RNA preparation. The WXY molecules, which are portions of the self-splicing group I intron from the isoleucine pre-tRNA in the purple bacterium *Azoarcus* (Reinhold-Hurek & Shub, 1992), were prepared by *in-vitro* transcription from DNA plasmid templates. The Z fragment was purchased from TriLink Biotechnologies (San Diego, CA) and was gel purified prior to use. All RNAs were resuspended in 1–10 μM solutions in 0.1 mM EDTA. For quantification, < 0.1 μM of the WXY-fragment was 5´-labeled with γ[^32]P]-ATP using OptiKinase (USB, Cleveland, OH).

Reaction assays. Reaction mixtures containing WXY (0.5 μM) of each genotype, equimolar Z (2 μM in 4-membered networks and 2.5 μM in 5-membered networks), and ^32^P-labeled WXY (≤ 0.003 μM) of each genotype, were heated to 80 °C for 2 minutes then cooled to 48 °C. Reactions were initiated with the addition of reaction buffer (100 mM MgCl₂ and 30 mM EPPS, pH 7.5). Time point samples were drawn and immediately quenched with quench solution (125 mM EDTA and 2X loading dye containing formamide and bromophenol blue) from 0 – 360 min. Rate comparison experiments were performed from 0 – 10 min. Samples were loaded on an 8% polyacrylamide/ 8M urea gel and WXY and WXYZ bands were separated. Visualization and quantification was possible via phosphorimaging on a Typhoon Trio+ variable mode phosphorimager (GE Healthcare) and accompanying ImageQuant software (GE Healthcare). A product
ratio was calculated by comparing the RNA in the product WXYZ band to the unreacted WXY band (fraction reacted = [reacted / (reacted + unreacted)]).

Tracking individual genotypes. In order to determine individual genotype WXYZ production, reaction mixtures were made as a master mix without $^{32}$P-labeled WXY. The master mix was then divided into four or five reaction tubes, depending on the number of genotypes being tracked, and each tube was doped with the appropriate $^{32}$P-labeled WXY genotype being followed.
Figure 19. Comparison of 5-membered overall network rates calculated from experimental rate constants vs. simplified Watson-Crick topology. A plot was created of all possible 5-membered networks, from the 16 genotypes in Table 1. The y-axis is the predicted overall network growth, calculated from the 16 autocatalytic rate constants (where cross-catalysis was estimated using autocatalytic rates as previously described in Yeates et al. (in press). The x-axis is the predicted overall network growth from W-C topology, calculated where W-C = 1 and all other connections = 0. All 5-membered combinations are in blue dots. The fully connected via W-C linkage 5-membered networks are in red dots. Plot and data analysis produced by Philippe Nghe and Harry Kemble (ESPCI ParisTech).
Figure 20. Slow vs. Fast 5-membered networks. (A) Schematic of the slow network (where genotype compositions was determined from Figure 19 analysis of all 5-membered W-C connected networks). Here, each arrow indicates a W-C link between genotypes (nodes). For example CU $\rightarrow$ GG indicates the IGS, GCG from CU, catalyzes the tag, CGU from GG. The slow network is connected through 6 W-C links. (B) Schematic of the fast network. The fast network is connected through 11 W-C links. (C) Plot of the production of total WXYZ by each network versus time. Fast network is in red and slow network is in blue.
Figure 21. Determination of best giver/receiver in a 5-membered network

Plot of the production of WXYZ of each genotype within the 5-membered slow network versus time. The production of WXYZ was determined by $^{32}$P differential radiolabeling. Genotype GG clearly demonstrates the highest WXYZ production rate and is thus the best receiver of catalytic activity.
Figure 22. Addition of links and removal of nodes in a 5-membered network. (A) Schematic of the slow network with a link addition between the CU and GG nodes. Links were “added” by increasing the concentration of the participating nodes by 2x. (B) Overall network rate results for each link addition within the slow network. The CU to GG link is in gray to highlight the best overall network rate with a link addition as compared to other link additions. (C) Schematic of the slow network with a CU node removal. (D) Overall network rate results for each node removed. Nodes CU and GG are highlighted in gray showing the largest detrimental affect to the network rate.
Figure 23. Competing cores - comparison of 4-membered networks. (Top)
Schematic of the 3-membered network with corresponding individual node WXYZ rates and the total network rate (µM/min). 4-membered networks. (A) Schematic of a 1-link connected CU node, where CU receives W-C catalytic activity from the 3-membered network. The network is connected through 4 total W-C links. Corresponding catalytic rates of each node, the 3-membered network, and the total network are tabulated on the right. (B) a 1-link connected GG node, where GG contributes W-C catalysis to the 3-membered network. Network W-C links = 4. (C) a 2-link connected AC node, where AC both gives and receives a W-C link with the network. Network W-C links = 5 (D) a 3-link connected GC node, where GC gives and receives a W-C link with the network, as well as produces an autocatalytic W-C link for its own genotype. Network W-C links = 6. (E) a 1-link
node, where CG is isolated from the network and the 1-link is an autocatalytic link for its own genotype. Network W-C links = 4. (F) a network containing 4 W-C links with collective IGS and tag sequences identical to network (E). Here, UC and GU are the same as the above network (E), and the tags of network (E) on nodes AA and CG are swapped to give network (F) nodes AG and CA.
Chapter 4 – Fluorescence anisotropy for the investigation of IGS and tag binding interactions in the recombinase ribozyme

Introduction.

In Chapters 2 and 3 we demonstrated a large diversity of RNA recombinase population outcomes dependent upon genotype interactions. The ability of the recombinase to exhibit such a variety of activity is due to only a 1- or 2-nucleotide (nt) change within the internal guide sequence (IGS) and/or tag sequences. We are further able to predict and engineer specific multi-interacting RNA population outcomes simply from the autocatalytic rate constants of these various IGS/tag sequences. Despite the significance of this nucleotide-pair triplet, the underlying contribution of this interaction is not fully understood, and yet it is precisely this thermodynamic interaction that forms the basis for the intermolecular (and inter-population) behavior described in the previous chapters. Thus in this chapter we will investigate more thoroughly a key facet of this triplet interaction: the binding strength, or $K_D$ analog.

In the in vivo system, the triplet IGS and tag recognition must be accurately selected from among many possible such nucleotide triplets. Failure to do so would result in insertions or deletions in the 5’ and 3’ splice sites of the tRNA precursor product. The two-splicing reactions required to remove the intron and ligate the exon utilize secondary interactions in the alignment of the G-U wobble
splice site at the IGS/tag interface as well as tertiary elements like correct folding of the intron and positioning of an exogenous guanosine in the G-binding site (Cech, 1990; Zaug et al., 1993; Adams et al., 2004).

The \textit{in vitro} recombinase system, as in the canonical tRNA splicing reaction, would also need to utilize these secondary and tertiary elements to perform catalysis. The difference between the two situations revolves primarily around the condition that \textit{in vivo} the ribozyme is performing an \textit{intramolecular} reaction, where here, the recombinase performs an \textit{intermolecular} reaction. Our mechanism primarily proceeds through a one-step process (Fig. 1C) (Draper et al., 2008), noticed first by Zaug et al. (1993), and is optimal at higher temperatures and magnesium concentrations than found in a living cell.

Understanding the \textit{in vivo} versus the \textit{in vitro} designed ribozyme reaction allows us to consider the factors that could impart differential activity among genotypes. Given that only a 1- to 2-nucleotide mutation is performed at the IGS/tag locations (sites that are not known as tertiary structural elements; see Denesyuk & Thirumalai, 2015), it is thought that the foundation for activity diversity lies within the secondary structural element of the IGS and tag recognition. As a control for global folding effects we did perform native gel experiments to search for any tertiary folding variability; however, no qualitative differences were detected among the genotypes (Fig. 24).
Typically enzyme substrate binding thermodynamics are measured through kinetic studies such as those outlined for the *Azoarcus* group I ribozyme reaction in Gleitsman & Herschlag (2014). For the autocatalytic nature of the recombinase system, however, the kinetic variable determined to best describe the activity of the genotypes is the autocatalytic rate constant (Hayden et al., 2008; von Kiedrowski et al., 1991). While this variable is an accurate description of the overall activity of the ribozyme, it is a simplification of the system that does not account for non-covalent complex concentrations and binding efficiencies. Therefore in an effort to characterize the contribution of the IGS/tag binding interaction, we use a fluorescence anisotropy (FA) method to further characterize the parameters that underlie our genotype/phenotype diversity.

Fluorescence anisotropy, also known as fluorescence polarization, is a tool for studying molecular interactions by monitoring an apparent size change in fluorescent or fluorescently-labeled molecules. The basic premise of FA is that when a small fluorescent molecule is excited with plane-polarized light, the emitted light is depolarized due to rapid tumbling in solution during its fluorescence lifetime. However if a large molecule binds the small molecule, the rotation of the small molecule is slowed and the emitted light remains polarized (Perrin, 1926). The bound and free states of the small molecule are characterized by a polarization or anisotropy value, a high value if bound and a low value in the
free-state. The polarization of the small molecule (or ligand) can be obtained by titrating in the large molecule (or receptor). The measured anisotropy value is a weighted average of the bound and free states thereby providing a direct measurement of the fraction of bound ligand.

FA is a technique that has been used for a variety of receptor ligand binding studies, including DNA-protein interactions (Kleigman et al., 2006) and RNA-RNA interactions (Schlax et al., 2001). Here we describe the development of its use for the characterization of the IGS/tag recognition interaction within a variety of our recombinase genotypes. Through FA binding observations, we find that in a handful of our genotypes we are able to gain further insight into the thermodynamics that lie at the foundation of the buildup of molecular networks and produce evolutionary outcomes.
Results.

Predicted Gibbs free energy ($\Delta G$). For an initial investigation into the IGS/tag binding interaction we used the “Nearest Neighbor Database”, secondary structure stability prediction tables, to predict the free energies of the triplet hybridization (Turner & Mathews, 2009). Here, the “Turner Rules” set parameters, typically based on melting point experiments, to predict the free energy of nucleic acid secondary structures. While our calculations illustrate the general trend that Watson-Crick interactions contain the lowest free energies, followed by the wobble base pair genotypes (Table 3), the predicted energies do not correlate well within the non-canonical base pair interactions when compared to our autocatalytic rate constant data. For example, there is no predicted difference in free energy between AC, UC, CC, and GA even though their catalytic activities differ greatly (Table 1). Thus we turned to experimental methods in an attempt to understand a correlation between the IGS/tag interaction and recombinase activities.

FA experimental design. In order to estimate experimentally the free energy of the IGS/tag interaction using FA, we utilized the WXYZ recombinase ribozyme as our receptor, and designed a small molecule fluorescently labeled ligand to act as a surrogate WXY-Z substrate region (Fig. 25A). The receptor was prepared using the same techniques as the kinetic assay preparations in Chapter 2. Each IGS genotype was prepared: GUG, GAG, GCG, and GGG, as well as an IGS-
less control where the first three 5’ nucleotides of the ribozyme were deleted (the IGS). It should be noted that here we will only be considering the recombinase ribozyme in its covalently-contiguous receptor form. The ribozyme is also active in the non-covalent WXYZ complex form, at a roughly 2-fold lower catalytic activity (Vaidya et al., 2012). The equilibrium binding constant ($K_D$) of the complex has previously been measured ($K_D = 8.6$ nM, Vaidya et al., 2012) and it would be interesting to compare the IGS/tag interactions of this complex utilizing the FA methodology at a later time as well.

The ligand, a designed WXY-Z substrate surrogate (where WXY-Z is notated with a hyphen to indicate the Y-Z junction is not covalently bound), was purchased as two fragments: a “WXY fragment” containing the tag sequence, CAU and CCU (only two tags have been purchased to date; future work could employ the other two) and a “Z fragment” that was 3’ labeled with a fluorescein derivative, 6-carboxyfluorescein (6-FAM). This surrogate WXY-Z junction was designed with the limitations of the FA method in mind. Here, our ligand needs to be small enough (ca. 10 kDa) to depolarize the emitted light within the lifetime of the fluorescein label ($\tau_F \approx 4$ ns, at this lifetime molecules > 30 kDa will result in highly polarized fluorescence) (Pope et al., 1999), prevent a catalytic reaction so the binding interaction is isolated, and contain enough base pairs in order to stay annealed at the 48 °C, 100 mM Mg$^{2+}$ assay conditions. To accomplish these requirements we designed a duplex with a 12-base paired GC rich stem. On the
“WXY fragment” the stem is followed by the last 4-nt of the 3’ end with the catalytically active uracil (U) replaced by a 3’deoxythymidine. On the “Z fragment” the GC stem is followed by the last 6-nt of the 5’ end (Fig. 25A). The duplex totals 35-nt, with a molecular weight of 11 kDa, and a predicted melting point ($T_m$) of 90 °C (as calculated by DINA melt, Markham & Zuker, 2005).

Prior to running FA assays, the two fragments of the ligand were annealed by heating to 90 °C for 5 min and then slowly cooling to room temperature. This was done with a slight excess of the “WXY fragment” to ensure all fluorescently labeled “Z fragment” was bound (10% higher concentration of “WXY fragment” as compared to “Z fragment”). Assay conditions utilized 0.03 µM ligand, in 100 mM MgCl$_2$, buffered with 30 mM EPPS at pH 7.5 at 48 °C. The receptor (WXYZ recombinase ribozyme) was titrated in great excess of the ligand (0.03 µM ligand to roughly 30 µM receptor) to avoid receptor depletion thereby simplifying calculations allowing the total receptor concentration to be used in place of the free receptor concentration. Polarization measurements were obtained with each titration addition until an equilibrium was reached (indicated by polarization values remaining constant) or the receptor supply was consumed.

*Experimental dissociation constant ($K_D$) and Gibbs free energy of binding ($\Delta G^°_{48 binding}$). Our FA results indicate a correlation between IGS/tag binding abilities with genotype catalytic activity. In Table 3, we see the dissociation constant ($K_D$)
and Gibbs free energy of binding ($\Delta G_{48}^{\circ}$ binding) increase as we move from high catalytic activity genotypes to low activity genotypes. This is an expected result where the W-C genotype (UA) binds with the highest affinity, followed by a wobble-like active genotype (AC), followed by a mediocre recombinase (CA), and a bulky, low activity, purine: purine interface (AA) has the lowest binding affinity (Table 3).

The binding plots for genotypes UA, AC, CA, and AA are found in Figures 26 – 29 and curves were fit to the equation $r = r_0 + (r_{\text{max}} - r_0)(K_a R / (1 + K_a R))$, where $r$ is the measured anisotropy, $r_0$ is the initial anisotropy, $r_{\text{max}}$ is the maximum anisotropy, $K_a$ is the association constant, and $R$ is the total receptor concentration. An interesting trend develops as we compare our genotype binding curves through decreasing binding affinity. The maximum anisotropy increases as binding affinity decreases and the curves for CA and AA do not reach an equilibrium. The $K_a$ for these two genotypes is a projected value and indicates a poor binding interaction between the IGS and tag sequences, which is a logical result. However, the increasing anisotropy and inability to reach equilibrium may indicate interactions that are increasingly characterized by non-specific binding. In order to evaluate the contribution of non-specific interactions we look to our IGS-less genotype.
**IGS-less control.** FA measurements for our IGS-less control demonstrates a dissociation constant and a binding free energy that falls between a W-C (UA) and a wobble-like (AC) genotype (Table 3, Fig. 30). The IGS-less receptor would be indicative of the non-specific binding contributions of the G-binding site and any tertiary stabilization of the ligand. It is noted in *Azoarcus* kinetic studies that despite the ribozyme forming only three base pairs in the IGS, its substrate affinity is similar to that of the group I intron *Tetrahymena*, which forms six base pairs in its corresponding IGS (Glei tsman & Herschlag, 2014). This is thought to be indicative of strong tertiary stabilization within the active site of the *Azoarcus* ribozyme, which is required for maintaining its biological specificity *in vivo*. This information, together with our FA data, could indicate that genotypes with binding constants higher than the IGS-less ribozyme have IGS/tag structures that interact or interfere with active site stabilization beyond the simple 3-nt base pair surfaces. Our catalytic activity may reside then, not in the IGS/tag interaction, but in additional structure stabilizations surrounding this site.

**Binding curves for genotypes GA and GC.** An FA analysis was additionally performed for the GA and GC genotypes (Fig. 31). Here, we found that no binding equilibrium was reached, despite the high receptor concentrations utilized in titration (ca. 70 µM receptor). This result is especially surprising within the high catalytic activity W-C genotype GC. In the case of the “GGG” IGS (as in GA and GC), there is some indication that the ribozyme is forming aggregates that do not
allow a receptor: ligand anisotropy equilibrium to be reached. If we examine the structure of the recombinase ribozyme, we can identify two loop regions that have complementary tag sequences to the IGS “GGG” where aggregate binding is potentially occurring (Fig. 32A, indicated by arrows). These loop regions correspond to the W-X and X-Y junctions that are utilized in the 4-piece recombinase reaction (Fig. 1). The complementary tag sequences found in these loops are CCC at the W-X junction and CCU at the X-Y junction. Additionally, within our “GGG” IGS kinetic assays we find conserved degradation bands that could potentially represent the reverse recombination reaction occurring at these loop regions (Fig. 32B). While the FA studies preclude any catalysis occurring (due to the 3’dT on the “WXY fragment” tag sequence) these observations do indicate that binding at these regions is highly probable.

We further performed a native gel analysis to compare the “GGG” ribozyme (corresponding to FA assays GA and GC) with the “GUG” ribozyme (corresponding to the FA assay UA) to see if any indication of ribozyme aggregation was occurring. Here we observed bands in the “GGG” native gels that correspond to complexes larger than the WXYZ receptor. As the concentration of the “GGG” ribozyme was increased so did the number of bands in the gel (Fig. 32C). These bands were not observed in the corresponding “GUG” native gels. It should be noted that the genotype AA does not demonstrate a binding curve approaching equilibrium (Fig. 29). Additional studies
will need to be performed to better understand low catalytic activity binding interactions (including titrations to higher receptor concentrations). However, we currently have no evidence for aggregates occurring here (i.e. conserved degradation bands within the “GAG” IGS kinetic assays) and thus, at this time, have treated this analysis as a poor binding interaction with non-specific interactions (correlating with the poor catalytic activity).

Discussion.

The FA method development and analysis have provided evidence towards some of the underlying thermodynamics that drive the catalytic abilities of our chemical genotypes. We see a general trend that correlates higher catalytic activity with higher binding ability indicating a potential thermodynamic mechanism for the catalytic phenotypes of our molecular genotypes. The FA analysis also alludes to active site perturbations that could be influencing non-canonical IGS/tag interactions as illustrated through comparison to the IGS-less control ribozyme.

Additionally, we may have found evidence as to why our GC genotype is not nearly as catalytically productive as would be expected from the predicted Gibbs free energy. Here, we note that the predicted GC binding interaction ($\Delta G_{48}^{\text{binding}} = -5.09 \text{ kcal/mol}$) is lower than that of CG ($\Delta G_{48}^{\text{binding}} = -4.58 \text{ kcal/mol}$), yet CG represents the strongest catalytic activity of our genotypes.
(Table 1). GC in contrast represents the weakest of the Watson-Crick genotypes. The FA data, native gel analysis, and kinetic assay degradation taken together indicate that alternative binding and activity may be hindering the capability of the GC genotype.

Further investigation into the binding of the remaining two tag sequences CGU and CUU is still needed, as well as additional probing into the lower-activity genotypes. However, the FA method has provided insight into some of the basic thermodynamic issues at play in the recombinase IGS/tag interactions. The kinetics and thermodynamics of interacting chemical species are ultimately the basis of chemical complexity, and this method and data provides us with an additional tool and look into the those mechanistic foundations.
Materials and Methods.

RNA preparation. The WXYZ molecules were prepared by in-vitro transcription from DNA plasmid templates. The ligand (WXY_{CN3dT} and Z-6FAM fragments) was purchased from TriLink Biotechnologies (San Diego, CA). All RNAs were resuspended in 1–10 µM solutions in 0.1 mM EDTA.

Native gel electrophoresis. For native gel electrophoresis WXY fragments of each of the 16 genotypes were diluted to 0.1 µM with 100 mM MgCl$_2$ and 30 mM EPPS, pH 7.5. Samples were either heated to 80 °C and cooled on ice for 2 min. prior to loading dye addition and gel loading or no heat/cool treatment was provided. A 6x native gel loading dye was added to samples (40% sucrose and 0.25% bromphenol blue) and samples were loaded onto a 10% polyacrylamide gel containing 4 mM MgCl$_2$ and 1 mM EDTA. Gels were run at a constant 48 °C temperature. Gels were stained with Sybr Green II and visualized with Typhoon Trio+ variable mode phosphorimager (GE Healthcare) and accompanying ImageQuant software (GE Healthcare).

FA assays. An LS 55 luminescence spectrometer (PerkinElmer) and accompanying FL Winlab software was used to measure fluorescence polarization at an excitation wavelength of 486 nM and an emission wavelength of 517 nM. All experiments were performed with 30 nM ligand, 100 mM MgCl$_2$, 30
mM EPPS, pH 7.5, at 48 °C. Buffer was pre-warmed prior to cuvette additions. The WXY-Z surrogate ligand was pre-annealed by heating to 90 °C for 5 min and cooling slowly to room temperature over 1 hr. The receptor (WXYZ ribozyme) was titrated into the cuvette, mixed well, and contents were allowed to equilibrate for 1 min. A total of four polarization values were obtained and the average value was converted to anisotropy (as shown below).

*Curve fitting, $K_D$, and $\Delta G^{\circ}_{48}$ binding calculations.* The binding curves for the calculation of $K_D$ and $\Delta G^{\circ}_{48}$ binding were created with Kaleidagraph software by plotting the concentration of receptor, WXYZ, versus the measured anisotropy (converted from polarization values through the equation $r = \frac{2 \times P}{3 - P}$, where $r$ is anisotropy and $P$ is the measured polarization). Curves were fit to the equation

$$r = r_o + (r_{max} - r_o)(K_a R / (1 + K_a R)),$$

where $r$ is measured anisotropy, $r_o$ is the initial anisotropy, $r_{max}$ is the maximum anisotropy, $K_a$ is the association constant, and $R$ is the total receptor concentration. Here the concentration of receptor (starting at 1 – 2 µM) far exceeds the concentration of ligand (30 nM) and receptor depletion is not considered. The dissociation constant $K_D$ is calculated as $1/K_a$. The Gibbs free energy of binding is calculated from $\Delta G^{\circ}_{48} = -RT \ln (K_d / 1M)$.
Figure 24. Native gels of the 16 genotypes of WXY. Native gel electrophoresis analysis was used to detect tertiary folding variability between WXY RNA genotypes. (Left) Native gel picture where no heat/cool preparation was used prior to RNA loading. This gel illustrates that some of the genotypes do contain alternative structures. (Right) Native gel picture where a heat/cool protocol was used prior to RNA loading. This demonstrates that the heat/cool protocol homogenizes the folding structures.
Table 3. Rate constants, predicted $\Delta G^{\circ}_{48}$, Dissociation rate constants ($K_D$), and, experimental $\Delta G^{\circ}_{48}$ binding for recombinase interactions.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>$k_a$ (min$^{-1}$)</th>
<th>$\Delta G^{\circ}_{48}$ predicted (kcal/mol)</th>
<th>$K_D$ (µM)</th>
<th>$\Delta G^{\circ}_{48}$ binding (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>0.0415</td>
<td>-0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>0.0319</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA</td>
<td>0.0197</td>
<td>0.77</td>
<td>3.7</td>
<td>-9.6</td>
</tr>
<tr>
<td>GC</td>
<td>0.0125</td>
<td>-1.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GU</td>
<td>0.0091</td>
<td>3.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.0069</td>
<td>5.48</td>
<td>7.8</td>
<td>-10.1</td>
</tr>
<tr>
<td>UG</td>
<td>0.0049</td>
<td>2.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UC</td>
<td>0.0038</td>
<td>5.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UU</td>
<td>0.0022</td>
<td>5.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.0020</td>
<td>4.75</td>
<td>40</td>
<td>-11.1</td>
</tr>
<tr>
<td>CC</td>
<td>0.0016</td>
<td>5.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>0.0006</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>0.0005</td>
<td>5.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>0.0004</td>
<td>4.86</td>
<td>55</td>
<td>-11.3</td>
</tr>
<tr>
<td>CU</td>
<td>0.0004</td>
<td>5.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>0.0001</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGS-less</td>
<td></td>
<td>6.2</td>
<td>-9.9</td>
<td></td>
</tr>
</tbody>
</table>
Figure 25. FA ligand and receptor design. (A) Design of the surrogate WXY-Z ligand. The “WXY fragment” (blue) contains the tag (orange) and a 3’deoxythymidine on the 3’ end prevents catalytic reaction. The “N” nucleotide represents an A or C nucleotide. The Z fragment (green) contains the 3’ fluorescent label 6-FAM. The ligand molecular weight is approximately 11 kDa with a $T_m = 90 \, ^\circ C$. (B) Schematic of the surrogate WXY-Z ligand binding to the WXYZ recombinase ribozyme (receptor) at the IGS (red) and tag (orange) interface. A binding interaction also occurs between the G on the 5’ side of the “Z fragment” and the G-binding site on the ribozyme (indicated by dotted lines). The IGS “M” can represent any of the four nucleotides and the IGS sequence within the IGS-less ribozyme is deleted. Molecular weight of the recombinase ribozyme is approximately 64 kDa.
Figure 26. UA binding curve. Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. Here the ribozyme IGS is GUG and the ligand tag sequence is CAU. Curve fitting equation is displayed where \( m_1 = r_o \) (initial anisotropy), \( m_2 = r_{max} \) (maximum anisotropy), and \( m_3 = K_a \) (association rate constant).
Figure 27. AC binding curve. Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. Here the ribozyme IGS is GAG and the ligand tag sequence is CČU. Curve fitting equation is displayed where $m_1 = r_o$ (initial anisotropy), $m_2 = r_{max}$ (maximum anisotropy), and $m_3 = K_a$ (association rate constant).
Figure 28. CA binding curve. Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. Here the ribozyme IGS is GCG and the ligand tag sequence is CAU. Curve fitting equation is displayed where m1 = $r_0$, (initial anisotropy), m2 = $r_{max}$ (maximum anisotropy), and m3 = $K_a$ (association rate constant)
**Figure 29. AA binding curve.** Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. Here the ribozyme IGS is GAG and the ligand tag sequence is CAU. Curve fitting equation is displayed where $m_1 = r_o$ (initial anisotropy), $m_2 = r_{\text{max}}$ (maximum anisotropy), and $m_3 = K_a$ (association rate constant).
Figure 30. IGS-less binding curve. Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. Here the ribozyme IGS is deleted and the ligand tag sequence is CAU. Curve fitting equation is displayed where $m_1 = r_o$, (initial anisotropy), $m_2 = r_{max}$ (maximum anisotropy), and $m_3 = K_a$ (association rate constant)
Figure 31. GA and GC binding curves. Plot of the measured anisotropy versus the concentration of WXYZ ribozyme. (Left) Ribozyme IGS is GGG and the ligand tag sequence is CAU. (Right) Ribozyme IGS is GGG and the ligand tag sequence is CCU. Curve fitting equation is displayed where $m1 = r_o$ (initial anisotropy), $m2 = r_{max}$ (maximum anisotropy), and $m3 = K_a$ (association rate constant).
Figure 32. “GGG” IGS aggregation evidence. Binding curves for both the GA and GC genotypes demonstrated the inability to reach equilibrium potentially due to recombinase aggregation. (A) Secondary and tertiary structure, denoting the W-X and X-Y junctions with arrows where sequences are complimentary to the IGS “GGG” (W-X and X-Y loop sequences are CCC and CCU, respectively). (B) An example gel electrophoresis picture of a “GGG” IGS kinetic assay. All genotype kinetic assays containing the “GGG” IGS display the same degradation bands below the WXY band. It is thought these represent the reverse reaction at the W-X junction and X-Y junction, due to the approximate migration of bands compared to the sequence lengths of the fragments. (C) Native gel electrophoresis comparing “GGG” IGS recombinase ribozyme and “GUG” IGS recombinase ribozyme. Additional bands are seen above the WXYZ band in both gels, however, an increase is seen in the “GGG” IGS with concentration increase.
Chapter 5 – Spatial heterogeneity in the recombinase system

Introduction.

Since the Big Bang, anisotropy and heterogeneity have characterized and developed the universe as we know it. In considering the origins of life, the environmental landscape of the early Earth would have played as much of a role in the outcome of life as the prebiotic agents themselves. The influence of spatial heterogeneity induced in a large variety of forms, from diffusion limitations and surface interfaces to protocells and compartments, has been extensively considered both theoretically and experimentally in abiogenesis. Our studies, typically performed in a homogeneous well-mixed system, are excellent starting points to characterize multiple interacting species. However, it seems prudent to also consider how environmental affects could impact the system.

Heterogeneous systems offer a wide range of life-facilitating phenomena. Physico-chemical factors such as adsorption on clays or dehydration/rehydration around lipids provide more facile routes to abiotic polymerization (Ferris, 1996; Rajamani et al., 2008; respectively). Crowding induced by inert molecules can lead to enhanced catalytic reaction rates (Nakano et al., 2009), stabilization of RNA structures (Kilburn et al., 2010), and can overcome RNA mutation destabilizations (Lee et al., 2015). These crowding benefits can also be induced at phase interfaces, such as aqueous two phase systems (ATPS) (Keating, 2012;
Strulson et al., 2012) or potentially at the air-water interface in aerosols (Woese, 1979; Dobson et al., 2000).

In the evolutionary setting, a homogeneous environment allows natural selection to favor defectors or parasites over cooperators. A spatially structured system can offer a barrier against parasitic invasion and allow for clustering of cooperators. When cooperators remain together the altruistic acts are localized and thus reciprocated among each other. In the well-mixed unstructured system these benefits are dispersed and diluted (Brogioli, 2010). This concept has been demonstrated theoretically in a variety of systems for example on two-dimensional spatial models with replicator populations in silico (Szabó & Szathmáry, 2002) and in hypercycle models (Boerlijst & Hogeweg, 1991; Sardanyes & Sole, 2006; 2007). This has also been demonstrated experimentally with yeast (van Dyken et al., 2013) and bacterial cells (Kerr et al, 2002). Interestingly, van Dyken et al. (2013) showed that a prisoner’s dilemma scenario could be engineered in budding yeast. In the homogeneous aqueous system defectors overtook the cooperator population; however on a solid surface media, cooperators were able to expand and even overtake defector sectors. Through these studies and many more it is thought that the limited interaction environment was a critical factor in the evolution of cooperation (Ohtsuki et al., 2006; Nowak, 2006; Fletcher & Doebeli, 2009).
An additional benefit of the heterogeneous environment is an increase in system diversity. By limiting interactions to local neighbors, beautiful complex patterns have been demonstrated to develop on graphs (Nowak & May, 1992) and empirically increase the diversity of bacteria populations (Chao & Levin, 1981; Kerr et al., 2002).

Whether simple spatial systems were formative in origins of life processes or whether protocells were always present and ushered in life is unknown. Ultimately, however, the current state of modern life is fully encapsulated. Compartments provide the above benefits of spatially distributed systems, as well as concentrate entities, allow for gradient potentials to exist, and physically link the genotype to the phenotype outcome (Deamer, 1997; Szostak et al., 2001).

Additionally, encapsulation creates a potential for intercompartmental competition. This type of group selection or multi-level selection allows for selection to act at the level of the group and can allow for individuals to persist that may have been selected against on an individual basis. An example of this type of phenomenon was originally shown in the stochastic corrector model (Szathmáry & Demeter, 1987) and later modeled using simpler replicator criteria (Markvoort et al., 2014). In these systems cooperators, which would be overcome by defectors in a homogeneous system, are able to evade extinction due to the
components inside compartments displaying higher fitness than those inside defector-containing compartments. When considering a network of entities initiating life, it is this very intercompartmental competition that is thought crucial for the possibility of evolvability (Vasas et al., 2012; Nghe et al., 2015).

Given the impact of spatial structures and encapsulation towards chemical and biological forces, here we investigate systems in which these variables can be characterized within our RNA recombinases. A diffusion-limited environment, surface-like scenario, artificial protocell, and encapsulation environments were developed and the catalytic ability of our recombinase is assayed in the varying environments. We initially used a simple 2-fragment construct (Fig. 33) of the recombinase for method development and perform follow-up studies with a 4-fragment (Fig. 1) and 3-membered cooperative cycle (Fig. 34). We not only demonstrated that the catalytic reaction is possible in these diverse environments, but also that there are physico-chemical effects at play that can alter genotype outcomes from the homogeneous environment.
Results.

**Diffusion limited environment.** In order to establish a diffusion limited environment, we assayed the 2-piece recombinase reaction (Fig. 33) in either 20% polyethylene glycol, molecular weight 8000 (PEG 8 K) or 20% Ficoll 400 K. Both polymers (PEG and Ficoll) are known crowding agents with demonstrated ability to stabilize RNA folding through excluded volume effects (Kilburn et al., 2013; Desai et al., 2014). Additionally we used a 20% dextran solution to evaluate the use of an aqueous two-phase system (ATPS). An ATPS can be created with PEG and dextran, which when mixed at specific ratios can result in biphasic micelles (Keating, 2012). The RNA partitions into the dextran phase of the dextran/PEG ATPS and the system has demonstrated the ability to increase the rate of ribozyme cleavage (Strulson et al., 2012).

Our preliminary diffusion limited environments demonstrate that the 2-piece reaction is functional among all three polymers (Fig. 35). Reaction products were separated on a denaturing polyacrylamide gel and stained for visualization using Sybr green II. When compared to an aqueous reaction there was no qualitative difference observed between the PEG and Ficoll; however, the dextran gel analysis was perturbed and optimization of this system is needed.

We next increased the complexity of the system by performing follow-up studies with the 4-piece recombinase (Fig. 1) in PEG and Ficoll at 20% using the same
methods as the 2-piece reaction. Here we see the reaction is greatly stunted in PEG and no product is detected. The activity in the Ficoll system was also depressed relative to the aqueous system with an estimated 44% WXYZ produced from W compared to 64% WXYZ in the aqueous system (Fig. 36). This result was surprising given that previous studies throughout literature indicate ribozyme reaction rates increase with crowding agents. However, it should be noted that the recombinase system is more complex than the classic substrate enzyme ribozyme reactions evaluated with crowding agents, in that the 4-piece reaction studied here requires multiple components to assemble to form an active enzyme.

In order to probe this reaction further we repeated this experiment with the more highly perturbed system, PEG, and evaluated a range of concentrations from 0%-15% PEG. Additionally, we quantitated reaction results by 5’ labeling the W-fragment with γ^{32}P-ATP. Results indicate that as the PEG concentration increases, the production of WXYZ ribozyme decreases (Fig. 37). This effect may be due to a few factors. First, the recombinase system is an intermolecular assembly and in this case requires four fragments to diffuse together, not a bimolecular substrate-enzyme complex as in previous literature. Second, the four fragments of the WXYZ recombinase need to correctly associate. It is known that crowding agents act by stabilizing RNA folding (Kilburn et al., 2010; 2013), and this stabilization is additionally able to overcome destabilizing mutations (Lee et
al., 2015). The increase in PEG concentrations may lead to an increase in non-native assemblies of the recombinase ribozyme and therefore activity may be decreased. Lastly, in Figure 37 we observe a band in the PEG reactions just above the WX fragment (101 nt) that is not observed in our aqueous reaction. This band could indicate the covalent assembly of a WY (110 nt) or WZ (118 nt) fragment and would represent a type of trapped assembly in that the full length ribozyme would not be able to be produced from these without first proceeding through a reverse reaction. These non-productive complexes have been previously observed (Vaidya et al., 2012), and reverse reactions have demonstrated the ability of the recombinase ribozyme to shuffle these non-productive assemblies back into productive fragments (Vaidya et al., 2013). While this demonstrates great robustness by the ribozyme system, any environment that stabilizes a variation in recombinase ensembles ultimately demonstrates the ability to change the genotype population outcome.

We next performed a 3-membered cooperative cycle assay (Fig. 34) in varying amounts of PEG (0%-15%). This cooperative network actually represents a simpler system than the 4-piece reaction above because each member is a 2-piece assembly reaction. However, as not every ribozyme in the chemical milieu is able to catalyze each genotype with high activity, this system requires the cooperation of all three members. In this cycle the first member is $\text{GUG}_{\text{CGU}} + \text{XYZ}$, the second is $\text{GAG}_{\text{WXCAU}} + \text{YZ}$, and the third is $\text{GCG}_{\text{WXYCUU}} + \text{Z}$ (Vaidya et
al., 2012). While the junctions vary in this 3-membered network, this system would be similar to the MN notation of Chapter 2 where UG acts on AA, which acts on CU, which acts on UG. In Vaidya et al. (2012), the third member \( GCGWXY_{CUU} + Z \) (the same as the CU genotype of chapter 2) was reacted alone as a control to demonstrate limited activity through autocatalysis. Here, we performed this same control in parallel with the cooperative cycle in PEG.

Results for the cooperative cycle assay in PEG again demonstrate a decrease in WXYZ production (Fig. 38). However, the decrease is not nearly as severe as in the 4-piece recombinase reaction. Interestingly, the isolated third member activity is much higher than would be expected for this genotype alone. In Vaidya et al. (2012), this reaction in aqueous solution produced 0.1% WXYZ after 8 hr and was 125-fold lower than the cooperative cycle production of WXYZ. While our assay here only allows a relative comparison of product formation (we used Sybr green for visualization of these results), we see over 30% of the isolated third member \( GCGWXY_{CUU} \) (CU) converted to WXYZ in 15% PEG after 8 hrs. This compares to the cooperative cycle in 15% PEG which produced ca. 40% WXYZ, only a 10% higher yield, which is not anywhere near a 125-fold higher activity. These results could indicate that the PEG is stabilizing the recombinase ribozyme structure allowing “mutated” genotypes (like a non-canonical IGS/tag pairing of CU) to recover activity.
Two-dimensional surface. An alternative approach to limit diffusion (or to manipulate the interaction environment) is possible using two-dimensional surfaces. Empirically this is typically seen as bacteria grown on plates compared to growth in liquid media (Chao & Levin, 1981; Kerr et al., 2002). Chemically, however, this has primarily been a theoretical endeavor (Szabó & Szathmáry, 2002; Boerlijst & Hogeweg, 1991; Sardanyes & Sole, 2006; 2007). As a “proof-of-concept” study we looked to see if we could spatially separate RNA on a 2-D plate surface and track the migration and reactions of our recombinase system.

In Figure 39, we see the passive diffusion of radiolabeled RNA through buffered agarose plates (0.2% agarose, 100 mM MgCl₂, 30 mM EPPS, pH 7.5) incubated at 48 °C from 2hrs. to 24hrs. This visualization demonstrates a 2-D surface where we can envision reactions at the edges of the diffusion patterns.

Using this agarose plate set-up, we spatially separated the recombinase fragments required for the 2-piece reaction at increasing distances (Fig. 40A). The fragments WX and YZ were then allowed to diffuse for 18 hr, after which time the agarose between the separated fragments was removed and assayed. The red lines in Figure 40A designate the excised agarose. The RNA was extracted from the agarose, reverse-transcribed, and amplified via PCR to determine if any full-length WXYZ was produced. Additionally a portion of agarose was removed as a negative control where there should be no RNA present (labeled “negative control” in Fig. 40A). Results are shown in Figure 40B.
and indicate that at 0 cm and 0.5 cm WXYZ production is detected; however, at the larger distances of 1.0 cm and 1.5 cm, no product is visualized.

When the agarose surface structure was evaluated using the more complex 4-piece recombinase reaction, no WXYZ product is observed even at the 0 cm distance (Fig. 41A and 41B). Given that the 4-piece reaction requires four fragments to first associate followed by the non-covalent complex associating with a second complex for three recombination reactions (one at each junction, W-X, X-Y, and Y-Z, Fig. 1), it seems probable that this spatial scenario is too diffuse for the multiple interaction requirements of the 4-piece reaction. The evaluation of the cooperative cycle in this 2-D scenario did produce positive WXYZ results (Fig. 42A and B). However, the ability to distinguish between PCR amplification of autocatalytic network member production or cooperative cycle production will require differential labeling and quantification techniques.

The demonstration of RNA migration and simple ribozyme reactions on a 2-D spatial structure does provide a tantalizing experimental scenario. The ability to spatially separate “cooperators” and “defectors” automatically endows the system with a protection from parasites and would therefore create alternative population outcomes. And, while our above techniques do satisfy the “proof of concept” idea of RNA populations on a surface, in order to produce a systematic study of genotype interactions on a surface we will need further development of the
labeling/quantification of genotypes, and we will need to optimize the agarose physical platform.

**Compartmentalization.** Encapsulation technology development has been incredibly fruitful over the last two decades. From artificial protells made of water-in-oil vortexed emulsions to highly precise microdroplets on a chip, the ability to study a vast variety of systems encapsulated has been made possible (Tawfik & Griffiths, 1998; Leman et al., 2015). Here we demonstrate the ability to encapsulate and study the recombinase system with a facile artificial protocell made in our own lab as well as in a micro- and nano-fluidics device with collaborators at ESPCI ParisTech.

A primitive version of an artificial protocell was previously demonstrated with the *Azoarcus* recombinase using the simple 2-piece reaction (Zenisek et al., 2007) In this study encapsulated RNA was produced by vortexing a small aqueous volume, containing the ribozyme components, in a mineral oil/surfactant solution (Fig. 43A). These experiments demonstrated not only that the 2-piece assembly could occur in compartments but that the fragments could be separately encapsulated such that a reaction would not occur (Fig. 43B). One other interesting find was the demonstration of recombinase activity when the Mg$^{2+}$ was separately encapsulated from the RNA, illustrating the idea that while RNA could
not migrate between compartments the Mg$^{2+}$ resource did have this capability (lane 7, Fig. 43B).

Using the experimental design from Zenisek et al., 2007, we evaluated the affects of compartmentalization on the 3-membered cooperative network (Fig. 34). Our initial experiments compared the cooperative cycle in an unencapsulated aqueous environment, a fully encapsulated network where all components were compartmentalized together, and an encapsulated network with the network members compartmentalized separately. In order to form separate compartments for the three different members of the cycle, $W + XYZ$, $WX + YZ$, and $WXY + Z$, three separate emulsions are created (one for each reaction) and then placed together in the same test tube with gentle mixing. The tube was then incubated either with or without continuous mixing. Results for the encapsulated cooperative cycle indicate that aqueous and encapsulated cycles produce comparative results and a small product band is detected in the separately encapsulated members (Fig. 44). This low activity is thought to be due to autocatalysis and most likely does not indicate that the network was able to establish itself.

To control for the autocatalytic reaction we paired non-matching member fragments within three separate emulsions ($WX + XYZ$, $W + Z$, and $WXY + YZ$) and ran an additional control of the third-member only, $I_3 (WXY + Z)$
encapsulated together (Fig. 45). We again see no appreciable amount of WXYZ product formed and the third-member control qualitatively appears to confirm our prior assumption that product formation in Figure 44 was due to autocatalysis, not network catalysis (Fig. 45). A surprising result from these two experiments was the inability to see product formation with continuous mixing. In Zenisek et al. (2007), the 2-piece recombinase reaction did not occur separately compartmentalized. However, with continuous mixing of the emulsion a comparable reaction occurred to the aqueous control (Fig. 43, lane 6 continuous mixing compared to lane 2 aqueous control). In our experiments with the 3-membered cycle, continuous mixing of the separately compartmentalized fragments produced similar results to those with no mixing of the emulsion. These results indicate the robustness with which primitive compartments are able to keep complex network arrangements separated and this technique may represent a more facile method and realistic prebiotic scenario for producing cooperator/defector barriers.

Lastly, we had the opportunity to collaborate with the Griffith’s lab at ESPCI ParisTech to test the feasibility of the Azoarcus recombinase in a microfluidics in vitro compartmentalization (IVC) scenario (Agresti et al., 2010). This system represents a similar compartment matrix as the above discussed system, an aqueous droplet in oil/surfactant phase. Here, however, the droplets are produced with much greater precision and reproducibility. The droplets are
approximately 15 pL in size and are produced at a rate of 2000/sec. A nanodroplet system was also tested in which a 60 fL droplet was formed. Our initial testing using these systems was very simple and used our best 2-piece recombinase reaction (GCGWXYGU + Z, CG genotype) and the cooperative cycle. Both reactions were performed with all RNA components encapsulated together and the Mg\(^{2+}\) was added as the droplet was formed to initiate catalysis. Results for all IVC experiments (microdropet 2-piece, nanodroplet 2-piece and cooperative cycle reactions) showed comparable catalytic activity between the IVC system and the homogeneous aqueous control reactions (Fig. 46). Further studies using IVC are currently being performed comparing the “slow” and “fast” network from Chapter 3 as well as random recombinase network formation. Here, very low numbers of RNA molecules are being encapsulated (5-10 molecules) and the effects of stochastic distribution on network outcomes are being evaluated. These data are forthcoming and will await development by a future student.
Discussion.

Our results here illustrate three different concepts for introducing heterogeneity, diffusion limitations through crowding, 2D surfaces, and compartmentalization. Each of these systems suggests the ability to alter genotype population outcomes. Spatial separation induced by diffusion limiting crowding agents produces the least amount of heterogeneity as compared to the physical barriers of surfaces and compartments. Yet, the crowding molecules themselves offer an additional physical-chemical effect on genotypes and that is the stabilization of structures with low activity. The results from the 4-piece assembly (indicating non-productive structure formations) and from the cooperative cycle (demonstrating an increase in autocatalytic reaction), suggest that the “rules of the game” in PEG are quite different from those in the aqueous scenario. The 2-D agarose surface represents more of a concept study, yet the migration patterns that emerge (Figure 39) hint at the ability to produce some fascinating population dynamics with some system optimization. Lastly, compartmentalization in both the rudimentary vortexed emulsion and an IVC high-throughput format allows for the complete separation of RNA genotypes to control who interacts with whom (or stochastically interacts with whom). These formats would demonstrate cooperator/defector population changes and would be ideal for studying how group selection dynamics could act in molecules.
It is still unclear how entities may have originally become associated with compartments or the extent that spatially distributed systems could build up chemical complexity. Development of the tools to compare a gradient of increasing spatial heterogeneity provides us with an ability to compare and contrast how these forces can alter the chemical dynamics of interacting molecular species. While we may never know the exact path life took some 4 billion years ago, we can characterize the basic chemical dynamics of ever increasing complexity with both molecules and the environment and potentially illuminate paths that could create life.
Materials and Methods.

RNA preparation. The W, X, Y, Z, WX, YZ, WXY, XYZ molecules were prepared by in-vitro transcription from double-stranded DNA derived from recursive gene synthesis as described previously (Hayden et al., 2005). All RNAs were resuspended in 1–10 μM solutions in 0.1 mM EDTA. Radiolabeled RNAs were 5’-labeled with γ[^32]P•ATP using OptiKinase (USB, Cleveland, OH).

Diffusion limited environment. Recombinase assays were performed with water, PEG, Ficoll, or dextran at the specified %wt/vol and each RNA fragment at 1 μM. Reactions were initiated with the addition of reaction buffer (100 mM MgCl₂ and 30 mM EPPS, pH 7.5) and incubated at 48 °C. Time point samples were drawn and immediately quenched with quench solution (125 mM EDTA and 2X loading dye containing formamide and bromophenol blue) from 0 – 8 hr. Negative controls were prepared with RNA only and no reaction buffer. Samples were loaded on an 8% polyacrylamide/8M urea gel and RNA bands were separated. Reactions without radiolabeled RNA were visualized via SybrGreen II stain and Visualization and quantification was possible via phosphorimaging on a Typhoon Trio+ variable mode phosphorimager (GE Healthcare) and accompanying ImageQuant software (GE Healthcare). A product ratio was calculated by comparing the RNA in the product WXYZ band to the unreacted substrate band (fraction reacted = [reacted / (reacted + unreacted)]).
Two-dimensional surface. Surface environments were simulated on agarose plates (0.2% agarose) containing reaction buffer (100 mM MgCl₂ and 30 mM EPPS, pH 7.5) and wells formed at the specified distances. RNA was pipetted into wells and allowed to passively diffuse for 18 hr at 48 °C. The agarose was excised between wells (as outlined in Figures 40 – 42) and RNA was eluted into 0.3 mM sodium acetate and 10 mM EDTA at 4 °C and subjected to RT-PCR amplification. Amplification of the RNA produces a WXYZ at ca. 230 nt. Negative control was agarose excised from the agarose plates as discussed in the above text. Positive and negative reverse transcription controls (+/- RT) were made performing the RT-PCR with known WXYZ (+ control) or no RNA (- control). Reaction PCR assays and controls were separated by gel electrophoresis on 2% agarose containing ethidium bromide. Samples were run alongside a 50 base pair (bp) ladder to verify product length.

Artificial protocell compartmentalization. Water in oil emulsion assays were performed as described previously (Zenisek et al., 2007; see also Fig. 43), using the cooperative cycle RNA fragments W, XYZ, WX, YZ, WXY, Z (Fig. 34) as illustrated in Figures 44 and 45. All samples were incubated at 48 °C for 4 hr, with or without mixing as specified. Samples were separated on 8% PA/8M urea gels and visualized using SybrGreen II staining as in Diffusion Limited Environment assays.
Drop-based microfluidics. Micro-droplet formation was performed as previously
described (Agresti et al., 2010). The recombinase reaction performed was
\( \text{GCGWXY}_{\text{CGU}} + Z \) and the cooperative cycle recombinase reaction. Reaction
mixtures contained RNA at 1 µM concentration and buffer was added
simultaneously in droplet formation (100 mM MgCl\(_2\) and 30 mM EPPS, pH 7.5).
Reaction containing droplets were collected in a tube containing mineral oil (to
prevent dehydration) on an ice block (to prevent reaction until incubation) and
incubated at 48 °C for 4 hr or 18 hr. Samples were quenched with 125 mM EDTA
and samples centrifuged to separate oil and water phases. Reactions were
separated on 8% PA/8M urea gels and stained for visualization.
**Figure 33. 2-piece recombinase reaction** $\text{GUGWX}_{\text{CAU}} + \text{YZ}$. Fragments $\text{GUGWX}_{\text{CAU}}$ and $\text{YZ}$ first associate via base pair interactions to form a non-covalent WXYZ molecule. The non-covalent WXYZ is then catalytically connected through one reaction at the X-Y junction by another molecule of non-covalently or covalently linked WXYZ recombinase.
Figure 34. Cooperative cycle recombinase reaction. Reaction scheme of the recombinase cooperative cycle adapted from Vaidya et al. (2012). In this cycle the first member is $GUGW_{CGU} + XYZ$, the second is $GAGWX_{CAU} + YZ$, and the third is $GCGWXY_{CUU} + Z$ (Vaidya et al., 2012). The fragments each first associate via base pair interactions to form a non-covalent $WXYZ$ molecule ($I_{in}$). The non-covalent $WXYZ$ (notated with the $I_{in}$ notation) is then catalytically connected through one recombination reaction at the $W-X$ junction for $I_1$, the $X-Y$ junction for $I_2$, and the $Y-Z$ junction for $I_3$. The covalent assembly is then notated $E_{in}$. While all genotypes have some catalytic ability, the strongest catalysis occurs when $I_1$ or $E_1$ catalyzes $I_2$, when $I_2$ or $E_2$ catalyzes $I_3$, and when $I_3$ or $E_3$ catalyzes $I_1$. 
Figure 35. 2-piece recombinase reaction $\text{GUGWXCAU} + \text{YZ}$ in aqueous, PEG, Ficoll, and Dextran environments. Bands are RNA fragments following assays, gel electrophoresis and staining via Sybr Green II. A comparison of the 2-piece recombinase reaction $\text{GUGWXCAU}$ and YZ in varying crowding agents over time (0 – 8 hr), separated on an 8% PA/8M urea gel and visualized using SybrGreen II. The aqueous, PEG, and Ficoll systems produce ca. 50% conversion of WX to WXYZ. The Dextran system was not quantitated.
Figure 36. 4-piece recombinase reaction $GUGW_{CAU} + X_{CAU} + Y_{CAU} + Z$ in aqueous, PEG, and Ficoll environments. A comparison of the 4-piece recombinase reaction in varying crowding agents over time (0 – 8 hr), separated on an 8% PA/8M urea gel and visualized using SybrGreen II. The aqueous reaction results in approximately 64% WXYZ as converted from W, no product was detected in PEG, and Ficoll resulted in approximately 44% WXYZ.
Figure 37. 4-piece recombinase reaction in varying concentrations of PEG. (Left) A comparison of the 4-piece recombinase reaction in PEG from 0% – 15% over time (0 – 8 hr). W fragments were 5’-radiolabeled with γ[^32P]-ATP and reaction products were separated on an 8% PA/8M urea gel. (Right) Quantitated reaction results.
Figure 38. Cooperative cycle recombinase reaction in varying concentrations of PEG. (left) A comparison of the cooperative cycle in PEG from 0% – 15% and I₃ reaction in 15% PEG (I₃ = GCGWXYCUU + Z). Reaction samples were drawn and quenched from time 0 – 8 hr, separated on an 8% PA/8M urea gel and visualized using SybrGreen II. (right) Approximated conversion of product for relative comparison between assays as the amount of WXY converted to WXYZ.
Figure 39. Diffusion of radiolabeled RNA fragments on 0.2% agarose plates. A comparison of passive diffusion of radiolabeled RNA 4-piece recombinase fragments W, X, Y, Z, and WXYZ from 2 hr – 24 hr, incubated at 48 °C. Plates were allowed to expose for 1 hr at 4 °C for visualization.
Figure 40. 2-piece recombinase reaction on a 2D agarose surface. (A) Schematic of the experimental setup for the 2-piece recombinase surface reaction. Circles represent wells that were made in the 0.2% buffered agarose surface in which RNA fragments were placed, as depicted. Red boxes indicate the excised agarose area for RNA extraction. Plates were incubated for 18 hr at 48 °C prior to RNA extraction. Eluded samples underwent RT-PCR to detect WXYZ product formation (B) Gel electrophoresis of RT-PCR samples and controls, visualized with ethidium bromide.
Figure 41. 4-piece recombinase reaction on a 2D agarose surface. (A)
Schematic of the experimental setup for the 4-piece recombinase surface reaction. Circles represent wells that were made in the 0.2% buffered agarose surface in which RNA fragments were placed, as depicted. Red boxes indicate the excised agarose area for RNA extraction. A similar experiment was performed where only the middle section of agarose was excised between wells with negative results (not pictured). Thus here, the full agarose area between wells was excised in order to determine if product formation occurred. Plates were incubated for 18 hr at 48 °C prior to RNA extraction. Eluded samples underwent RT-PCR to detect WXYZ product formation (B) Gel electrophoresis of RT-PCR samples and controls, visualized with ethidium bromide.
Figure 42. Cooperative cycle recombinase reaction on a 2D agarose surface. (A) Schematic of the experimental setup for the cooperative cycle recombinase surface reaction. Circles represent wells that were made in the 0.2% buffered agarose surface in which RNA fragments were placed, as depicted. Red boxes indicate the excised agarose area for RNA extraction. Plates were incubated for 18 hr at 48 °C prior to RNA extraction. Eluded samples underwent RT-PCR to detect WXYZ product formation (B) Gel electrophoresis of RT-PCR samples and controls, visualized with ethidium bromide.
Figure 43. 2-piece recombinase reaction in artificial protocells. This figure is taken directly from Zenisek et al. (2007) for comparison to data in Figs. 44 & 45. (A) Schematic of the protocell system. RNA aqueous solutions are vortexed into a mineral oil surfactant mixture for 5 min at 4 °C to form emulsification. Reactions are incubated at 48 °C for 4 hr with or without mixing as specified. (B) Gel electrophoresis of protocell recombinase reactions and controls. Lane 1, negative control containing RNA fragments, WX and YZ in aqueous solution with no buffer (Mg²⁺). Lane 2, unencapsulated aqueous control containing the WX and YZ fragments with buffer in aqueous solution. Lane 3, encapsulated WX, YZ, and buffer encapsulated together. Lanes 4 – 6, WX and YZ fragments encapsulated separately with buffer under increasing mixing conditions. Lane 7, WX and YZ fragments encapsulated separately from buffer.
Figure 44. Cooperative cycle recombinase reaction in artificial protocells.

Gel electrophoresis of protocell cooperative cycle recombinase reactions and controls. Lane 1, WXYZ size control. Lane 2, negative control containing RNA fragments in aqueous solution with no buffer (Mg$^{2+}$). Lane 3, unencapsulated aqueous control containing all fragments with buffer in aqueous solution. Lane 4, all fragments encapsulated together with buffer. Lanes 4 – 5, cooperative cycle fragments that produce I$_1$, I$_2$, and I$_3$ encapsulated separately with buffer without and with mixing.
Figure 45. Cooperative cycle recombinase reaction in artificial protocells with alternative encapsulation. Gel electrophoresis of protocell cooperative cycle recombinase reactions and controls. Lane 1, WXYZ size control. Lane 2, negative control containing RNA fragments in aqueous solution with no buffer (Mg$^{2+}$). Lane 3, unencapsulated aqueous control containing all fragments with buffer in aqueous solution. Lane 4, all fragments encapsulated together with buffer. Lanes 4 – 5, cooperative cycle fragments encapsulated in alternative arrangements. Fragments are encapsulated separately as follows: WX and XYZ, W and Z, WXY and YZ, each is encapsulated with buffer and are incubated without and with mixing. Lane 7, I$_3$ fragments only encapsulated with buffer.
Figure 46. Compartmentalization using droplet based microfluidics system.

(A) Gel electrophoresis results of the 2-piece recombinase reaction $GCGWXY_{CGU} + Z$. The negative control contains RNA fragments without buffer. “Batch” assay indicates an aqueous recombinase reaction with no compartmentalization, incubated at 4 hr and 18 hr. “Encapsulated” assays are recombinase reactions performed inside of micro-droplets, incubated at 4 hr and 18 hr. (B) Gel electrophoresis results of the same 2-piece recombinase reaction where the encapsulated reaction is performed in nano-droplets. (C) Gel electrophoresis results of the cooperative cycle recombinase reaction under the same conditions as (A).
References.


Kliegman, J. I., Griner, S. L., Helmann, J. D., Brennan, R. G., & Glasfeld, A. The Structural Basis for the Metal Selective Activation of the Manganese Transport Regulator of *Bacillus subtilis,* *Biochemistry, 45*(11), 3493–3505 (2006).


Rushdi, A.I. and Simoneit, B. Lipid formation by aqueous Fischer-Tropsch type synthesis over a temperature range of 100 to 400°C. *Origins of Life and Evolution of Biospheres*, 31, 103–118 (2001).


Appendix - Mathematical modeling

**Derivation of the kinetic equation**

To describe the dynamics of the serial dilution experiment, we derive a simple ODE model. We consider a contest with two strategies, A and B. The payoffs are given by the matrix:

\[
\begin{array}{cc}
A & B \\
A & a & b \\
B & c & d \\
\end{array}
\]

In our context all entries are positive: \(a, b, c, d > 0\).

We use the following variables: \(x_0\) is the concentration of the A precursor (WXY), \(y_0\) is the concentration of the B precursor (WXY), \(z\) denotes the concentration of the Z molecule, \(x_r\) is the complex formed between the A precursor and Z, \(y_r\) is the complex formed between the B precursor and Z. The variables \(x\) and \(y\) denote the concentration of the A and B molecule (WXYZ), respectively.

In the dilution experiment, the two precursor molecules and the Z molecule are provided at constant level, and the complex is formed in a reversible chemical reaction

\[
\begin{align*}
\text{x}_0 + z & \rightleftharpoons x_1 \\
\text{y}_0 + z & \rightleftharpoons y_1
\end{align*}
\]

Thus, \(x_r\) and \(y_r\) are also provided at constant level, and they give rise to the respective WXYZ molecule according to the catalyzed reactions

\[
\begin{align*}
x_1 & \rightarrow x \\
x_1 & \rightarrow x \\
y_1 & \rightarrow y \\
y_1 & \rightarrow y
\end{align*}
\]

Therefore, the kinetic equation (that would occur in a flow reactor) is

\[
\begin{align*}
\dot{x} & = (ax + by)x_1 - \phi x \\
\dot{y} & = (cx + dy)y_1 - \phi y,
\end{align*}
\]

Here \(\phi\) is a parameter chosen such that the concentration of A and B is constant, \(x + y = C\) and \(\dot{x} + \dot{y} = 0\). Without loss of generality, we can set \(C=1\) (we only need to replace the variables \(x\) and \(y\) by the transformed variables \(x/C\) and \(y/C\)). In that case, \(\phi\) can be calculated as

\[
\phi = (ax + by)x_1 + (cx + dy)y_1.
\]
Moreover, since \( x_1 \) and \( y_1 \) are provided at equal concentrations, we may set \( x_1 = y_1 = 1 \). This may lead to a change of the time scale, but it leaves the trajectories of Eq. [3] unchanged. Thus, the dynamical equation simplifies to

\[
\begin{align*}
\dot{x} &= ax + by - \phi x \\
\dot{y} &= cx + dy - \phi y,
\end{align*}
\]

with \( \phi = (a + c)x + (b + d)y \).

**Properties of the dynamical equation**

In the following, we list a few interesting properties of the dynamical equation.

**Property 1.**

The dynamical equation [1] has a unique and globally stable equilibrium \((\hat{x}, \hat{y})\) with \(0 < \hat{x}, \hat{y} < 1\) and \(\hat{x} + \hat{y} = 1\). In the generic case that \(a + c \neq b + d\), the equilibrium frequency of \(\hat{x}\) is given by

\[
\hat{x} = \frac{a + b - d + \sqrt{(a - d)^2 + 4bd}}{2(a + c - b - d)}.
\]

Otherwise, if \(a + c = b + d\), the equilibrium frequency is \(\hat{x} = b / (b + c)\).

**Proof.** Since \(x + y = 1\), the first equation in [1] can be written as

\[
\dot{x} = ax + b(1 - x) - \phi x = b + (a - 2b - d)x - (a + c - b - d)x^2 =: f(x).
\]

The function \(f(x)\) has the unique zero \(\hat{x}\) in the unit interval \((0,1)\). Since \(f(0) = b > 0\), it follows that \(f(x) > 0\) when \(0 \leq x < \hat{x}\), whereas \(f(x) < 0\) when \(\hat{x} < x \leq 1\). Therefore, for any given initial frequency \(x\), orbits converge towards \(\hat{x}\). \(\square\)

Let us next explore how the position of the equilibrium is affected by the entries of the payoff matrix. The following results follow directly from Eq. [2].

**Property 2.**

1. The equilibrium frequency \(\hat{x}\) is strictly increasing in \(a\) and \(b\), and strictly decreasing in \(c\) and \(d\).
2. All other parameters unchanged, \(a \to \infty\) or \(b \to \infty\) implies \(\hat{x} \to 1\), whereas \(c \to \infty\) or \(d \to \infty\) implies \(\hat{x} \to 0\).
3. The equilibrium frequency satisfies \(\hat{x} = 1/2\) if and only if \(a + b = c + d\). Similarly, it satisfies \(\hat{x} > 1/2\) if and only if \(a + b > c + d\).

As a consequence of the previous result, we can also draw the following connection between the equilibrium frequency and the type of game considered.

**Property 3.**

1. In a Dominance or Counter-Dominance scenario, \(\hat{x} > 1/2\) if and only if it is \(A\) that dominates \(B\).
2. In a Selfish scenario, $\hat{x} > 1/2$ if and only if $A$ risk-dominates $B$.
3. In a Cooperation scenario, $\hat{x} > 1/2$ if and only if $A$ is also played with higher frequency in the symmetric Nash equilibrium.

**Dynamical equation for $n \times n$ contests**

To describe the dynamics of the rock-paper-scissors contest, we generalize the previous dynamical equation to arbitrary $n \times n$ contests. Let $M = (m_{ij})$ be the payoff matrix of such a contest, and let $x = (x_1, \ldots, x_n)^T$ be the vector that gives the frequency of each WXYZ molecule, such that $x_1 + \cdots + x_n = 1$. Then the $n$-strategy analogue of Eq. [1] is

$$\dot{x} = Mx - \phi x,$$

with $\phi = \sum_{i,j} m_{ij} x_j$. Equation [4] can be considered as a slightly generalized version of the quasi-species equation. It has a unique fixed point in the interior of the state space, which is globally stable. The fixed point can be found by solving the eigenvector problem $Mx = \lambda x$, where $\lambda$ is the largest eigenvalue of $M$. The theorem of Perron and Frobenius for positive matrices guarantees that the corresponding normalized eigenvector $x$ is unique, and that all entries of $x$ are positive.

In the following, let us summarize a few simple properties of the dynamical equation [4].

**Property 4.**

1. The unit simplex $\Delta = \{ x \in \mathbb{R}^n \mid x \geq 0, x_1 + \cdots + x_n = 1 \}$ is invariant under the dynamical equation [4]; that is, if the initial state $x(0) \in \Delta$ then $x(t) \in \Delta$ for all times $t$.
2. The edges of the unit simplex are not invariant under the dynamics in [4]; if $x_i = 0$ then $\dot{x}_i > 0$.

**Proof.** The sum of all entries of $x$ does not change over time, due to our choice of $\phi$. Moreover, if $x_i = 0$, then it follows from Eq. [4] that $\dot{x}_i = (Mx)_i - 0 > 0$.

The previous result points to an important difference between the dynamical equation [4] and replicator dynamics (see also Fig. S6). Replicator dynamics is *non-innovative* – if a strategy is initially absent, then the evolutionary dynamics does not introduce this strategy at some later time point. In contrast, the kinetic dynamics described in [4] predicts that absent WXYZ strategies are introduced immediately, due to the catalytic effect of the other WXYZ molecules (provided that the required precursor WXY for the absent strategy is available).
Next, let us describe the relationship between the Nash equilibria of a contest with payoff matrix $M$, and the unique equilibrium of the kinetic equation [4].

**Property 5.**

Let $M = (m_{ij})$ be the payoff matrix of an $n$-strategy contest, and let $x = (x_1, \ldots, x_n)^T$ be the unique equilibrium of Eq. [4]. Then the following are equivalent:

1. $x$ is a Nash equilibrium.
2. The equilibrium is in the center of the simplex, $x = (1/n, \ldots, 1/n)^T$.
3. The row sums of $M$ coincide, $m_{i1} + \cdots + m_{in} = m_{j1} + \cdots + m_{jn}$ for all $1 \leq i, j \leq n$.

**Proof.**

1 $\Rightarrow$ 2. As $x$ is a Nash equilibrium in the interior of the state space, all strategies yield the same expected payoff,

$$(Mx)_i = (Mx)_j \text{ for all } 1 \leq i, j \leq n.$$  

Moreover, since $x$ is the fixed point of Eq. [4], and hence the eigenvector of $M$ corresponding to some real eigenvalue $\lambda > 0$, it follows that

$$\lambda \cdot x_i = (Mx)_i = (Mx)_j = \lambda \cdot x_j.$$  

In particular, $x_i = x_j$ for all $1 \leq i, j \leq n$, and thus $x = (1/n, \ldots, 1/n)^T$.

2 $\Rightarrow$ 3. As $Mx = \lambda x$ for some $\lambda > 0$ and for $x = (1/n, \ldots, 1/n)^T$, it follows that

$$\frac{1}{n} (m_{i1} + \cdots + m_{in}) = \lambda/n \text{ for all } 1 \leq i \leq n.$$  

Since the right hand side does not depend on $i$, neither does the left hand side. Therefore,

$$m_{i1} + \cdots + m_{in} = m_{j1} + \cdots + m_{jn} \text{ for all } 1 \leq i, j \leq n.$$  

3 $\Rightarrow$ 1. If the row sums of $M$ coincide, it is easy to check that $x = (1/n, \ldots, 1/n)^T$ is the unique fixed point of Eq. [4]. In this fixed point, $(Mx)_i = (Mx)_j$ for all $1 \leq i, j \leq n$, and hence $x$ is a Nash equilibrium.

The previous result shows that in general, the kinetic equilibrium of Eq. [4] is not a Nash equilibrium – the only exception occurs when all rows of the payoff matrix sum up to the same value.