מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Doctor of Philosophy

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

By
Omer Markovitch

מאת
עומר מרקוביץ'

כימיה-מערכתית כגישה לחקר אבולוציה פרֶהביוטית

Systems-Chemistry approach to prebiotic evolution

Advisor:
Doron Lancet

מנחה:
דורון לנצט

30 March 2014

כ"ח אדר ב' תשע"ד

Contents

1. ABSTRACT 4

2. LIST OF KEY SYMBOLS AND ABBREVIATIONS 6

3. INTRODUCTION 7

3.1. Origins of life scenarios 8

3.1.1. The RNA world 8

3.1.2. The metabolism-first scenario 9

3.2. Mutual catalysis 10

3.3. The lipid world 10

3.4. The GARD model 12

3.5. GARD Composomes and compotypes 13

3.6. Spontaneous chiral symmetry breaking in early molecular networks 14

4. METHODS 17

4.1. Computer simulations 17

4.2. Compotypes 17

4.3. Selection 18

4.4. Population dynamics 19

4.5. Logistic equation and fit to population data 20

5. RESULTS 22

5.1. Excess mutual catalysis is required for effective evolvability 22

5.1.1. Network motifs and their effect on selection in GARD 47

5.1.2. A completely selfish β 49

5.2. Chemistry biased fission 51

5.2.1. Chemistry biased fission algorithms 51

5.2.2. Analyzing GARD's fission behavior 52

5.3. A physiochemical realistic GARD model 54

5.4. Ecological dynamics of GARD population 58

5.4.1. GARD population exhibit natural-like dynamics 59

5.4.2. Compotype intrinsic molecular repertoire 59

5.4.3. Compotype replication fidelity 62

5.4.4. Takeover 64

5.4.5. On the choice of specific logistic formulation 68

5.4.6. Broadening the analysis of selection in GARD 71

5.4.7. Network motifs and their effect on ecological dynamics 73

5.4.8. Discussion 77

5.5. Compotypes are quasispecies 78

5.5.1. Sequential vs. compositional information 79

5.5.2. A compotype is an attractor in compositional space 80

5.5.3. GARD operates near its error threshold 81

5.5.4. GARD’s steady state is correlated with that of the quasispecies equation 82
5.5.5. GARD’s error-catastrophe resembles that of the quasispecies equation 84
5.5.6. The time dependent dynamics of the quasispecies equation resembles that of GARD 84
5.5.7. How the β eigenvector was calculated..87
5.5.8. How the compositional space was sampled and Q and A constructed 88
5.5.9. Conclusions 90
5.6. Is there an optimal level of open-endedness in prebiotic evolution? 91
5.6.1. Universe-GARD...94
5.6.2. Early results...95
6. DISCUSSION 97
6.1. Compotypes as quasispecies...97
6.2. GARD is a minimally living system...97
6.3. Evolution towards lower entropy of compotypes..98
6.4. Lack of GARD experiments..98
7. REFERENCES...101
8. PUBLICATIONS STEMMING FROM THIS THESIS..111
9. DECLERATION...111
10. FUNDING...112

1. ABSTRACT

The puzzle of the origin of life is grand. A major challenge is to understand the transition from a mixture of molecules into an entity with basic life faculties, such as a protocell, capable of self-replication and inheritance. Two major schools tackle this problem: the genetic, or replicator-first approach, and the metabolism-first approach. The replicator-first approach focuses on a single self-perpetuating informational biopolymer, e.g., RNA, as the first step, and it is thus often referred to as the "RNA world". In contrast, the metabolism-first approach focuses on a network of chemical reactions among simpler chemical components that became endowed with some reproductive characteristics as the first step that led to a protocell. The lipid world scenario, largely initiated by our laboratory, delineates a specific example of metabolism first. It suggests that spontaneously forming assemblies of relatively simple molecules, such as mutually interacting lipids, that resemble primitive metabolism, are capable of storing and transmitting information similar to sequence-based polymeric RNA, except that in this case it is compositional information that is at work.

This thesis is about further exploration of the lipid world scenario, showing in more detail how a relatively simple chemical system can acquire features such as selection and evolution. This was accomplished by studying dynamical aspects of the graded autocatalysis replication domain (GARD) computer-simulation lipid world model, previously developed at our laboratory. GARD simulates the homeostatic growth of a compositional amphiphile assembly by reversible accretion from a buffered heterogeneous external pool. This process is governed by a network of mutually catalytic reactions, and exhibits quasi-stationary compositional states termed compotype, that may be regarded as GARD species.

I have demonstrated that that such GARD species exhibit positive as well as negative selection, an important prerequisite of a minimally living system. I further showed that when the catalytic network becomes dominated by mutual catalysis, as opposed to self-catalysis, selection is enhanced. When studying the dynamics of large populations of GARD assemblies under constant population conditions, I rewardingly found that they exhibit dynamics similar to natural ecosystem populations, e.g. similes of competition or predator-prey dynamics. I was able to establish relationships between a compotype's internal molecular parameters (e.g. its molecular diversity) and population ecology behavior. In a separate vein, I have developed a new approach towards observing open-ended evolution, which enables asking whether there is an optimal level of open endedness in prebiotic evolution. Finally, I was able to show clear similarities between GARD compotypes and quasispecies in the Eigen-Schuster model for evolution, further underlining GARD's capacity as an alternative to RNA World. Taken together, these results uncover quantitative aspects of the GARD model which in turn contribute towards our understanding of the origin of life via the lipid world scenario.

חידת ראשית החיים הינה מן הגדולות. אתגר עיקרי הוא להבין את המעבר מתערובת של מולקולת לישות בעלת כושר חיים, כגון הקדם-תא (protocell), המסוגלת לשכפול עצמי והורשה. שני גישות תוקפות בעיה זו: הגנטית או השכפול-תחילה, והמטבוליזם-תחילה. תרחיש השכפול-תחילה מתמקד בביופולימר יחיד המעתיק עצמו, כגון רנ"א (ולכן לעיתים קרובות נקראת "עולם הרנ"א"). לעומת זאת, תרחיש המטבוליזם-תחילה מתמקד ברשתות של ראקציות כימיות בין מרכיבים כימיים פשוטים שפיתחו יכולות העתקה עצמית כצעד הראשון שהוביל לקדם-תא. עולם הליפידים, שפותח ברובו במעבדתנו, מתאר דוגמא ספציפית של מטבוליזם-תחילה. הוא מציע שיצירה ספונטנית של צברים המורכבים ממולקולת פשוטות, כגון ליפידים העוברים אינטראקציות הדדיות, הדומים למטבוליזם קדמוני, מסוגלים ולהעביר מידע בדומה למידע רצפי של רנ"א, בהבדל אחד - שבמקרה כזה מדובר במידע הרכבי.

תזה זו נסבה על פתוח מחקרי נוסף של תרחיש עולם הליפידים, ומראה בייתר פירוט כיצד מערכת כימית פשוטה יכולה לפתח מאפיינים כגון סלקציה ואבולוציה. זה נעשה על ידי חקירת היבטים דינמיים של המודל הממוחשב גארד (GARD) לעולם הלפידים, שפותח בעבר בקבוצתנו. גארד מדמה גדילה משמרת של צברים-הרכביים המורכבים ממולקולות אמפיפיליות המתאספות מסביבה אחידה בעלת רבגוניות כימית. תהליך זה נשלט על ידי רשת של יחסי גומלין הדדיים, ומראה מצבים כמו-קבועים המכונים קומפוטייפים (compotypes), ואשר ניתן להתייחס אליהם כאל דמויי מינים ביולוגיים במודל.

במחקרי הראיתי שהמינים של גארד מראים סלקציה חיובית וכן שלילית, שהיא תנאי הכרחי לחיים מינימאליים. יתר על כן, הראיתי שככל שהרשת נשלטת יותר על ידי קטליזה הדדית, בניגוד לקטליזה עצמית, סלקציה זו מוגברת. כאשר חקרתי את הדינמיקה של אוכלוסיות של צברי גארד בעלות גודל אוכלוסייה קבוע, למרבית הסיפוק גיליתי שאוכלוסיות אלה מראות דינמיקה הדומה לזו של מערכות אקולוגיות בטבע, מעין דינמיקת תחרות או טורף-נטרף. הצלחתי לבסס קשר בין פרמטרים מולקולריים פנימיים של קומפוטייפ (כגון מגוון מולקולרי) להתנהגות שלו באוכלוסייה. במסלול מחקרי אחר, פיתחתי גישה חדשה לבחינה של אבולוציה פתוחה (open ended evolution), המאפשרת לשאול האם יש רמה אופטימלית של פתיחות באבולוציה פרהביוטית. לבסוף, הצלחתי להראות דמיון ישיר בין הקומפוטייפים של גארד לכמו-מינים (quasispecies) שבמודל אייגן-שוסטר לאבולוציה. במבט כולל, תוצאות אלה מגלות היבטים כמותיים במודל הגארד, באופן התורם להבנה משופרת של תרחיש עולם הליפידים לתיאור ראשית החיים.

2. LIST OF KEY SYMBOLS AND ABBREVIATIONS

LUCA	Last universal common ancestor
GARD	Graded autocatalysis replication domain
Composome	A set of consecutive faithfully replicating assemblies
Compotype	Composome type
β	Network of rate-enhancement values
QSS	Quasi stationary state
N_C	Compotype count
N_G	Size of environmental molecular repertoire
N_{max}	Assembly pre-fission size
H	Compositional similarity
N_{mol}	Compotype intrinsic molecular repertoire size
r	Compotype intrinsic growth rate
K	Compotype carrying capacity
α	Competition parameter
V_β	The eigenvector of β, with the largest real eigenvalue

3. INTRODUCTION

The conundrum of how life began has captivated mankind for ages. The question is how life could arise from non-living matter, which might be one of the most important questions in science, still unanswered. The origin of life field attempts to answer this question, by combining supramolecular and prebiotic chemistry with theoretical biology and complex systems research, otherwise known as a systems chemistry approach. Thus, the origin of life is perhaps the most exhaustive systems chemistry "experiment" [61, 104, 112].

The origin of life is about the emergence of the first entity with minimal life faculties, and can be delineated to have occurred along the following timeline. At the accretion of planet Earth, some 4.5 billion years ago, it was a hot molten body incapable of sustaining life. Once the planet cooled and hydrated by cometary infall some 4.0 billion years ago, the conditions for the origin of life were in place [146, 150]. Jumping forward, the oldest widely agreed upon cellular fossil record is dated to about 3.5 billion years ago [117]. The exact time is disputed, as older cell-like fossils have been reported [15], which seem to resemble cyanobacteria by size and appearance [91]. However, there is no concrete knowledge on the molecular consistency of such fossils.

A term often used in the context of life's origin is the "last universal common ancestor" (LUCA), an organism which is at the common base of the phylogenetic tree of life, and possessing molecular machinery fundamentally similar to present-day life, i.e. genome, genetic code and ribosome-like translation apparatus, as well as proteins, including enzymes that control an elaborate biosynthetic metabolism [39, 62]. As looking at fossils does not reveal the inner structure, it is extremely difficult to use the cellular fossil record to time the emergence of LUCA. It is however absolutely obvious that LUCA must have emerged by a lengthy evolutionary process in a continuum way, passing through intermediate forms likely to have been different, and much simpler than LUCA. Such intermediate LUCA ancestors are often loosely referred to as protocells [17, 105, 130, 135] (Figure 1).

Concepts about life's origin strongly depend on life's definition. A widely accepted definition of minimal life comes from NASA: life is a self-sustaining system capable of undergoing Darwinian evolution [9], and other definitions are often similar [139]. This definition is general, hence a minimally living entity needs not be a cell as we know it, i.e. LUCA, but could be a much simpler protocell, i.e. container with some necessary molecular content. In this thesis I consider a NASA-consistent entity that is even more primitive than a protocell, such as a compositional lipid micelle or very small vesicle, without any content.

The state of earth surface during the origins of life is often referred to as "primordial (or prebiotic) soup", a term coined by Oparin [94]. This represents a body of water with organic chemical building blocks. The source of earth's water could be adsorption during the planet's accretion [26], or from icy comets [75]. The source of organics could be comet infall, which has been suggested to occur at a significant rate during early earth [18]. Alternatively, organics could be synthesized on early earth from inorganic compounds [20]. In a seminal experiment, Miller showed how in an environment similar to the assumed earth's early atmosphere, a plethora of organics, including amino acids, could form [54, 74, 86]. So, most researchers agree that prebiotic earth contained water teaming with simple carbon-based molecules. The big question is how the transition from prebiotic soup to a functioning protocell occurred, which is elaborated in the next sections.

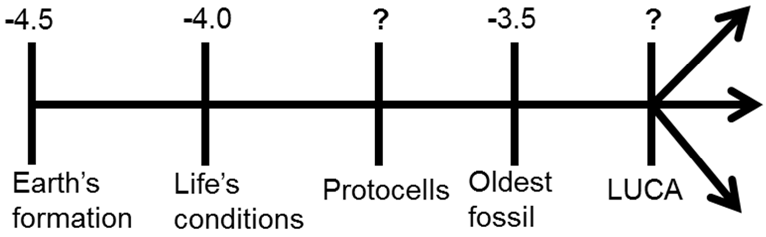

Figure 1: A general timeline of life on Earth. Time is given in billions of years from the present (year 2014).

3.1. Origins of life scenarios

The path from the organic mixtures in the primordial soup to life- has been dominated by two views: "replicator-first", influenced by the present day genetic machinery, and "metabolism-first", stemming from a molecular network similar to present day metabolism. While the first scenario necessitates the early appearance of long and relatively complex information carrier, i.e. self-copying or self-perpetuating biopolymer, the second can go a long way with early chemistry that includes only mutually interacting simple chemical components such as carbohydrates, amino acids, peptides and lipids.

3.1.1. The RNA world

A detailed and widely accepted example of replicator-first is the RNA world. It assumes that a molecule identical or very similar to present day RNA played the role of the self-perpetuating biopolymer [36, 37, 43, 55]. The "free-floating" or surface-adsorbed mixture of such molecules is assumed to have later evolved both a metabolic network and an encompassing container. The wide appeal of RNA as a precursor molecule is understandable, as RNA is capable of both information storage and propagation and the manifestation of certain catalytic activities typical of metabolism. One of the earliest supports for the RNA world was Spiegelman's experiment, which demonstrated that RNA can be copied in vitro, aided by a simple viral enzyme, Qβ RNA replicase [87, 88, 131]. Later, a key finding supporting the RNA World concept was that concrete demonstration that RNA can manifest certain catalytic activities. Such catalytic RNA is termed ribozyme [66]. The first ribozymes discovered were capable of self-splicing and cleavage [16, 41, 110]. This finding was so surprising because up until then it was believed that only proteins were capable of catalytic activities, and indeed such a finding led to the awarding of the 1989 Nobel prize in chemistry[1]. Recent experiments demonstrate more and more elaborate features of ribozymes. In one example, two R3C RNA ligases with complementary sequences were modified such that each was able to catalyze the other's synthesis in a potentially self-sustained manner [76]. In another example, the self-replicating Azoarcus ribozyme was modified in two ways, one where tagged copies of itself cross- catalyze other copies and another where the tagged copies catalyze their own replication [45, 140]. It was found that the system with the mutually interacting ribozymes outperforms the selfish ones. Thus, the wide appeal of the RNA-world is understandable, though such experiments are best put in perspective using Spiegelman's own words: "When you create a living object the presumption is that the object didn't exist before. This I did not do. Working with simple chemical compounds, I take a primer of a living object and I generate many living objects from it"[2].

The "holy grail" of the RNA world is a ribozyme being able to replicate itself from a pool of its constituting nucleotide monomers. This has not been attained yet. A criticism of the RNA world scenario is that demonstrating the formation of nucleotide monomers under abiotic conditions is challenging. This is because nucleotide synthesis requires the binding together of phosphate, sugar and a nitrogenous base, thought recent studies show that it may be synthetically possible [2, 103]. Another synthetic challenge is the polymerization of such nucleotides to form long hetero-polymers, which also recently has been suggested to be synthetically plausible [14, 50]. So, the RNA world is not without problems, which can be generally put as requiring complex initial conditions [11, 96]. The metabolism-first notion attempts to overcome this

3.1.2. The metabolism-first scenario

The metabolism-first scenario suggests that the very first life precursors are likely to have been relatively elaborate molecular networks of simple organic molecules [4, 27, 79, 118, 123]. The reverse citric acid cycle (reverse Krebs cycle), during which carbon compounds are formed from carbon dioxide and water, is one example. Krebs cycle is a common mode of oxidative degradation in eukaryotes and prokaryotes, which uses eight different enzymes and some of the steps are catalyzed by the interim products [145]. Thus, demonstrating the reverse citric acid cycle under plausible prebiotic conditions is important for understanding the origin of life, not

[1] The Nobel Prize in Chemistry 1989 was awarded jointly to Sidney Altman and Thomas R. Cech "*for their discovery of catalytic properties of RNA*".

[2] Taken from: Profiles in Science by the National Library of Medicine (http://profiles.nlm.nih.gov).

just because it represents an autocatalytic cycle (see below), but also because it is a source of carbon molecules necessary for a cells' survival [21, 42, 114]. Another example is the formose reaction, during which sugars are formed from formaldehyde. It has been shown to occur under diverse prebioticaly plausible conditions [19, 63, 108, 109, 122]. Oparin, one of the first to suggest a possible chemical pathway for the emergence of life, proposed that it could be manifested by the molecular reactions of relatively simple organic molecules in the primordial soup, interacting with each other to spontaneously form colloidal molecular assemblies (coacervates) [73, 92, 93].

3.2.Mutual catalysis

Regardless of the specific details of the replicator-first and the metabolism-first, both scenarios acknowledge the need for reliable information storage and transfer, assisted by self-replication. It was Orgel who highlighted the relationship between molecular replication and the concept of autocatalysis or self-catalysis [95]. Kauffman [59] defined a set of mutually catalytic compounds as "collectively autocatalytic" if within this set, the reaction producing any of the set's components can be catalyzed by at least one member of the set. Thus the entire set is self-sustaining and may be considered as undergoing self-reproduction, as long as input of energy and molecular building blocks is provided [49, 59, 60]. The notion that mutual catalysis (cross-catalysis) is an important facet of self-replication draws from these aforementioned ideas. Collectively autocatalytic systems resemble present-day living cells, which harbor self-catalytic polynucleotides as well as a plethora of mutual catalysts that constitute the metabolic pathways. This is exemplified by the famous hypercycle, a set of self-replicating polynucleotides, coding for and acted upon by catalytic enzymes [30]. Likewise, Autopoiesis [142] and the Chemoton [35] are example models of collective autocatalysis that also harbor self-replicators.

3.3. The lipid world

An example of metabolism first is the lipid world, where specific types of small molecules, i.e. lipids, are assumed to form catalytic networks, with the advantage that such molecules spontaneously accrete to form kinetically-controlled distinct supramolecular structures. Our laboratory was one of the pioneers of the lipid world scenario for the origin of life, attempting to generate a synthesis between the replicator first and metabolism first approaches [118, 120] (Figure 2). The main point of strength of this scenario is that it suggests an entity that can undergoe self-reproduction of a set of relatively simple molecules, without any self-templating biopolymer. This happens via a specific mechanism resembling that at work in collectively autocatalytic systems.

The lipid world scenario considers non-covalent reversible accretion of amphiphiles, e.g. lipids, to form assemblies such as micelles and vesicles. Importantly, these assemblies are considered to store information in the form of non-random molecular compositions. These are passed to progeny via homeostatic growth accompanied by fission. It is suggested to draw an analogy between the transmission of compositional information to the copying of sequential information by templating biopolymers. The importance of amphiphiles in this origin of life scenario derives from concepts similar to those of Oparin's coacervates. This is because lipids and similar amphiphiles spontaneously form distinct assemblies due to hydrophobic and hydrophilic interactions. These assemblies, in a way, combine properties of container structure, metabolism and information transfer [119].

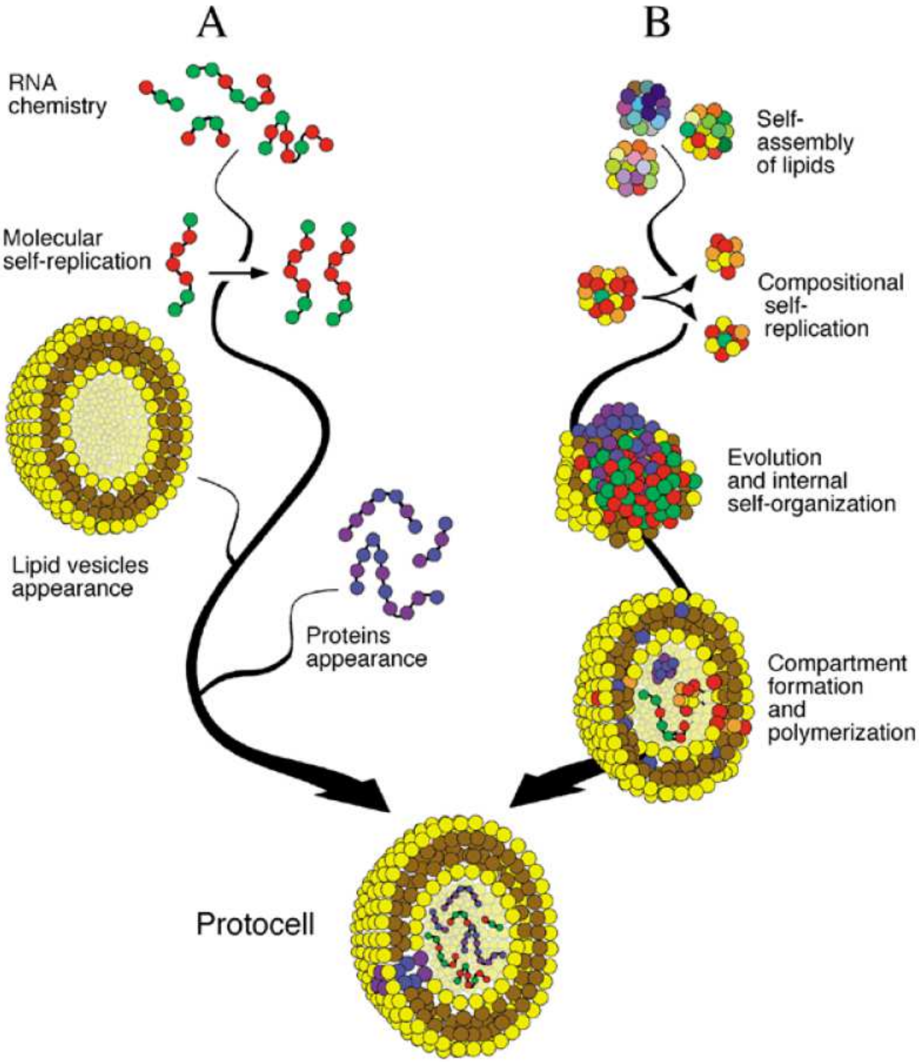
A
B
RNA chemistry
Molecular self-replication
Lipid vesicles appearance
Proteins appearance
Self-assembly of lipids
Compositional self-replication
Evolution and internal self-organization
Compartment formation and polymerization
Protocell

Figure 2: The lipid world attempts to generate a synthesis between the replicator first and metabolism first approaches. Figure taken from [120].

As for the question of where the monomer building blocks come from, prebiotic syntheses have been shown to include the formation of lipid-like amphiphilic molecules with long-chain hydrocarbons [40, 44, 101, 112]. In parallel, lipids are found in carbonaceous meteorites, and it has even been shown experimentally that such infalling amphiphiles are capable of forming vesicle-like boundary structures [24, 100]. Finally, a rudimentary role of membrane composition was recently shown experimentally, where it was inherited by daughter vesicles and affected daughter fission [3], supporting the plausibility of the lipid world.

3.4.The GARD model

The graded autocatalysis replication domain (GARD) model quantitatively describes the details of the lipid world [119]. GARD is a systems-chemistry kinetic model which entails supramolecular assembly of amphiphiles and elaborates some of its evolution-related attributes, with an implied route to minimal protocells [52, 57, 71, 124, 125, 126, 127, 128]. The model is based on a catalytic network whose nodes and edges respectively represent molecular types and catalytic events, including autocatalysis (self-catalysis) and cross (mutual) catalysis. In Equation 2 below these catalytic terms are respectively represented as the diagonal and off-diagonal terms of a matrix β, hence the term "β network". The model assumes that molecules from a buffered environment join and leave an assembly in a reversible manner. Once an assembly reaches a pre-defined maximal size, N_{max}, a random fission action is applied to produce two progenies of same size which can grow again and again in growth-fission cycles (Figure 3). Importantly, the system is kept away from thermodynamic equilibrium by assembly fission. GARD's dynamics displays species-like quasi-stationary states (QSS) in compositional space called composomes [119].

GARD is thus a kinetic model which describes the growth and fission of a molecular assembly [80, 119].

The composition of an assembly is given by the vector *v*:

$$v = \left\{ n_1 \ldots n_i \ldots n_{N_G} \right\}$$

Equation 1

Where N_G is the number of molecular types (environmental molecular repertoire) and n_i ($i=1..N_G$) is the current (time dependent) count of molecular type i within the assembly. Assembly growth is controlled by its molecular composition and the dynamics are described by a set of ordinary differential equations:

$$\frac{dn_i}{dt} = \left(k_f \rho_i N - k_b n_i \right) \left(1 + \sum_{j=1}^{N_G} \beta_{ij} \frac{n_j}{N} \right)$$

Equation 2

Where n_i is as in Equation 1, ρ_i is its environmental concentration (equal for all molecule types) and β_{ij} is the catalytic rate-enhancement exerted by an assembly molecule of type j on incoming or outgoing molecule of type i. k_f and k_b are the basal forward and backward rate constants (joining and leaving, respectively) and N is current assembly size. β_{ij} values are based a lognormal distribution, drawing from the experimentally derived receptor affinity distribution [69, 70, 119]. The model does not assume a priori relation between β_{ij} and β_{ji} values. It was previously found that when β_{ij} values obey such a distribution, faithful transfer of information to progeny is augmented [121]. Different randomly drawn β networks may be viewed as representing different environmental chemistries.

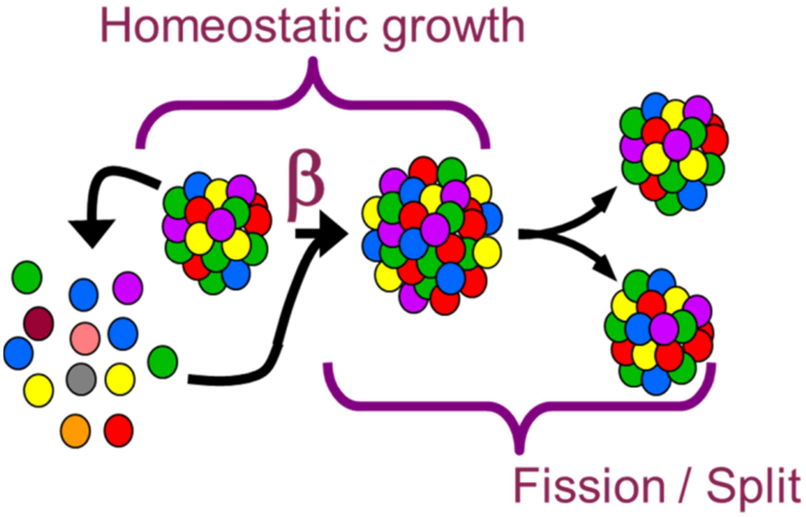

Figure 3: A cartoon representation of the GARD model cell-cycle. Molecules from the environment form and accrete to an assembly, biased by the β network (matrix). Once an assembly reaches a pre-determined maximal size it undergoes fission. Different colors represent different molecular types.

3.5.GARD Composomes and compotypes

The similarity between two assemblies, at generations χ and δ, was defined as the dot product of their compositions vectors [119], typically calculated at assembly size N_{max}:

$$H(v^{\chi},v^{\delta}) = \frac{v^{\chi} \cdot v^{\delta}}{|v^{\chi}| \cdot |v^{\delta}|}$$

Equation 3

A faithfully replicating assembly was previously defined as an assembly which is highly similar to predecessor and successor (H>0.9 for generations χ-1 and χ+1) [126]. A set of subsequent faithfully replicating assemblies is termed *composome* (a term originally derived from compositional genome) [119]. A composome is a QSS in the N_G-dimensional compositional space, when the trajectory of GARD dynamics is followed. The dynamics can also be presented as a ‘carpet’: a two-dimensional matrix showing H values for all assemblies encountered during a simulation [119] (Figure 4). A composome appears as a dense area with high H near the main diagonal in a carpet, signifying consecutive generations where a composition was transferred with high fidelity. A *compotype* (composome type) is subsequently defined as one of several clusters computed out of all assemblies that belong to any of the composomes in a simulation [126]. This is as contrasted with "drift" – assemblies that belong to no composome.

GARD's composomes (more specifically – compotypes) are treated as its species. This is because composomes are made out of a series of assemblies that share similar composition and faithfully replicate, making them persistent in time and on average appearing more than other compositions (i.e. drifts). The idea that that a compositional assembly of lipids (i.e. a vesicle) is a distinct species with distinct properties is supported by experiments that show that different binary and ternary composition of vesicles show different features such as permeability [82] or boundary structure [144].

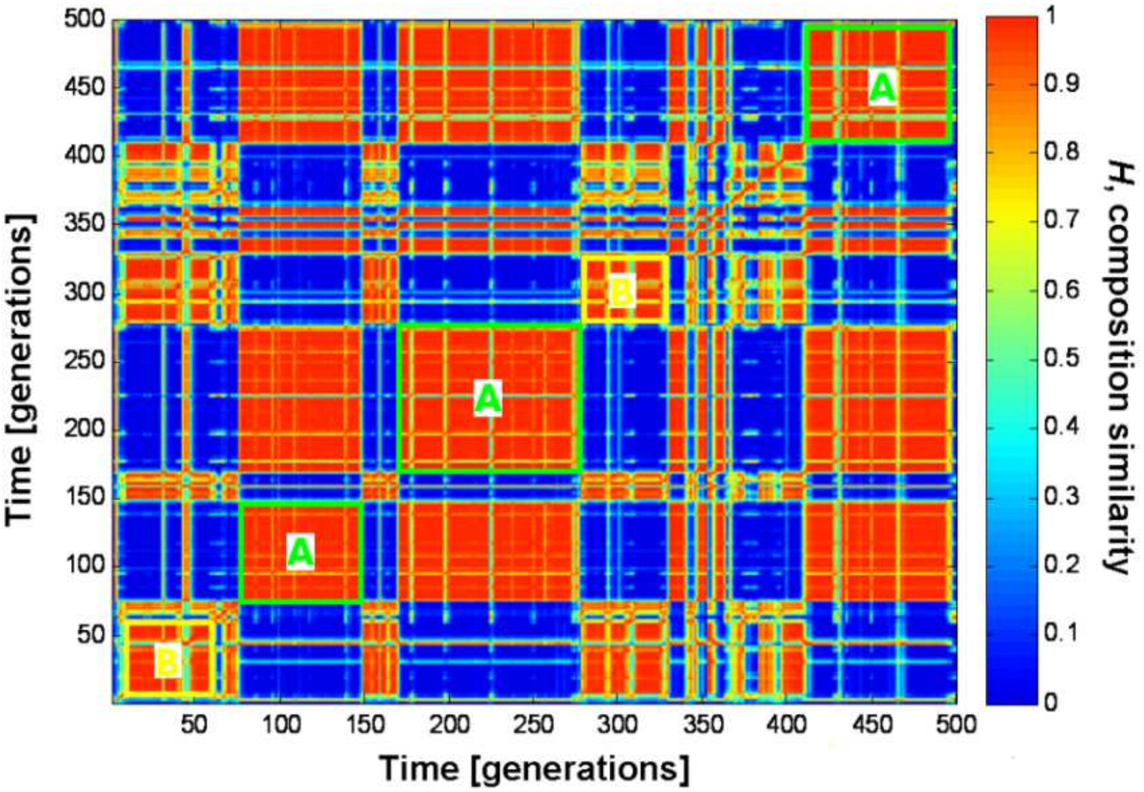

Figure 4: Similarity 'carpet' shows the degree of compositional similarity (H, Equation 3) between all assemblies in a GARD simulation. Red is high similarity. Composome appear as a dense red area near the main diagonal.

3.6.Spontaneous chiral symmetry breaking in early molecular networks

This work was done in collaboration with Dr. Ran Kafri from Harvard Medical School, and was published as a M.Sc. thesis [56] and a peer-reviewed paper [58].

An important facet of early biological evolution is the selection of chiral enantiomers for molecules such as amino acids and sugars. The origin of this symmetry breaking is a long-standing question in molecular evolution [13]. A more general kinetic formalism for early enantioselection, based on the GARD model, has been developed (Chiral-GARD, C-GARD [58]). The key is applying symmetry constraints to β, by considering an environment with asymmetric molecules in a racemic mixture. All 2×N_G molecular types are treated as different compounds with different kinetic parameters, keeping in mind that they actually constitute N_G enantiomer pairs (Figure 5).

The ensuing dynamics shows spontaneous chiral symmetry breaking, with transitions towards composomes enriched with one of the two enantiomers for some of the constituent molecule types (Figure 6). A global analysis of the dependence of weak-enantioselection on molecular enantiodiscrimination finds that increasing the latter enhances the probability of assemblies to have high enantioselection, yet even for the highest enantiodiscrimination studied here there is an almost even chance for assemblies to show high or low enantioselection (Figure 7). This may indicates a stochastic effect: high enantiodiscrimination is necessary, but not sufficient to lead to symmetry breaking.

It follows that chiral selection may be an emergent consequence of early catalytic molecular networks rather than a prerequisite for the initiation of primeval life processes.

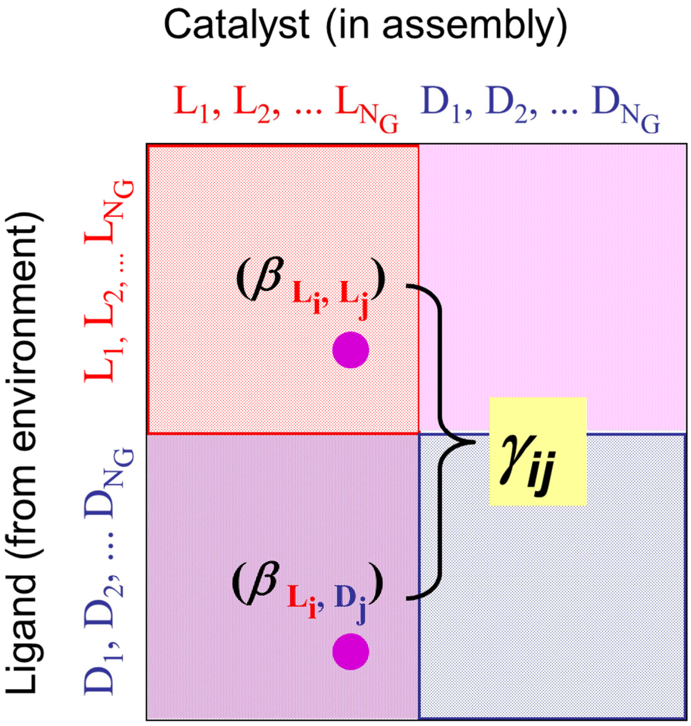
Catalyst (in assembly)
$L_1, L_2, \dots L_{N_G}$
$D_1, D_2, \dots D_{N_G}$
Ligand (from environment)
$L_1, L_2, \dots L_{N_G}$
$D_1, D_2, \dots D_{N_G}$
(β_{L_i, L_j})
(β_{L_i, D_j})
γ_{ij}

Figure 5: An illustration of a 2N_G×2N_G β and the value of enantiodiscrimination (β_{ij}). Note that the two blocks along each diagonal have identical values of the affinities (β_{LL}=β_{DD} and β_{LD}=β_{DL}).

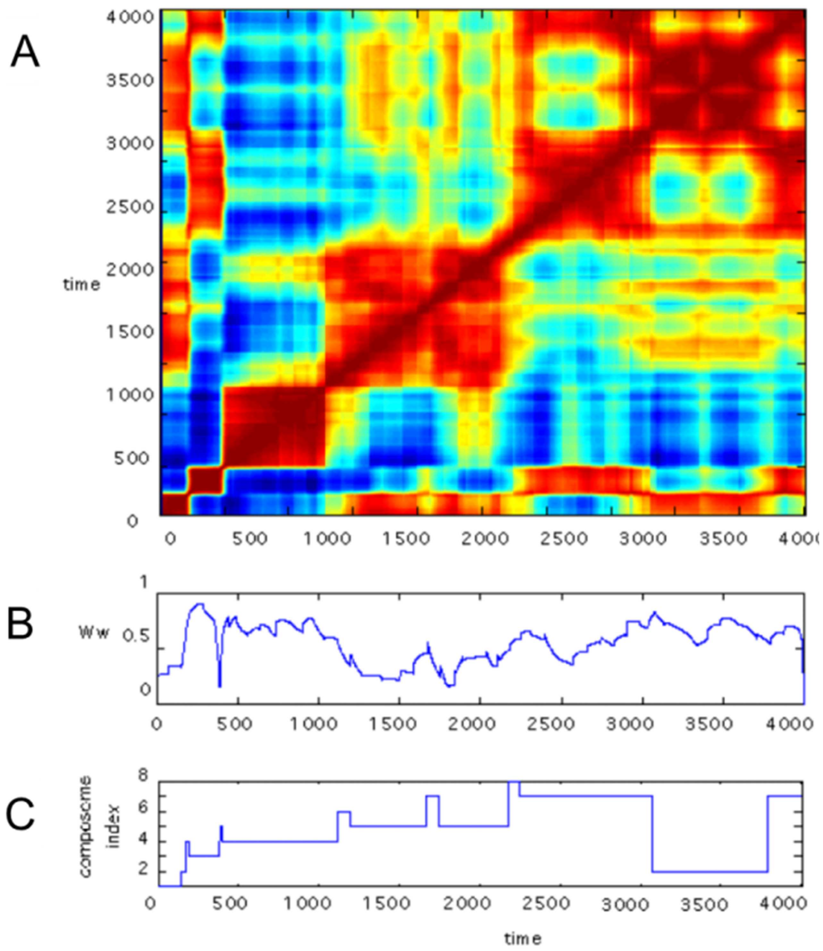

Figure 6: Example of a C-GARD simulation. (A) Similarity carpet (red is high similarity). (B) Weak enantiomeric selection during this simulation ($Ww = \sum_{i=1}^{N_G} \left| n_i^L - n_i^D \right| / N$). (C) Compotype assignments for the assemblies during this simulation.

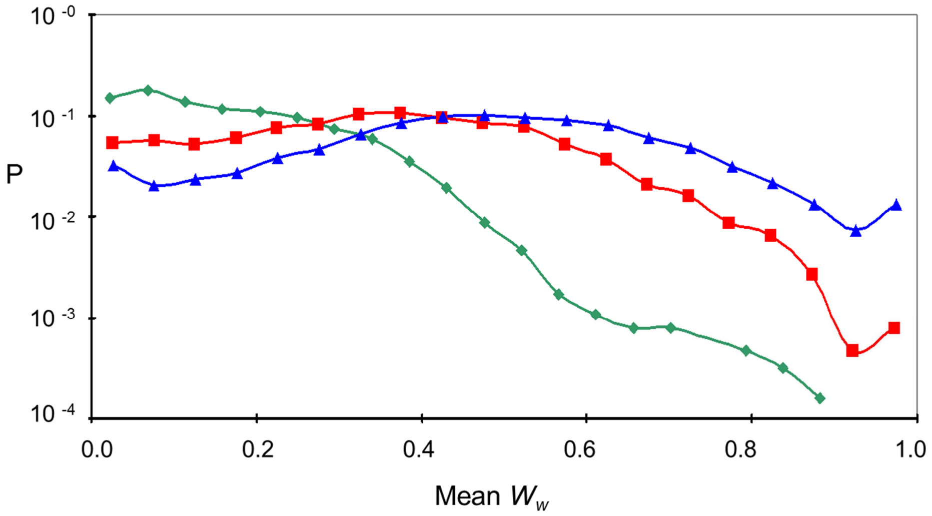

Figure 7: Probability distribution of Ww (see Figure 6 legend) at different values of enantiodiscrimination-related parameter (green=low, red=intermediate, blue=high).

4. METHODS

4.1. Computer simulations

Computer simulations were run using MATLAB versions 7.6-7.13. The GARD10 code package was prepared to be delivered upon request [80]. Different simulations were run using different underlying β matrices, which were generated by the MATLAB Mersenne-Twister pseudorandom number generator with different seeds. For each β, a set N_G^2 random numbers (labeled Z) was drawn from a normal distribution mean=0 and standard deviation=1.0, and converted to a set Q of lognormally distributed number by the following transformation: $Q=\exp(\mu+\sigma Z)$, where μ and σ are respectively the lognormal mean and standard deviation. GARD is subjected to a kinetic Monte Carlo simulation based on Gillespie's algorithm [38]: in each iteration, a set of $2N_G$ rate values is generated based on Equation 2 (the forward and backward parts in Equation 2 are treated separately) and then one reaction is randomly picked and executed, where the chance of picking a reaction is directly proportional to the reaction rate. It is assumed that the time passed is the inverse of the rate of the reaction picked. This is repeated for each assembly until its size reaches N_{max} (or 0, and then the simulation terminated) and then random fission applied. Fission is performed stochastically, whereby one progeny was created by selecting, one by one, molecules from the parent and placing them in this progeny. The chance to select a molecule of type i is proportional to its current count in the parent assembly, and this is continued until the size of this progeny is $N_{max}/2$ and the other progeny assumes the remainder of the parent.

Unless otherwise mentioned, the parameters used in this thesis are given in Table 1 [80].

N_G	100
N_{max}	N_G
k_f	10^{-2}
K_b	10^{-4}
ρ_i	$1/N_G$
μ	-4
s	4
L_{pop}	1000

Table 1: Simulations parameters used in this thesis.

4.2. Compotypes

Compotypes were found by K-means clustering algorithm, using 1-minus-cosine type of silhouette [126, 132]. Clustering was repeated for k=2,3,… number of clusters and the k with the highest silhouette was picked as the compotype count (N_C) of this simulation (for k=1, the silhouette is calculated as the average H between all assemblies in the simulation). A compotype is represented by a compositional vector constituting the center of mass of all its member assemblies.

In chapter 5.4 the size of a compotype's intrinsic molecular repertoire (N_{mol}), its replication fidelity (F_{rep}) and time (t_{rep}) are measured and employed. These parameters result from β and the GARD dynamics, as they are reflected in each compotype. N_{mol} is calculated as the number of molecule types (out of the total N_G types) whose fractional counts in a compotype center of mass are bigger than 1.0. F_{rep} and t_{rep} are assessed using the following method: an assembly with exactly the same composition as in the compotype's center of mass (rounded to nearest integer) is used as parent. This parent than undergoes split and each of the two progeny is grown according to its idiosyncratic composition (Equation 2) until it reaches N_{max} size. This is repeated for 4,000 times, each time beginning with the same parent, giving a total of 8,000 fully grown progeny of this compotype. F_{rep} of a compotype is than defined as the average H between the fully grown progeny to the parent. This is an extension to a previous analysis, where the fidelity was assessed based only on the split action [121]. Each event of a molecule joining (or leaving) the assembly has a rate, and the total growth time of each progeny is the sum of 1 over each of these rates. t_{rep} of a compotype is than defined as the average growth time of all grown progeny who are highly similar to this compotype.

4.3.Selection

This part mainly relates to chapter 5.1. Selection performed by applying a selection pressure towards the center of mass of a specific target compotype, T [80]. This was done by biasing the growth of an assembly towards the target (Equation 5) via a growth bonus parameter:

$$G_b = sH\left(v^{\chi}, T\right)$$

Equation 4

Manifested as a temporary enhancement of the corresponding β_{ij} values, as suggested [143], where s>1 is a fitness gain, embodying a selective advantage, and for consistency with a previous work [143] Equation 4 is calculated at assembly size N_{min}, that is, at the beginning of the growth cycle.

The modified β_{ij}' is obtained at each generation according to:

$$\beta_{ij}^{'} = \begin{cases} \beta_{ij} & i \text{ or } j \notin v^{\chi} \\ G_b \cdot \beta_{ij} & i \text{ and } j \in v^{\chi} \end{cases}$$

Equation 5

Where i and j are indices of molecular types, and β_{ij} modification is effected for all i,j (and j,i) pairs contained within the current assembly v^{χ}. Thus, the network is perturbed only at edges present within the current assembly according to how similar to current assembly is to the target. The selection excess (SE) is defined as the ratio between the frequencies of the target in a simulation performed with and without selection (Equation 6), while the probability of allele fixation is proportional to 1 over the total population size [97].

$$SE = \frac{f_{T'}}{f_T}$$

Equation 6

4.4.Population dynamics

This part mainly relates to chapter 5.4. The chemical dynamics of L_{pop} GARD compositional assemblies in a reactor under constant population conditions are simulated in a buffered environment (Figure 8). Each simulation starts by seeding the reactor with L_{pop} random assemblies, each at size N_{min}. Assembly growth is controlled by its molecular composition (Equation 2). Each time an assembly reached N_{max} a random fission was applied and one of the progeny replaced the parent while the other replaced a random assembly among the other L_{pop}-1, thus keeping the population size constant. This protocol is based on the classical Moran process [89]. Each simulation is performed for 50,000 split events in the reactor, typically sufficient to reach steady state, and data is saved every 10 split events. The composition of each assembly at each time point is assessed as belonging to one of the N_C compotypes characterizing a specific β (H>0.9 to the compotype center of mass) or to drift (Figure 9).

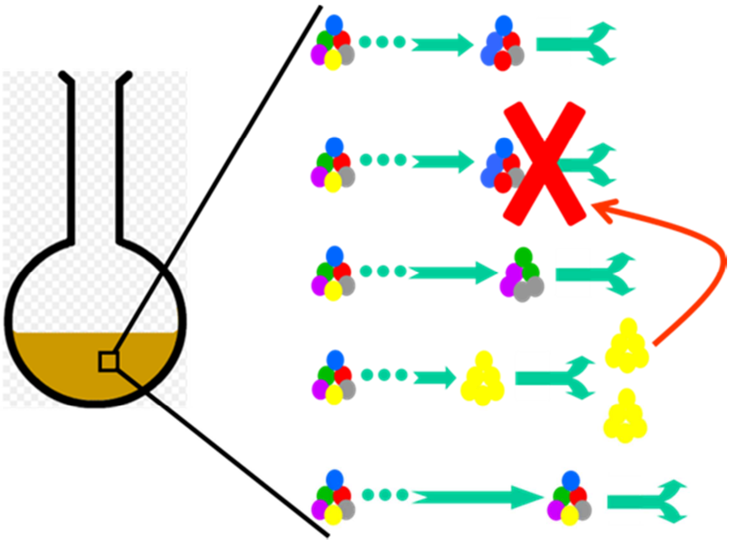

Figure 8: A heuristic representation of population dynamics, under constant population condition. One an assembly reaches N_{max} size, it undergoes a split, where one progeny replaces the parent while the second progeny replaces a random assembly.

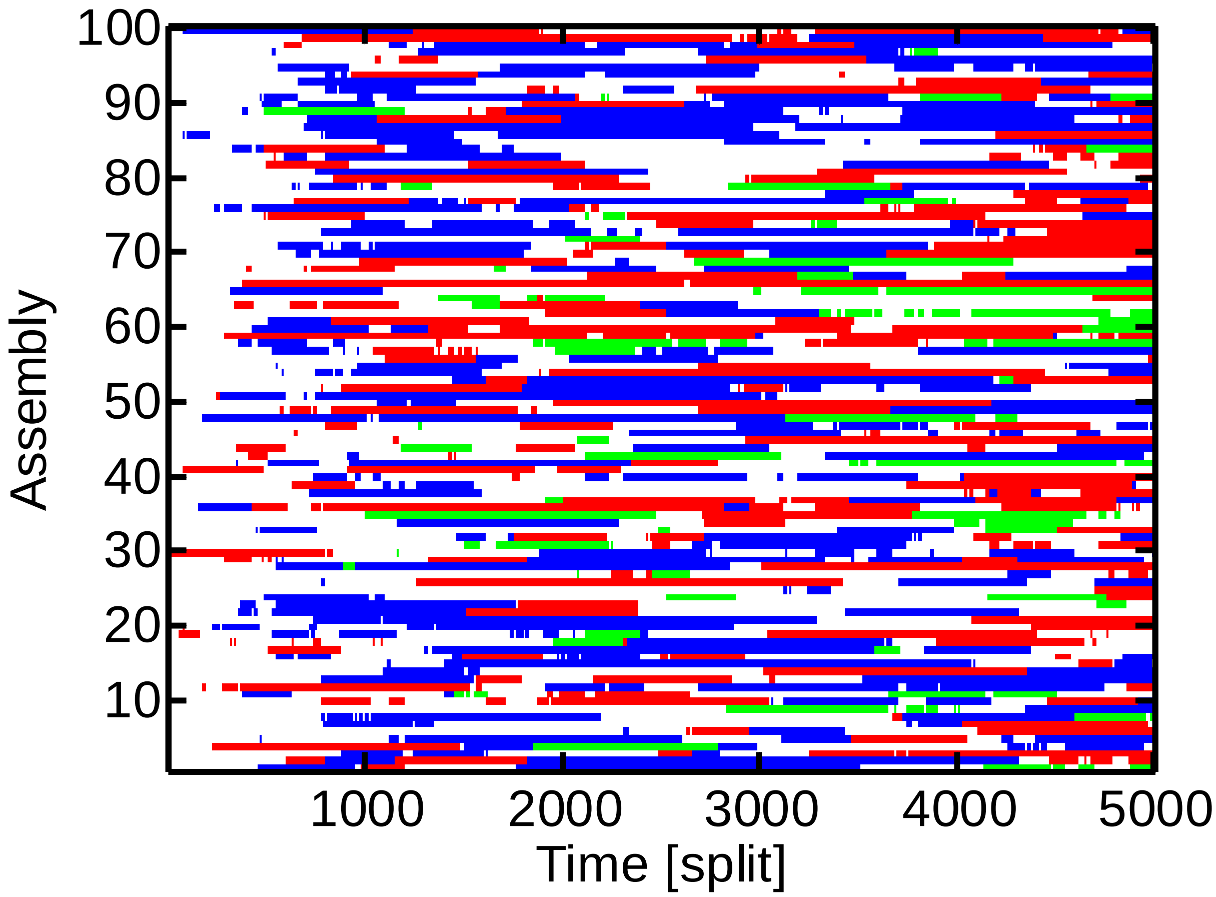

Figure 9: An example of GARD's population dynamics (for brevity only 100 assemblies are shown). In this example N_C=3. The color of an assembly represents it compotype assignment (white means drift).

4.5.Logistic equation and fit to population data

The fractional counts of assemblies belonging to each compotype (C_i) in a population are plotted and analyzed according to the multi species logistic model (r-K or Lotka-Volterra competition model) [34, 133, 141]:

$$\frac{dC_i}{dt} = r_i C_i \left(\frac{K_i - C_i - \sum_{i \neq j}^{N_C} \alpha_{ij} C_j}{K_i} \right)$$

Equation 7

Where t is time, measured in the number split events that occurred in the population. For compotype species i, r_i is its intrinsic growth-rate, K_i its carrying capacity and α_{ij} is the extent of competition exerted by compotype j on compotype i. The entire carrying capacity of a given environment is $\sum K_i$. The entire set of α_{ij} values of a given simulation represents a quantitative food-web network [138], whose nodes and edges are respectively compotypes and α_{ij} values.

Fitting between the time dependent frequencies and the logistic equation is performed using MATLAB's *lsqcurvefit* and *ode* functions for least-squares fitting and numerical integration of ordinary differential equations, respectively. The fit procedure is as follows:

1. $C_i(t)$ data of each compotype is smoothed 100 times by a 5-point moving average.
2. In order to avoid over sampling of the long times over the short times, the fit time window is until twice the time the variance of the data (for each time point, the variance is calculated until that point) drops below half its maximal size, plus 100 points along the tail in equal intervals. This is calculated for each compotype individually and then the largest window is picked.
3. Compotypes with $<C_i>$ <0.01 are ignored and their assemblies classified as drift.
4. For simulation with $N_C=1$, if the time curve exhibits a plateau lower than the maximum by more than 20%, then this simulation is ignored.
5. MATLAB *lsqcurvefit* is used to perform least-squares curve fitting with the following function parameters (the rest are at their default values): TolFun=1e-10; TolX=1e-10; MaxFunEval=200*N_C*(N_C+1); MaxIter=1000.
6. *ode45* and *ode15s* ordinary differential equation solvers are used to numerically solve Equation 7, and the fit with the lowest residuals is considered. The following function parameters are used and the rest are at their default values: AbsTol=1e-10; RelTol=1e-10.
7. Initial parameter guesses are: K_i=max(C_i); a_{ij}=0.1; $r_i = \sum_{t=1}^{100} dC_i/100$ (dC_i is approximated by 5th order numerical differentiation); $C_i(0)$=mean[C_i(1..100)].
8. Constraints are: $r_i>0$, $0\le K_i\le 1.0$, $0\le a_{ij}\le 10.0$, $0\le C_i(0)\le \max(C_i)$.

For each simulation, the quality of the fit was assessed using root-mean-square-difference (RMSD):

$$RMSD=\sqrt{\left\langle \sum_{i=1}^{N_C}\left[f_i(t)-C_i(t)\right]\right\rangle^2}$$

Equation 8

Where f_i is the fitted curve and <…> denotes an average over all time-points in that simulation.

5. RESULTS

The results presented in this thesis show how the lipid world scenario and the GARD model have several characteristics of living systems, and elaborate some advanced features of the GARD model.

5.1. Excess mutual catalysis is required for effective evolvability

An important prerequisite from any living system is to be able to respond to selection and thus undergo evolution [9, 139]. In GARD, it is of special importance to demonstrate this attribute, as it is a non-standard model and it is not obvious why and how its species (compotypes) should respond to a selection pressure. This chapter analyzes the selection behavior of compotypes in order to address this issue. This is done by considering the change in the abundance of a compotype, in a given simulation, as a mimic to selection. It is found that GARD's compotypes can indeed portray selection. Further, a fundamental relation between the general structure of β and this selection behavior is discovered: the higher the mutual catalysis level in β is, the stronger the selection response portrayed, i.e. a bigger change in the abundance. The argument that GARD's selection should be studied with respect to compotypes is addressed in chapter 5.4.6.1, which also helps understand the difference in results with a recent erroneous criticism against GARD's evolvability [143].

Excess Mutual Catalysis Is Required for Effective Evolvability

Omer Markovitch**
Weizmann Institute of Science

Doron Lancet*,**
Weizmann Institute of Science

Abstract It is widely accepted that autocatalysis constitutes a crucial facet of effective replication and evolution (e.g., in Eigen's hypercycle model). Other models for early evolution (e.g., by Dyson, Gánti, Varela, and Kauffman) invoke catalytic networks, where cross-catalysis is more apparent. A key question is how the balance between auto- (self-) and cross- (mutual) catalysis shapes the behavior of model evolving systems. This is investigated using the graded autocatalysis replication domain (GARD) model, previously shown to capture essential features of reproduction, mutation, and evolution in compositional molecular assemblies. We have performed numerical simulations of an ensemble of GARD networks, each with a different set of lognormally distributed catalytic values. We asked what is the influence of the catalytic content of such networks on beneficial evolution. Importantly, a clear trend was observed, wherein only networks with high mutual catalysis propensity (p_{mc}) allowed for an augmented diversity of composomes, quasi-stationary compositions that exhibit high replication fidelity. We have reexamined a recent analysis that showed meager selection in a single GARD instance and for a few nonstationary target compositions. In contrast, when we focused here on compotypes (clusters of composomes) as targets for selection in populations of compositional assemblies, appreciable selection response was observed for a large portion of the networks simulated. Further, stronger selection response was seen for high p_{mc} values. Our simulations thus demonstrate that GARD can help analyze important facets of evolving systems, and indicate that excess mutual catalysis over self-catalysis is likely to be important for the emergence of molecular systems capable of evolutionlike behavior.

Keywords
Evolution, networks, selection, composomes, metabolism

A version of this paper with color figures is available online at http://dx.doi.org/10.1162/artl_a_00064. Subscription required.

1 Introduction

The fundamental question of how primitive life emerged on the prebiotic Earth has drawn considerable scientific attention throughout the centuries [2, 5, 14, 15, 22, 42, 59, 64]. The path from organic mixtures (i.e., the primeval soup) to reproducing lifelike protocells has traditionally been dominated by two different views: the genetic, or *replicator-first*, approach, and the *metabolism-first* approach [2, 42]. Both

[*] Contact author.
[**] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: omermar@weizmann.ac.il (O.M.); doron.lancet@weizmann.ac.il (D.L.)

acknowledge the need for reliable information storage and transfer, assisted by self-replication. The replicator-first approach suggests that life began with a single self-perpetuating biopolymer (e.g., RNA) [14, 15, 18, 19, 22, 37, 42, 64], which later evolved into multimolecular networks under the replicator's control. Orgel [41] has highlighted the relationship between molecular replication and the concept of autocatalysis or self-catalysis. The metabolism-first approach suggests that the very first life precursors must have been relatively complex molecular networks arising via spontaneous accretion of simpler organic molecules [3, 9, 24, 25, 34, 48, 51, 53, 60]. In this scenario, it is further proposed that faithful reproduction directly stems from certain network attributes. Therefore, one should better understand the network properties of the implicated molecular assemblies [1, 47, 57, 66] if one can merge the two seemingly conflicting scenarios for prebiotic evolution.

One embodiment of the metabolism-first view is the lipid world scenario, which considers noncovalent assemblies of amphiphiles, such as micelles and vesicles formed by lipids [8, 39, 48, 50, 53, 69]. These are assumed to store information in the form of nonrandom molecular compositions, and pass it to progeny via homeostatic growth accompanied by fission [49]. The graded autocatalysis replication domain (GARD) kinetic model for prebiotic evolution quantitatively describes the details of such a process. It elaborates some of its evolution-related attributes [10–12, 27, 30, 44, 50, 58, 62, 67, 68], with an implied route to minimal protocells [8, 45, 49, 63, 65]. The model is based on a catalytic network, usually presented in the form of a matrix β with autocatalysis (self-catalysis) and cross (mutual) catalysis terms. Importantly, the system is kept away from thermodynamic equilibrium by assembly fission [49]. Key in GARD dynamics are compotypes—clusters of replication-prone quasi-stationary states (composomes, a term derived from the notion of compositional genomes [49]), proposed to play a crucial role in the GARD's evolutionary behavior. Introducing substantial inhibition in β is expected to result in net catalysis because an inhibitor of an inhibitor is an activator [20].

Catalysis, the enhancement of reaction rate by an external chemical component, was recognized as early as 1836 by Berzelius, and Ostwald applied the term *autocatalysis* in 1890 to reactions that gain speed as they proceed [26, 44]. In the genetic approach to life's origin, researchers invoke one or several autocatalytic molecules as the core of a prebiotic entity. This is exemplified by the *hypercycle*, a set of self-replicating polynucleotides, coding for and acted upon by enzymes [10, 30, 58]. In the metabolism-first domain, autopoiesis [67] and the chemoton model [12] are examples of collective autocatalysis [25].

Collectively autocatalytic systems feature a central role not only for self-catalysis, but also for mutual catalysis. In this, they arguably resemble present-day living cells, which harbor self-catalytic polynucleotides as well as a plethora of mutual catalysts that constitute metabolic pathways. Here we utilize a metabolism-first simulator to examine the relative importance of the two catalytic modes (self- and mutual catalysis). Previously [11], an abstract chemistry model has been used to demonstrate that self-maintaining organizations arise only once self-catalysis is completely inhibited [11, 62]. We attempt to extend such results in the realm of the GARD kinetic model, asking what features of the β network contribute to the evolution of the ensuing compositional assemblies. It is shown that excess mutual catalysis is a necessary, though not sufficient, condition for displaying several evolutionlike characteristics, including a high number of composome types, higher evolvability scores, and a significant response to selection.

Recently, it has been argued that collectively autocatalytic metabolic networks, such as the GARD, do not allow for fitter compositional genomes to be maintained by selection. Vasas et al. [68] compared the frequency ranking of random GARD compositional assemblies before and after selection, and found that the relative ranks changed only slightly. This was taken as evidence for an inherent evolutionary limitation of metabolism-first scenarios. Here it is demonstrated, based on a large number of simulations, that when quasi-stationary composomes rather than arbitrary compositions serve as selection targets, GARD networks *are* capable of a significant response to selection. Importantly, this can happen chiefly when a high proportion of mutual catalysis is present in a GARD network. The results highlight the potentially important role of mutual catalysis, as compared to self-catalysis, in the emergence of early lifelike systems.

2 Model and Methods

2.1 GARD Formalism

The regular GARD formalism describes the time-dependent dynamics of a molecular assembly, by following the fate of a compositional vector whose elements are the molecular counts n_i within the assembly:

$$v = \{n_1, n_2, \ldots, n_{N_G}\} \qquad (i = 1, \ldots, N_G) \tag{1}$$

The vector dynamics is governed by mutually catalytic interactions among the invariable number of constituent molecule types, N_G. The assembly grows by accretion of environmental molecules, and once a limiting size $N_{\max}$ is attained, random fission is applied, producing two progeny of the same size, $N_{\min} = N_{\max}/2$, one of which grows again, generating growth-fission cycles of consecutive generations. GARD dynamics is described by a set of ordinary differential equations

$$\frac{dn_i}{dt} = (k_f \rho_i N - k_b n_i)\left(1 + \sum_{j=1}^{N_G} \beta_{ij} \frac{n_j}{N}\right), \qquad N = \sum_{i=1}^{N_G} n_i, \tag{2}$$

where dn_i/dt is in units of the individual reaction rates at which the counts of elements are changing [49], and k_f and k_b are respectively the basal forward and backward rate constants (joining and leaving the assembly). Typically $k_f \gg k_b$, reflecting a high equilibrium constant k_f/k_b for spontaneous amphiphile accretion (Table 1). Here ρ_i is the buffered concentration of molecule type i in the environment (assumed here to be equal for all i values), N is the assembly current size, and β_{ij} is

Table 1. Simulation parameters. N_G is the number of molecular types (repertoire size); $N_{\max}$ is the assembly pre-fission size; k_f and k_b are the respective basal forward and backward rate constants; ρ_i is the buffered environmental concentration of molecule type i; μ and σ are the respective mean and standard deviation of the lognormal distribution of β_{ij} values (Appendix A.1, Equation 12); *GEN* is the duration of a simulation; Lognormal random seeds is the range of random seeds used for simulations; L_{pop} is the constant size of the population in the population GARD.

N_G	100
N_{max}	N_G
k_f	10^{-2}
k_b	10^{-4}
ρ_i	$1/N_G$
μ	−4.0
σ	4.0
GEN	5,000
Lognormal random seeds	1–10,000
L_{pop}	1,000

the non-negative matrix element signifying the rate enhancement exerted by an assembly molecule of type j on an incoming or outgoing molecule of type i

$$i_{\text{out}} + j_{\text{in}} \underset{k_{\text{f}} \times (1+\beta_{ij})}{\overset{k_{\text{b}} \times (1+\beta_{ij})}{\leftrightarrows}} i_{\text{in}} + j_{\text{in}} \tag{3}$$

The chemical reaction in Equation 3 embodies the notion that molecular catalysis equally affects the forward and the backward rates, obeying the constraint that a catalyst may not change the equilibrium constant of the reaction it affects. This means that even under catalytic action, the relationship $k_{\text{f}} \gg k_{\text{b}}$ prevails.

The matrix β represents a network of self-catalytic (diagonal elements) and mutually catalytic (off-diagonal elements) catalytic interactions (Figure 1), with self-catalysis represented by the case $j = i$ (Appendix A.1, Equation 13). The matrix elements are randomly drawn from a lognormal distribution (Appendix A.1 and Equation 12) [49].

2.2 GARD Simulations

The model is subjected to a kinetic Monte Carlo simulation based on Gillespie's algorithm [16, 17, 51] using parameter values similar to those employed in previous studies (Table 1). Simulations are run using MATLAB versions 7.6–7.10 (the GARD10 code is available upon request). A set of 10,000 GARD simulations is generated, all with the same parameters, and each with a different matrix β generated by the MATLAB pseudorandom number generator with seeds 1–10,000. The validity of the conclusions drawn here is ascertained by repeating the simulations with smaller data sets, with seeds 1–2,000 and 2,001–4,000, striving to verify that the entire 10,000-strong data set adequately represents the GARD simulation space. The random sampling of β values may be perceived as representing different possible GARD environmental chemistries.

The relative mutual catalysis power

$$p_{\text{mc}} = \frac{\sum_{i=1}^{N_G} \sum_{j=1}^{N_G} \beta_{ij}}{\sum_{q=1}^{N_G} \beta_{qq}} \cdot \frac{N_G}{N_G^2} \tag{4}$$

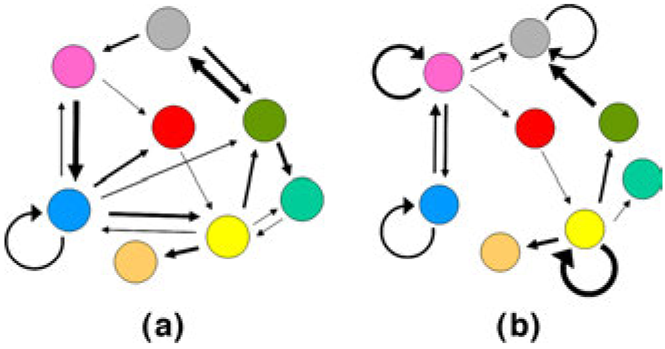

Figure 1. Network representation of GARD's β matrix. Two cartoon networks are shown, one with excess mutual catalysis (a) and the other with excess self-catalysis (b). In the electronic version, colored circles represent different molecular types, and arrow thickness represents catalysis strength (Equation 3). Self-catalysis is the shortest closed loop, containing one molecular type (see Appendix A.1, Equation 13).

is defined as the sum of all rate enhancements divided by the sum of self-catalysis rates (diagonal β elements). Because there are only N_G diagonal elements and the total number of elements is N_G^2, an appropriate correction is introduced. Thus, the excess of mutual catalysis is represented by $p_{\mathrm{mc}} > 1$, while the excess of self-catalysis (or autocatalysis) is portrayed by $p_{\mathrm{mc}} < 1$.

2.3 Compositional Similarity and Compotypes

The similarity between the compositions ν^{χ} and ν^{δ} of the respective assemblies at generations χ and δ is defined as the dot product H (see Equation 5) of their composition vectors [49], typically calculated at assembly size $N_{\max}$ (end of the growth cycle).

$$H(\chi, \delta) = H(\nu^{\chi}, \nu^{\delta}) = \frac{\nu^{\chi} \cdot \nu^{\delta}}{|\nu^{\chi}| \cdot |\nu^{\delta}|} \tag{5}$$

GARD dynamics is visually portrayed by a *similarity carpet*, showing H between any pair of parent assemblies during a simulation (e.g., Figure 10 in Appendix A.4). Composomes, appearing as dense areas with high similarity near the main diagonal, are defined as any two consecutive generations where $H(\chi, \chi + 1) \geq 0.9$ [56]. Inter-composome similarity is viewed by off-diagonal examination. The time duration of different generations (Equation 2) is different due to different growth pathways; hence a certain level of selection is already achieved by the matrix β causing composomes to appear more frequently than random compositions [49].

All the compositions belonging to composomes in the entire simulation undergo *k*-means clustering [56, 61], and the centers of mass of the resulting clusters are defined as compotypes.

2.4 Similarity Autocorrelation

The similarity autocorrelation function, $c(\Delta t)$, akin to a Fourier transform of the compositional similarity time series, is defined by

$$c(\Delta t) = \langle H(\chi, \chi) \cdot H(\chi, \delta) \rangle = \langle H(\chi, \delta) \rangle \tag{6}$$

where $\langle \cdots \rangle$ denotes averaging over all generation pairs fulfilling $\delta - \chi = \Delta t$. This function is history independent, that is, no conditions are imposed on the events occurring between generations χ and δ.

$c(\Delta t)$ is fitted with a single exponential with parameters τ and H_0 using a least squares fit (see Appendix A.2, and Figure 12 in Appendix A.4):

$$c(\Delta t) = (1 - H_0)\, \exp\left(-\frac{\Delta t}{\tau}\right) + H_0 \tag{7}$$

The parameters τ and H_0 are used to define a measure of evolvability (Section 3).

2.5 Selection in GARD

For each simulation, the most frequent compotype is chosen as a target, T. A selection-GARD simulation is then run, whereby the growth of an assembly at generation χ is biased toward T via a growth bonus parameter

$$G_b = s \cdot H(\nu^{\chi}, T) \tag{8}$$

manifested as a temporary enhancement of the corresponding β_{ij} values, as suggested [68], where $s > 1$ is a fitness gain, embodying a selective advantage, and for consistency with previous work [68] $H(\nu^{\chi}, T)$ is calculated at assembly size $N_{\min}$, that is, the beginning of the growth cycle.

The modified matrix element β'_{ij} is obtained at each generation according to

$$\beta'_{ij}(\chi) = \begin{cases} \beta_{ij}, & i \text{ or } j \notin \nu^{\chi} \\ G_b \cdot \beta_{ij}, & i \text{ and } j \in \nu^{\chi} \end{cases} \tag{9}$$

where i and j are molecular type indices, and β_{ij} modification is effected for all i,j (and j,i) pairs contained within the current assembly. Thus, the network will be perturbed only at edges present within the current assembly according to how similar the current assembly is to the target. In the selection-GARD simulation, a compotype T' is identified as having the highest H value with respect to T. An unambiguous identification of T' is afforded by the fact that the mean similarity between T and T' in the entire data set is $H = 0.9933 \pm 0.0217$. The selection excess is subsequently defined as

$$SE = \frac{f_{T'}}{f_T} \tag{10}$$

where $f_{T'}$ and f_T are the fractions of generations belonging the respective compotype (before and after selection). Selection excesses ≥1.05 and ≤0.95 are respectively taken to represent positive and negative target selection; the rest are taken to signify no selection.

2.6 Selection Dynamics in a Population of Compositional Assemblies

An initially random population of a fixed number of assemblies, L_{pop}, is allowed to simultaneously grow according to Equations 1 and 2 and its idiosyncratic composition. When one of the assemblies reaches the limiting size $N_{\max}$, it divides by random fission, and a randomly chosen assembly from among the other $L_{\text{pop}} - 1$ assemblies is removed, thus keeping the population size constant. This is repeated for GEN splits (Table 1). This protocol is based on the classical Moran process [36, 68, 70], and to some degree reflects an earlier attempt to simulate GARD populations [38].

The frequency of the target in each population is defined as the number of assemblies that are highly similar ($H \geq 0.9$) to the target compotype taken from regular GARD for the same β network (Figure 13 in Appendix A.4). Selection is exerted by performing a simulation with the same parameters, biasing the growth of assemblies toward a target compotype as for regular GARD (Equations 8 and 9). The selection excess is defined as in Equation 10, where $f_{T'}$ and f_T are respectively the fractions of assemblies within the population belonging to the target compotype before and after selection.

3 Results

3.1 Selection in GARD

We used GARD simulations to ask what is the selection response of compositional assemblies. A value for the selection excess was obtained for each of 10,000 simulations, using a modest value of the fitness gain, $s = 1.1$, in line with previous work [68]. Figure 2a shows the correlation between the frequencies of the target compotype with and without selection (examples of regular GARD carpets before and after selection are given in Figure 14 in Appendix A.4). An overall skew is seen here toward positive selection. The figure also demonstrates that significant positive selection, as well as negative, occurs over most of the range of f_T.

Figure 2b shows the distribution of selection excess values for the entire data set (Equation 10). Importantly, a considerable percentage of the simulations (33%) show positive selection, with a mean selection excess of 1.38 for selection excess >1.05, and as much as 10% shows selection excess >1.5. Interestingly, 31% of the cases showed negative selection, with a mean selection excess of 0.775 for selection excess <0.95, and about 36% were neutral to the selection pressure. Similar to the skewness in Figure 2a, there is a slight bias in favor of positive selection, as indicated by an overall mean selection

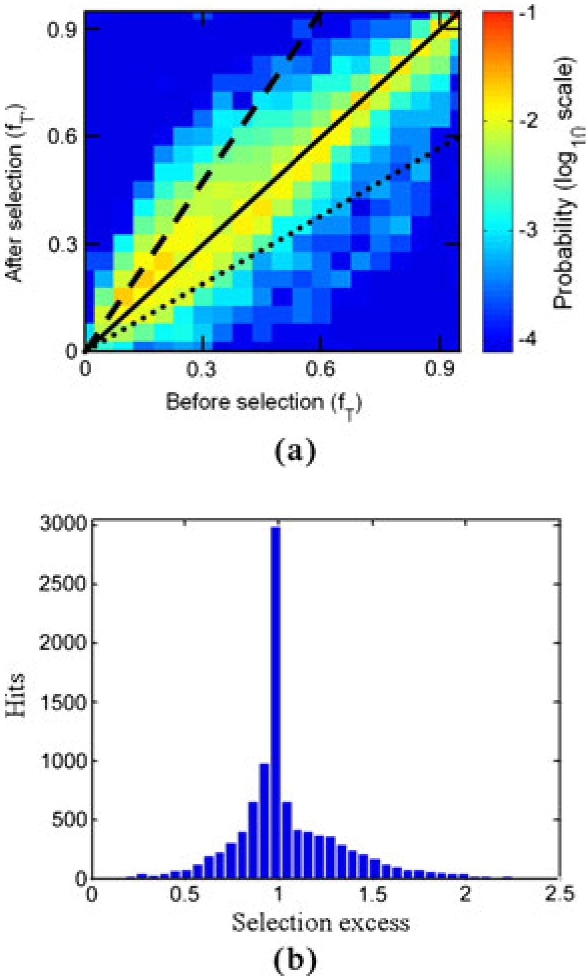

Figure 2. Selection in GARD. (a) The correlation between the frequencies of the target compotype in the basal simulation (f_T) and its frequency after applying selection ($f_{T'}$). In the electronic version, color represents probability out of the entire data set of 10,000 simulations, and positive and negative selection are respectively seen above and below the diagonal (selection excess = 1.0, solid black line). The dashed and dotted lines respectively mark selection excesses of $\frac{3}{2}$ and $\frac{1}{2}$. (b) Selection excess histogram for the entire data set. Simulation parameters are given in Table 1.

excess equal to 1.05. Notably, higher mean selection values positively correlate with the number of other compotypes coexisting with the target compotype in a given system (Figure 15 in Appendix A.5).

GARD simulations are used to see how attributes of the catalytic network embodied in the matrix β govern the evolution-related dynamics of compositional assemblies. It is asked how the mutually catalytic power p_{mc} (Equation 4) influences the selection response. A clear trend appears here, whereby strong positive or negative selection is found almost entirely for p_{mc} higher than 1 (Figure 3b).

The main trends appear also at lower simulation counts, barring small-number fluctuations at high p_{mc} (Figure 3a). For example, for the range of $p_{\mathrm{mc}} > 100$, a meaningful p-value with 5% significance level is achieved only after performing more than 2,500 simulations (Table 3 in Appendix A.6). The other two evolution-related parameters withstand similar scrutiny (below).

3.2 Populations of GARD Assemblies

The foregoing simulations of the regular GARD model addressed the case in which at each time point only one GARD assembly is considered. To enhance the capacity to draw conclusions about

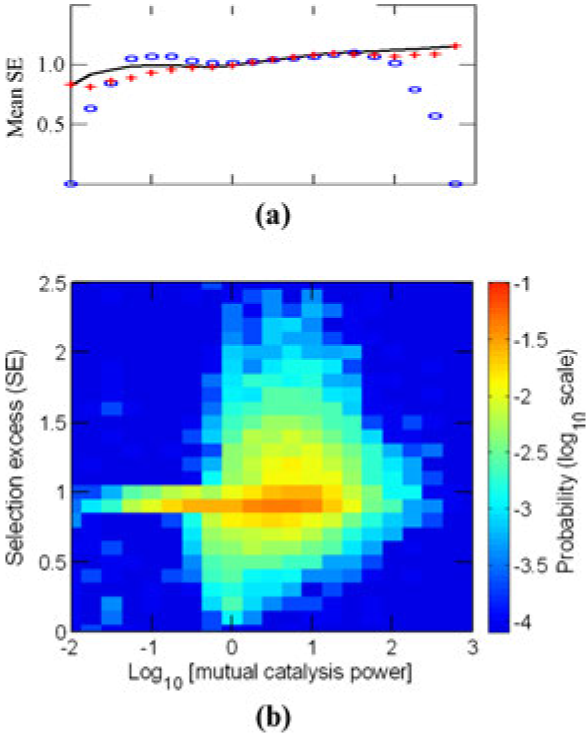

Figure 3. The dependence of selection excess (*SE*) on mutual catalysis power (p_{mc}). (a) Mean *SE* versus $\log_{10}p_{mc}$, collected from 10,000 GARD instances (solid black line, smoothed) or from two subsets of 2,000 instances, random seeds 1–2,000 (ovals) and 2,001–4,000 (crosses). (b) Density plot of *SE* versus $\log_{10}p_{mc}$. In the electronic version, color represents probability of finding instances with specific (*SE*, p_{mc}) values in all 10,000 GARD instances. Data is the same as in Figure 2.

selection in GARD, 1,000 simulations were performed, each for a population of 1,000 assemblies, under the constant population conditions. Figure 4 shows an example of the dynamics for one of the networks. Starting from a population of random assemblies, the population frequency of the target compotype gradually grows over the first 10,000 split events, reaching a plateau with fluctuations, signifying the compositional preference imposed by the matrix β towards this compotype. When selection toward this compotype is applied (Equation 9), this general behavior is retained, with a faster

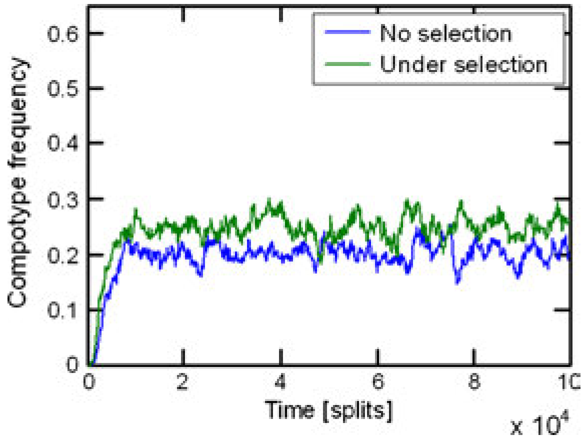

Figure 4. An example of the development of a compotype in population dynamics, without and with selection. This figure shows the fraction of assemblies in the population that are highly similar to a given compotype (see Section 2 and Figure 13 in Appendix A.4) over a large number of splits. Simulation parameters are lognormal seed = 3, *GEN* = 100,000, and the rest are given in Table 1.

growth and a higher plateau, that is, showing positive response to selection. Similar to Figure 3b, strong positive or negative selection is much more prevalent for p_{mc} values higher than 1 (Figure 5b).

The effect of selection pressure on the frequency of the target compotype for all 1,000 networks is presented in Figure 5a. Similarly to Figure 2a, an overall skew toward positive selection is seen (about 50% of cases), with some cases of negative (about 15% of cases) or no response to selection, and with a mean selection excess of 1.254 ± 0.804. Significantly, the ratio of the number of simulations showing positive selection to that showing negative selection increased more than threefold, from 1.06 in the regular GARD to 3.33 in the population-GARD. In line with previous work [68], the growth bonus was calculated when the assembly size was N_{min} (Equation 8). When the bonus was calculated for all time points between N_{min} and N_{max} (for a smaller set of 100 population-GARD simulations), the overall selection response seems to become even more positive (70% of cases), with a higher selection excess value of 1.399 ± 0.997.

3.3 Compotype Diversity

The influence of p_{mc} on one of the attributes of GARD diversity, the mean number of different compotypes appearing in a simulation, is now analyzed. It is found that as p_{mc} increases, so does the mean

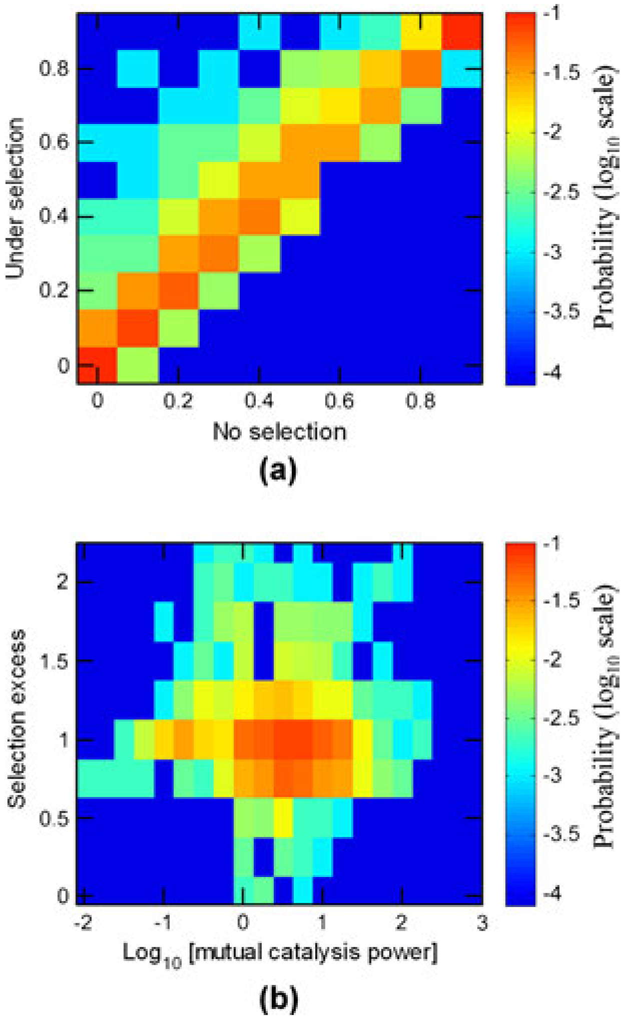

Figure 5. Selection in population-GARD. Figure details for (a) and (b) are as in Figures 2a and 3b, respectively. Data set is 1,000 population-GARD simulations, whose parameters are collected in Table 1.

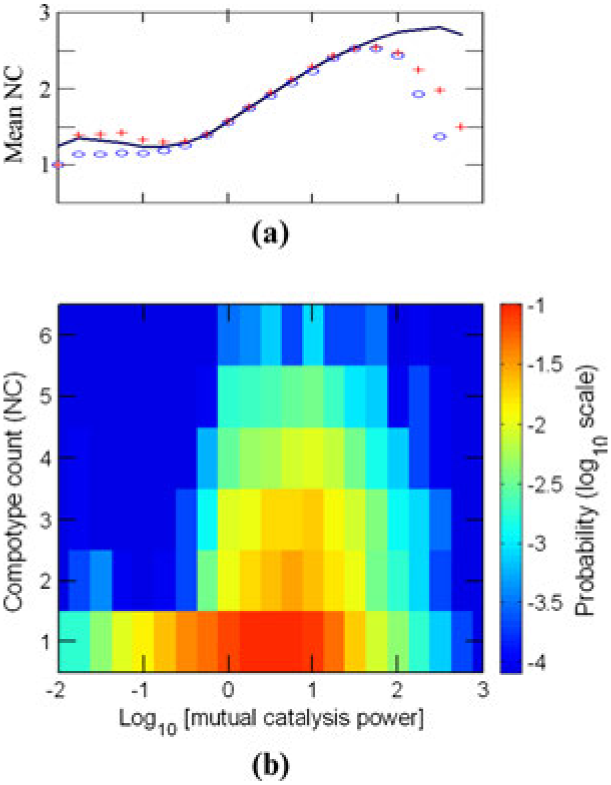

Figure 6. The dependence of compotype count (*NC*) on p_{mc}. Details are as in Figure 3.

number of compotypes, reaching a maximal value of nearly 3 at $p_{mc} = 100$ (Figure 6a). Furthermore, in the realm of excess self-catalysis ($p_{mc} < 0.5$), one compotype appears in an overwhelming majority of the cases (91%) (Figure 6b). In contrast, compotype counts between 2 and 6 are almost entirely confined to the domain of excess mutual catalysis ($p_{mc} > 2$). Curiously, even among the ~5,000 simulations that show only one compotype, a large majority have $p_{mc} > 2$, suggesting that high mutual catalysis is a necessary but not sufficient condition for a high number of compotypes.

3.4 GARD Evolvability

The similarity autocorrelation function (Equation 6) and its derived parameters (Equation 7) are employed to obtain information on the evolutionlike dynamics of GARD assemblies. One possible interpretation of the value of τ is a depiction of the whole-simulation average of the assembly compositional lifetime. Longer τ may be taken to represent better average maintenance of compositional similarity between consecutive GARD generations, symbolizing better reproduction fidelity. Likewise, $1/\tau$ may be thought of as related to the compositional mutation rate. Indeed, effective compositional preservation is implicated by the most frequent number of generations, $\tau \approx 3$, with a non-negligible probability for $\tau \geq 10$ (Figure 7a). Note that τ does not represent the composomal lifetime. In fact, the most probable target compotype lifetime (taking for simplicity the maximal time from each simulation) is 30, and the average is 434 generations (Figure 7c). The other similarity autocorrelation parameter, H_0, is interpreted here as showing the residual compositional similarity among assemblies along many generations in the entire simulation. Thus, $1 - H_0$ is taken as proportional to the overall compositional diversity of assemblies across the entire simulation. Note that H_0 is not strongly correlated with the compotype count (Figure 16 in Appendix A.5, correlation coefficient -0.049, $r^2 = 0.89$) and therefore constitutes a rather independent diversity assessment attribute. The most probable H_0 value is ~0.5, with a smaller probability peak at $H_0 \approx 1$. The latter stems from simulations in which a single compotype tends to dominate.

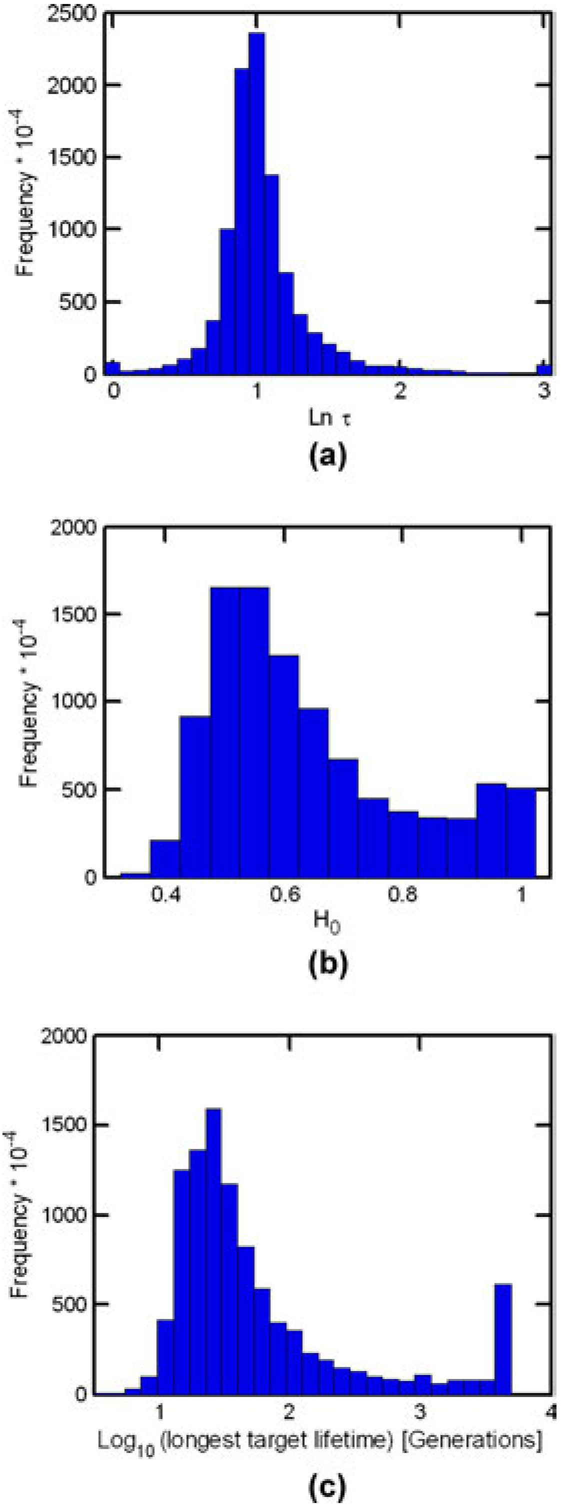

Figure 7. Distributions of τ, H_0, and composome duration. (a) A histogram of τ: unit is number of generations, and the rightmost bin represents all data with $\ln\tau > 3$. (b) A histogram of H_0, unitless. (c) Distribution of the longest appearance of target compotypes. Data in panels is the same as in Figure 2.

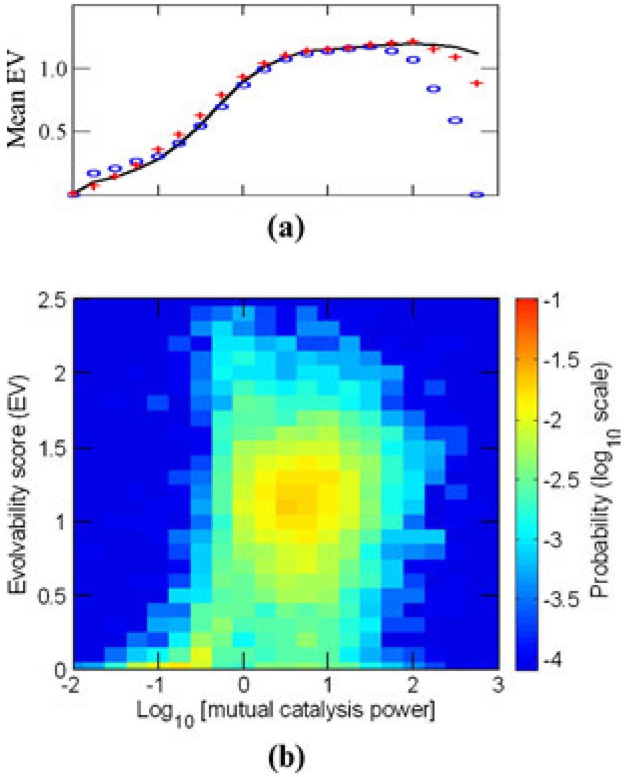

Figure 8. The dependence of the evolvability score (*EV*) on p_{mc}. Details are as in Figure 3.

A score is defined, which could arguably assess a GARD system's evolvability:

$$EV = \tau(1 - H_0) \tag{11}$$

A larger evolvability score will typically arise when the system concomitantly displays appreciable trans-generation compositional preservation and higher overall compositional diversity. This compound

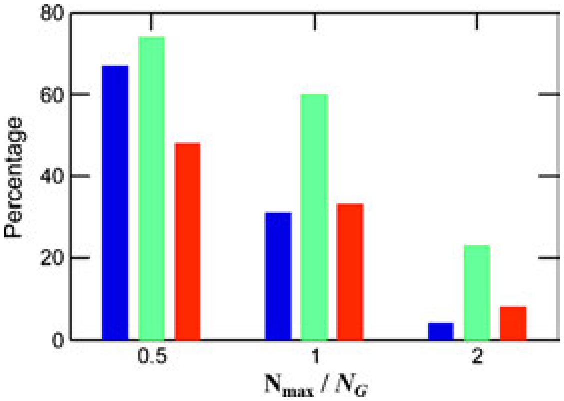

Figure 9. The percentage of regular-GARD instances exhibiting extreme evolution-related parameters as a function of maximal assembly size (N_{max}). In the electronic version, the values taken are: compotype count >2 (blue), evolvability score >1 (green), and selection excess >1 (red). All parameters, except N_{max}, are as in Figure 3b. Full histograms and their related data are given in Appendix A.6 (Figure 17 and Table 3).

measure reflects similar definitions of evolvability [7, 43]. Similar to the selection excess and number of compotypes, a clear trend appears, whereby high evolvability scores are much more prevalent for p_{mc} values higher than 1 (Figure 8).

3.5 The Effect of Assembly Size

The effect of the assembly pre-fission size $N_{\max}$ on GARD's evolutionlike behavior was studied by performing two additional sets of 10,000 simulations, each with the same parameters as in Table 1 except for $N_{\max} = N_G/2$ and $2N_G$ (Figure 9; Appendix A.6 Figure 17 and Table 3). While for the smaller $N_{\max}$ value, the percentage of beneficial outcomes seems to be even higher than for the nominal $N_{\max} = N_G$, a larger $N_{\max}$ value appears to have a disruptive effect because the system is nearing the equilibrium steady state [23]. This is especially seen for the compotype count and the evolvability score.

4 Discussion

4.1 The Significance of Mutual Catalysis

One of the dominant concepts in prebiotic evolution research is the replicator-first scenario [10, 32, 40]. Based on the concept that molecular replication is related to self-catalysis [41], such views may be perceived as related to the RNA-first scenario, positing that life began with a unique self-replicating polyribonucleotide. In this realm, it is argued that more complex interaction networks have arisen only at later stages, as when precursors for the autocatalytic molecule have been exhausted [31]. Our simulation results demonstrate an advantage for a network-first scenario, in which a large number of molecular components mutually interact. While arising from a metabolism-related framework, such results may be taken as relevant to the question of whether life's early precursors were a set of replicators or a metabolic network. Note that the present work makes a direct comparison between a metabolic network with frequent self-catalytic interactions and a metabolic network with frequent mutually catalytic interactions, and therefore has only indirect relevance to the question of the validity of replicator models. It is conceivable that future work incorporating templating biopolymers together with mutually catalytic networks will better resolve this issue.

A widespread argument against metabolism-like entities being the first seed of life is the assertion that metabolic networks cannot store and propagate information. The GARD model may be viewed as a counterexample, as it is endowed with a (limited) capacity to store and propagate compositional information. This has implications for a set of previously proposed models involving networks of molecular interactions. Two of the earliest relevant concepts are Gánti's chemoton [12, 13, 63] and Maturana and Varella's autopoietic systems [35, 67]. Autopoiesis characterizes a spatially confined network of molecular components, whose mutual interactions continuously regenerate the network itself. The chemoton is described as a system of three subnetworks: metabolite generation, template copying, and membrane synthesis. We prudently suggest that GARD may be viewed as a special case of autopoietic-chemoton-like models, where template copying and compartmentation are embodied in one entity, and a continuous supply of metabolites is afforded by the spontaneous accretion of lipids from the buffered environment.

4.2 The Effect of Mutual Catalysis on GARD Diversity and Evolvability

An important result of this work is that networks within a certain range of kinetic parameters, namely those that exhibit excess mutual catalysis, lead to enhanced diversity and evolvability of GARD compotypes. The compotype count is a direct indication of the degree of composomal diversity. This result is related to an important aspect of early evolution: Self-catalysts tend to propagate their own identity and suppress processes essential for the increasing complexity necessary for transitions from early seeds of life toward systems resembling present-day life. The presently demonstrated importance of mutual catalysis echoes the notion of systems prebiology [21, 57], whereby it is suggested that life began its trajectory from complex chemical mixtures obeying network behavior similar to that of metabolism in present-day cells.

4.3 Compotypes as Selection Targets

One of the unique corollaries of the GARD model is the emergence of composomes, dynamic states of compositional assemblies that embody both metabolism-like characteristics and a rudimentary capacity to store and propagate molecular information [49]. Composomes may be considered as forming bridges between seemingly disparate views of the early seeds of life: metabolism first and replicators first. Compotypes are further defined as centers of mass of composome clusters, which may be regarded as analogous to species or quasi species [6]. This is due to the fact that a compotype is a distinct entity, with distinct physical properties and hence fitness encoded in its compositional information, different from those of other compotypes but still harboring considerable internal variability of constituents. Therefore, compotypes are considered as natural targets of selection, as compared to randomly chosen compositions, as previously pursued [68]. Note that here we have a measure of selection inherently present in the GARD model even in the absence of external selective pressure, due to the fact that different composomes have different average growth rates. This is seen in the present population GARD simulations, which are seeded with a random population, but show a gradual increase of the population frequency of a specific compotype even in the absence of externally imposed selection. This increase comes at the expense of other compositions because of the constant population condition.

4.4 Selection in a GARD

The present results show that GARD assemblies can exhibit positive or negative selection toward a compotype target, as well as no selection at all. While in regular GARD the overall average selection excess is merely 1.05, it is noteworthy that as many as 10% of the simulations show high selection excess, >1.5. Importantly, these general results are borne out both in simulations of the regular model and in simulations involving populations of assemblies. Previously, GARD population dynamics has been studied by addressing various emergent properties, including a comparison of finite and infinite chemical environments [38]. Another study [70] showed that compositional inheritance also emerges in the GARD model variants involving assembly populations and spatial proximity interaction effects, and that it emerges in both a thermodynamic and a kinetic interaction regimen.

Analyzing GARD, both positive and negative selections can be observed in practice only when the underlying network exhibits mutual catalysis excess. This conclusion is strengthened by its demonstration in two different simulation modes: in the regular model and in populations. Notably, positive selection is observed appreciably more often in population GARD simulations, perhaps reflecting the advantage of addressing populations of competing entities with different reproductive rates. Furthermore, this selection response tends to be augmented as the number of coexisting compotypes increases in a given simulation, which may indicate a capacity of selective forces to provide an edge to the target compotype in inter-compotype competition. Further in-depth analyses (currently underway) of the ultrastructure of the β network, as well as subnetworks (quasi compartments [68]), could lead to a better understanding of the influence of p_{mc} and the compotype count on selection.

The present method for biasing the growth rate of a GARD target composition is in principle similar to that used previously [68]. In both cases, modifications are in effect introduced to β matrix elements. However, the previous analysis utilizes an interim formalism, the Eigen equation, for replication-mutation dynamics [10], and the selection-related modification is exerted by multiplying the growth rate by fH, defined in the same way as in Equation 8. The method utilized here involves direct modification (Equation 9), a possible explanation for the discrepant results obtained by the two reports. There are, however, additional significant differences between the two studies: (a) a prefission value $N_{\max} = 100$ used here, as compared to $N_{\max} = 6$ used previously, an obligatory small value required for the realistic application of Eigen's formalism with the available computing power; (b) a large difference in repertoire size ($N_G = 100$ here versus $N_G = 10$ in the earlier study); (c) the performance here of 10,000 random simulations, considered essential for proper statistical rigor, as compared to only a single simulation done previously. Both points (b) and (c) provide a significant edge to the present simulations in sampling the β interaction space, which allows drawing conclusions with higher certainty. In the future it will be interesting to consider additional methodologies to exert external selection. One could be a variant of the presently used method, whereby the β network will be biased by a constant factor and not employing target similarity-oriented bias. Another could be biasing the environmental concentration ρ_i (Equation 2) by a constant factor based on the molecules that are contained in the target compotype.

5 Conclusion

The GARD model embodies the inheritance of compositional information in the realm of a lipid world scenario for early evolution [20, 21, 23, 27, 48, 49, 51, 55–57]. The GARD has recently been pursued in several additional publications [20, 39, 68, 70] and has been chosen as an archetypal metabolism-first realization [68]. This suggests that despite being a simulated toy model, the GARD has sufficient complexity to shed light on some important questions in the field of prebiotic origins. In the present work an attempt is made to shed further light on some of the GARD's evolutionary features. It is expected that the present insights will become instrumental in further efforts to extend the GARD beyond the monomer world [54], as has been preliminarily explored [55]. This might be necessary to reveal the capacity of the GARD model to capture the much-needed open-ended attributes of natural selection and evolution.

Acknowledgments
We thank Raphael Zidovetzki and Tsviya Olender for enlightening comments, Aron Inger for assistance in coding, and Ágnes Tóth Petróczy, Azhar Ali Shah, Hugues Bersini, Leong Ting Lui, and Natalio Krasnogor for discussions. This work is partly supported by EU-FP7 project MATCHIT and by the Crown Human Genetics Center at the Weizmann Institute of Science.

References

1. Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461.
2. Anet, F. A. (2004). The place of metabolism in the origin of life. *Current Opinion in Chemical Biology*, *8*(6), 654–659.
3. Bachmann, P. A., Luisi, P. L., & Lang, J. (1992). Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature*, *357*, 57–59.
4. Barabas, B., Toth, J., & Palyi, G. (2010). Stochastic aspects of asymmetric autocatalysis and absolute asymmetric synthesis. *Journal of Mathematical Chemistry*, *48*(2), 457–489.
5. Bedau, M. A. (2010). An Aristotelian account of minimal chemical life. *Astrobiology*, *10*(10), 1011–1020.
6. Biebricher, C. K., & Eigen, M. (2006). What is a quasispecies? *Current Topics in Microbiology and Immunology*, *299*, 1–31.
7. Brookfield, J. F. Y. (2009). Evolution and evolvability: Celebrating Darwin 200. *Biology Letters*, *5*(1), 44–46.
8. Chen, I. A., & Walde, P. (2010). From self-assembled vesicles to protocells. *Cold Spring Harbor Perspectives in Biology*, 2/7/a002170.
9. Dyson, F. J. (1982). A model for the origin of life. *Journal of Molecular Evolution*, *18*(5), 344–350.
10. Eigen, M., & Schuster, P. (1977). Hypercycle—Principle of natural self-organization. A. Emergence of hypercycle. *Naturwissenschaften*, *64*(11), 541–565.
11. Fontana, W., & Buss, L. W. (1994). What would be conserved if "the tape were played twice"? *Proceedings of the National Academy of Sciences of the U.S.A.*, *91*(2), 757–761.
12. Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units—Chemotons. *Biosystems*, *7*, 15–21.
13. Gánti, T. (1997). Biogenesis itself. *Journal of Theoretical Biology*, *187*(4), 583–593.
14. Gesteland, F. R., Cech, R. T., & Atkins, F. J. (1999). *The RNA world* (p. 709). Cold Spring Harbor, MA: Cold Spring Harbor Laboratory.
15. Gilbert, W. (1986). Origin of life—The RNA world. *Nature*, *319*, 618–618.

16. Gillespie, D. T. (1976). General method for numerically simulating stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434.

17. Gillespie, D. T. (1977). Master equations for random walks with arbitrary pausing time distributions. *Physics Letters A*, *64*(1), 22–24.

18. Hayden, E. J., von Kiedrowski, G., & Lehman, N. (2008). Systems chemistry on ribozyme self-construction: Evidence for anabolic autocatalysis in a recombination network. *Angewandte Chemie—International Edition*, *47*(44), 8424–8428.

19. Hughes, R. A., Robertson, M. P., Ellington, A. D., & Levy, M. (2004). The importance of prebiotic chemistry in the RNA world. *Current Opinion in Chemical Biology*, *8*(6), 629–633.

20. Hunding, A., Kepes, F., Lancet, D., Minsky, A., Norris, V., Raine, D., Sriram, K., & Root-Bernstein, R. (2006). Compositional complementarity and prebiotic ecology in the origin of life. *Bioessays*, *28*, 399–412.

21. Inger, A., Solomon, A., Shenhav, B., Olender, T., & Lancet, D. (2009). Mutations and lethality in simulated prebiotic networks. *Journal of Molecular Evolution*, *69*(5), 568–578.

22. Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature*, *418*, 214–221.

23. Kafri, R., Markovitch, O., & Lancet, D. (2010). Spontaneous chiral symmetry breaking in early molecular networks. *Biology Direct*, *5*(38). doi: 10.1186/1745-6150-5-38

24. Kaneko, K. (2002). Kinetic origin of heredity in a replicating system with a catalytic network. *Journal of Biological Physics*, *28*(4), 781–792.

25. Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. Oxford, UK: Oxford University Press.

26. Laidler, K. J. (1986). The development of theories of catalysis. *Archive for History of Exact Sciences*, *35*(4), 345–374.

27. Lancet, D., Kafri, R., & Shenhav, B. (2002). Compositional genomes: Pre-RNA information transfer in mutually catalytic assemblies. *Geochimica et Cosmochimica Acta*, *66*(15A), A429–A429.

28. Lancet, D., Kedem, O., & Pilpel, Y. (1994). Emergence of order in small autocatalytic sets maintained far from equilibrium—Application of a probabilistic receptor affinity distribution (RAD) model. *Berichte der Bunsen-Gesellschaft—Physical Chemistry Chemical Physics*, *98*(9), 1166–1169.

29. Lancet, D., Sadovsky, E., & Seidemann, E. (1993). Probability model for molecular recognition in biological receptor repertoires—Significance to the olfactory system. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(8), 3715–3719.

30. Lee, D. H., Severin, K., Yokobayashi, Y., & Ghadiri, M. R. (1997). Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature*, *390*, 591–594.

31. Lifson, S. (1997). On the crucial stages in the origin of animate matter. *Journal of Molecular Evolution*, *44*, 1–8.

32. Lifson, S., & Lifson, H. (1999). A model of prebiotic replication: Survival of the fittest versus extinction of the unfittest. *Journal of Theoretical Biology*, *199*(4), 425–433.

33. Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience*, *51*(5), 341–352.

34. Luisi, P. L., Walde, P., & Oberholzer, T. (1999). Lipid vesicles as possible intermediates in the origin of life. *Current Opinion in Colloid & Interface Science*, *4*(1), 33–39.

35. McMullin, B. (2000). Remarks on autocatalysis and autopoiesis. *Annals of the New York Academy of Sciences*, *901*(1), 163–174.

36. Moran, P. A. P. (1958). Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, *54*, 60–71.

37. Muller, U. F. (2006). Re-creating an RNA world. *Cellular and Molecular Life Sciences*, *63*(11), 1278–1293.

38. Naveh, B., Sipper, M., Lancet, D., & Shenhav, B. (2004). Lipidia: An artificial chemistry of self-replicating assemblies of lipid-like molecules. In *9th International Conference on the Simulation and Synthesis of Living Systems (ALIFE9)* (pp. 466–471).

39. Norris, V., Hunding, A., Kepes, F., Lancet, D., Minsky, A., Raine, D., Root-Bernstein, R., & Sriram, K. (2007). The first units of life were not simple cells. *Origins of Life and Evolution of Biospheres*, *37*(4–5), 429–432.

40. Orgel, L. (2000). Origin of life—A simpler nucleic acid. *Science*, *290*, 1306–1307.

41. Orgel, L. E. (1992). Molecular replication. *Nature*, *358*, 203–209.

42. Orgel, L. E. (2004). Prebiotic chemistry and the origin of the RNA world. *Critical Reviews in Biochemistry and Molecular Biology*, *39*(2), 99–123.

43. Pigliucci, M. (2008). Opinion—Is evolvability evolvable? *Nature Reviews Genetics*, *9*(1), 75–82.

44. Plasson, R., Brandenburg, A., Jullien, L., & Bersini, H. (2010). Autocatalyses. In H. Fellermann, M. Dorr, M. Hanczyc, L. Laursen, S. Maurer, D. Merkle, P. Monnard, K. Stoy, & S. Rasmussen (Eds.), *Twelfth International Conference on the Synthesis and Simulation of Living Systems* (pp. 4–11). Cambridge, MA: MIT Press.

45. Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., & Stadler, P. F. (Eds.). (2009). *Protocells: Bridging nonliving and living matter* (p. 684). Cambridge, MA: MIT Press.

46. Rosenwald, S., Kafri, R., & Lancet, D. (2002). Test of a statistical model for molecular recognition in biological repertoires. *Journal of Theoretical Biology*, *216*(3), 327–336.

47. Schuster, P., & Stadler, F. (2003). Networks in molecular evolution. *Complexity*, *8*(1), 34–42.

48. Segre, D., Ben-Eli, D., Deamer, D. W., & Lancet, D. (2001). The lipid world. *Origins of Life and Evolution of the Biosphere*, *31*(1–2), 119–145.

49. Segre, D., Ben-Eli, D., & Lancet, D. (2000). Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4112–4117.

50. Segre, D., & Lancet, D. (2000). Composing life. *Embo Reports*, *1*(3), 217–222.

51. Segre, D., Lancet, D., Kedem, O., & Pilpel, Y. (1998). Graded autocatalysis replication domain (GARD): Kinetic analysis of self-replication in mutually catalytic sets. *Origins of Life and Evolution of the Biosphere*, *28*(4–6), 501–514.

52. Segre, D., Shenhav, B., Kafri, R., & Lancet, D. (2001). The molecular roots of compositional inheritance. *Journal of Theoretical Biology*, *213*(3), 481–491.

53. Shapiro, R. (2006). Small molecule interactions were central to the origin of life. *Quarterly Review of Biology*, *81*(2), 105–125.

54. Shapiro, R. (2007). A simpler origin for life. *Scientific American*, *296*(4), 46–53.

55. Shenhav, B., Bar-Even, A., Kafri, R., & Lancet, D. (2005). Polymer GARD: Computer simulation of covalent bond formation in reproducing molecular assemblies. *Origins of Life and Evolution of the Biosphere*, *35*(2), 111–133.

56. Shenhav, B., Oz, A., & Lancet, D. (2007). Coevolution of compositional protocells and their environment. *Philosophical Transactions of the Royal Society B—Biological Sciences*, *362*, 1813–1819.

57. Shenhav, B., Solomon, A., Lancet, D., & Kafri, R. (2005). Early systems biology and prebiotic networks. *Transactions on Computational Systems Biology*, *1*, 14–27.

58. Silvestre, D. A. M. M., & Fontanari, J. F. (2008). The information capacity of hypercycles. *Journal of Theoretical Biology*, *254*(4), 804–806.

59. Sole, R. V., Rasmussen, S., & Bedau, M. (2007). Introduction. Artificial protocells. *Philosophical Transactions of the Royal Society B—Biological Sciences*, *362*, 1725–1725.

60. Stadler, P. F. (1991). Dynamics of autocatalytic reaction networks. 4. Inhomogeneous replicator networks. *Biosystems*, *26*(1), 1–19.

61. Steinley, D. (2006). *K*-means clustering: A half-century synthesis. *British Journal of Mathematical & Statistical Psychology*, *59*, 1–34.

62. Szathmáry, E. (1995). A classification of replicators and lambda-calculus models of biological organization. *Proceedings of the Royal Society B—Biological Sciences*, *260*(1359), 279–286.

63. Szathmáry, E., Santos, M., & Fernando, C. (2005). Evolutionary potential and requirements for minimal protocells. *Topics in Current Chemistry*, *259*, 167–211.

64. Szathmáry, E., & Smith, J. M. (1997). From replicators to reproducers: The first major transitions leading to life. *Journal of Theoretical Biology*, *187*(4), 555–571.

65. Thomas, J. A., & Rana, F. R. (2007). The influence of environmental conditions, lipid composition, and phase behavior on the origin of cell membranes. *Origins of Life and Evolution of Biospheres*, *37*(3), 267–285.

66. Ullrich, A., Rohrschneider, M., Scheuermann, G., Stadler, P. F., & Flamm, C. (2011). In silico evolution of early metabolism. *Artificial Life*, *17*(2), 87–108.

67. Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187–196.

68. Vasas, V., Szathmáry, E., & Santos, M. (2010). Lack of evolvability in self-sustaining autocatalytic networks constrains metabolism—First scenarios for the origin of life. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1470–1475.

69. Weber, A. L. (2000). Sugars as the optimal biosynthetic carbon substrate of aqueous life throughout the universe. *Origins of Life and Evolution of the Biosphere*, *30*(1), 33–43.

70. Wu, M., & Higgs, P. G. (2008). Compositional inheritance: Comparison of self-assembly and catalysis. *Origins of Life and Evolution of Biospheres*, *38*(5), 399–418.

Appendix

A.1 Distribution and Sampling of the GARD Matrix β

While not much is known about the values of the rate enhancement between prebiotic molecules, there is a need to consider such values by a physically reasonable method. β_{ij} values are randomly generated based on a lognormal distribution

$$P(\beta_{ij}) = \frac{1}{\beta_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln \beta_{ij} - \mu)^2}{2\sigma^2}\right) \tag{12}$$

where μ and σ are the mean and standard deviation, respectively, which can be considered as a "natural" distribution [33], in accordance with the receptor affinity distribution formalism [28, 29, 46], and it was also shown that a lognormal β increases the reproduction fidelity over the normal β in GARD [52]. Each randomization of the β network may be thought of as representing the relative rates of the N_G molecules as they might ensue from different possible GARD environments.

Self-catalysis in GARD is represented by

$$i_{\text{out}} + i_{\text{in}} \xrightarrow{\beta_{ii}} i_{\text{in}} + i_{\text{in}} \tag{13}$$

Often self-catalysis is written as [4]

$$X + Y \xrightarrow{\beta_{XY}} Y + Y \tag{14}$$

The seeming dichotomy between the notations β_{ii} and β_{XY} is clarified on noting that in the GARD, molecules have two states, *in* and *out*, which behave as distinct chemical species. While it is possible that more complex pathways would also be autocatalytic [44], this work refers to self-catalysis as the simplest closed subnetwork of the β network, containing one element (Figure 1).

A.2 Fitting the Similarity Autocorrelation Function

The fitting procedure is as follows: (1) Calculate H_0 as the mean of $c(\Delta t)$ in the interval $[GEN/4, GEN/2]$. (2) Guess $\tau^\ddagger$ as the first instance $c(\Delta t)$ drops below H_0. (3) Smooth the $c(\Delta t)$ tail by forcing: $c(\Delta t > \tau^\ddagger) = H_0$. (4) Fit an exponential (Equation 7) to the smoothed $c(\Delta t)$, using nonlinear least squares with a tolerance of 10^{-5}.

Examples are given in Figure 12 in Appendix A.4.

A.3 *p*-Values

See Table 2.

Table 2. Student's *t*-test statistical analysis for the selection excess of β networks exhibiting $p_{mc} > 100$ (Equations 4 and 10). Test was run using MATLAB function *ttest*, against the null hypothesis that the data are a random sample from a normal distribution with mean 1.0, per specific ranges of lognormal random seeds.

Random-seed range	$p_{mc} > 100$*	Selection excess†	*p*-Value
50–300	3	0.973 ± 0.0395	3.57×10^{-1}
300–800	5	1.311 ± 0.389	1.49×10^{-1}
1,000–3,500	40	1.100 ± 0.325	5.95×10^{-2}
5,000–10,000	70	1.119 ± 0.248	1.45×10^{-4}
1–10,000	143	1.105 ± 0.272	8.03×10^{-6}

*The number of networks exhibiting high p_{mc} value.

†Mean and standard deviation of the selection excess of these networks (under regular GARD simulations).

A.4 Examples

See Figures 10–14.

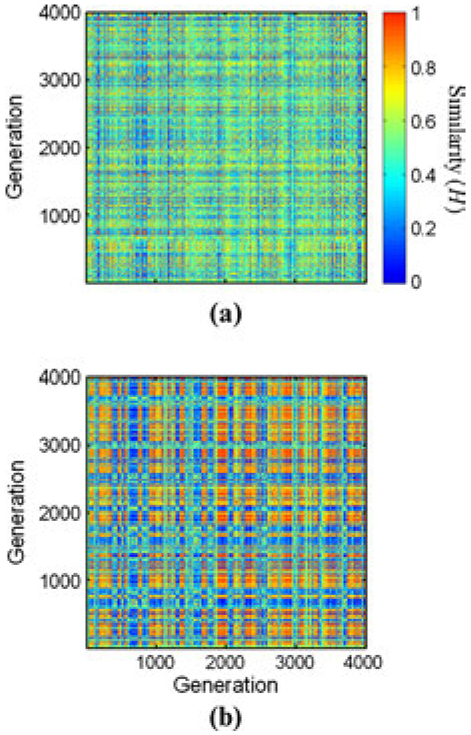

Figure 10. Example of carpets from two regular-GARD simulations with lognormal seeds 42 and 41 (a and b, respectively) and the rest of the parameters as in Table 2. Compotype counts are 4 and 2, respectively. β matrices are presented in Figure 11, and functions $c(\Delta t)$ in Figure 12.

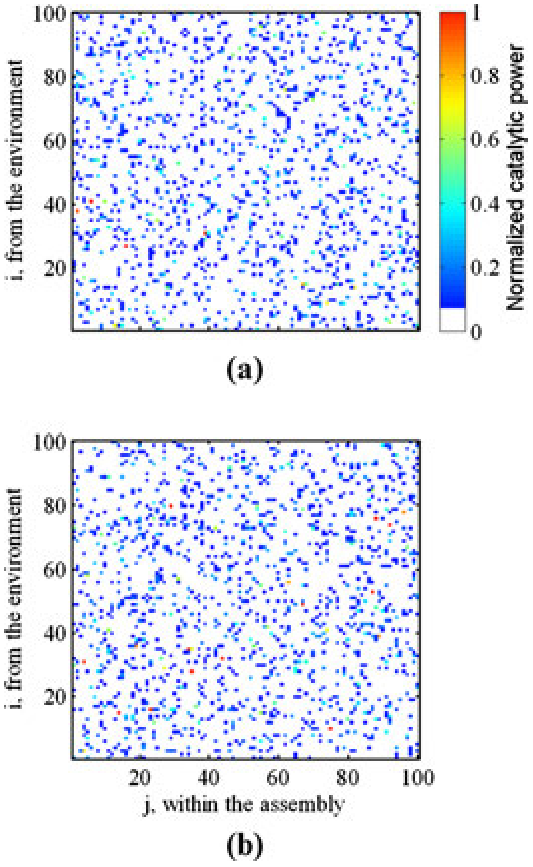

Figure 11. β matrices for the two simulations in Figure 10. p_{mc} values are 1.98 and 0.81, respectively. To better express the richness of the β matrix, catalytic values are scaled according to $\beta_{ij} = 2^{\log 10\ \beta^0_{ij} - 4}$ (values of β^0_{ij} are generated according to Equation 12).

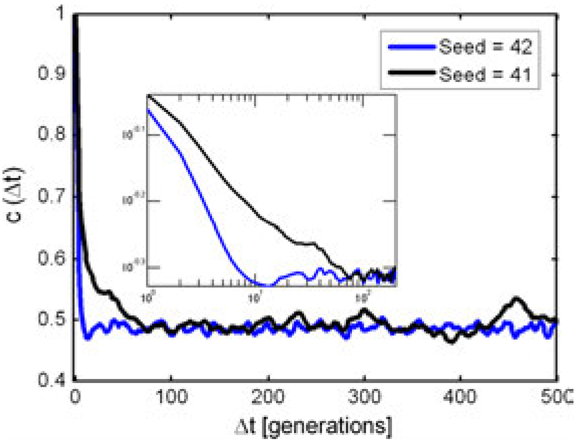

Figure 12. Functions $c(\Delta t)$ for the two simulations in Figure 10. Insert shows initial decay on a log-log scale. Fitted parameters for Equation 7 are $\tau = 2.57$, $H_0 = 0.49$ (seed = 41), and $\tau = 6.32$, $H_0 = 0.50$ (seed = 42).

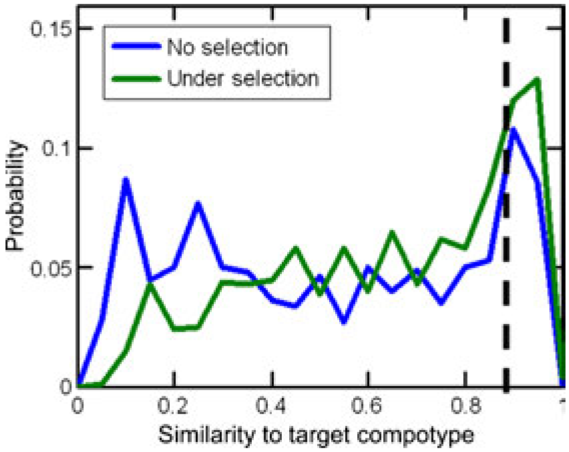

Figure 13. Example histograms of similarity between the target compotype from a regular GARD simulation, for a population of 1,000 assemblies, with and without selection. A cutoff of $H \geq 0.9$ (dashed line) is imposed to identify the frequency of the compotype in the GARD population. Simulation details are lognormal seed = 3, GEN = 5,000, and the rest as in Table 2.

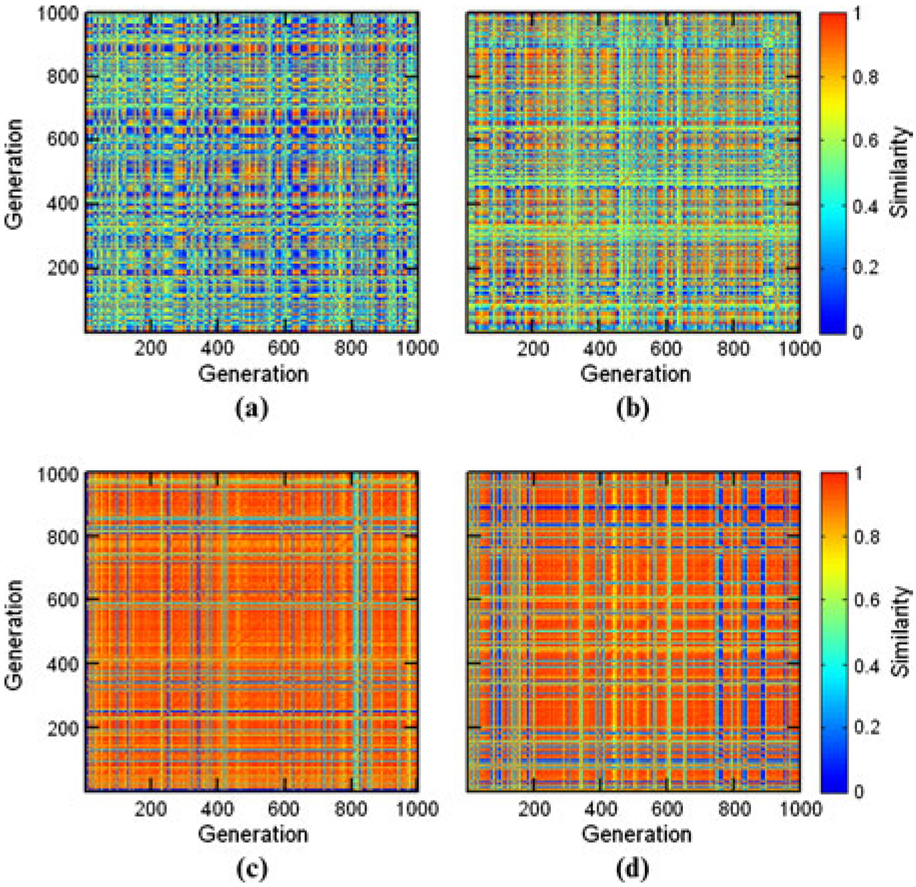

Figure 14. Regular-GARD similarity carpets before and after selection. (a) Similarity carpet of the GARD instance generated with lognormal seed 114. The frequency of the target compotype is $f_T = 0.27$. (b) The same carpet as (a), after applying selection pressure, whereby the new frequency of the target is $f_{T'} = 0.34$. (c) Similarity carpet of the GARD instance generated with lognormal seed 168. Here $f_T = 0.82$. (d) The results after applying selection pressure. Here $f_{T'} = 0.74$. Simulation parameters are in Table 1.

A.5 Selection Excess and the Number of Compotypes

See Figures 15–16.

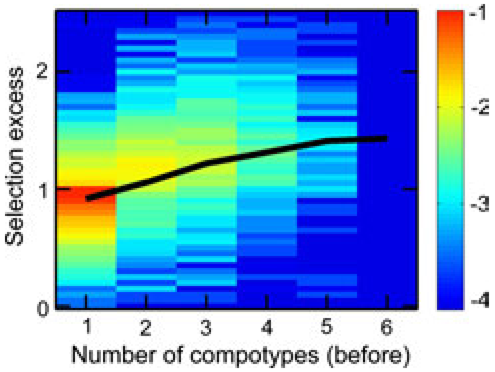

Figure 15. The dependence of selection excess on the number of compotypes (before selection). Black solid line plots the average selection excess per compotype count. Figure details are as in Figure 2a.

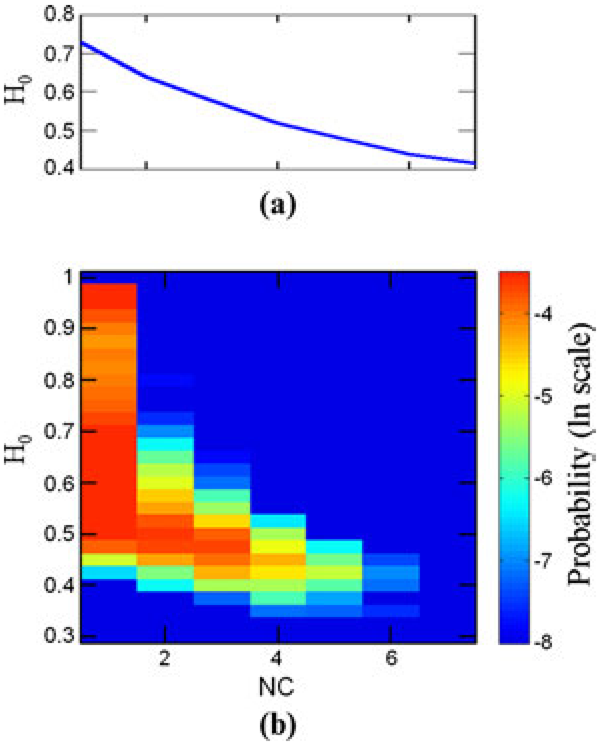

Figure 16. The weak dependence of H_0 on the number of compotypes (*NC*). (a) Average H_0 versus *NC* after 5-point moving-average smoothing. Fitting the smoothed data to a linear curve gives a slope of –0.0485 with $r^2 = 0.89$. (b) Density plot of the probability to have a simulation with a pair of H_0 and *NC* values. In the electronic version, the color represents the normalized probability to find a network with such a pair (in scale; red means that about 300 simulations fall in this bin). Simulation parameters are as in Figure 15.

A.6 Assembly Size

See Figures 17 and Table 3.

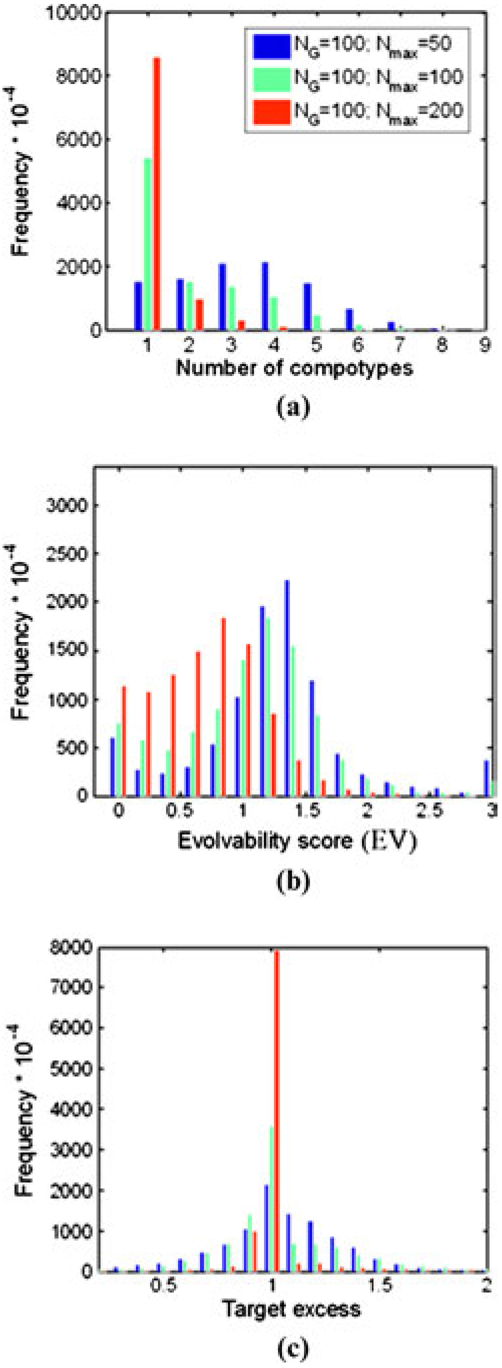

Figure 17. Histograms of the three evolution-related parameters, per $N_{\max}$ values. (a) Number of compotypes (*NC*). In the electronic version, blue bars are with $N_{\max} = N_G/2$, green bars are with $N_{\max} = N_G$, and red bars are with $N_{\max} = 2N_G$. The rest of parameters are as in Figure 15. (b) Evolvability score (*EV*). (c) Selection excess (*SE*).

Table 3. Mean values collected from Figure 17. Number in parenthesis refers to the percentage of simulations that show positive or negative selection.

	Mean value		
Number	$N_{\mathrm{max}} = 2N_G$	$N_{\mathrm{max}} = N_G$	$N_{\mathrm{max}} = N_G/2$
NC	1.20	2.03	3.38
EV	0.72	1.11	1.35
SE	1.01	1.05	1.04
SE > 1.05	1.36 (8%)	1.38 (33%)	1.28 (48%)
SE < 0.95	0.85 (13%)	0.77 (31%)	0.71 (30%)

5.1.1. Network motifs and their effect on selection in GARD

This work was done with collaboration with Prof. Uri Alon and Dr. Avi Mayo from the Weizmann Institute.

The result above, that mutual catalysis excess is a required condition for effective evolvability, demonstrates an advantage for a network first theme in the origin of life, and calls for a more detailed analysis of the inner structure of β.

To this end, β is analyzed under the scope of network motifs - basic interaction patterns that recur throughout biological and other networks [1]. The goal is to understand how the spectrum of motifs in a given β affects the selection behavior observed.

5.1.1.1. Binarizing β

As motifs are typically considered in binary networks while β is graded (an edge in the former can only have a weight of 0 or 1, whereas in the latter it can acquire a range of β_{ij} values), β is binarized using a cutoff which determine the minimal β_{ij} value to be considered:

$$\beta_{ij} = \begin{Bmatrix} 1 & \beta_{ij} \geq cutoff \cdot \sum_{i,j} \beta_{ij} \\ 0 & \beta_{ij} < cutoff \cdot \sum_{i,j} \beta_{ij} \end{Bmatrix}$$

Equation 9

In each β, the counts of the 13 commonly studied triplet motifs [1] are found by finding which molecular types exhibits which triplet motif (with a total of $N_G/3!(N_G-3)!$ possible triads for cutoff=0), and the count of a given motif in a β is standardized according to:

$$\text{standard score } (\mathrm{x}, \beta) = \frac{\text{count}(\mathrm{x}, \beta) - \mu(\mathrm{x})}{\sigma(\mathrm{x})}$$

Equation 10

Where x is the motif index (x=1..13) and μ and σ respectively represents the mean and standard deviation of its counts across all β networks studied. This extends an earlier study on the structure of β [128].

5.1.1.2. The motif spectrum of β

Figure 10 shows the overall motif counts for different cutoff values, which is the underlying motif spectrum. The shape of the spectrum does not depend on the exact cutoff used. The relative counts of the different motifs, especially those with the same number of edges, can be understood when considering the graded to binary transition (Figure 11). When there is no cutoff, any three nodes will be maximally connected, that is motif #13, because β_{ij} values are always positive as they are picked from a lognormal distribution. As the cutoff gradually increases, removal of any single edge from motif #13 will lead into #12. By removing another edge (e.g., by increasing the cutoff further), #11, #10, #8 or #6 have equal probability to form, depending on the exact edge removed. This can be continued further, giving rise to shape of the motif spectrum. Thus, the general structure of binary β is constant on average, and the counts of motifs depend on the exact value of the cutoff.

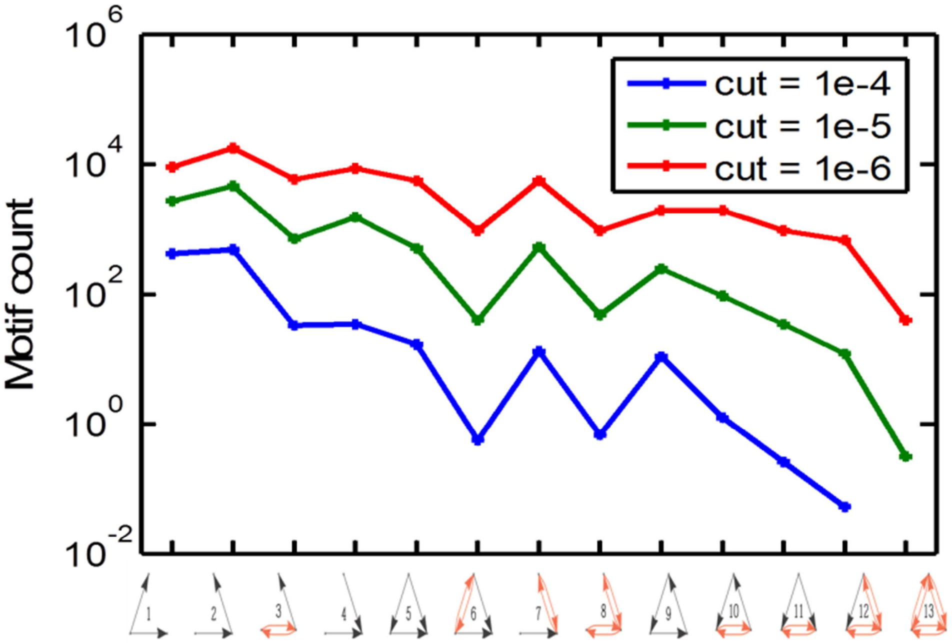

Figure 10: The underlying network motif spectrum of β, under different cutoffs (Equation 9). Data is averaged over 1,000 networks. Default cutoff used = 1e-5.

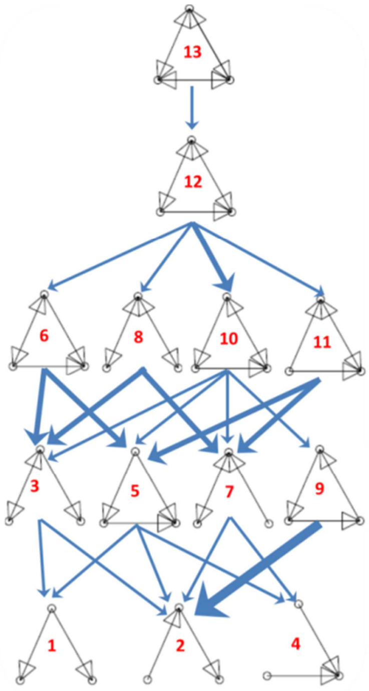

Figure 11: Network motifs hierarchy. When no cutoff is applied, any three molecules exhibit motif #13 by definition (see text).

5.1.1.3. Motifs in selection

In order to understand how does the motif spectrum effect selection, networks are grouped according to their selection outcome (positive, none or negative) and the average motif spectrum is calculated for each group (Figure 12). Networks which exhibit positive selection show over representations of two out of the three doubly-connected motifs (i.e. motifs with only two edges), #1 and #2, while deficient with the third (motif #4). An opposite behavior is observed with networks exhibiting negative selection. Presumably, a network with more of motifs #1 and #2 is more connected, i.e. existence of more paths connecting distant parts a network, because each edge in these motifs points *to* a different molecular type, whereas #4 reduces the connectivity because the two edges point to the same molecular type. It is suggested that a highly connected network exhibit positive selection because it is possible to have an increased flux towards the compotype when applying the selection pressure. Additionally, networks with the average spectrum exhibit no selection (Figure 12).

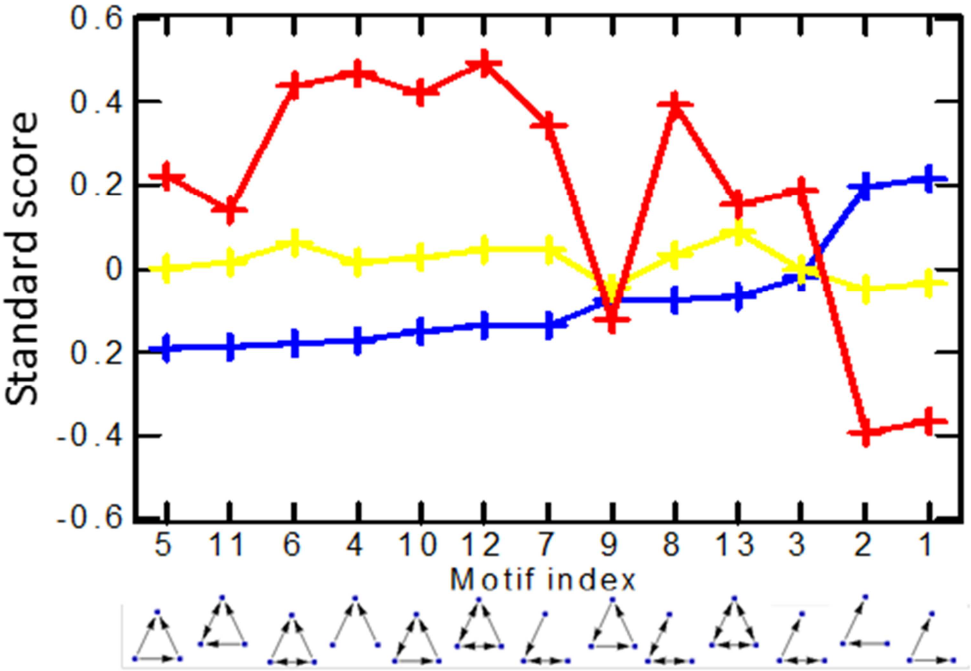

Figure 12: Average motifs standard scores (relative to data in Figure 10), collected for simulations showing positive, none and negative selection (blue, yellow and red, respectively). Positive selection is defined as selection-excess>1.05 (SE, Equation 6), negative selection as SE<0.95 and the rest is no selection.

5.1.2. A completely selfish β

It is interesting to do a thought experiment, as to the outcome of a simulation based on a completely selfish β, e.g. for i≠j β_{ij}=0 and β_{ii} values drawn from a lognormal distribution. In such a case, the value of the mutual-catalysis-power parameter will assume its lowest value = N_G^e (p_{mc}. Equation 4 in [80]). It is suggested that in such a case the most frequent compotype in the simulation will be composed almost entirely out of the molecular type with the highest β_{ii} value as it will act as an exponential replicator. Additional compotypes may appear composed majorly out of a single molecular type yet with different levels of additional molecular types depending on the ratios of β_{ii} values, due to the stochastic nature of GARD. Similar cases have indeed been encountered (the left most bin in Figure 3b in [80], for example)

5.2. Chemistry biased fission

GARD's fission is a stochastic process, on average creating two equal progeny, without regard to mutual molecular interactions within the splitting assembly. This fission behavior may not be true to reality. Therefore, it was examined how deviating from this rule effects GARD behavior.

A new fission action proposed, with aim to allow a more realistic budding-like split in GARD by invoking a process with dynamics analogous to that of assembly accretion. This type of fission is referred to as "chemistry biased fission".

A major drawback of random fission is that it could destroy the composition the parent assembly transfers to progeny, thus hampering its ability to faithfully replicate. The idea of the new fission was guided by the notion that if a molecule of type i was drawn into an assembly by molecules of type j (as directed by the value of β_{ij}) then i and j will also favorably interact within the assembly and are likely to be located spatially close within the assembly towards fission. This involve a concept derived from the study of present-day cellular membranes, namely rafts, i.e. membrane microdomains that are more ordered and tightly packed than their surrounding bilayer [10, 77]. Rafts are related to the membrane function as they influence membrane fluidity and trafficking [64, 99] and even relate to signal transduction [33]. In GARD, chunks from a parent assembly, strongly connected by a network of interaction in β, are budded together during chemistry biased fission and transfer as a single unit, akin to a raft, to a progeny.

Two types of new split action developed and studied: competitive and non-competitive biased fission. In competitive biased fission, molecular types which are connected by a strong β_{ij} rate enhancement value have a higher propensity to be in the same progeny. This means that fission is governed by a process analogue to how assembly grows out of the environment. The assembly is treated as the (non-buffered) environment and a progeny is grown out of the parent according to a modified version of Equation 2, were at each step a molecule is picked out of the parent and placed into a progeny, until the parent diminishes and the size of each progeny reaches N_{min}. Thus, the two progenies simultaneously grow and compete on the same set of limited resources, i.e. the parent. In contrast, non-competitive biased fission describes a case whereby only one progeny is grown out of the parent as best as it can, where the second progeny is left with the 'leftovers'. These are in contrast to random fission, in which a progeny is created by selecting, one by one, molecules from the parent and placing them in one of the progeny. The chance to select a molecule of type i is proportional to its current count in the parent assembly, and this is continued until the size of the progeny is N_{min}.

5.2.1. Chemistry biased fission algorithms

The algorithm for chemistry-biased-fission is:

1. Select a random molecule from the parent assembly for each progeny (labeled childA & childB).
2. Continue until the parent assembly has 0 molecules, perform for childA:
    2.1. Select 1 molecule from the parent assembly by using Equation 2 with k_b=0. ρ_i is the current concentration of molecular type i in the parent assembly, n_j is the current count of molecular type j in childA and N is the current size of childA.
    2.2. Update the parent and childA according to 2.1 above (decrease and increase by 1 the relevant molecular type from the parent and in childA, respectively).
    2.3. Perform 2.1-2.2 above for childB.

The algorithm for non-competative-biased-fission is as chemistry-biased-fission, with the following changes:

2. Continue until the size of childA assembly reaches N_{min}.
    2.3. This operation is canceled.
3. ChildB receive the remaining composition of the parent.

5.2.2. Analyzing GARD's fission behavior

The effect of different fission actions on the similarity autocorrelation (Equation 7 in [80] and chapter 5.1) was studied. The similarity autocorrelation is akin to a Fourier transform of the compositional similarity time series. Its derived parameters, τ and H_0, were employed to obtain information on the evolution like dynamics of GARD assemblies. τ depicts whole-simulation average of compositional lifetime, where longer τ represents better average maintenance of compositional similarity between consecutive generations, hence $1/\tau$ is related to the compositional mutation rate. H_0 depicts the residual compositional similarity among assemblies along many generations in the entire simulation, thus a lower H_0 represents a higher overall compositional diversity.

Figure 13 presents τ and H_0 distributions for the three fission actions. H_0 peaks around 1.0 with competitive biased fission and is narrower than random fission, while the second peak is larger and shifted to smaller values. With non-competitive biased fission (while always choosing the preferred progeny) the first peak is almost two times higher than competitive biased fission, and its second peak is about one third than that of competitive biased fission. This suggests that competitive biased fission increases assembly diversity, compared both to random fission and non-competitive biased fission. For the latter it is interesting to note that despite having an overall lower probability for a high diversity, it does allow in general for slightly higher diversity than random fission. τ distribution is practically identical with random fission and competitive biased fission, with essentially the same power-law tail (Table 2). With non-competitive biased fission, τ distribution peaks at smaller values than competitive biased fission and has a power-law factor of about three fourths of competitive biased fission. Competitive biased fission τ distribution is very similar to that of random fission. Non-competitive biased fission distribution is different than competitive biased fission, peaking at slightly lower τ and is below competitive biased fission until τ~50, after which non- competitive biased fission probability is always higher than competitive biased fission and random fission.

Thus, removing the competition during fission results in a lower number of composomes, some of which can reproduce more faithfully than if competition existed, yet including competition during fission significantly contributes to increasing diversity.

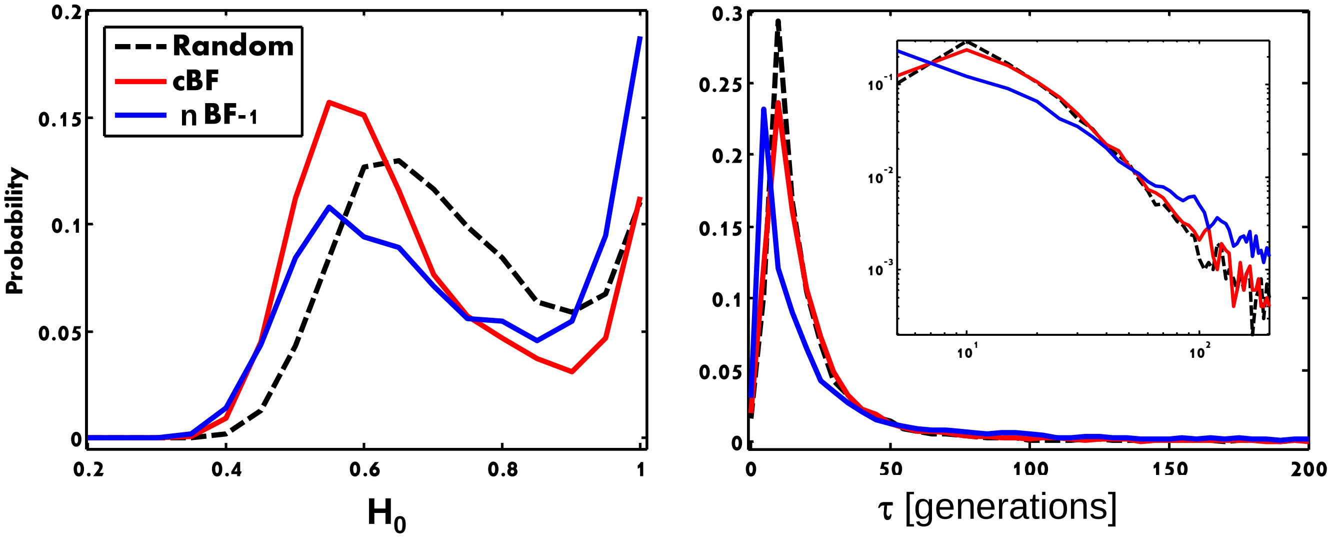

Figure 13: Probability distributions of H_0 (left) and τ (right) for the three fission actions. Abbreviations are: random fission, RF; competitive biased fission, cBF; non-competitive biased fission, nBF-1. Right panel insert shows data on a log-log scale.

Fission	A	B	R^2
Random	144	2.435	0.96
cBF	96.1	2.306	0.96
nBF-1	9.51	1.683	0.92

Table 2: Parameters of fitting the data of Figure 13 right panel to: $P = A\tau^{-B}$, for the three fissions actions. Fit interval is 10<τ<250.

5.3.A physiochemical realistic GARD model

This work is done with collaboration with Prof. Raphael Zidovetzki and Dr. Don Armstrong from University of California at Riverside, while Rafi visited the Weizmann Institute of Science.

Understanding how vesicles replication rates depends on physicochemical parameters can open a new direction in the study of prebiotic vesicles and lay the groundwork for experimentally studying the lipid world. To this end, a new model, real-GARD (R-GARD), was developed [6]. In the new model, replication and thus evolution of lipid vesicles is based on semi-empirical foundation using experimentally measured kinetic values of selected extant lipid types, comprising present-day animal cell. The concept of R-GARD draws from the regular-GARD as it models molecular accretion rates.

Four lipid families were considered: phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine and sphingomyelin, and cholesterol (respectively PC, PE, PS, SM and CHOL, see Figure 14). The physiochemical properties considered for each lipid family were seven carbon chain lengths (12-24 carbons) with five possible degrees of unsaturation (0-4 double bonds). This gives a total environmental repertoire size of N_G=141 (=4×7×5+1) molecular types. R-GARD rate equation (Equation 11) is different than GARD's equation (Equation 2), as in the former the rate also depends on the average properties of the vesicle and in the latter it is based on interactions between individual molecules.

$$\frac{d[C_{vi}]}{dt} = k_{fi} K_f^{adj} [C_{mi}] S - k_{bi} K_b^{adj} [C_{vi}]$$

Equation 11

C_{vi} and C_{mi} respectively are the concentration of molecule type i in the current vesicle and environment, k_{fi} and k_{bi} respectively are the forward and backward rate constant of molecule type i, S is vesicle surface area, and K_f^{adj} and K_b^{adj} are the respective forward and backward functions that are a function of vesicle physical properties. Composomes are found to emerge in R-GARD, similarly to GARD (Figure 15). This supports the possibility of experimentally observing faithful replication of lipid vesicles. The fact that composomes appear weaker in R-GARD than in GARD is attributed to the usage of realistic molecules, suggesting that in reality differences in mutual catalysis values are expected to be distributed more uniformly than in a lognormal distribution as usually employed in GARD.

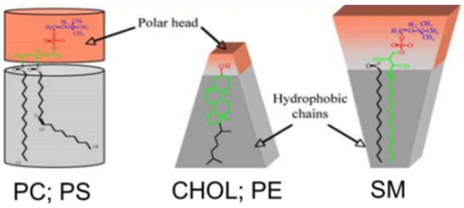

Figure 14: Schematics of the lipids structure used in R-GARD.

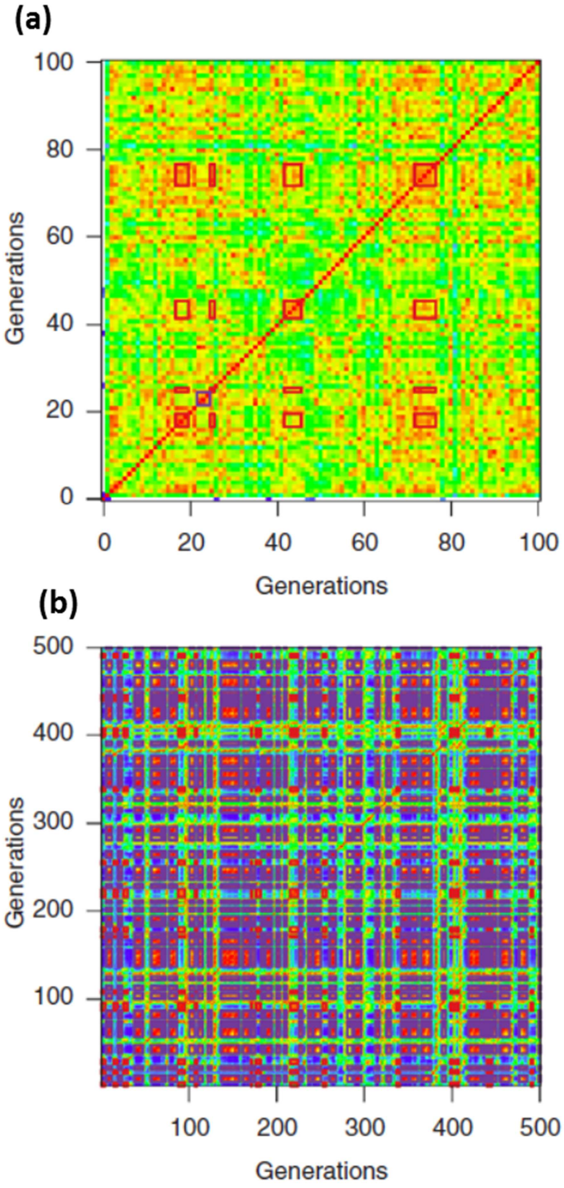

Figure 15: (a) R-GARD ‘carpet’ [6]. Composomes are marked by square boxes on the diagonal; similar composomes are colored with the same color. As a similarity measure, the Euclidean distance between the property-vector is used (as opposed to the composition vector, Equation 1). (b) GARD ‘carpet’ with the standard similarity measure between the composition vectors (H, Equation 3). Color code for both panels is: Red marks most similar (values of 0 in R-GARD and 1 in GARD), blue marks least similar (values of 3 in R-GARD and 0 in GARD) and white marks being outside of the range [0, 3] in R-GARD. Both simulations were run with N_G=20 and N_{max}=100.

It is further found that vesicle replication rate is largely influenced by variations in the chain length, unsaturation and relative environmental concentrations of molecular types (Figure 16), and as expected, the initial vesicle composition does not affect the final vesicle composition. The decrease of replication time with chain length or unsaturation is a consequence of the defects created in the plane of the bilayer by the mismatch in these characteristics, as expressed in its rate equations. These defects allow for molecules to be accreted into the vesicle. Interestingly, the effect of lipid variation on vesicle's replication rate can be considered as a further support to the previous finding that the higher N_{mol} of a compotype is the faster it replicates (chapter 5.4.2), as a higher N_{mol} represents a higher molecular variation. A correlation of environmental concentration of lipid types with average replication time is best seen in the case of PC, a consequence of the larger surface area of PC which increases vesicle surface size for a given number of lipids in the vesicle coupled with the effect of additional PC concentration reducing the concentration of the other species.

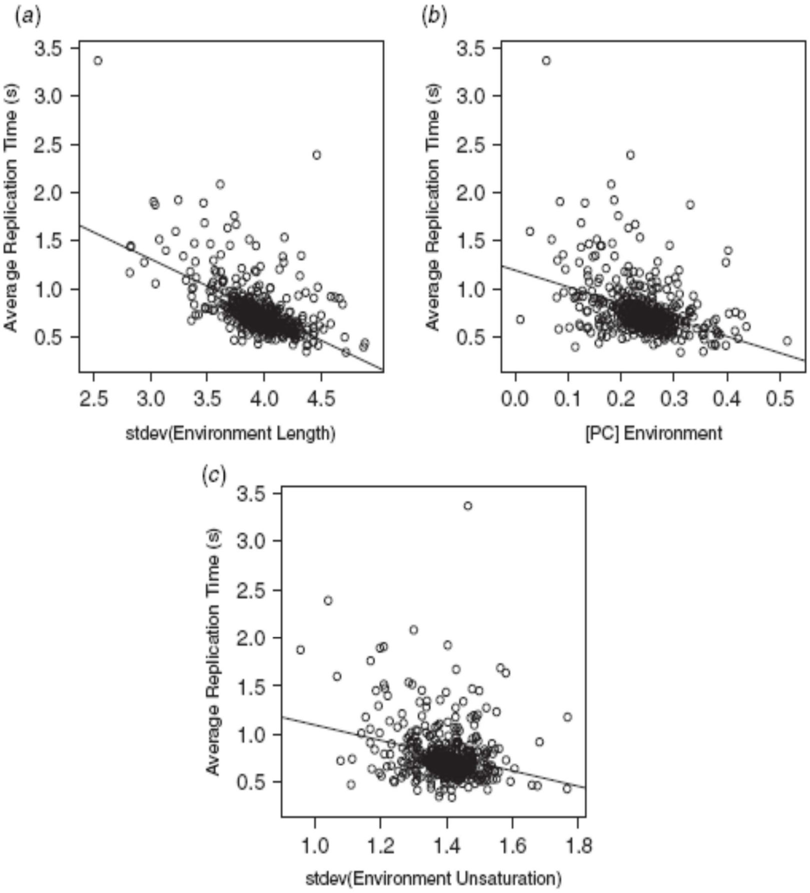

Figure 16: Dependence of average replication time (ART) on various properties and lipid species [6]. Circles represent simulations with starting vesicle composition with binomial distribution and environmental composition of gamma distribution with sizes from 10^2 to 10^5 molecules and 20 to 141 species. (a) ART vs. the standard deviation of the length of species. (b) ART vs. the standard deviation of the unsaturation of species. (c) ART vs. the environmental molar concentration of PC. Linear correlations for the

three panels produce: (a) R=-0.61, slope=-0.56; (b) R=-0.39, slope=-1.73; (c) R=-0.28, slope=-0.80.

An ongoing investigation with this model, studies populations under a limited environment, in contrast to the simulations performed in chapter 5.4. In this investigation, the system is seeded with a single vesicle and the reactor contains enough molecules to form 4096 vesicles (i.e. 12 generations). The rest of parameters are identical to those employed [6]. Interestingly, it is found that vesicles of newer generations show a bigger difference between growth-time of same generations (Figure 17B). This suggests an evolution of vesicles towards an optimal composition which enables for faster growth, i.e. higher fitness. An overall increase in the growth-time is due to the decreasing environmental concentration of molecules (Figure 17A). This phenomenon is currently under study.

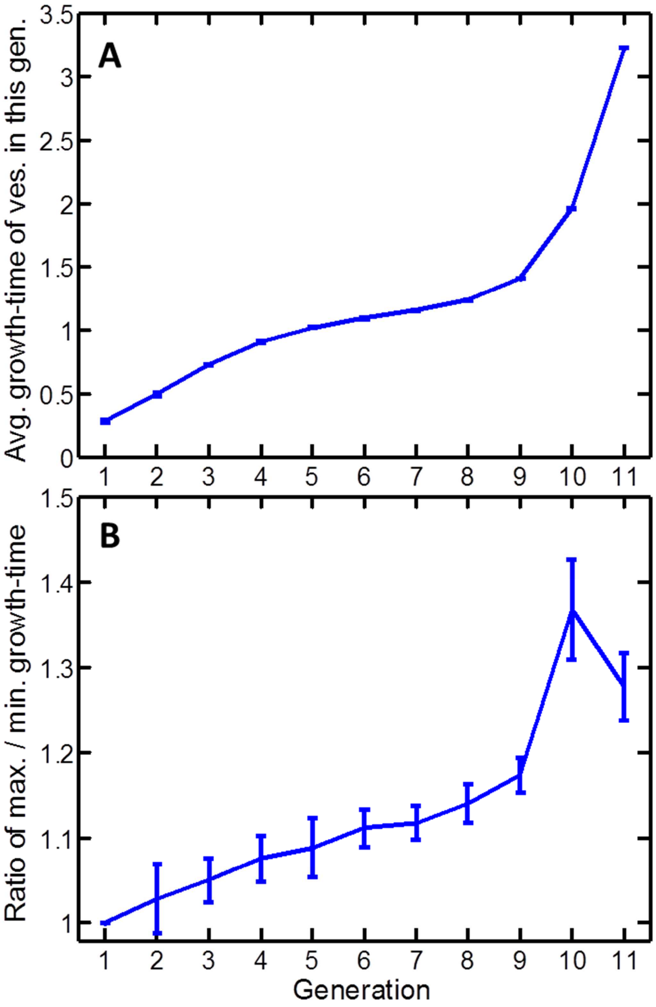

Figure 17: Vesicles growth time vs. vesicle generation, under a limited environmental concentration of molecules. (A) Average growth-time of all vesicles that belong to each generation. (B) The ratio of the fastest to slowest growing vesicles in each generation. The first vesicle that was seeded in the system is generation=1. This figure is based on 20 simulations.

5.4.Ecological dynamics of GARD population

This chapter describes work which has been accepted for publication at the Journal of Theoretical Biology.

Thus far, GARD was typically studied by focusing on individual assemblies. The present paper studies GARD population in a rigorous way by using standard ecological tools. When simulating populations of GARD assemblies, compotypes exhibit competition similarly to that seen for natural species, and it analyzable by a common multi species logistic formulation. With that, it becomes possible to relate compotypes' chemical parameters to population ecological behaviors, and to predict GARD's behavior based on compotypes' structure. Further, GARD's assemblies until now have been treated as information carriers, positing the lipid world as an alternative to the RNA world. This paper adds another layer to GARD, by considering β as a rudimentary metabolic network and a compotype as a rudimentary organism.

Prebiotic models have often focused on evolution in populations of self-replicating molecules, without explicitly invoking the intermediate molecular-to-supramolecular-to-ecology transition. Of note, a similar transition has been studied in a model of RNA-like replicators, in which supramolecular entities (traveling waves) were found to play a role in the ecology and evolution of replicators [136]. Present life portrays a two-tier phenomenology: molecules compose self-replicating supramolecular structures such as cells or organisms, which in turn portray population behavior, including selection, evolution and ecological dynamics. Thus, understanding how molecular mixtures gave rise to evolving entities which in turn gave rise to simple ecological niches will greatly contribute to our understanding of the origin of life and to a degree akin to the on-going pursuit to understand and predict the dynamics of ecological populations from the, often complex, metabolic or genetic networks of the underlying species [7].

An admitted shortcoming of GARD is the paucity of experimental verification of many of its predictions. A proof-of-principle experiment should address the question of whether vesicles are capable of homeostatic growth and even rudimentary transfer of compositional information to fission-generated progeny. Such experiments would require complex setups, accurate compositional monitoring of individual amphiphile assemblies, not yet fully elaborated. A promising lead would be the recent experimental exploration of multi-component vesicles [82, 144]. Another critique of GARD asserts that it simulates abstract molecules without specified chemical properties. This point has been recently addressed in an extension of the simulated model to incorporate realistic physicochemical properties of amphiphilic molecules, showing that a measure of compositional heredity may be observed ([6] and chapter 5.3).

GARD simulations are used in this chapter to quantitatively follow population-ecological dynamics of composmal species. In the foregoing analyses a multivariate logistic equation is used to relate systems chemical parameters of GARD assemblies, including chemical diversity, replication fidelity and compositional similarity to specific ecological measures such as the carrying capacity, the intrinsic growth rate and the competition parameters.

5.4.1. GARD population exhibit natural-like dynamics

Different simulations with different underlying β networks result in widely different dynamic behaviors, such as delayed growth, different plateaus and "takeover" of a fast-rising compotype by a slower one (Figure 18). Such dynamics are typical of natural ecosystems that harbor multiple species with competition or predator-prey relationships. The resulted dynamics are analyzed by a multi species logistic model for population ecology (r-K or Lotka-Volterra competition model, Equation 7) [34, 133, 141].

That equation has a steady state $C_i^{ss}=K_i-\sum\alpha_{ij}C_j^{ss}$ and can be solved analytically only for the case of a single species ($N_C=1$). For each simulation, the logistic parameters for all N_C compotypes are obtained by least square fitting and numerical integration, as detailed in chapter 4.5. Notably, an adequate fit to such equation was observe for practically all GARD simulations performed, with average root mean square difference=0.019±0.011 for the entire set of 1,000 simulations performed. In contrast, several other models with similar overall characteristics gave an inferior fit (chapter 5.4.5). Next, analyses are performed, aimed at relating the chemistry-base molecular parameters of GARD to the ecology-related parameters of the logistic equation.

5.4.2. Compotype intrinsic molecular repertoire

Each GARD compotype contains a subset (N_{mol}, chapter 4.2) of the total N_G molecular types present in the environment. Such repertoire restriction emerges as a result of the intermolecular catalytic interactions in β and in the present simulations an average of $N_{mol}=16\pm5$ is observed (Figure 19). The effect of this chemical diversity parameter, N_{mol}, on the K_i and r_i values of individual compotypes is examined (Figure 20). It is found that K values are inversely correlated with N_{mol}, and in contrast, r values show a weak positive correlation. Thus, compotypes with a large N_{mol} will tend to have a larger growth rate and a smaller carrying capacity. For cases with negligible competition parameters this will amount to a steeper ascent and a relatively low plateau in cases of large N_{mol}.

The dependence of K and r on N_{mol} may be explained considering the random nature of the processes involved and the fact that external concentrations of all molecular types are equal. A higher value of N$_{mol}$ increases the probability that a randomly impinging molecule will be part of the compotype's intrinsic molecule repertoire, enhancing homeostatic growth rate.

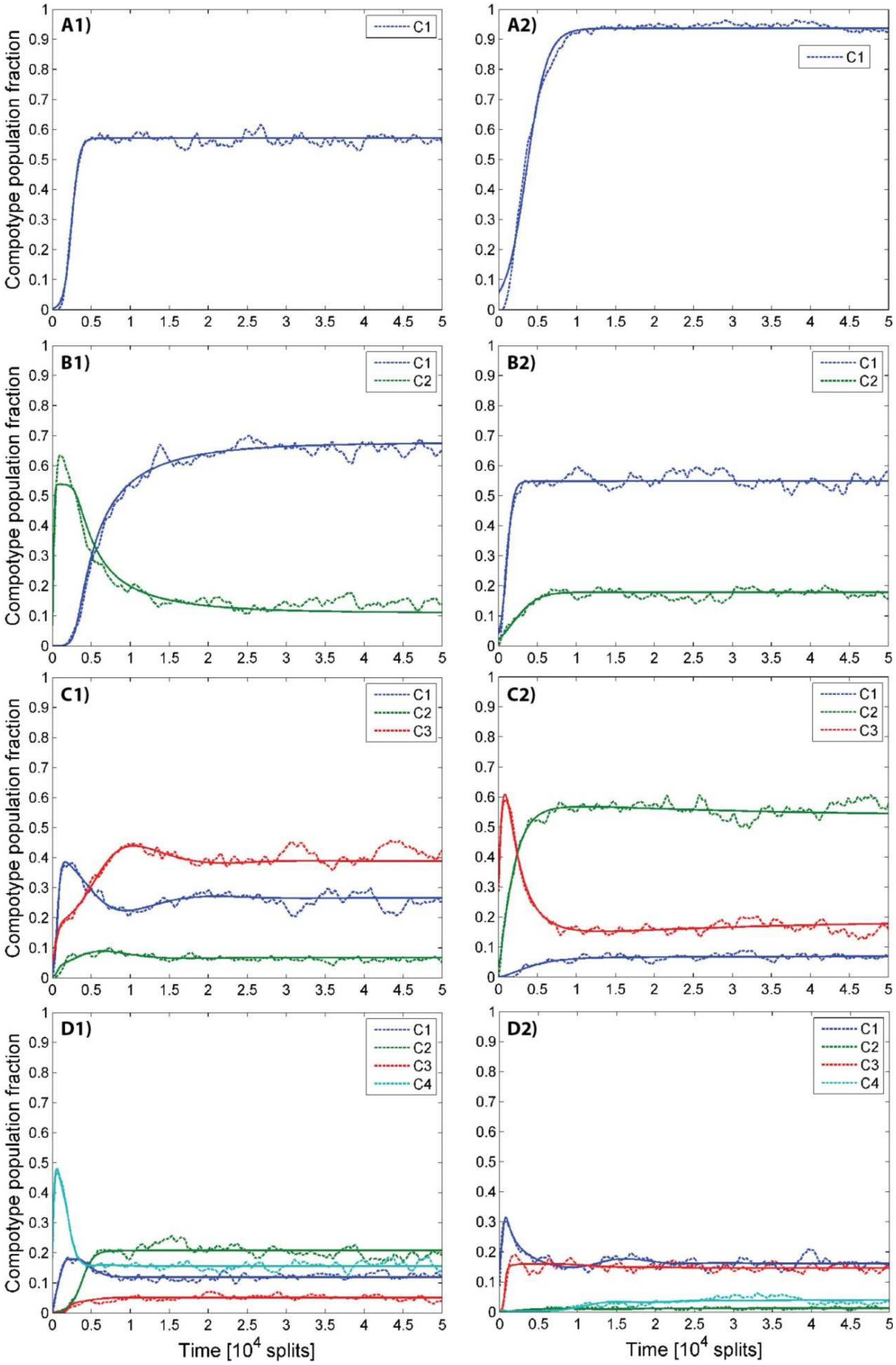

Figure 18: Examples of GARD simulations and fit to logistic growth. Simulation data is broken line and fit is solid line. Fitted parameters are collected in Table 3.

Panel	β seed		K	r [splits^{-1}]	α_{ij}				C(0)
					j=1	j=2	j=3	j=4	
A1	18	i=1	0.572	2.06E-03					3.58E-03
A2	30	i=1	0.937	7.70E-04					5.68E-02
B1	170	i=1	0.843	8.48E-02		1.518			7.86E-11
		i=2	0.538	1.28E-02	0.633				6.89E-02
B2	261	i=1	0.548	2.40E-03		0.000			4.15E-02
		i=2	0.273	1.00E-03	0.170				2.25E-02
C1	45	i=1	0.555	4.90E-03		1.487	0.486		1.94E-02
		i=2	0.494	4.30E-03	0.788		0.557		3.10E-03
		i=3	0.724	2.00E-03	1.260	0.000			5.46E-02
C2	7	i=1	0.462	4.80E-03		0.601	0.359		1.00E-03
		i=2	0.734	5.30E-03	1.018		0.663		3.50E-02
		i=3	0.828	3.60E-03	0.000	1.919			3.13E-01
D1	149	i=1	0.348	7.10E-03		0.274	2.094	0.414	1.57E-02
		i=2	0.448	3.20E-03	0.585		1.914	0.456	1.40E-03
		i=3	0.113	4.51E-02	0.281	0.038		0.130	1.59E-11
		i=4	0.581	6.20E-03	1.292	0.000	5.277		2.15E-01
D2	133	i=1	0.518	2.40E-03		7.198	1.814	0.000	1.46E-01
		i=2	0.342	1.06E-02	0.002		2.034	0.791	1.91E-06
		i=3	0.187	8.50E-03	0.101	0.885		0.334	1.20E-03
		i=4	0.341	2.70E-03	1.266	4.141	0.304		4.10E-03

Table 3: Fitted parameters of the simulations given in Figure 18.

Conversely, low N_{mol} means that on average every molecular type exists inside the compotype in higher counts, so when split occurs there is a better chance that a progeny will contain the same composition as the parent.

One may ask, whether there could be a parallelism for any of these results in present day life. An interesting analogous trend was observed in experimental data for 113 bacteria, whereby a negative correlation was seen between measured doubling time and metabolic network size [32]. However, direct comparison between compotype dynamics and present-day metabolism might not be possible, as the latter is controlled by a genome, centralized informational entity acting via a complex hierarchy of interactions [46, 116].

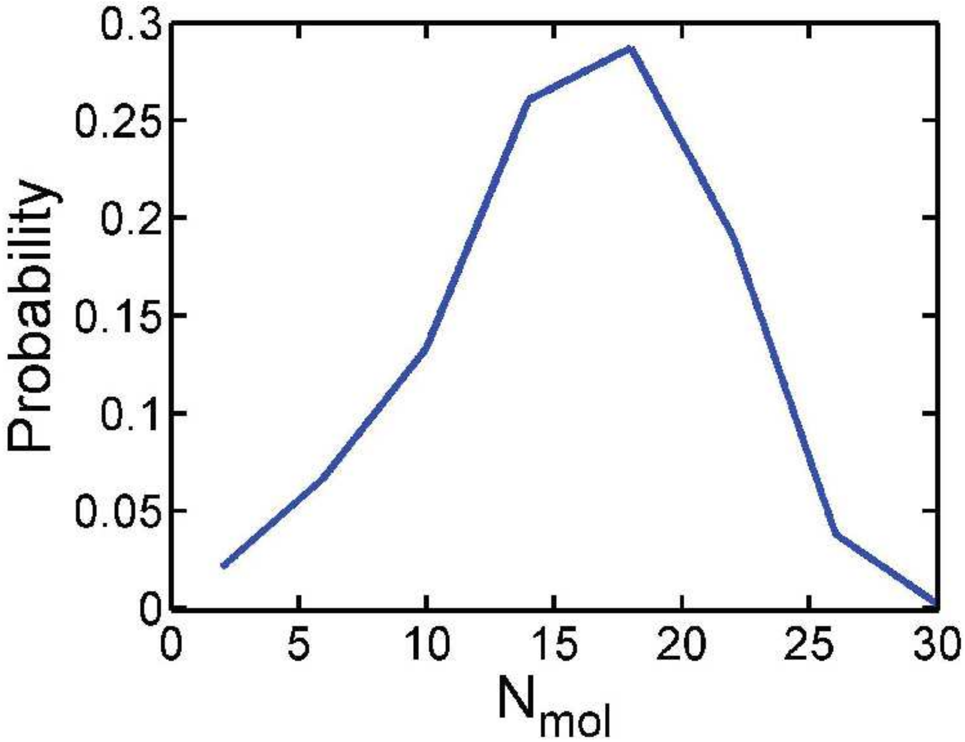

Figure 19: Distribution of intrinsic molecular repertoire size (N_{mol}) values for all compotypes.

5.4.3. Compotype replication fidelity

The relation between K and the replication fidelity (F_{rep}, chapter 4.2) is analyzed in this section. F_{rep} measures the average degree of compositional similarity between an assembly representing a compotype and its progeny, both in fully-grown state. K, the carrying capacity, represents the maximal number of individuals of a given species that may be sustained in an ecological niche. In the original Verhulst formalism, death was introduced to counter the Malthusian exponential growth. Later, the r-K logistic formalism defined K=birth/death [34]. In GARD, a positive correlation between K and F_{rep} is observed (Figure 21). Unfaithful replication (low F_{rep}) means that the progeny has lost its compotype state, either to another compotype species or to drift, somewhat comparable to death of the species in question. This may rationalize the somewhat unexpected positive correlation between an emergent molecular parameter such as F_{rep} and an ecological one – the carrying capacity. Other relationships explored, between r and F_{rep} and between K and r and the replication-time (t_{rep}) showed no appreciable correlations (Figure 22 and Figure 23).

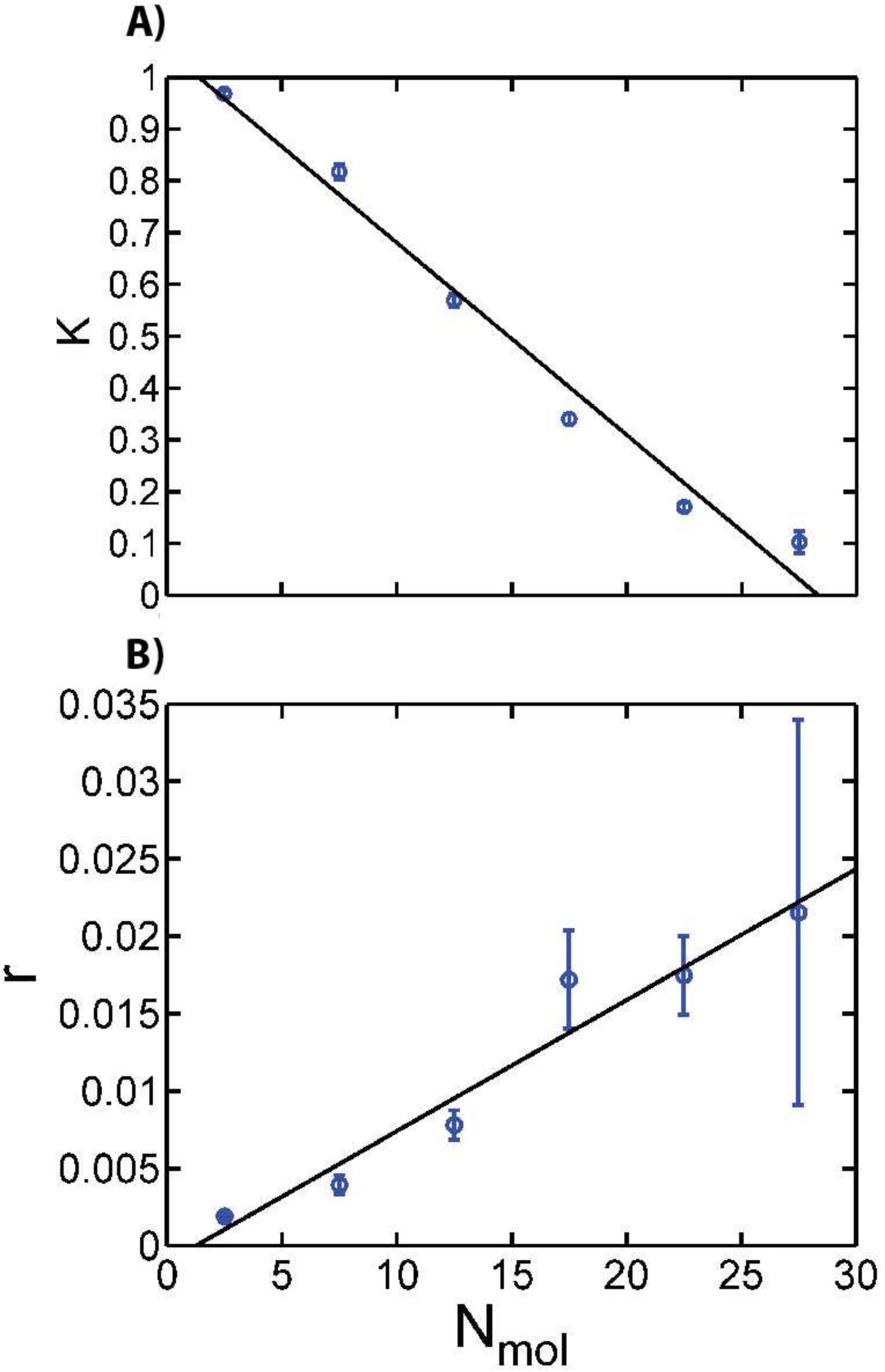

Figure 20: The dependence on N_{mol}. Data are binned according to N_{mol} values and vertical lines represent standard error of the mean. Black solid line is a linear fit. (A) K vs. N_{mol}. Linear fit: $K=–0.0371*N_{mol}+1.052$, $R^2=0.978$; (B) r vs. N_{mol}. Linear fit: $r=8.46*10^{-4}*N_{mol}–1.06*10-3$, $R^2=0.946$.

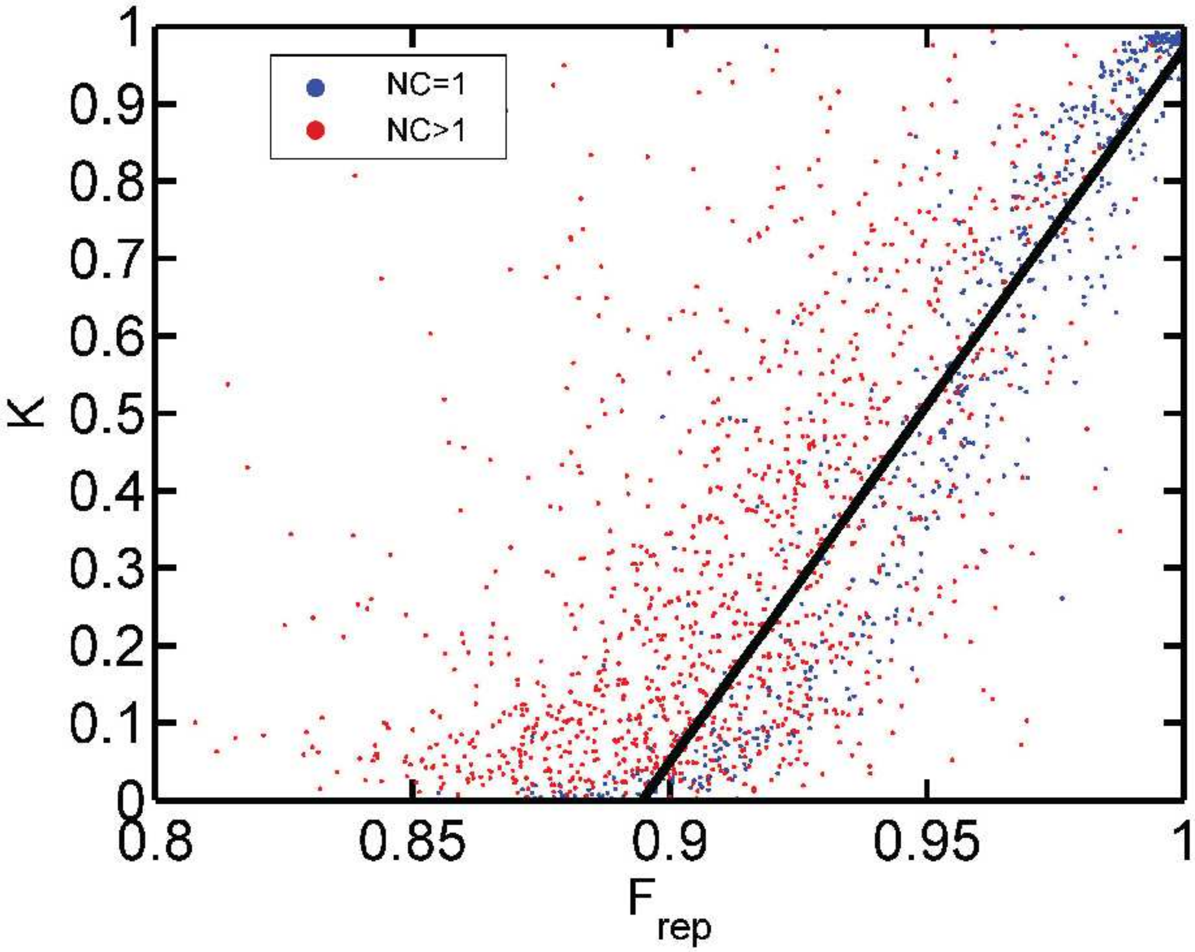

Figure 21: K vs. F_{rep}. Blue and red dots are compotypes taken from simulations exhibiting N_C=1 and N_C>1, respectively. Black solid line is linear fit to N_C=1 data: $K=9.23*F_{rep}-8.23$, $R^2=0.876$.

5.4.4. Takeover

In this section, the molecular mechanism behind the ecological takeover is addressed. This phenomenon is exemplified in panel B1 of Figure 18, by the observation that compotype C_2 shows a much faster ascent, reaching a 538 fold excess over C_1 at time 990. Subsequently, C_1 increases substantially, becoming 5.82 fold more abundant than C_2 at steady state. This was examined by analyzing an extended set of 316 β networks that exhibit N_C=2. Two parameters, MP and PP, were defined to quantify takeover behavior: MP=Max(C_{low})/Plateau(C_{low}), PP=Plateau(C_{high})/Plateau(C_{low}), where C_{low} is the compotype with the lower plateau (Figure 24). Two subgroups were examined: one showing clear takeover, with MP>2 and PP>5, and another in which no takeover occurs, with MP<1.5 & PP<4 (control). The inter-compotype compositional similarity for pairs that exhibit takeover is found to be significantly lower than for control pairs (Figure 25). These two behavior types are also seen to be partially segregated in a principle component analysis of the 6 fitted logistic parameters (r_1, r_2, K_1, K_2, α_{12} and α_{21}) (Figure 26 and Figure 27). Intriguingly, the majority of the variance in this plot is contributed by the competition parameter α_{21} (Figure 27). Work is underway to study the molecular mechanism that governs across-compotype competition.

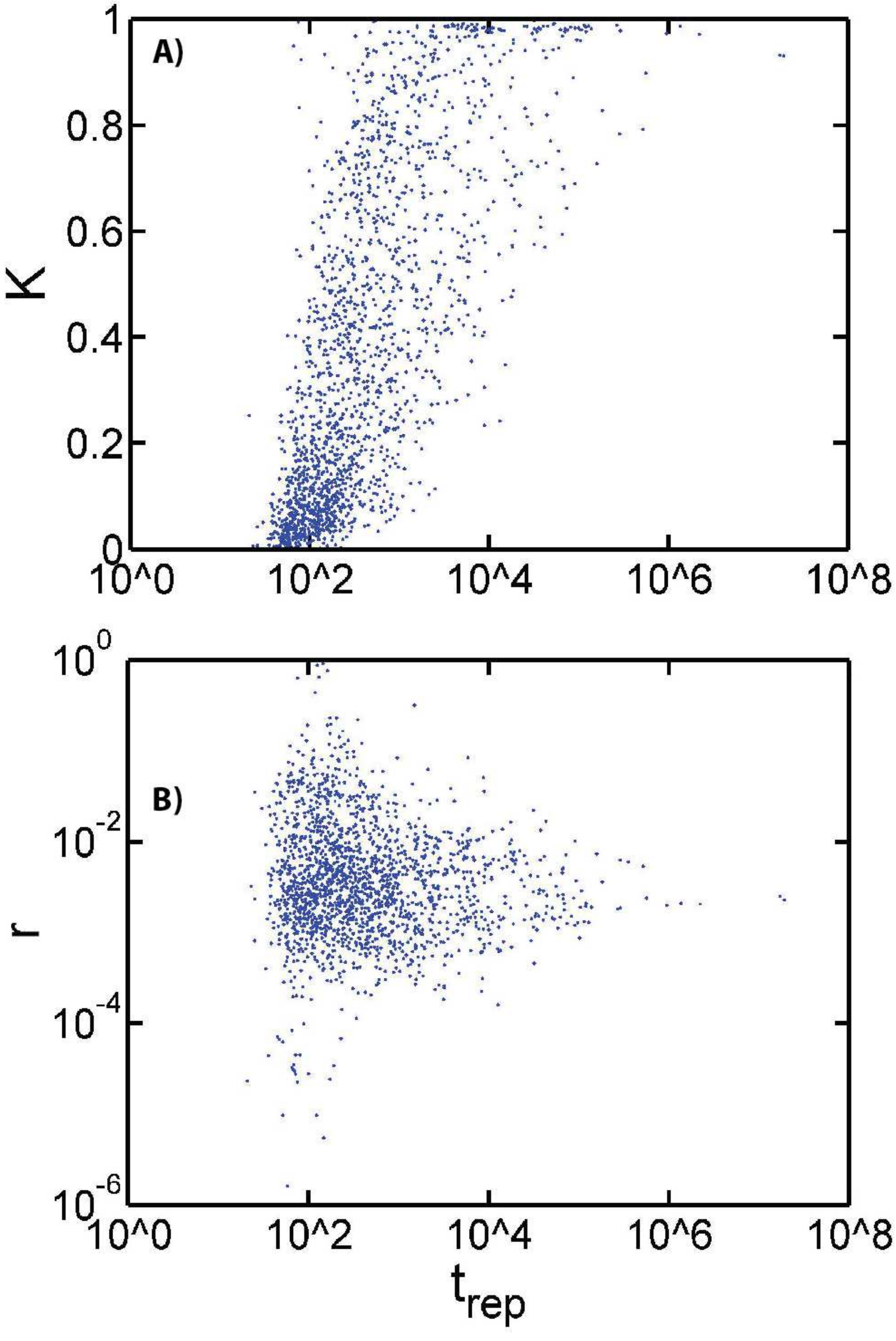

Figure 22: K and r vs. t_{rep} of all compotypes (A and B, respectively). Linear fit: K vs. log10(t_{rep}) gives R^2=0.48; log10(r) vs. log10(t_{rep}) gives R^2=0.007.

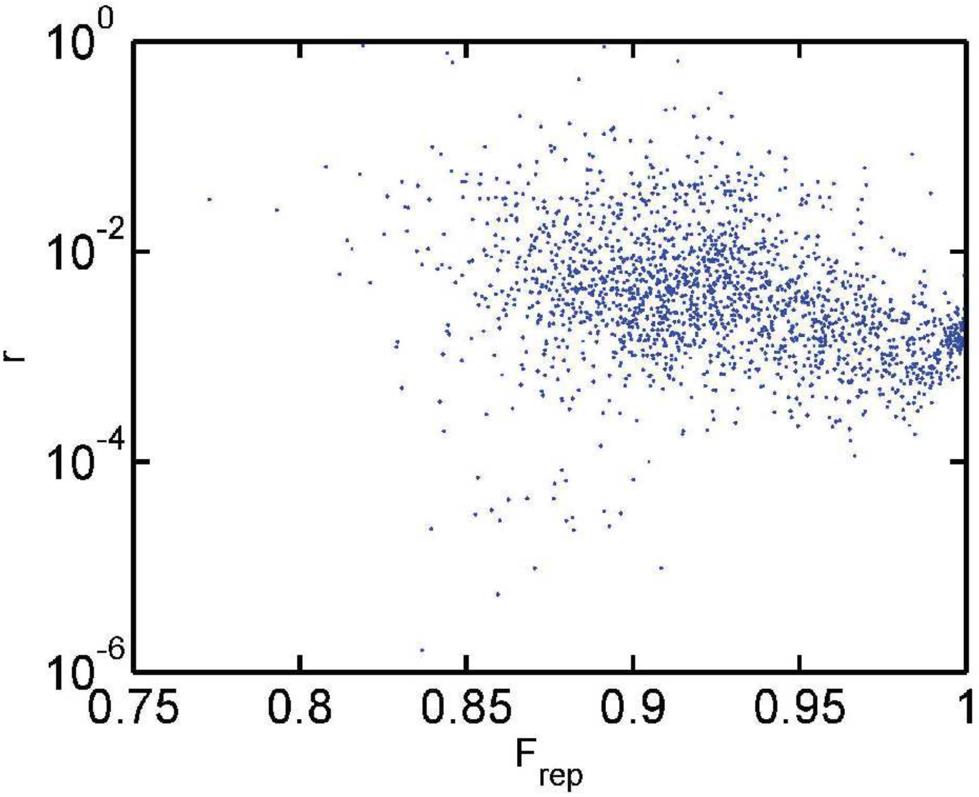

Figure 23: r vs. F_{rep} of all compotypes. Linear fit: log10(r) vs. F_{rep} gives R^2=0.10.

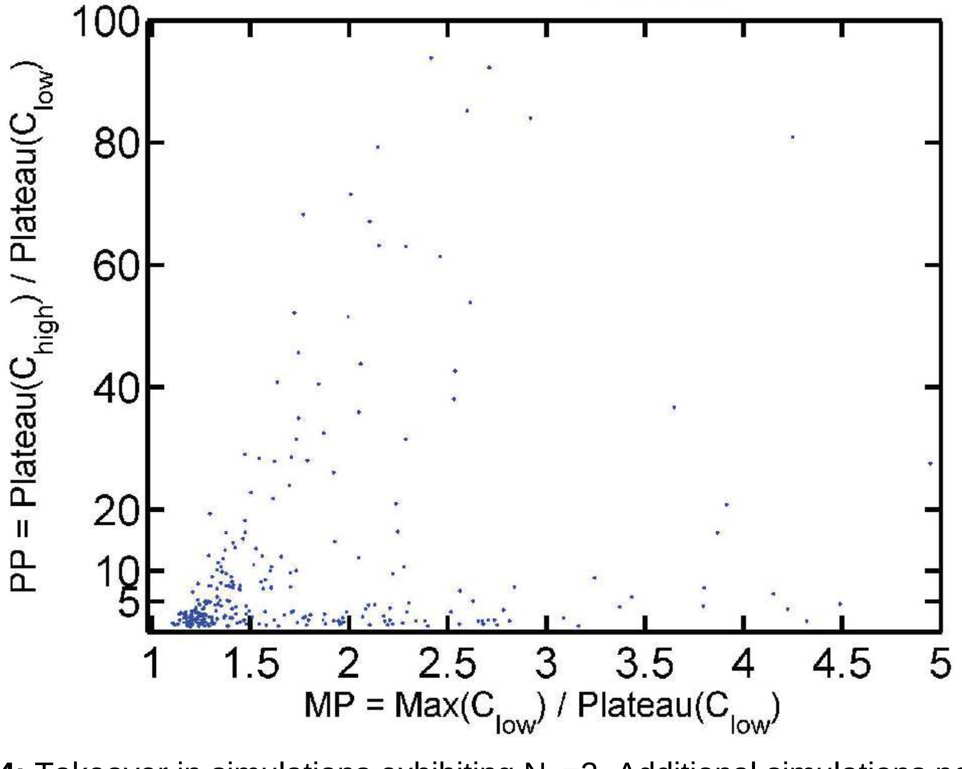

Figure 24: Takeover in simulations exhibiting N_C=2. Additional simulations performed for this part, giving a total of 316 such simulations. Takeover is represented by MP>2 & PP>5 and control is MP<1.5 & PP<4 (each group consists of 40 and 87 simulations, respectively). Max(C) is the value at the highest point and Plateau(C) is the average value along the plateau.

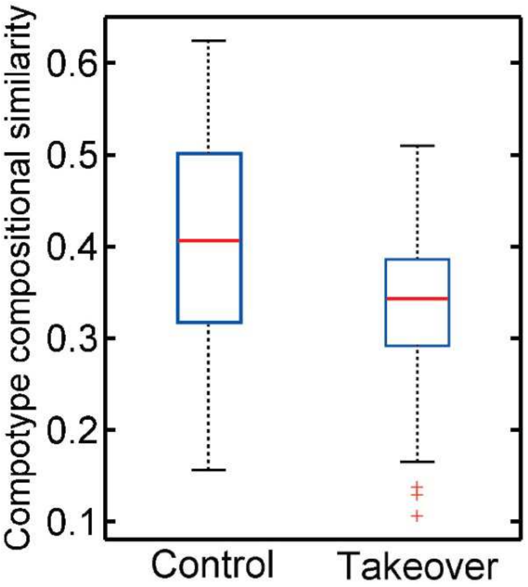

Figure 25: Compotype similarity in the takeover and control groups. Average cross-compotype similarity (H) for the takeover group is lower than for the control group (0.33±0.09 vs. 0.40±0.12; supported by a two-sample Kolmogorov-Smirnov test with p-value=1.3*10^{-3}).

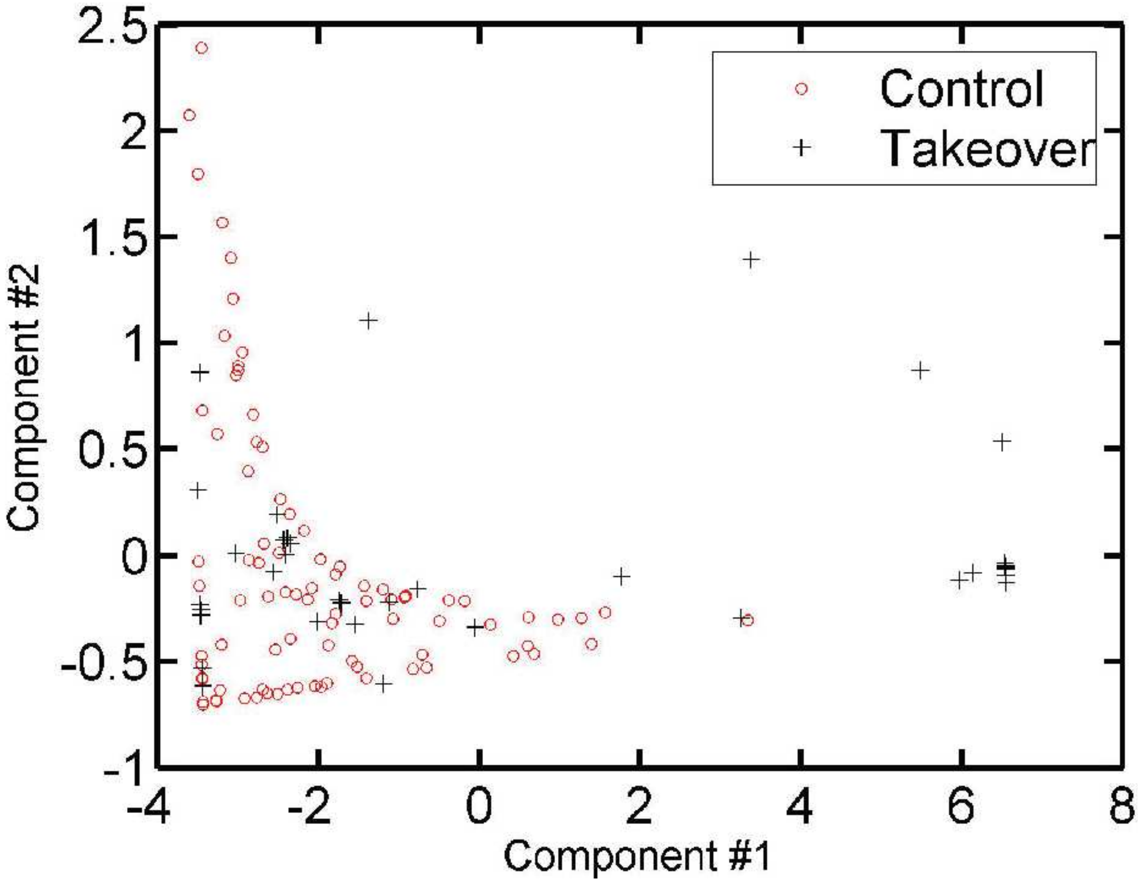

Figure 26: Principle component analysis (PCA) of the 6 fitted parameters (r_1, r_2, K_1, K_2, α_{12} and α_{21}), performed using MATLAB *princomp* routine. The first two components are responsible for 96% and 3% of the variance in the data, respectively.

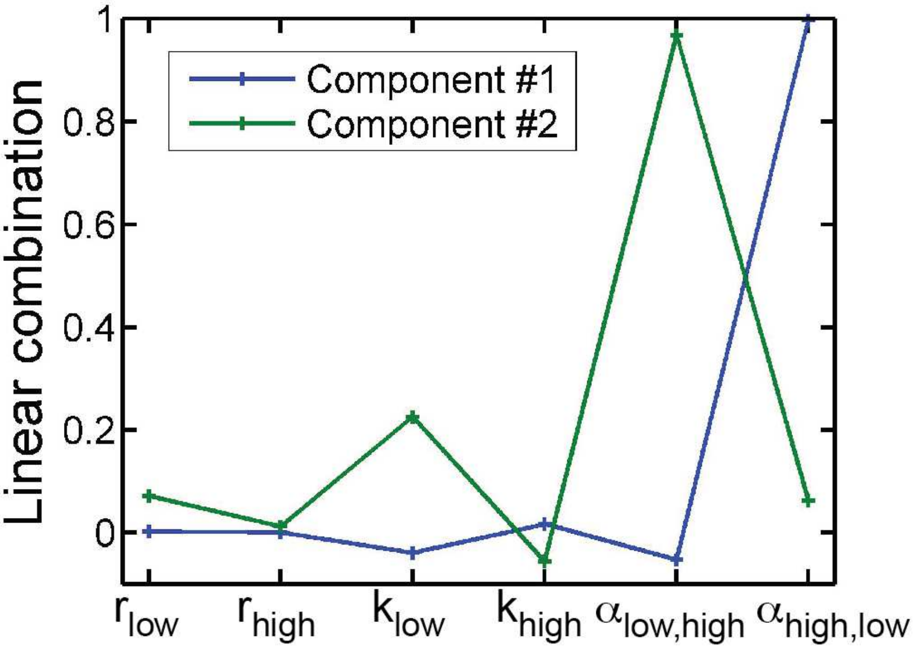

Figure 27: Linear combination of the first two components found by the PCA. A two sample Kolmogov-Smirlov test found that the takeover group have higher $\alpha_{high,low}$ values than the takeover (p-value=$3.2*10^{-4}$).

5.4.5. On the choice of specific logistic formulation

It was found above that the fit between the standard multi species logistic equation to GARD's population dynamics data is very good. However, it is unclear if this is due to a profound similarity between the GARD model dynamics and real-world ecology. It may very well be that the good fit is due to the fact that this equation has many parameters, as even von Neumann is said to say: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk"[3] [28]. To address this, several other multi species logistic formulations were tested, asking whether they can similarly fit the data.

Any equation tested must include cross species coupling aswell (i.e. the existence of a C_j term in dC_i/dt), because when multiple compotypes exist in a population they sometime exhibit a coupled behavior such as takeover. The subset of 141 simulations with a compotype count N_C=2 was used, being the simplest multi-species case.

Table 4 shows the additional equations tested here, together with the original equation (marked V0). V5a is taken from [68] and V5b is a variation of it. V2a and V4a are manually generated based on V0. These four equations are fitted with identical procedure to that of which the original logistic equation was fitted (chapter 4.5), while they are similarly written in the original form (form I in Table 4). The preferential status of the specific logistic formulation used in this

[3] Demonstrated in: http://www.johndcook.com/blog/2011/06/21/how-to-fit-an-elephant. The 5th parameter it time.

thesis is apparent when the average root-mean-square-deviations (RMSD, Equation 8) from the fit to V0 is smaller than the in the fit to the other 4 variations (Figure 28).

Additional equations were also tested. When V0 is written in a polynomial-like form (form II in Table 4), it is possible to systematically vary it while keeping a power of 2 present (e.g. C_i^2 or C_iC_j). This gives variations V3a-V3e. These were fitted in this polynomial-like form, with exactly the same procedure as before, yet the fitting routine was unable to converge in these 5 cases. No convergence was obtained also when looser convergence criteria were used. As this is potentially alarming, the fitting code functionality was verified, by comparing the ecological parameters obtained for V0 when it is fitted in both form I and II. As expected, this gave identical parameters values in both forms (for example: $Z_1=\alpha_{12}r_1/K_1$). The origin of this issue requires further investigation[4].

Overall, these results support the usage of the original logistic formulation, which is commonly used to describe natural phenomena, thus contributing to recognizing GARD as a model of a natural phenomenon.

Table 4 (below): Additional differential equations tested against the population data, written for N_C=2. V0 is the original equation. **a)** Grey cell represents that this form has been fitted (see text). **b)** Number of free parameters for N_C=2 (in brackets for any N_C value). **c)** Mean RMSD (Equation 8), when fitted against the dataset of 141 simulations with N_C=2.

	Form I^{a}	Form II^{a}	Parameters^{b}	<RMSD>^{c}
V0	$\frac{dC_1}{dt}=r_1C_1\left(\frac{K_1-C_1-\alpha_{12}C_2}{K_1}\right)$ $\frac{dC_2}{dt}=r_2C_2\left(\frac{K_2-C_2-\alpha_{21}C_1}{K_2}\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_1^2-Z_1C_1C_2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_2^2-Z_2C_1C_2$	6 (N_C^2+N_C)	0.0215± 0.0125
V2a	$\frac{dC_1}{dt}=r_1C_1\left(\frac{K_1-C_1+\alpha_{1D}C_D}{K_1}\right)$ $\frac{dC_2}{dt}=r_2C_2\left(\frac{K_2-C_2+\alpha_{2D}C_D}{K_2}\right)$ $C_D=L_{pop}-C_1-C_2$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_1^2-Z_1C_1C_D$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_2^2-Z_2C_1C_D$ $C_D=L_{pop}-C_1-C_2$	6 ($3N_C$)	0.0291± 0.0165

[4] When I asked colleagues in Israel and around the globe should I be surprised when get such a good fit only to one specific variation of the logistic equation, every single one told me that I should not, though none was able to provide further information as to the reasoning behind this…

V4a	$\frac{dC_1}{dt}=r_1C_1\left(\frac{K\ -C_1-\alpha_{12}C_2}{K}\right)$ $\frac{dC_2}{dt}=r_2C_2\left(\frac{K\ -C_2-\alpha_{21}C_1}{K}\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_1^2-Z_1C_1C_2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_2^2-Z_2C_1C_2$	5 (N_C^2+1)	0.0414± 0.0278
V5a	$\frac{dC_1}{dt}=\left\{b_1\left(\frac{C_1}{C_1+\alpha_{12}C_2}\right)-d_1\right\}C_1-h_1C_1^2$ $\frac{dC_2}{dt}=\left\{b_2\left(\frac{C_2}{C_2+\alpha_{21}C_1}\right)-d_2\right\}C_2-h_2C_2^2$		8 (N_C^2+2N_C)	0.0644± 0.0586
V5b	$\frac{dC_1}{dt}=\left\{b_1-d_1\left(\frac{C_1+\alpha_{12}C_2}{C_1}\right)\right\}C_1-h_1C_1^2$ $\frac{dC_2}{dt}=\left\{b_2-d_2\left(\frac{C_2+\alpha_{21}C_1}{C_2}\right)\right\}C_2-h_2C_2^2$		8 (N_C^2+2N_C)	0.0331± 0.235
V3a	$\frac{dC_1}{dt}=C_1\left(X_1-Y_1\frac{C_2}{C_1}-Z_1C_2\right)$ $\frac{dC_2}{dt}=C_2\left(X_2-Y_2\frac{C_1}{C_2}-Z_2C_1\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_2-Z_1C_1C_2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_1-Z_2C_1C_2$	6 (2N_C^2-N_C)	-
V3b	$\frac{dC_1}{dt}=C_1\left(X_1-Y_1\frac{C_2^2}{C_1}-Z_1C_2\right)$ $\frac{dC_2}{dt}=C_2\left(X_2-Y_2\frac{C_1^2}{C_2}-Z_2C_1\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_2^2-Z_1C_1C_2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_1^2-Z_2C_1C_2$	6 (2N_C^2-N_C)	-
V3c	$\frac{dC_1}{dt}=C_1\left(X_1-Y_1C_1-Z_1\frac{C_2^2}{C_1}\right)$ $\frac{dC_2}{dt}=C_2\left(X_2-Y_2C_2-Z_2\frac{C_1^2}{C_2}\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_1^2-Z_1C_2^2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_2^2-Z_2C_1^2$	6 (N_C^2+N_C)	-
V3d	$\frac{dC_1}{dt}=C_1\left(X_1-Y_1C_1-Z_1\frac{C_2}{C_1}\right)$ $\frac{dC_2}{dt}=C_2\left(X_2-Y_2C_2-Z_2\frac{C_1}{C_2}\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_1^2-Z_1C_2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_2^2-Z_2C_1$	6 (N_C^2+N_C)	-
V3e	$\frac{dC_1}{dt}=C_1\left(X_1-Y_1\frac{C_2}{C_1}-Z_1\frac{C_2^2}{C_1}\right)$ $\frac{dC_2}{dt}=C_2\left(X_2-Y_2\frac{C_1}{C_2}-Z_2\frac{C_1^2}{C_2}\right)$	$\frac{dC_1}{dt}=X_1C_1-Y_1C_2-Z_1C_2^2$ $\frac{dC_2}{dt}=X_2C_2-Y_2C_1-Z_2C_1^2$	6 (2N_C^2-N_C)	-

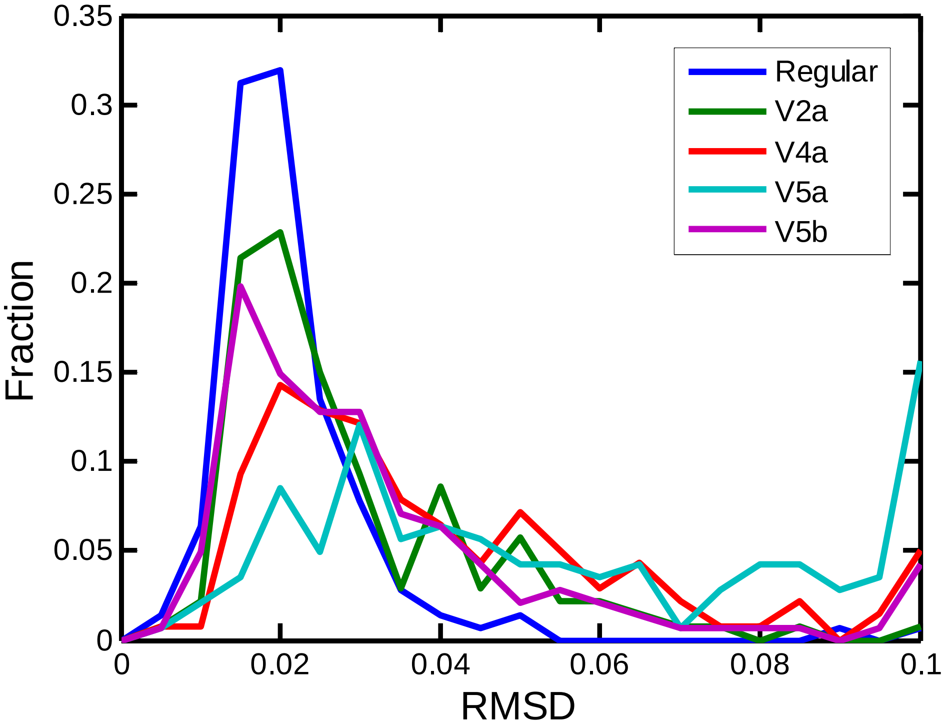

Figure 28: Distribution of RMSD values of 4 variations of the logistic equation and of the regular form (V0, blue color).

5.4.6. Broadening the analysis of selection in GARD

GARD populations were considered in two papers: in the earlier one it was an addition to the main selection and mutual catalysis analyses ([80] and chapter 5.1), and in the later one, where ecology was studied, it was a natural part of the work (the present chapter). The selection behavior focused on only the most frequent compotype and in this section and the following one the selection behavior when focusing on other assemblies and compotypes is addressed.

5.4.6.1. Testing on drift assemblies

When selection was studied, for the first time in GARD [80], only compotypes were considered as they are treated as GARD's species (and in chapter 5.5 their quasispecies nature is revealed). In this section, the selection response of drift assemblies is analyzed and compared to that of compotypes, in order to further understand the nature of selection in GARD. Additionally, it was argued that GARD lacks evolvability by studying the selection response of a particular GARD system and drift assemblies [143]. One of our arguments against those conclusions is that those authors did not designate a compotype as the selection target, and studying the selection of drift (non-compotype) assemblies resolves this point.

Drift assemblies selection is studied similarly to the selection of compotypes, whereby the change in abundance of a given target is considered a mimic to selection (chapter 4.3). In each simulation under a given β, the drift assembly which is the least similar to all of the N_C

compotypes centers of mass is picked as the target for selection. This is done in order to make proper comparison, as drift assemblies analyzed must be different from compotypes and must be related to their own chemical environment characteristics (represented by β). The reason for picking the drift from the simulation and not generating one at random is that it may very well be that the structure of drifts is somehow related to β. The reason for picking the assembly least similar to the compotypes is that it may very well be that composome and compotype identification algorithm is not perfect, so there are assemblies that might be considered as compotypes, if only they were slightly more similar to them (i.e. similarity threshold slightly reduced). For each of 1,000 β used in this section, under regular-GARD simulation, the average similarity between the target drift picked to the N_C compotypes centers of mass is 0.220±0.129, signifying that a substantial difference indeed exists between compotypes and the drifts picked.

The superior selection of compotypes over drifts is apparent, where only 34% of cases exhibited negative selection (selection excess<0.95, SE, Equation 6) vs. 76% in drifts (Figure 2). Compotypes also exhibit a higher tendency for positive selection (SE>1.05), where 29% of cases exhibiting it vs. 21% in drifts.

Thus, designating compotypes as selection targets proves to be a central point when addressing GARD's selection.

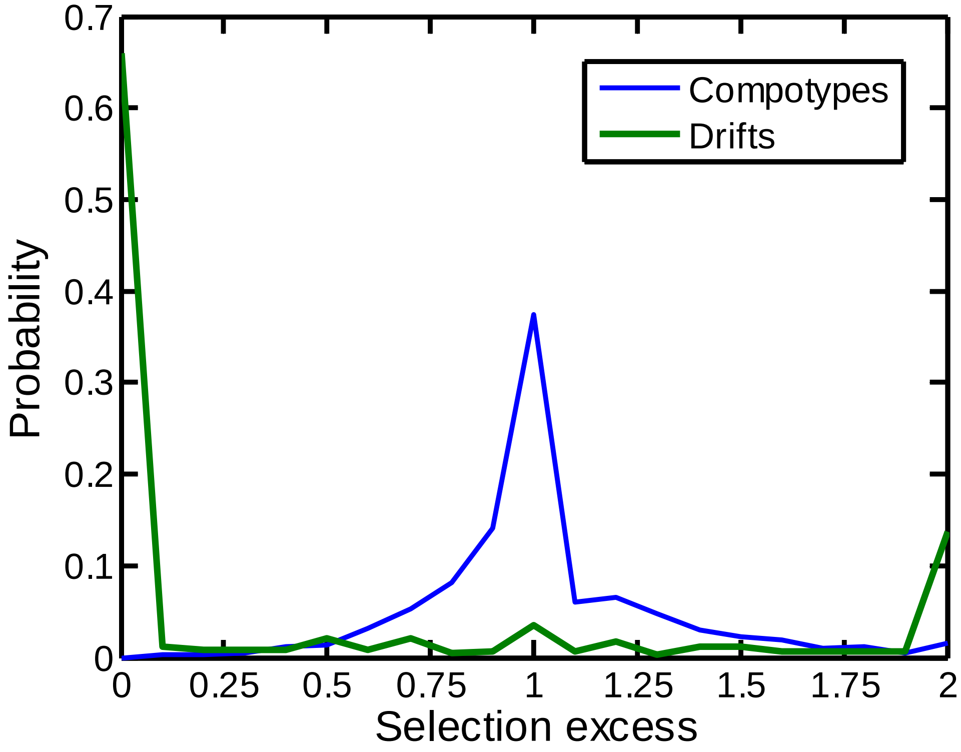

Figure 29: Selection excess (SE, Equation 6) of compotypes and drifts. The rightmost bin collects SE>2. Compotype data is taken from Figure 2b in [80].

5.4.6.2. Testing on every compotype in a population

Previously, the selection response of only a single compotype out of each simulation, under a given β, was studied. This section will briefly present post-selection dynamics when all of the compotypes in a given simulation are addressed, one at a time. Further, in light of the results that GARD population exhibit ecological behavior similar to natural systems (chapter 5.4), this section will focus on the changes in population dynamics as a result of the selection. As a 'proof of principle' only a single case was analyzed, and is described in this brief section.

The β studied exhibits N_C=3. Therefore, four simulations performed: one with no selection pressure ("wild type") and three were in each a different compotype was designated as selection target. Each simulation's data is fitted with the logistic equation (chapter 4.5) and the ecological parameters are extracted. Additionally, α_{ij} values are used to construct a food-web network whose nodes and edges are respectively compotypes and α_{ij} values [138].

Selection alters the dynamics of the entire population rather than just of the targeted compotype (Figure **30**). This was quantified by examining the values of the ecological parameters before and after selection (Figure 31). Interestingly, when selecting for a compotype, its own K and r values typically did not change with respect to the wild type. When selecting for C1, K_2 increased by almost 100% and r_3 by more than 50%. When selecting for C2, its maximum increased by about 50% and its plateau by almost 100%, thought K_2 and r_2 did not change. This is indicative for a change in the food-web as a result of the selection (Table 5). Interestingly, the food-web does not change when selecting for C3 (Table 6). Selecting for C1 alters the food-web the most, and it is more similar to the food web when selecting for C2 than for the wild type.

Thus, applying selection towards a compotype in the population alters the food web but keeps this compotype's parameters constant (r and K). This is an interesting result that requires further inspection.

5.4.7. Network motifs and their effect on ecological dynamics

When the inner structure of β was decomposed into network motifs (see chapter 5.1.1.1), it was found that they are useful in predicting GARD's selection response. In this section, it will be asked if the motif spectrum can also be used to understand some of the ecological behaviors.

When β's are grouped according to their N_C values, it is found that the more compotypes a β exhibits, the lower are the counts of all motifs it shows. This can be understood when first considering that the size of a compotype intrinsic molecular repertoire (N_{mol}) was found always be much smaller than the environmental one (N_{mol}<N_G. See Figure 19). This means that a compotype is a distinct subpart of β. Therefore, the higher the counts of all motifs a β exhibits, the more it is connected and thus the lesser the chance it will exhibit more separated subparts, i.e. a lesser chance for a higher N_C value.

The structure of the β can also be used to predict cross-compotype competition dynamics in GARD populations. This is seen when α_{ij} values are presented as food-web, and binarized in the same way as β. For brevity, this analysis focuse on cases with N_C=2 and 3 only, because the food-webs in such cases is simple and exhibits only a single mode of interaction – only a single motif (Figure 33). In general, when the competition between pairs is one sided (i.e. $C_i \rightarrow C_j$), β exhibits less motifs. Interestingly, for N_C=3, when all three pairs exhibit reciprocal competition (i.e. $C_i \leftrightarrow C_j$) β shows less motifs than when only two pairs are reciprocal and the third is one-sided. The scores are overall negative because simulations with N_C>1 show negative scores to begin with (Figure 32).

Thus, the connectivity of β affects both the number of species and their ecological interactions.

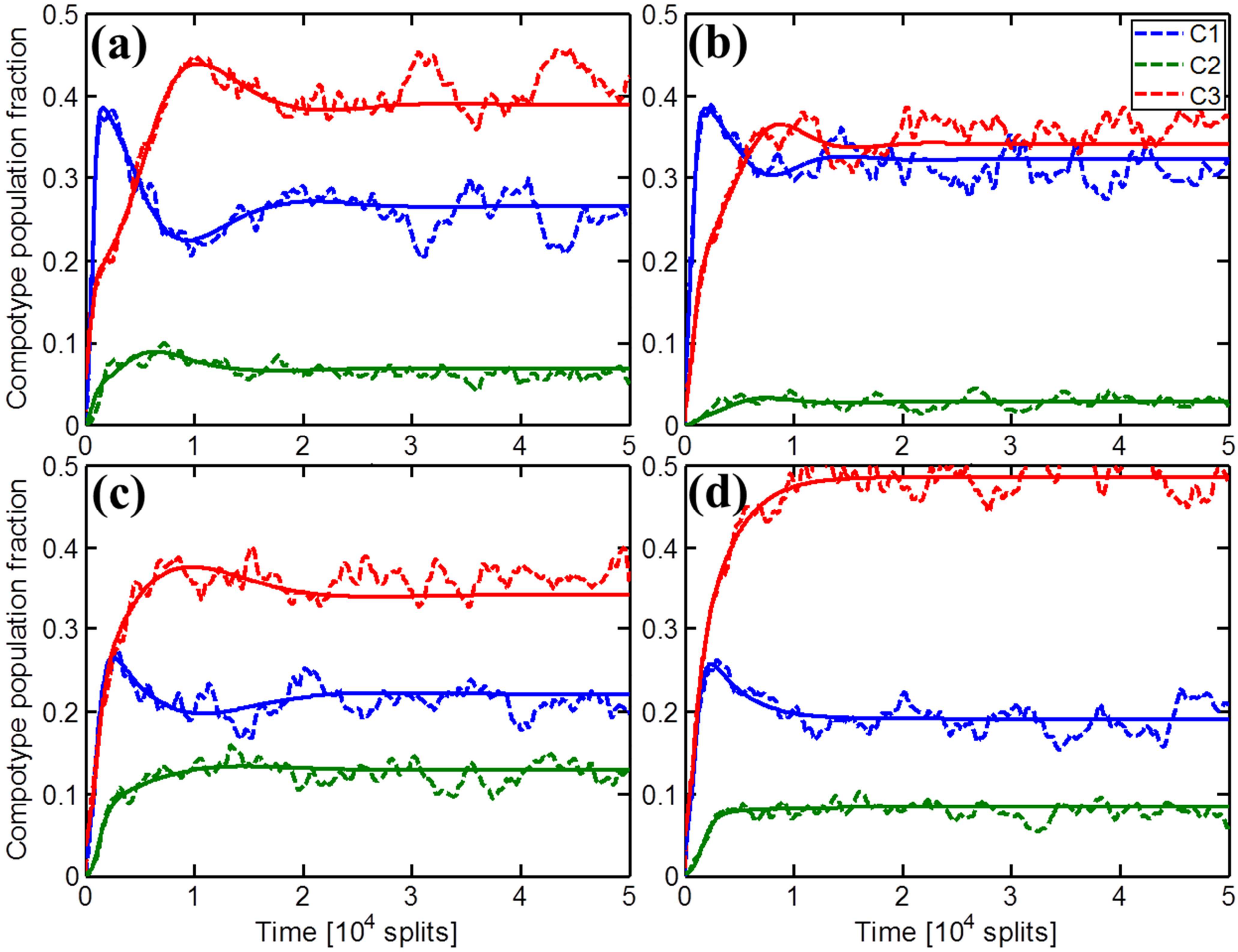

Figure 30: Example of selection in a population with multiple compotypes (lognormal random seed=45) and the fit to logistic growth (Equation 7). (a) Wild type population dynamics when no selection applied, (b-d) Population dynamics when selection target is C1, C2 and C3, respectively.

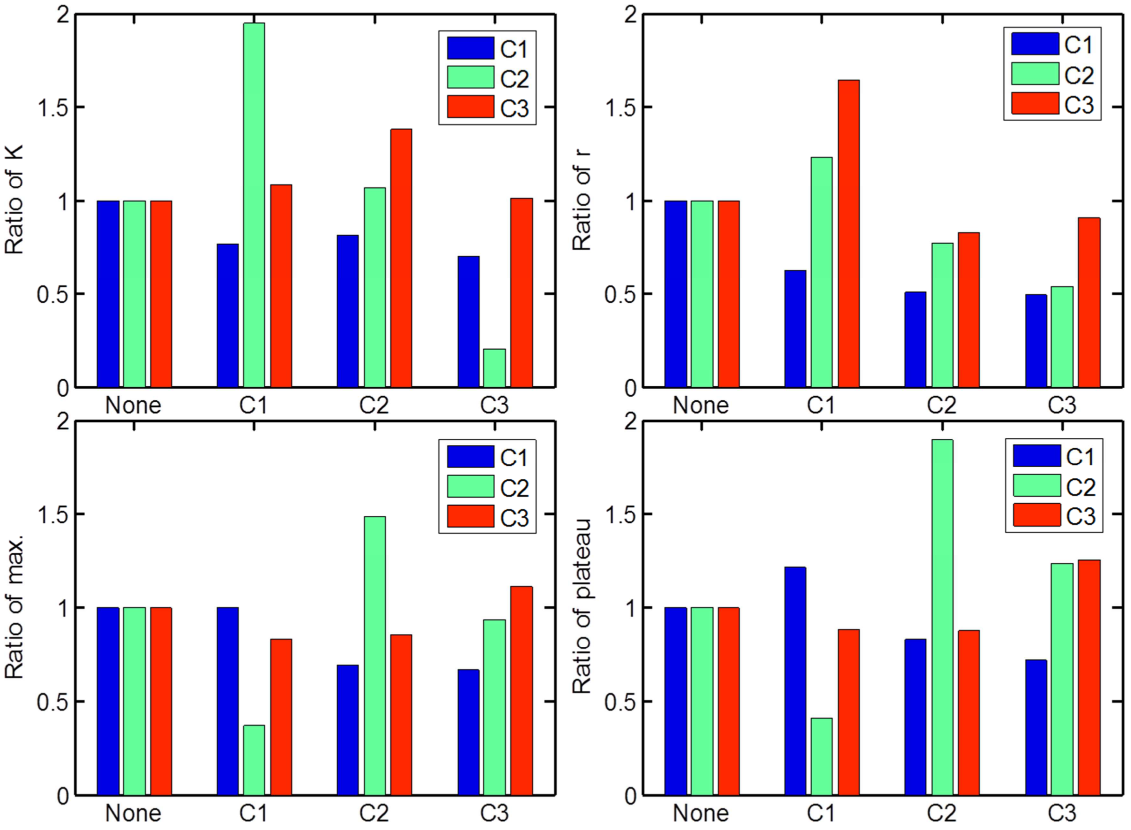

Figure 31: Ratios of values (carrying capacity, K; growth rate, r; maximum; plateau) after selection, compared to no selection (wild-type). A ratio >1.0 represent an increase in the value due to selection with the target stated on the X axis.

None	i=1	i=2	i=3	C1	i=1	i=2	i=3
α_{ij} (j=1)	-	0.79	1.26	α_{ij} (j=1)	-	0.48	1.10
α_{ij} (j=2)	1.49	-	0.00	α_{ij} (j=2)	0.00	-	10.0
α_{ij} (j=3)	0.49	0.56	-	α_{ij} (j=3)	0.51	0.25	-
C2	**i=1**	**i=2**	**i=3**	**C3**	**i=1**	**i=2**	**i=3**
α_{ij} (j=1)	-	0.97	1.55	α_{ij} (j=1)	-	1.25	1.35
α_{ij} (j=2)	0.00	-	2.46	α_{ij} (j=2)	2.79	-	0.14
α_{ij} (j=3)	0.68	0.54	-	α_{ij} (j=3)	0.00	0.83	-

Table 5: Values of competition parameters (α_{ij}) before and after selection for different targets.

	None	C1	C2	C3
None		0.10	0.47	0.95
C1	10.1		0.85	0.11
C2	2.90	7.58		0.37
C3	1.50	10.3	3.72	

Table 6: Similarity between the food-webs (Table 5) under different selections. Upper diagonal shows H (Equation 3) and lower diagonal shows Euclidean distance.

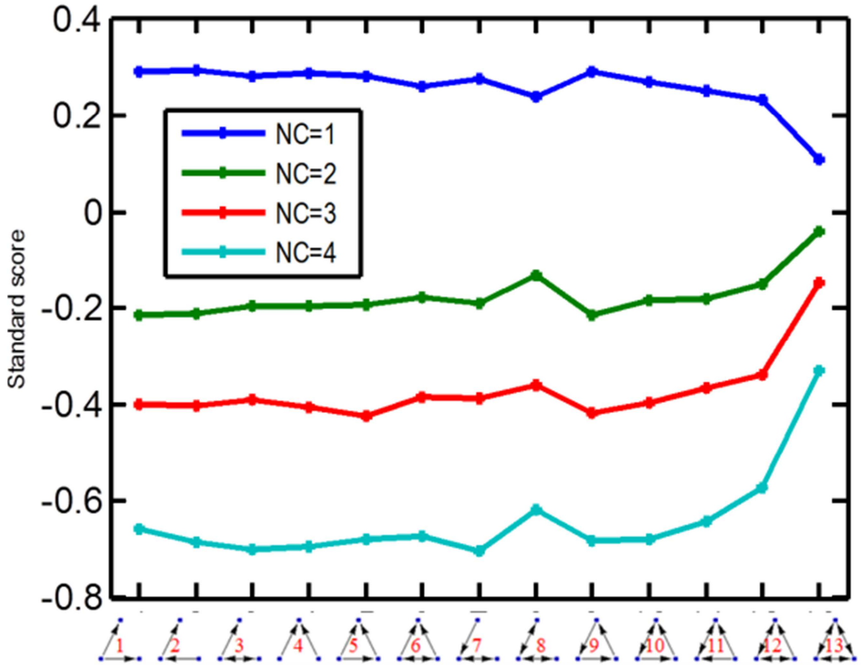

Figure 32: Average motif spectrum (relative to data in Figure 10), collected for simulations with different N_C values.

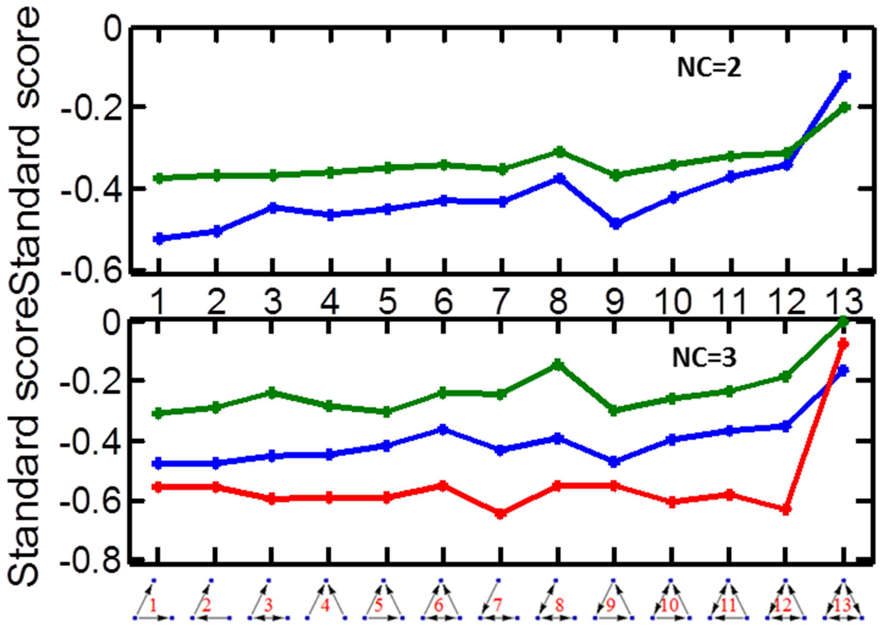

Figure 33: Average motif spectrum, collected for different food-webs. (top) All simulations with N_C=2. One-sided (i.e. C_i→C_j) and reciprocal (i.e. C_i↔C) competitions are respectively blue and green lines. (bottom) All simulations with N_C=3. Blue line is when all three compotype pairs show reciprocal competition (i.e. motif #13). Green line is when only two pairs are reciprocal and the third is one-sided (i.e. motif #12). Red line is when only one-sided competition exists (C1→C2→C3→C1).

5.4.8. Discussion

In a majority of the cases the logistic (or Lotka-Volterra) formalism is used for cases such as predator-prey or inter-species competition for resources. Should one use such formalism in the case of GARD populations?. These populations are characterized by a different dynamics, whereby species interconvert into each other, a situation resembling macro-evolutionary dynamics. Indeed, there are reports of utilizing Lotka-Volterra equations for such systems [22]. In GARD analyses, some of the ecological parameters are thus interpreted differently than in classical ecology. The carrying capacity (K) is related to the chance that a compotype will produce progeny belonging to the parent's compotype (and not to drift, or another compotype). Hence, K is related to the replication fidelity of a given species, independent of environmental parameters. Specifically, in the simulations presented here there is no competition for resources as the environment is buffered. GARD's α_{ij} parameter measures the extent of species inter-conversion, made possible by the fact that every compotype is a sub-network of the global β network. Thus, the forgoing results could seed a better understanding of early evolution, whereby protocellular entities were sufficiently malleable so as to reveal aspects of evolutionary ecology.

We utilize here the logistic equation to fit the dynamic behavior of GARD compotypes. This equation can show oscillations for certain parameter ranges [83, 106, 107]. Notably, in 1000 different sets of fitted logistic parameters here no oscillatory behavior observed. It is important, though, that such parameters are derived from chemical rate-enhancement values embodied in beta. Future analyses could give insights into the conditions for the existence of stationary states vs. oscillations.

5.5. Compotypes are quasispecies

This chapter describes work done by Renan Gross, a summer student I closely supervised, in order to answer a key question I posed regarding GARD's behavior and its similarity to a widely used evolutionary model, the quasispecies model.

Here, the GARD model dynamics are compared to that stemming from Eigen's quasispecies theory, seeking to unravel the quasispecies nature of GARD's compositional assemblies, as an archetype of a system replicating and evolving without a hierarchical genome. Sequential and compositional information are intrinsically different, which makes it appealing to study the quasispecies nature of compositional replicators. To the best of our knowledge, this has not been done before.

Eigen's quasispecies theory was first proposed to describe error-prone replication of primitive macromolecules carrying information at the origin of life [30]. It referred to information carriers that undergo self-replication with errors, and extended the classical concept of a species to include not only the main replicating sequence, but also its closely connected mutants [12, 29, 30]. A quasispecies is a steady-state population of variants ("cloud") around the master-sequence, which are linked through mutations and collectively contribute to the characteristics of this cloud. The master-sequence is referred to as the sequence with the highest fitness and it is thus the dominant sequence amongst the distribution within the cloud. It is typically the wild-type sequence from whose erroneous replication gave rise to other mutants.

Perhaps the best example in nature of quasispecies is RNA viruses, which have low replication fidelity with measured high mutations rates [25, 48, 72, 115, 147]. While in the past it was argued that RNA viruses' evolution does not follow the quasispecies theory [53], this is largely disputed [72, 111]. Two additional examples of natural quasispecies, are the genome of Chinese hamster ovary which has genetic diversity due to non-standardized cloning [149] and catalytic RNA molecules [5, 67].

The quasispecies equation describes a population of self-replicating genotypes (Equation 12) [12, 29, 30]. Due to replication errors, a genotype produces not only offspring of its own kind, but might also produce offspring of other genotypes. This is represented by the transition matrix (Q) which denotes the probability that a certain genotype will produce an offspring of another genotype. Thus, the growth of a particular genotype is governed not only by its own replication rate, but also by the replication rate of the other genotypes. The is written as:

$$\frac{dx_i}{dt} = \left(A_i Q_{ii} - \tilde{E}(t)\right)x_i + \sum_{j \neq i} A_i Q_{ij} x_j$$

Equation 12

Where for a genotype i, x_i is its time dependent concentration, A_i is its replication rate (as it reflects its fitness [29]) and Q_{ij} is the probability of genotype j mutating into i (with Q_{ii} being the probability of self-replication). $\breve{E}(t)=\sum x_i A_i$ is termed “average excess rate” and serves to keep the total population size constant ($\sum x_i=1$ at all time points). A steady-state solution to this equation is obtained as the eigenvector with largest eigenvalue of the matrix W={QA}, in accordance to Perron-Frobenius theorem [29, 104]. This eigenvector holds the steady-state distribution of the concentrations of phenotypes, which is the quasispecies.

Using the quasispecies equation, it is possible to quantify an error threshold, which relates the amount of information a replicating system can store at a given mutation rate to its single digit error rate (e.g. the chance to insert a wrong nucleotide in a specific location) [12, 30, 147]. The error threshold is defined as the minimum accuracy of replication which is required in order to preserve the information of the selected state of the system. For optimal selection, the required precision of information transfer has to be adjusted to the amount of information to be transferred, and if the mutation rate is increased beyond this limit the population structure breaks down [30]. As RNA viruses replicate with relatively high mutations rates [115], they are susceptible to a treatment by mutagenic drugs which increase their mutation rates to push them beyond the error catastrophe [23, 129, 134]. This not only supports the quasispecies nature of RNA viruses, but is also an example of a relation between modeling and experiments.

5.5.1. Sequential vs. compositional information

There are inherent differences between sequential and compositional information. The differences between two binary *sequences* with *same length* are not like between two *compositions* with same *total number of molecules* (typically the former is represented by a string and the latter by a vector). Consider the two binary sequences: S1=01 and S2=11. The difference between them can be pinpointed to the first location mutated to “1”, i.e. a Hamming distance of 1. However, when considering the two compositions: C1=[1A,1B] and C2=[2A,0B] (where A and B are two different molecule types), their differences are that one molecule of type B is missing in C2 and an extra molecule of type A is in C1, i.e. an Euclidean distance of $\sqrt{2}$ (Equation 13). Another property of compositions, is that one composition can be a direct multiplication of the other, as in the case of C3=[2A,1B] and C4=[4A,2B], which in fact both hold the same composition (C4=2×C3). For such cases the similarity measure has been applied [119] (H, Equation 3).

Another difference between sequence and composition is related to the probability of a back mutation (revertants). Consider the sequence of length=100: S3=11…1 (each digit is set to 1). A single mutation at the first location will produce the sequence S4=01…1. Now, the probability of a back mutation from S4 back to S3 is 1/100 (the sequence length), when neglecting any other factors for simplicity. A point mutation in a composition is when 1 molecule of a certain type is replaced by a molecule of a different type. With compositions, the probability of back mutations depends on the composition itself, as shown in the next two examples. When the composition C5=[100A,0B] mutates into C6=[99A,1B], the probability of C6 mutating back to C5 is 1/100 (when neglecting any other factors for simplicity). However, the probability of C7=[51A,49B] mutating into C8=[50A,50B] is 51/100, as changing any A molecule into B will suffice.

Additional difference relates to the genome length. The longer a polymer is the bigger is the number of genes it can potentially code for. In GARD, however, having a bigger assembly size (N_{max}) leads to decrease in the number of compotypes because the system is nearing the equilibrium steady state (Figure 17a in [80]). Therefore, the equivalence of a longer genome in GARD may necessitate an increase a constant N_{max}/N_G while N_{max} is increased.

5.5.2. A compotype is an attractor in compositional space

Each simulation was performed with identical parameters yet with a different β. For each simulation, the most frequent compotype (FC) was identified, and Q (Equation 12) was constructed by sampling assemblies in different distances around the center of mass of FC (see chapter 5.5.8). Figure 34 shows the average Q over the entire set of 1,000 simulations performed. The most striking feature is that for all but the smallest distances, replication occurs towards FC. For distance>40, replication always occurs towards FC, while for intermediate distances between 20 and 40 replication can occur towards and away FC, and for distance<15 replication is usually away from FC. In other words, the progeny of any parent assembly located far (distance>40) from FC will always grow to be closer to FC than the parent, while for a parent located very close (distance<15) to the FC will typically grow to be slightly further away from the FC. Thus, a compotype is an attractor in the compositional space.

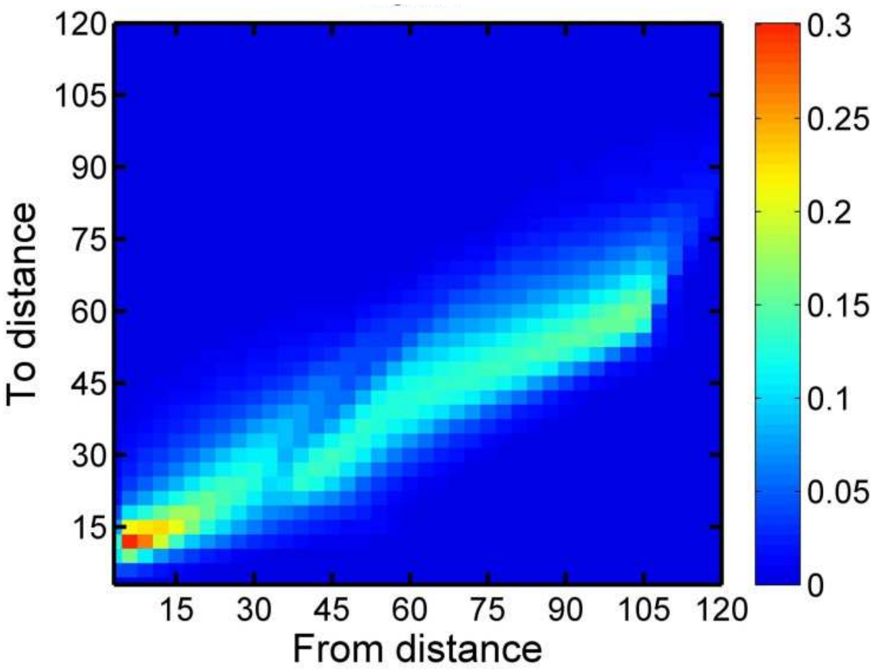

Figure 34: GARD's transition matrix (Q) with respect to the most frequent compotype (FC) center of mass. Q_{ij} is the probability that a parent at distance j (=X axis) gave rise to a progeny at distance i (=Y axis). Data in figure is averaged over 1,000 measurements, each with a different β network (see Figure 41 for specific examples).

5.5.3. GARD operates near its error threshold

For each β, the eigenvector with the highest eigenvalue was calculated and marked as V_β (see chapter 5.5.7). In each simulation, overall replication accuracy was assessed by comparing the degree of H between the center of mass of FC and V_β (Figure 35A). This was done for decreasing values of k_f rate constant (Equation 2) – the basal molecular joining rate, which is found to be a proxy analogue the single digit error. High k_f value contributes to overall faster assembly growth and shifts the k_f/k_b equilibrium towards a higher value. Decreasing k_f by more than a factor of 2, results in FC becoming more and more dissimilar to V_β and the former's frequency significantly diminishes, hinting to an error catastrophe (Figure 35B). Even in the highest k_f value examined, there is still a difference between the FC and V_β, which is likely caused by the stochastic nature of the model and perturbations caused by the assembly cell cycle (i.e. growth-split cycle), with $H(k_f=0.01)=0.917\pm0.161$. As k_f is reduced, especially below 0.005, V_β and FC become more and more dissimilar, seeming to asymptotically reaching $H\rightarrow0$ as $k_f\rightarrow0$ (the lowest value explored here $H(k_f=0.00035)=0.488\pm0.210$). Interestingly, when the relative frequency of FC is compared to its frequency with $k_f=0.01$, a slightly lower k_f values actually increases the frequency by almost 50% at $k_f=0.005$ (Figure 35B). FC frequency reaches 0 as the value of k_f is lowered. The increased FC frequency in intermediate k_f levels is suggested to occur due to an improved ability of assemblies to explore new regions in the compositional space around FC. Thus, the replication of GARD's compositional assemblies seems to be near an error threshold. This draws for a quasispecies analysis, which will be presented in the next section.

5.5.4. GARD's steady state is correlated with that of the quasispecies equation

As GARD operates near its error threshold, it is susceptible to be analyzed in accordance to the quasispecies theory. Figure 36 shows examples of the steady state distribution around the FC center of mass, when predicted based on the quasispecies equation and when measured from GARD's population dynamics (such as those presented in Figure 18). Strikingly, the two distributions share similar features, including the distance span and number and location of peaks. The differences between the two distributions might result from the grouping of assemblies according to their distance, which can take different assemblies who occupy different locations in the compositional space and exhibit different replication rates and directions, and assign them with average properties.

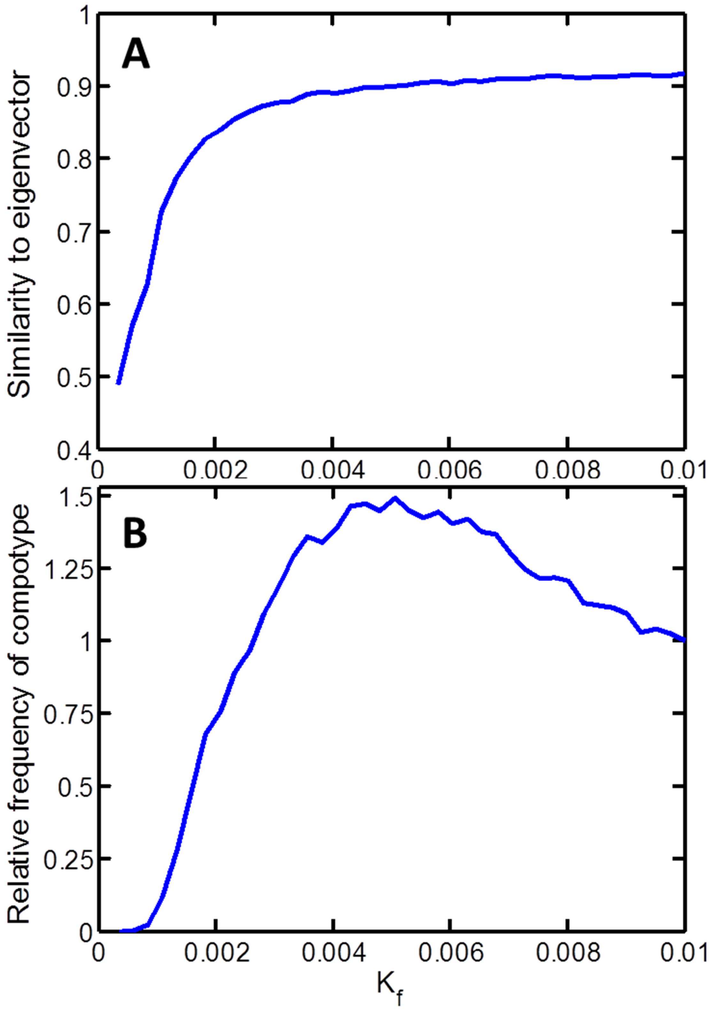

Figure 35: GARD's error threshold. (A) Average compositional similarity (H) between the center of mass most frequent compotype (FC) to the β eigenvector (V_β) as a function of k_f. (B) Frequency of FC as a function of k_f (relative to the frequency at k_f=0.01). Dataset is based on 1,000 simulations, each with same parameters except for

different β networks. The default value typically used in simulations in the past is k_f=0.01. Each simulation was ran for 2,000 generations.

The effect of decreasing k_f is further seen here, where when the distributions were calculated based on simulation with low k_f the steady state distribution is shifted towards substantially greater distances. A widespread agreement between the steady states of GARD and quasispecies equation is observed when comparing the entire dataset of 1,000 different simulations (Figure 37). These results support the description of a population of compositional assemblies around a central compotype (master compotype) as a quasispecies.

An important point – is whether the choice of FC as the master compotype is justified. To this end, the steady state distributions with respect to an assembly randomly generated were compared (Figure 38), to find a complete lack of correlation. This further supports the choice of a compotype as a master compotype.

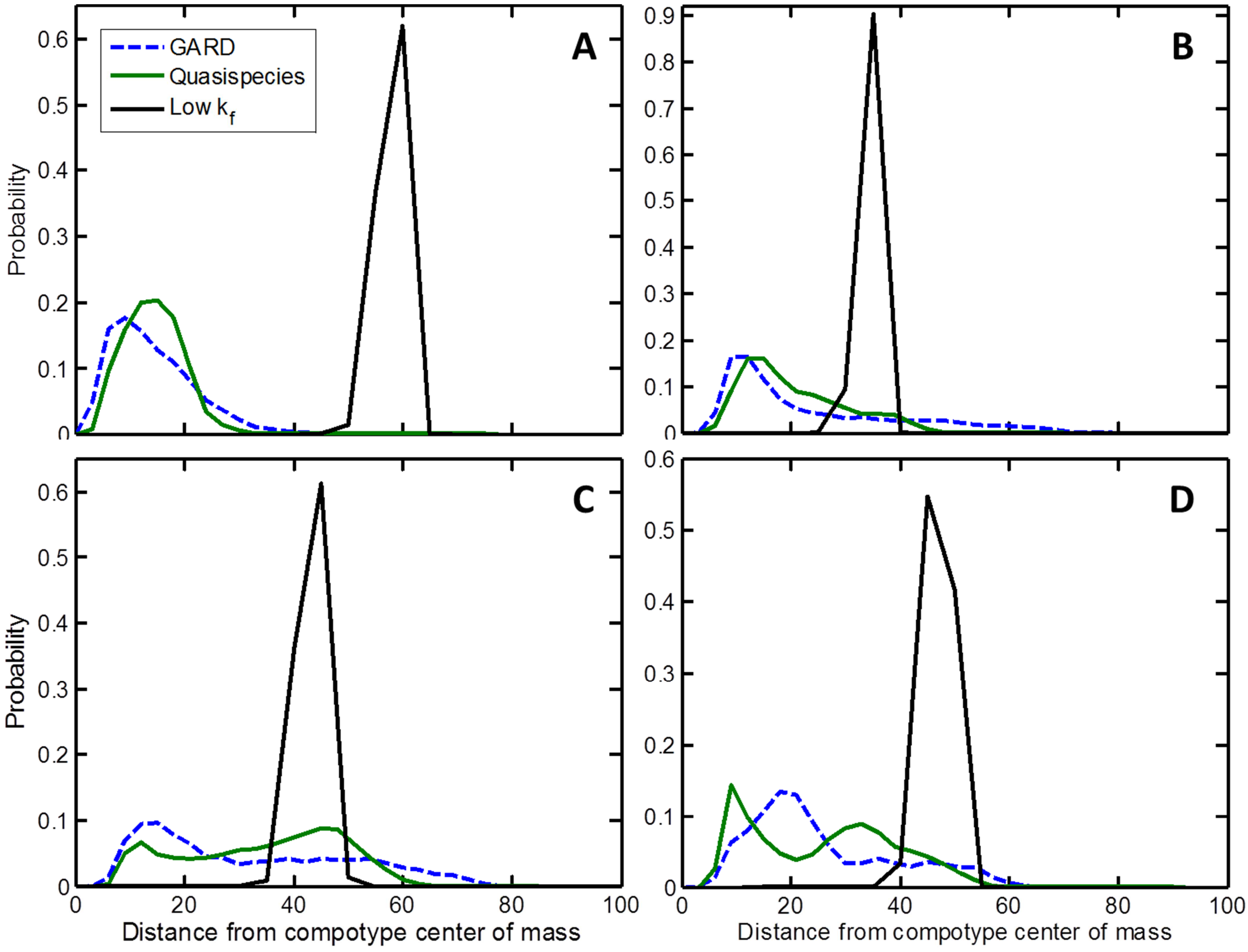

Figure 36: Examples of the steady state from population-GARD and from the quasispecies equation (Equation 12). The distributions of distances around FC are shown from GARD (blue broken line) and from the quasispecies equation (green solid

line) with k_f=0.01. Black solid line shows GARD's steady state distribution for k_f=0.00035. β random seeds are 1, 79, 45 and 90 for panels A-D respectively. Q_{ij} and A_i are presented in Figure 41 and Figure 42.

5.5.5. GARD's error-catastrophe resembles that of the quasispecies equation

GARD's population shows similar behavior to sequences, when approaching the error threshold. Figure 39 shows an example of the occupancy of different distance shells around a compotype's center of mass, when the value of k_f is reduced. While the frequency of the compotype only slightly reduces at first, it is seen that the occupancy shifts from shells of smaller distances to larger distances. Only when the occupancy shifts towards shells of relatively high distances the frequency of the compotype quickly diminishes towards 0. Importantly, a similar transition was observed in simulated population of replicating polymers, where the concentration of the average sequence (consensus) remains constant as the single digit error probability is reduced, while sequences with multiple mutations at larger and larger Hamming distances was observed [139]. When the error threshold was reached, the concentration of the average sequences exhibits a first-order phase-transition and drops to 0.

5.5.6. The time dependent dynamics of the quasispecies equation resembles that of GARD

Lastly, an example of the time dependent evolution of the distance distribution is compared between population-GARD and the quasispecies equation, when both started from the same initial conditions, and allowed to propagate until steady state. Strikingly, in the case tested the time development of the quasispecies equation resembles that of GARD, where the formation of a peak around distance ~20 at the expanse of the peak around distance ~60 is exhibited in both (Figure 40). This further support the description of composomes around a compotype as a quasispecies, suggesting that this description applies also to the full dynamics.

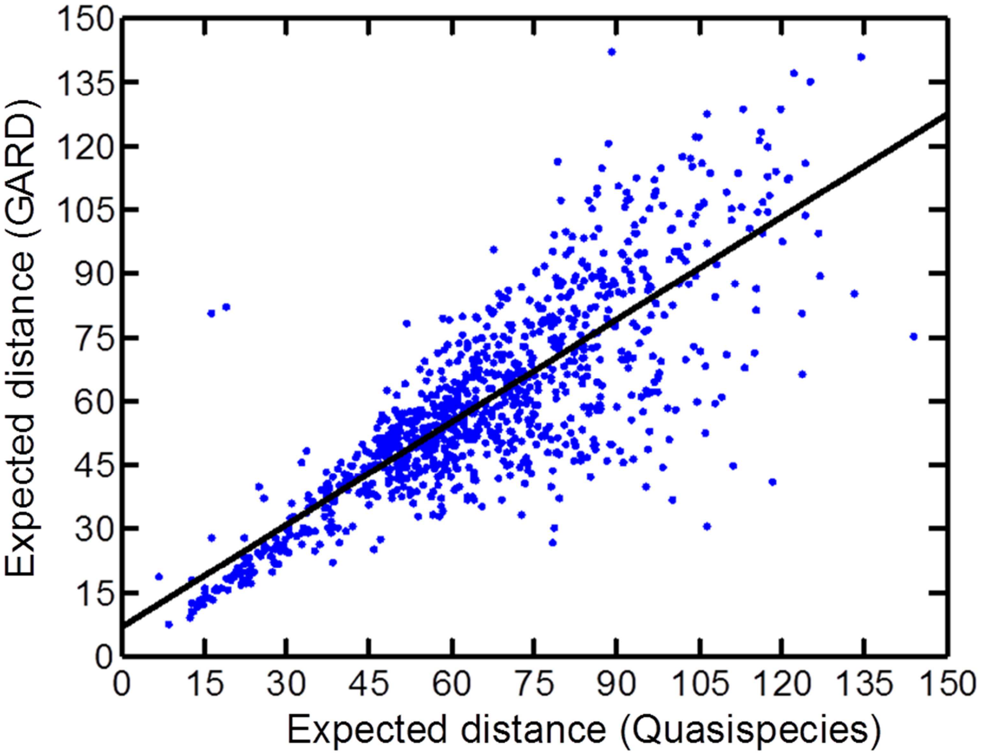

Figure 37: Steady-state comparison of GARD's vs. the quasispecies equation, with respect to FC center of mass. Data shows the expected distances (=∫p(r)dr) from the steady-state distributions (Figure 35). Black solid line is a linear fit: y=0.804*x+2.27, R^2=0.60.

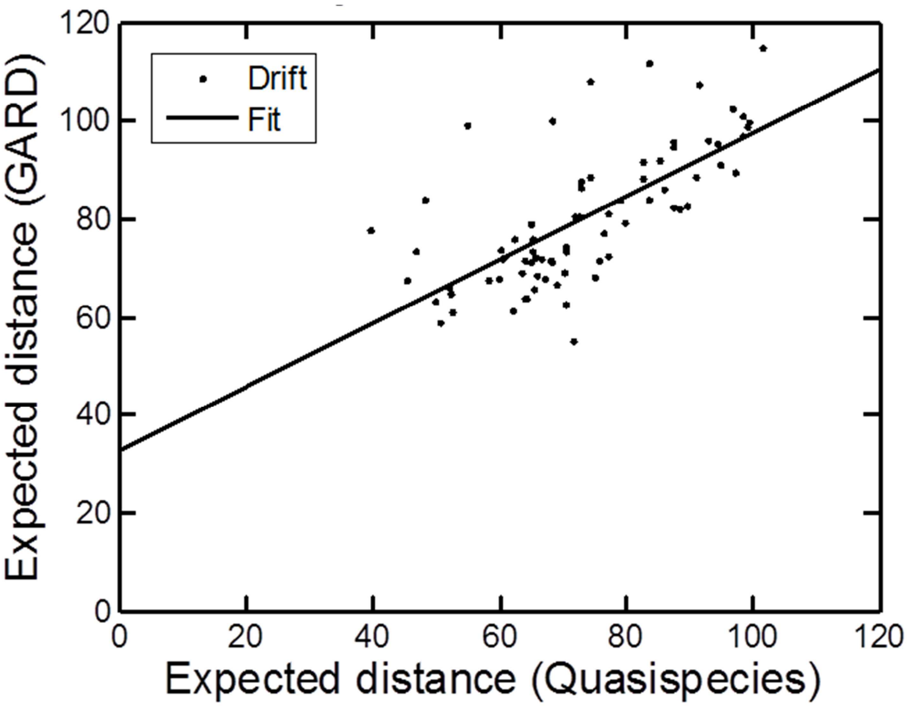

Figure 38: Comparison the steady-states, with respect to a random assembly. Figure details are similar to Figure 37, except for a set of 75 simulations instead of 1000. Linear fit gives: y=0.65*x+33, R^2=0.49, whereas performing the fit of Figure 37 with the same 75 simulations gives: y=0.88*x+0.66, R^2=0.80**.**

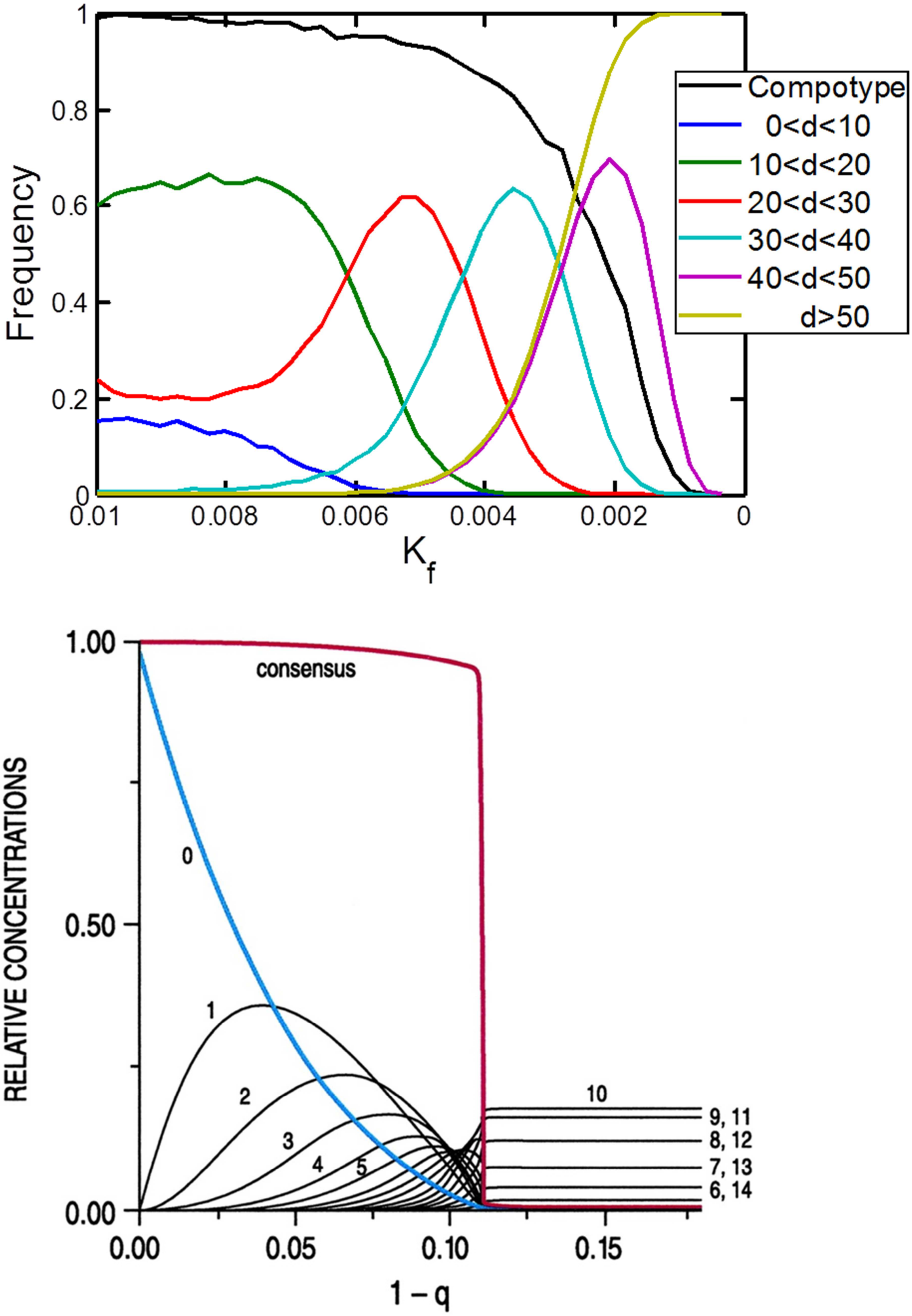

Figure 39: Error-catastrophe. (top) An example of the change in compotype frequency and occupancy of distance shells in GARD as k_f is reduced. (bottom) The change in the concentration of the average (consensus) sequence and sequences at different Hamming distances from the master-sequence as the single digit error probability is reduced; Taken from [139].

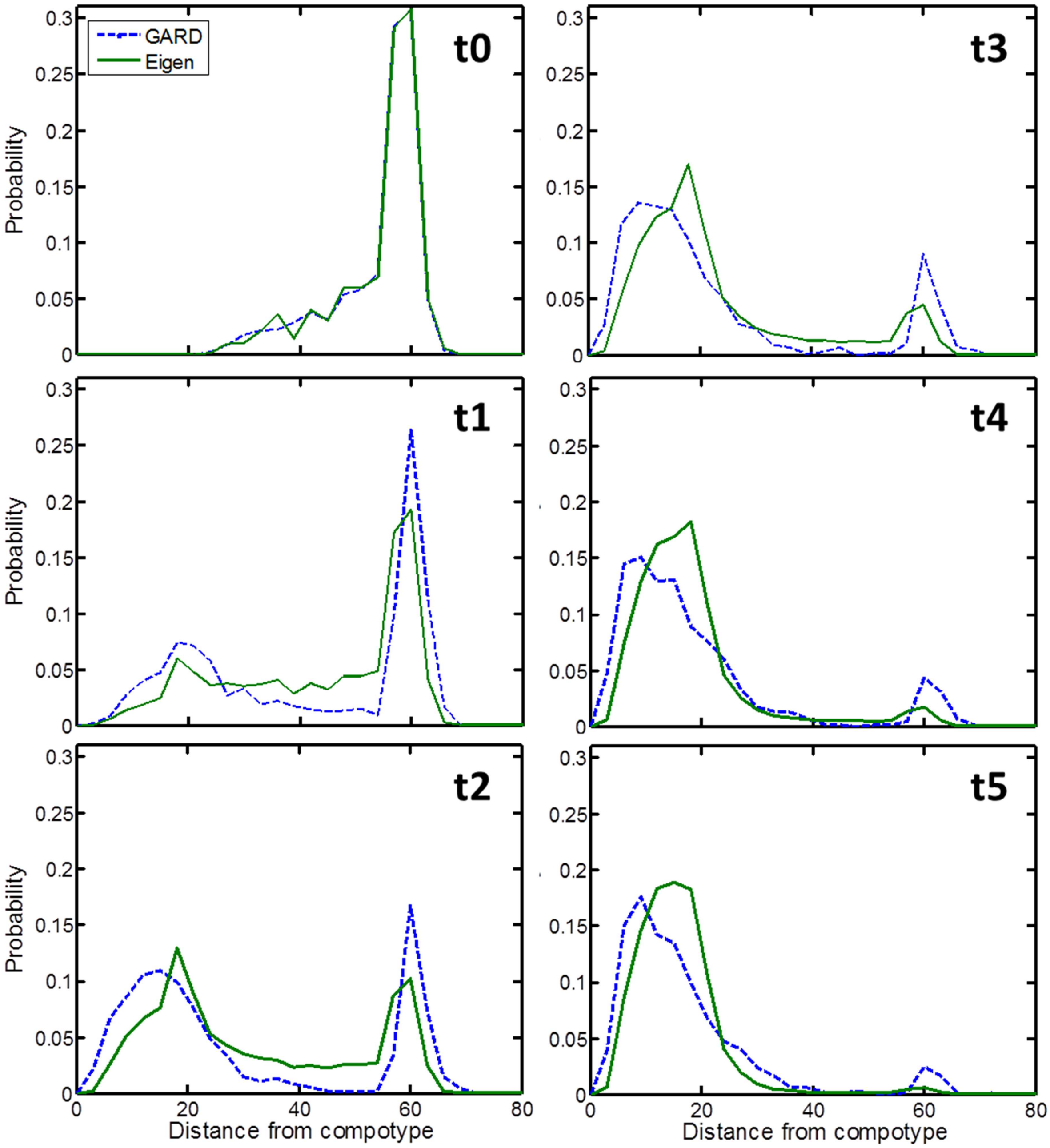

Figure 40: Example of GARD and QSEQ dynamics towards steady-state. Distance distribution at different time points until steady state, for population-GARD (blue broken line) and QSEQ (green solid line). Time points are arbitrary and monotonic (t0<t1<t2<t3<t4<t5). β random seed = 1. The time dependent behavior of QSEQ was obtained by numerical integration using MATLAB routine ode45.

5.5.7. How the β eigenvector was calculated

When β is represented as a matrix, it is positive as each of its β_{ij} values are sampled from a lognormal distribution [121]. According to the Perron-Frobenius theorem, such a matrix has a unique largest real eigenvalue with a corresponding all positive real eigenvector [104]. This eigenvector is treated as a compositional assembly and marked V_β (see chapter 5.5.3).

5.5.8. How the compositional space was sampled and Q and A constructed

The large size of the compositional space, particularly given the values used in this work, N_G=100 and N_{max}=100, makes direct calculation of Q matrix computationally impossible. Therefore, assemblies were grouped according to their distance from the center of mass of a compotype (the FC) and the molecular space was divided into shells of constant thickness. This is similar to how genetic sequences are grouped according to their Hamming distance from the master-sequence [61].

The Euclidean distance between two assemblies is calculated as:

$$D(V^1,V^2)=\sqrt{\sum_{i=1}^{N_G}\left(n_i^1-n_i^2\right)^2}$$

Equation 13

Where n_i^1 is the count of the i'th molecular type in assembly V^1 (Equation 1).

Assemblies in the same distance shell were grouped together and the relevant properties (i.e. Q and A) of each shell were averaged over the assemblies contained in this shell. The compositional space was sampled in the following manner for each simulation:

10,000 assemblies were generated at random, each by randomly picking a molecular type and adding a random count of this type until the assembly size reaches N_{max}. Another 10,000 assemblies were generated by conducting 10,000 random walk step pairs starting from the FC, where in each step a molecule is randomly removed from the FC and a random one is added to it. Another 10,000 assemblies were generated by random walk starting from the V_β, similarly to the FC. This gave rise to a total of 30,000 parent assemblies. Each of these was then split and its progenies grown until they reach N_{max}, this was repeated 10 times for each parent. Q_{ij} is than the probability that a parent at distance j gave rise to a progeny at distance i, and A_i is the average growth rate of progeny at distance i. Examples of Q and A are given in Figure 41 and Figure 42.

The sampling of population-GARD steady-state distribution was done by collecting the entire population along the population steady state (time=4.9-5.0×10^4 with time intervals of 0.1×10^4. See for example Figure 18).

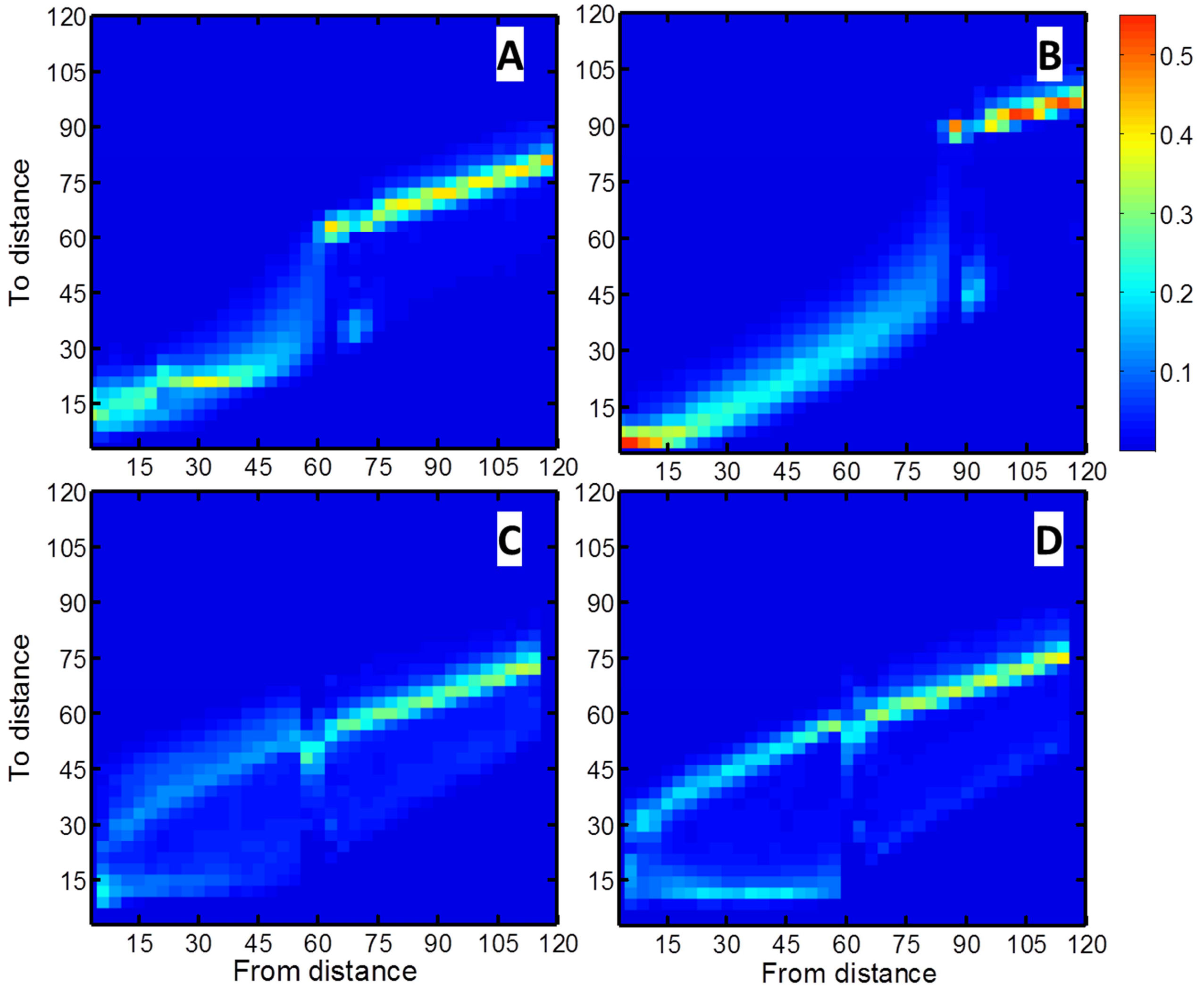

Figure 41: Examples of Q. Figure details are as Figure 34. Panels A-D respectively corresponds to Panels A-D in Figure 36.

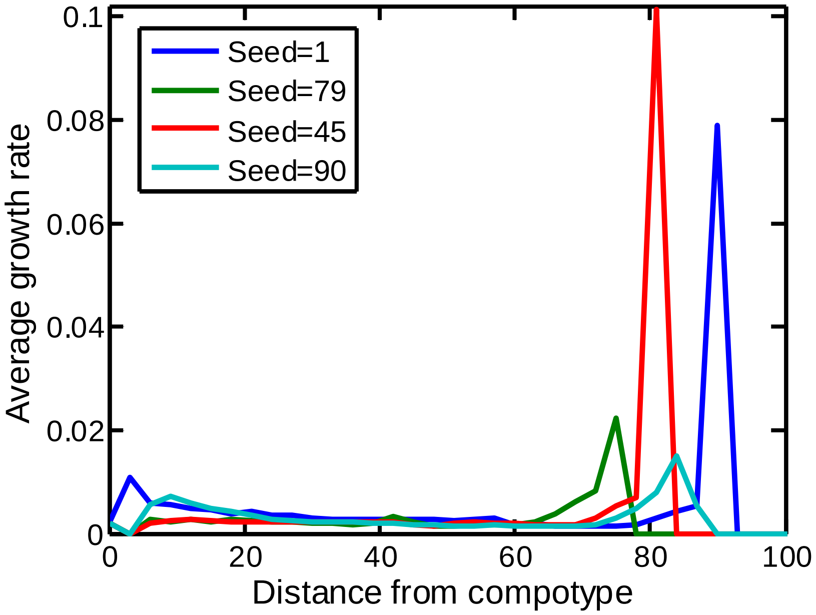

Figure 42: Examples replication rates (A). Lines respectively correspond to Panels A-D in Figure 36.

5.5.9. Conclusions

While the quasispecies is a general concept, not specific to any sort of replicators, in the past it has always been applied to the only replicators we know – sequential polymers. The present demonstration of the quasispecies nature of GARD assemblies, which hold compositional information as a group rather than sequential information as an individual molecule, thus supports the generality of the QSEQ. Importantly, as GARD was developed and is often used to study the lipid world scenario for the origin of life, the present results, together with experiments demonstrating the quasispecies nature of catalytic RNA [5], further points to the role of quasispecies in the origin of life [31].

Lastly, in a separate vein, if one agrees that it is not wrong to represent a cell's transcriptome, which holds the composition of the different types of RNA molecules [78, 98, 148], as a compositional vector, then it is further suggested that regardless a cell's genome, the transcriptome is a quasispecies.

5.6.Is there an optimal level of open-endedness in prebiotic evolution?

This work was done with collaboration with Prof. Natalio Krasnogor from Newcastle University, and was published as an extended abstract [81].

Open ended evolution is considered an important feature of life. GARD simulations cycle between composomes (Figure 4) and typically exhibits only a few compotypes (Figure 9 in [80] and chapter 5.6.1), which impends on its ability to portray open ended evolution. Therefore, modifying GARD to show open ended evolution will greatly contribute to its acceptance as an evolutionary model. This chapter deals with this issue exactly and further suggests a new method to quantify open ended evolution in a way that will enable comparing different systems and models on the same scale.

Open ended evolution may be thought of as referring to a “system in which components continue to evolve new forms continuously, rather than grinding to a halt when some sort of `optimal' or stable position is reached” [137]. Notably, open-ended evolution does not necessarily imply evolutionary progress or complexification. Yet, a system in which complexity increases along the evolutionary time axis fulfills a sufficient (even if not necessary) condition for open-endedness.

Indeed, evolution of complexity and the related concept of open ended evolution have been a topic of scientific enquiry since Darwin and Wallace introduced the Theory of Evolution by Natural Selection. There is no doubt that a complexification process took place over the extended evolutionary time frame, with some endosymbiotic events [90]. With the advent of powerful computational tools in which one could seamlessly run “what-if” scenarios about the origins of life, the questions of how complexity emerges from evolution-like processes and how open-ended the emergent processes have gained renewed impetus [8]. Researchers have proposed multiple definitions of both open-ended evolution and (pre)biotic complexity and have applied these measures to several, more or less convoluted, “Artificial Life” and prebiotic systems. Common definitions of open-ended evolution consider an increase in the internal complexity of species [84, 113] or species occupying ever more diverse niches of the natural design space [65, 85]. A difference between the two approaches is apparent when considering the case of a few species in an ecosystem becoming more and more complex vs. the emergence of multiple species that might each be relatively simple but overall occupy a relatively large portion of the natural possible design space. The former definition can be viewed as “species-centric” whereas the latter as more “system-centric” [65].

Korb and Dorin discuss at length the various attempts made at measuring open-endedness and suggest a two-part measure based on message minimum length required for conveying (or encoding) information (MML). They propose that a measure that considers the complexity of events of species evolving (part 1) and of the related hypothesis (part 2) would be a better measure for evolutionary complexity as it is a relative measure that takes into account not only the end product but the context in which these are produced. Building on this concept, we address the question of whether there is an optimal set up for a putative prebiotic universe that leads to greater open ended evolution of the species evolving within it. We define an index of the excess-complexity of species (event, E) in relation to the universe in which they evolve (U), as a proxy for open ended evolution. This index is ec(P(E|U), P(U)), that relates the probability of observing events (P(E|U)) to the probability of the initial conditions (P(U)). Our index (Equation 14), like suggested by Korb and Dorin (2011), is a two part index but with the additional advantage of having the following properties:

1. $P(E|U)$ and $P(U)$ are normalized such that $0 \leq P(E|U), P(U) \leq 1$.
2. $P(E|U) \rightarrow 0$ and $\rightarrow 1$ respectively represents improbable and probable outcomes unraveling from the initial conditions U. Similarly, $P(U) \rightarrow 0$ and $\rightarrow 1$ represents improbable and probable initial universe conditions.
3. $ec \geq 0$ and can grow arbitrarily large. The larger the value of ec, the more complex are the unfolding events in relation to a given universe.
4. $\lim_{P(E|U)\rightarrow 1, P(U)\rightarrow 1,} ec\big(P(E|U), P(U)\big) = 0$, that is, probable initial conditions that lead to probable events receive the lowest rank (i.e. no surprises can be expected from this universe under the given initial conditions). This is marked as Scenario A in Fig. 1.
5. $\lim_{P(E|U)\rightarrow 1, P(U)\rightarrow 0} ec\big(P(E|U), P(U)\big) = K\,;\ K > 0$, that is, improbable initial conditions that lead to probable events are ranked slightly higher than 0 (Scenario B).
6. $\lim_{P(E|U)\rightarrow 0, P(U)\rightarrow 0} ec\big(P(E|U), P(U)\big) = L\,;\ L > K > 0$, that is, an improbable initial state that leads to improbable events ranks even higher as this clearly represents an unexpected observation emerging from an unexpected initial condition (“Garden of Eden”, scenario C).
7. $\lim_{P(E|U)\rightarrow 0, P(U)\rightarrow 1} ec\big(P(E|U), P(U)\big) = M\,;\ M > L > K > 0$, that is, a probable set of initial conditions throws out surprising outputs thus ranking at the top of the scale (“*Elegant* Garden of Eden”, scenario D).

We now define ec with exactly the above characteristics:

$$ec(E,U) = -\frac{Log_2[P(E\,|\,U)\cdot P(U)]}{2} + Log_2\frac{Max[P(E\,|\,U), P(U)]}{2}$$

Equation 14

where the first part is an embodiment of MML and is scale invariant, and the second part is different than zero only when added value in the complexity of events has occurred (i.e. U is simpler than E). Increase in ec during a simulation will serve as a proxy to identify open ended evolution, as increase in complexity is generally considered to be indicative of open ended evolution. Moreover, following the previous discussion, ec(E,U) is a species-centric measure of open ended evolution but can easily be “system-shifted” by encompassing all outcomes E.

In Figure 43 we identify the four extreme ec values a simulated universe might receive. As one moves from A, to B, to C and finally to D the level of ec increases thus open ended evolution is observed. In fact, as the likelihood of the initial conditions increases (P(U)→1) and the likelihood of the events decreases (P(E|U)→0) the level grows, potentially without limit.

A major challenge lays in exactly defining and measuring P(U) and P(E|U). To this end, it is suggested to build on the fact that not only nucleic acid repertoire underwent evolution, but also the genome [47, 51, 102]. The earliest genetic code is proposed to encode for fewer amino acids than the present-day one from one side, and from the other side the phylogenetic tree of life has too been suggested to expand during evolution. P(U) can, be related to the probability of a specific code out of the entire possible codes. P(E|U) might be related to the observed size of the tree at different evolutionary stages, given that this code led to that tree, compared to the sizes of the rest of the trees.

Lastly, I would like to submit that life can be considered as being about under achievement. Looking at the huge diversity of life all around us, it is understandable why some may, mistakenly, think that life is about giving rise to many species or the potential to give rise to infinite more, which is often referred to as open ended evolution. However, as suggested here, it is not enough to consider just the genome size or the actual number of species (i.e. the *output* of the system). One has to consider the **actual number of species with respect to the potential** (which is a function of the *input*). Doing so, it becomes apparent, for example, that the ratio of actual species to genome size has actually decreased in the course of evolution. Thus, life is about **underachievement**.

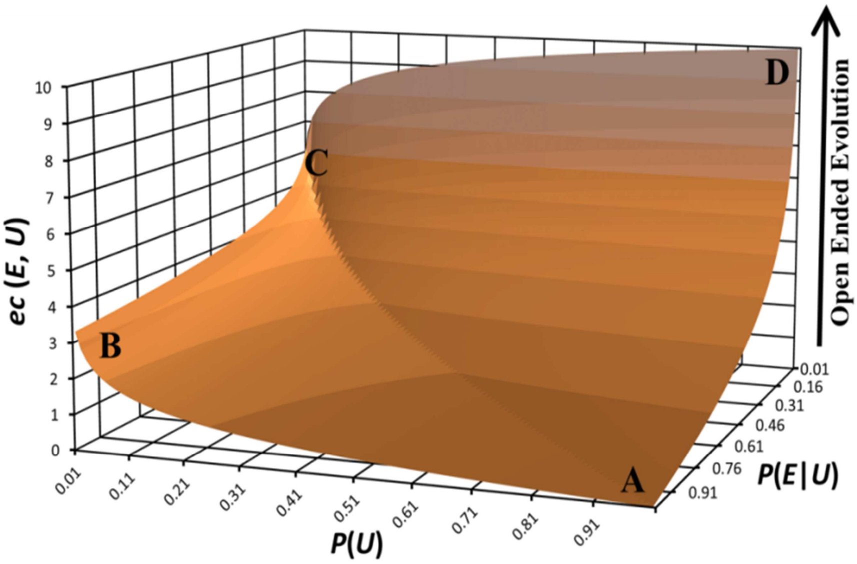

Figure 43: Excess-complexity (ec, Equation 14) as a proxy for open-ended-evolution. A, B, C and D mark different scenarios (see text).

5.6.1. Universe-GARD

In order to address the question of whether there is an optimal set up for a putative prebiotic "universe" with events unfolding inside such that open ended evolution is observed, the GARD model was extended. Typically up to eight different compotypes are being cycled and they can be identified early on in a simulation, suggesting that GARD does not appear to display high levels of ec to begin with, which impends on its open ended evolution (Figure 44). The new proposed model, termed universe-GARD (U-GARD), will allow systematically studying the tradeoff between the initial conditions of the universe and the emerging compotypes (i.e. map the ec surface). In U-GARD, the immediate environment is embedded in a larger "universe" with N_U ($\geq N_G$) different molecular types, instances of which are continually being diffused in and out of the immediate environment (Figure 45). This is physiochemically feasible, as exemplified by the immediate environment being absorbed to a mineral surface, contained in a mineral pore or constituting an ineffectively stirred sub-region of a larger prebiotic aquatic body. As a compotype constitutes a set of molecules that function better as a whole in their particular environment and thus faithfully replicate, the organization of a compotype is also assumed to protect its constituting molecular types from being diffused out to the larger universe.

The simulation will be run for sufficiently long time course and P(U) and P(E|U) will be measured along time intervals of fixed length. For each interval, P(U) is the probability of randomly picking the set of N_G types observed during the interval out of N_U (assuming that the diffusion rate is much slower than the accretion rate so that N_G is relatively constant in this interval) and P(E|U) is the probability of finding *new* compotypes in this interval. Open ended evolution will be identified when a universe will exhibit an increase in ec over time. Different universes with different N_U, N_G and β parameters could be compared by using the expected value of ec:

$$\left\langle ec\{N_U, N_G, \beta\}\right\rangle = \sum\left[ec(E,U)\cdot P(ec(E,U))\right]$$

Equation 15

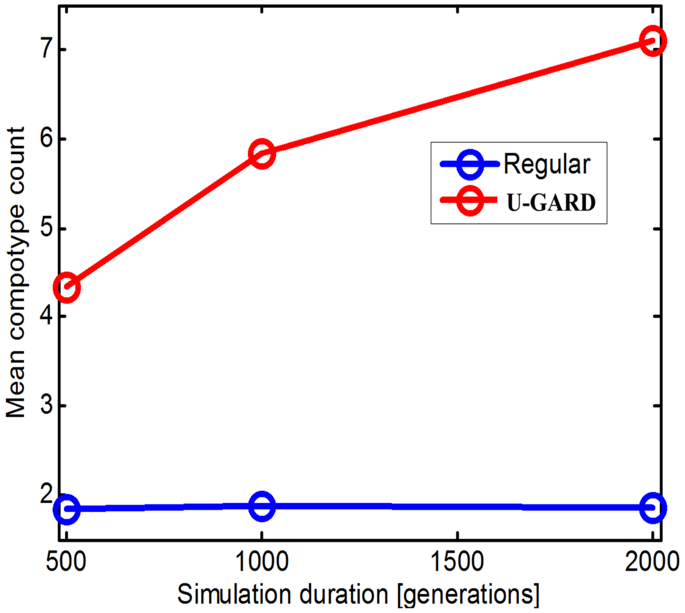

Figure 44: Average N_C value per simulation duration, for regular-GARD and U-GARD.

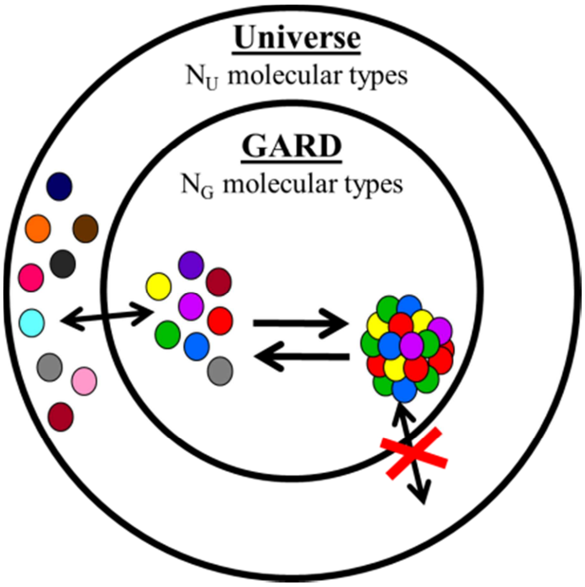

Figure 45: Open-ended GARD (Universe-GARD, U-GARD). Assemblies undergo growth-fission cycles in the immediate environment, obeying the GARD dynamics. At any given time step diffusion of molecular types in and out of the GARD environment occurs for molecules not included in compotypes.

5.6.2. Early results

Figure 46 shows examples of a preliminary implementation of the U-GARD model, which shows the emergence of new compotypes even in later stages of the simulation, unlike in regular-GARD (see Figure 4 for example). Indeed, when the mean N_C value is compared between U-GARD and regular-GARD, it becomes apparent that not only that the latter exhibits a higher average N_C value but also that this value tends to increase with simulation length (Figure 44). This suggests that GARD can be tweaked into showing open ended evolution.

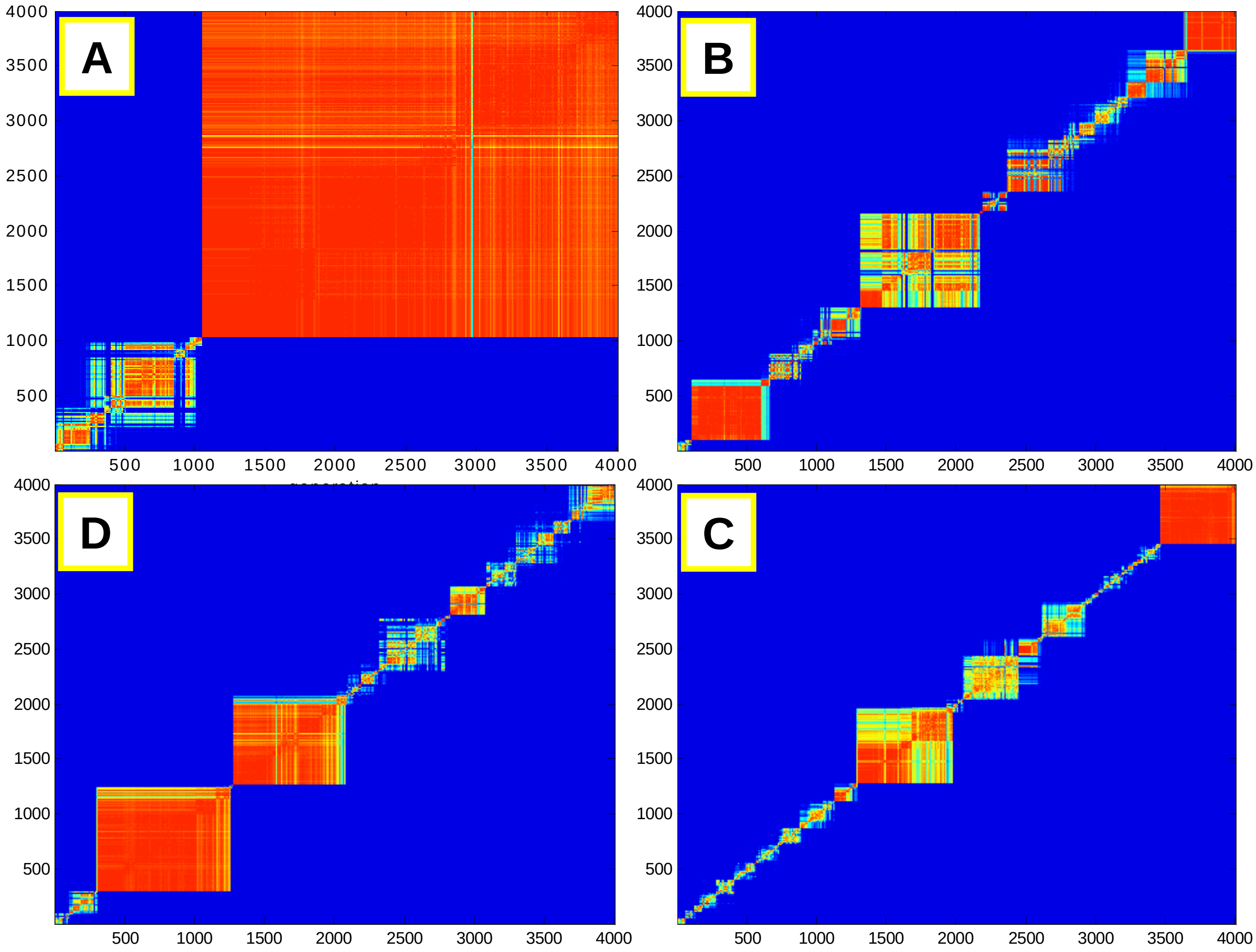

Figure 46: Examples of early U-GARD simulation with $N_U=\infty$. Diffusion/exchange rates are increasing from A to D.

6. DISCUSSION

6.1. Compotypes as quasispecies

One of the unique corollaries of the GARD model is the emergence of quasi-stationary compositional states – composomes, which embody both metabolism-like characteristics and a capacity to store and propagate molecular information. These composomes often interchangingly mutate towards a central point – compotype, which is an attractor in the compositional space. This is analogue to the quasispecies concept, where polymers mutate around a master entity in the sequential space – the master sequence. The good agreement between the steady state of the quasispecies equation to that of GARD supports the description of a cloud of assemblies around a compotype as a quasispecies, similarly to how the cloud of sequences around a master sequence is. Critically, this description does not hold for any assembly, only for compotypes. A point of difference between compositional- and sequential- quasispecies is between the master-sequence and the compotype. A compotype is represented by the center-of-mass of its member-assemblies and as such does not have to represent an assembly encountered during the simulation, while a master-sequence is the one with the highest fitness and therefore exist in a substantial proportion inside the cloud.

6.2. GARD is a minimally living system

The finding that GARD's species, i.e. compotypes, can exhibit Darwinian evolution has an important implication directly related to the origin of life and the definition of minimal life. Early on, the ability of GARD's assemblies to faithfully replicate was demonstrated [119]. GARD is now viewed as an autopoietic-chemoton system[5], where template copying and compartmentation are embodied in one entity, and a continuous supply of metabolites is afforded by the spontaneous accretion of lipids from the environment [80]. Thus, a compotype is self-sustaining, which is an important prerequisite for life [9]. The other prerequisite comes from the accepted definition of minimal life: "Life is a self-sustaining system capable of undergoing Darwinian evolution" [9]. Thus, showing that compotypes can indeed evolve is critical if one wants to consider GARD as a minimally living system. Further, demonstrating that the distribution of assemblies inside a compotype cloud agrees with the quasispecies equation strengthen this point, as the quasispecies model describes the process of evolution of replicating entities [30]. Thus, a non-biological system, i.e. devoid of information carrying polymers, can exhibit (minimal) life.

[5] There is disagreement with some other scholars, who points that Autopoiesis requires actual metabolic *production* in the sense of chemical covalent bonds modifications. However, I am treating production in the broader sense of the word.

It should be noted, that the span of evolution in GARD is likely to be limited as the effective number of compotype species is always below 8 hence the model does not portrays full-fledged open ended evolution (though the relative number of species with respect to the theoretical one is greater in GARD than in nature). This is probably because the underlying β network, which represents a chemical environment, is constant throughout each simulation. This is un-realistic as environments change over the course of evolution and in fact this change is intertwined with evolution. Addressing this point, the universe-GARD model has been formulated which changes the environment by systematically changing β. Indeed, preliminary results suggest the new model exhibits a larger number of compotype species and that this number is increasing with time, which hints for open ended evolution.

6.3.Evolution towards lower entropy of compotypes

Population ecology typically involves complex organisms, for which relating biochemical parameters to organismal behavior is extremely difficult. The GARD model, governed by mutually-catalytic networks, analyses supramolecular assemblies that are uniquely positioned at the interface between systems chemistry on the one hand and population dynamics on the other. This allows presenting a direct and quantitatively analyzable link between individual molecules and ecology.

The N_{mol} analysis might advocate a prebiotic scenario initiated by fast-replicating assemblies with a high molecular diversity, evolving into more faithful replicators with narrower molecular repertoires. This is not unlike the transition from prebiotic “random chemistry” to the relatively restricted repertoire of small molecules (monomers) seen in present-day living [119]. Such a transition might be considered as a change from a compotype with higher entropy into a lower one, as a composition with a lower N_{mol} has lower entropy (under a given N_G and N_{max}).

6.4.Lack of GARD experiments

While GARD is based on realistic physical and chemical considerations, an experiment demonstrating homeostatic growth and faithful inheritance of a vesicle composed out of several molecular types is lacking. Such experiment is likely to be affected by the choice of participating amphiphiles, as lipids tend to be selfish and often form homogenous vesicles. However, as was shown here, mutually interacting groups of lipids can show better behaviors in terms of selection response and growth-rate, which can give them an advantage over selfish lipids. Such an experiment would also require accurate compositional monitoring, not yet elaborated. A difficulty may arise from the size of vesicles. The minimal diameter of vesicles and micelles is in the order of dozens of nanometers and contains many hundreds of molecules, which could make it difficult to differentiate between assemblies of different compositions. A further difficulty is that GARD usually employs small assembly size and larger sizes result in diminishing compotype diversity. A point in favor of performing such experiments, is that recent studies of vesicles with multiple components have demonstrated that vesicles with different compositions show different distinct features such as permeability [82] (Figure 47) or boundary structure [144] (Figure 48).

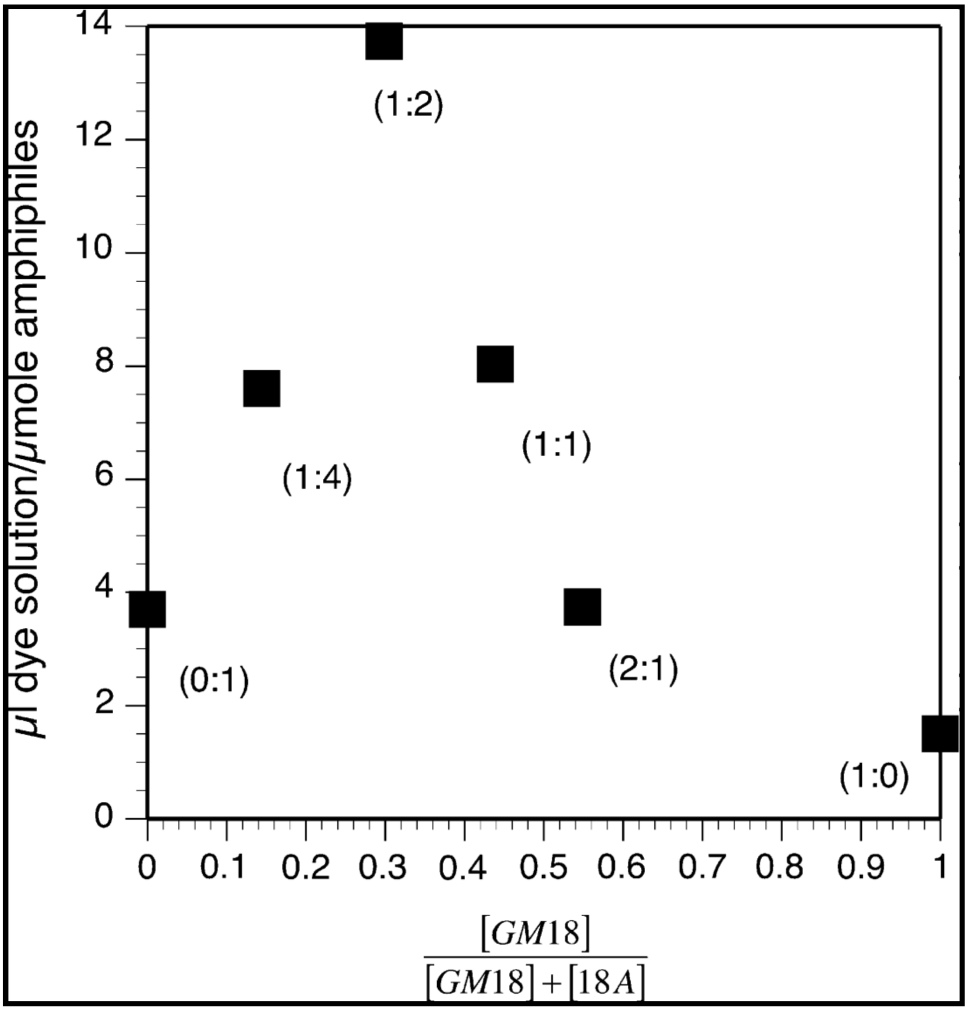

Figure 47: Effect of bilayer composition on the encapsulation of pyranine. A 1:2 ratio between amphiphiles shows the highest encapsulation efficiency, defined as the ratio of volume of dye solution to concentration of lipid. GM18 and 18A stands for glycerol monoacyl and lauric acid amphiphiles, both with 18 carbons. Figure taken from [82].

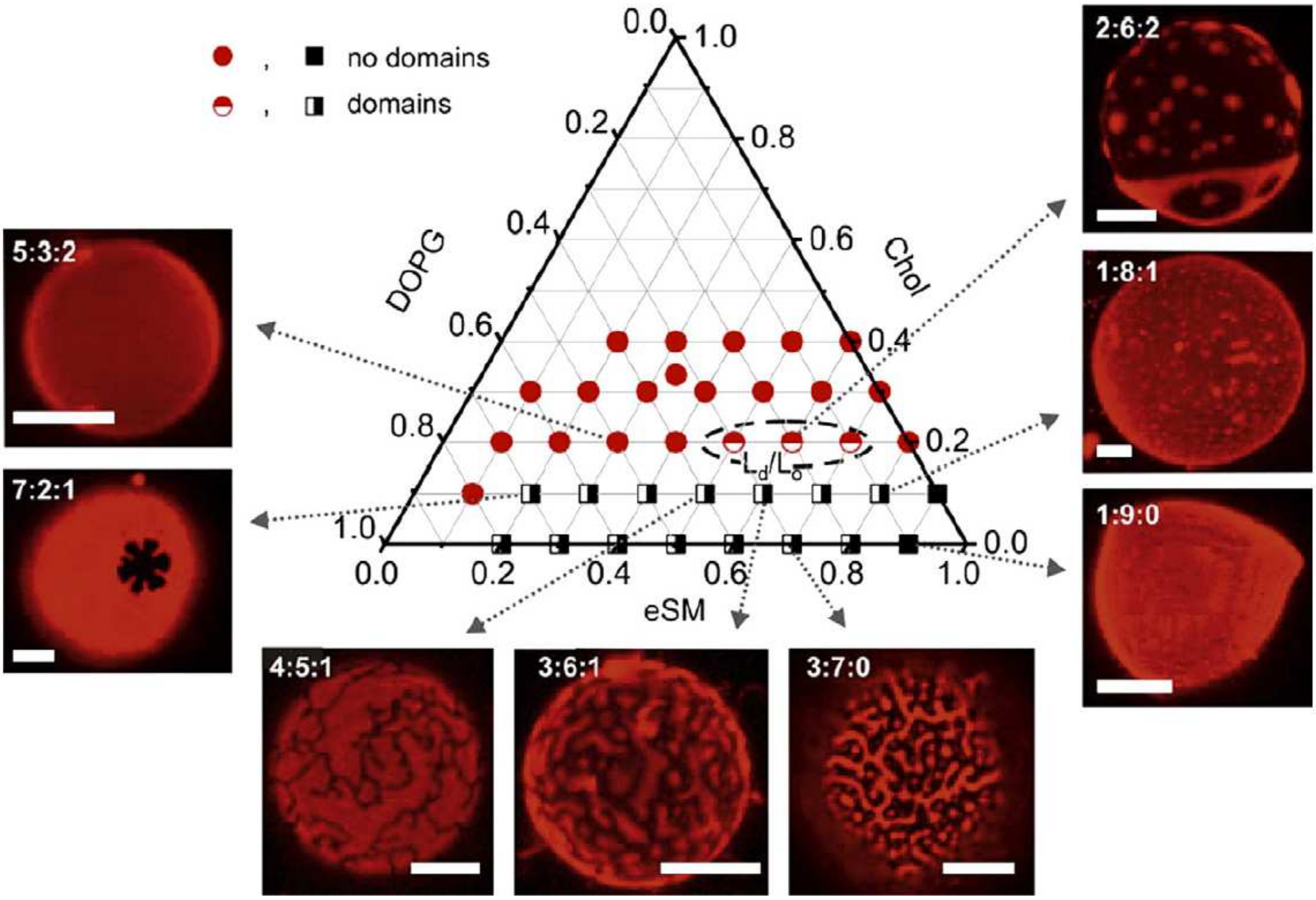

Figure 48: Typical confocal microscopy images of giant unilamerllar vesicles composed out of a mixture of dioleoylphosphatidylglycerol (DOPG), egg sphingomyelin (eSM) and cholesterol (Chol). Black scale bar correspond to 10 micrometer. Figure taken from [144].

7. REFERENCES

1. Alon, U., *Network motifs: theory and experimental approaches* Nature Reviews Genetics, 2007. **8**(6): p. 450-461.
2. Anastasi, C., F.F. Buchet, M.A. Crowe, M. Helliwell, J. Raftery, and J.D. Sutherland, *The search for a potentially prebiotic synthesis of nucleotides via arabinose-3-phosphate and its cyanamide derivative.* Chemistry, 2008. **14**(8): p. 2375-88.
3. Andes-Koback, M. and C.D. Keating, *Complete Budding and Asymmetric Division of Primitive Model Cells To Produce Daughter Vesicles with Different Interior and Membrane Compositions.* J Am Chem Soc, 2011. **133**(24): p. 9545-9555.
4. Anet, F.A., *The place of metabolism in the origin of life.* Current Opinion in Chemical Biology, 2004. **8**(6): p. 654-659.
5. Arenas, C.D. and N. Lehman, *Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube.* Bmc Evolutionary Biology, 2010. **10**(1): p. 80.
6. Armstrong, D.L., O. Markovitch, R. Zidovetzki, and D. Lancet, *Replication of simulated prebiotic amphiphile vesicles controlled by experimental lipid physicochemical properties.* Physical Biology, 2011. **8**(6).
7. Bailey, J.K., A.P. Hendry, M.T. Kinnison, D.M. Post, E.P. Palkovacs, F. Pelletier, L.J. Harmon, and J.A. Schweitzer, *From genes to ecosystems: an emerging synthesis of eco-evolutionary dynamics.* New Phytologist, 2009. **184**(4): p. 746-749.
8. Bedau, M.A., J.S. McCaskill, N.H. Packard, S. Rasmussen, C. Adami, D.G. Green, T. Ikegami, K. Kaneko, and T.S. Ray, *Open problems in artificial life.* Artificial Life, 2000. **6**(4): p. 363-376.
9. Benner, S.A., *Defining life.* Astrobiology, 2010. **10**(10): p. 1021-1030.
10. Berkowitz, M., *On the nature of lipid rafts: Insights from molecularly detailed simulations of model biological membranes containing mixtures of cholesterol and phospholipids.* Computational Modeling of Membrane Bilayers, 2008. **60**: p. 257-279.
11. Bernhardt, H.S., *The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others).* Biol Direct, 2012. **7**: p. 23.
12. Biebricher, C.K. and M. Eigen, *What is a quasispecies?* Curr Top Microbiol Immunol, 2006. **299**: p. 1-31.
13. Bonner, W.A., *The origin and amplification of biomolecular chirality.* Orig Life Evol Biosph, 1991. **21**(2): p. 59-111.
14. Bowler, F.R., C.K. Chan, C.D. Duffy, B. Gerland, S. Islam, M.W. Powner, J.D. Sutherland, and J. Xu, *Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation.* Nature Chemistry, 2013. **5**(5): p. 383-9.

15. Brasier, M.D., O.R. Green, A.P. Jephcoat, A.K. Kleppe, M.J. Van Kranendonk, J.F. Lindsay, A. Steele, and N.V. Grassineau, *Questioning the evidence for Earth's oldest fossils.* Nature, 2002. **416**(6876): p. 76-81.

16. Cech, T.R., A.J. Zaug, and P.J. Grabowski, *In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence.* Cell, 1981. **27**(3 Pt 2): p. 487-96.

17. Chen, I.A. and P. Walde, *From Self-Assembled Vesicles to Protocells.* Cold Spring Harbor Perspectives in Biology. **2**(7).

18. Chyba, C.F., P.J. Thomas, L. Brookshaw, and C. Sagan, *Cometary Delivery of Organic-Molecules to the Early Earth.* Science, 1990. **249**(4967): p. 366-373.

19. Cleaves, H.J., *The prebiotic geochemistry of formaldehyde.* Precambrian Research, 2008. **164**(3-4): p. 111-118.

20. Cleaves, H.J., *Prebiotic Chemistry: Geochemical Context and Reaction Screening.* Life, 2013. **3**(2): p. 331-345.

21. Cooper, G., C. Reed, D. Nguyen, M. Carter, and Y. Wang, *Detection and formation scenario of citric acid, pyruvic acid, and other possible metabolism precursors in carbonaceous meteorites.* Proc Natl Acad Sci U S A, 2011. **108**(34): p. 14015-14020.

22. Coppex, F., M. Droz, and A. Lipowski, *Lotka-Volterra model of macro-evolution on dynamical networks.* Computational Science - Iccs 2004, Proceedings, 2004. **3039**: p. 742-749.

23. Crotty, S., C.E. Cameron, and R. Andino, *RNA virus error catastrophe: Direct molecular test by using ribavirin.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6895-6900.

24. Deamer, D.W., *Boundary Structures and the Nonpolar Organic-Components of the Murchison Carbonaceous Chondrite.* Origins of Life and Evolution of the Biosphere, 1986. **16**(3-4): p. 363-364.

25. Domingo, E., *Quasispecies theory in virology.* Journal of Virology, 2002. **76**(1): p. 463-465.

26. Drake, M.J., *Origin of water in the terrestrial planets.* Meteoritics & Planetary Science, 2005. **40**(4): p. 519-527.

27. Dyson, F., *Origins of life*. 2nd ed. 1999, Cambridge: Cambridge University. 100.

28. Dyson, F., *A meeting with Enrico Fermi.* Nature, 2004. **427**(6972): p. 297.

29. Eigen, M., J. Mccaskill, and P. Schuster, *Molecular Quasi-Species.* Journal of Physical Chemistry, 1988. **92**(24): p. 6881-6891.

30. Eigen, M. and P. Schuster, *Hypercycle - Principle of Natural Self-Organization. A. Emergence of Hypercycle.* Naturwissenschaften, 1977. **64**(11): p. 541-565.

31. Eigen, M. and P. Schuster, *Stages of Emerging Life - 5 Principles of Early Organization.* J Mol Evol, 1982. **19**(1): p. 47-61.

32. Freilich, S., A. Kreimer, E. Borenstein, N. Yosef, R. Sharan, U. Gophna, and E. Ruppin, *Metabolic-network-driven analysis of bacterial ecological strategies.* Genome Biology, 2009. **10**(6): p. R61.

33. Futerman, A.H. and Y.A. Hannun, *The complex life of simple sphingolipids.* EMBO Rep, 2004. **5**(8): p. 777-82.

34. Gabriel, J.R., F. Saucy, and L.F. Bersier, *Paradoxes in the logistic equation?* Ecological Modelling, 2005. **185**(1): p. 147-151.

35. Gánti, T., *Organization of Chemical-Reactions into Dividing and Metabolizing Units - Chemotons.* Biosystems, 1975. **7**(1): p. 15-21.

36. Gesteland, F.R., R.T. Cech, and F.J. Atkins, *The RNA world.* 1999, Cold Spring: Cold Spring Harbor Laboratory. 709.

37. Gilbert, W., *Origin of Life - the RNA World.* Nature, 1986. **319**(6055): p. 618-618.

38. Gillespie, D.T., *General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions.* Journal of Computational Physics, 1976. **22**(4): p. 403-434.

39. Glansdorff, N., Y. Xu, and B. Labedan, *The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner.* Biol Direct, 2008. **3**.

40. Goto, K., M. Kinjo, K. Hashimoto, and M. Ishigami, *Synthesis of Hydrocarbons under Presumed Prebiotic Conditions Using High-Frequency Discharge.* J Mol Evol, 1986. **23**(2): p. 113-118.

41. Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman, *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**(3 Pt 2): p. 849-57.

42. Guzman, M.I. and S.T. Martin, *Prebiotic Metabolism: Production by Mineral Photoelectrochemistry of alpha-Ketocarboxylic Acids in the Reductive Tricarboxylic Acid Cycle.* Astrobiology, 2009. **9**(9): p. 833-842.

43. Hanczyc, M.M., S.M. Fujikawa, and J.W. Szostak, *Experimental models of primitive cellular compartments: encapsulation, growth, and division.* Science, 2003. **302**(5645): p. 618-22.

44. Hargreaves, W.R., S.J. Mulvihill, and D.W. Deamer, *Synthesis of Phospholipids and Membranes in Prebiotic Conditions.* Nature, 1977. **266**(5597): p. 78-80.

45. Hayden, E.J. and N. Lehman, *Self-assembly of a group I intron from inactive oligonucleotide fragments.* Chemistry & Biology, 2006. **13**(8): p. 909-918.

46. Heinemann, M. and U. Sauer, *Systems biology of microbial metabolism.* Curr Opin Microbiol, 2010. **13**(3): p. 337-43.

47. Higgs, P.G., *A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code.* Biol Direct, 2009. **4**: p. 16.

48. Holland, J.J.d., J. De La Torre, and D. Steinhauer, *RNA virus populations as quasispecies*, in *Genetic Diversity of RNA Viruses*. 1992, Springer. p. 1-20.

49. Hordijk, W., J. Hein, and M. Steel, *Autocatalytic Sets and the Origin of Life.* Entropy, 2010. **12**(7): p. 1733-1742.

50. Huang, W. and J.P. Ferris, *One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis.* J Am Chem Soc, 2006. **128**(27): p. 8914-9.

51. Ilardo, M.A. and S.J. Freeland, *Testing for adaptive signatures of amino acid alphabet evolution using chemistry space.* Journal of Systems Chemistry, 2014. **5**(1): p. 1-9.

52. Inger, A., A. Solomon, B. Shenhav, T. Olender, and D. Lancet, *Mutations and Lethality in Simulated Prebiotic Networks.* Journal of Molecular Evolution, 2009. **69**(5): p. 568-578.

53. Jenkins, G.M., M. Worobey, C.H. Woelk, and E.C. Holmes, *Evidence for the non-quasispecies evolution of RNA viruses.* Molecular biology and evolution, 2001. **18**(6): p. 987-994.

54. Johnson, A.P., H.J. Cleaves, J.P. Dworkin, D.P. Glavin, A. Lazcano, and J.L. Bada, *The Miller volcanic spark discharge experiment.* Science, 2008. **322**(5900): p. 404.

55. Joyce, G.F., *The antiquity of RNA-based evolution.* Nature, 2002. **418**(6894): p. 214-221.

56. Kafri, R., *Kinetic Enantio-selection by mutually catalytic networks within molecular assemblies - An Origin of Life Scenario*, in *Unpublished M.Sc. Thesis*2002, Weizmann Institute of Science: Rehovot.

57. Kafri, R. and D. Lancet, *Probability rule for chiral recognition.* Chirality, 2004. **16**(6): p. 369-378.

58. Kafri, R., O. Markovitch, and D. Lancet, *Spontaneous Chiral Symmetry Breaking in Early Molecular Networks.* 2010.

59. Kauffman, S.A., *Autocatalytic Sets of Proteins.* Journal of Theoretical Biology, 1986. **119**(1): p. 1-24.

60. Kauffman, S.A., *The origins of order: Self organization and selection in evolution*. 1993: Oxford University Press, USA.

61. Kiedrowski, G., S. Otto, and P. Herdewijn, *Welcome Home, Systems Chemists.* Journal of Systems Chemistry, 2010. **1**(1): p. 1.

62. Koonin, E.V., *Comparative genomics, minimal gene-sets and the last universal common ancestor.* Nat Rev Microbiol, 2003. **1**(2): p. 127-36.

63. Kopetzki, D. and M. Antonietti, *Hydrothermal formose reaction.* New Journal of Chemistry, 2011. **35**(9): p. 1787-1794.

64. Korade, Z. and A.K. Kenworthy, *Lipid rafts, cholesterol, and the brain.* Neuropharmacology, 2008. **55**(8): p. 1265-1273.

65. Korb, K.B. and A. Dorin, *Evolution unbound: releasing the arrow of complexity.* Biology & Philosophy, 2011. **26**(3): p. 317-338.

66. Kruger, K., P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech, *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**(1): p. 147-57.

67. Kun, A., M. Santos, and E. Szathmary, *Real ribozymes suggest a relaxed error threshold.* Nature Genetics, 2005. **37**(9): p. 1008-1011.

68. Kuno, E., *Competitive-Exclusion through Reproductive Interference.* Researches on Population Ecology, 1992. **34**(2): p. 275-284.

69. Lancet, D., O. Kedem, and Y. Pilpel, *Emergence of Order in Small Autocatalytic Sets Maintained Far from Equilibrium - Application of a Probabilistic Receptor Affinity Distribution (Rad) Model.* Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics, 1994. **98**(9): p. 1166-1169.

70. Lancet, D., E. Sadovsky, and E. Seidemann, *Probability Model for Molecular Recognition in Biological Receptor Repertoires - Significance to the Olfactory System.* Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3715-3719.

71. Lancet, D., A. Solomon, R. Kafri, and B. Shenhav, *The simplest cellular life.* Origins of Life and Evolution of the Biosphere, 2006. **36**(3): p. 213-214.

72. Lauring, A.S. and R. Andino, *Quasispecies Theory and the Behavior of RNA Viruses.* PLoS Pathog, 2010. **6**(7).

73. Lazcano, A., *Historical Development of Origins Research.* Cold Spring Harb Perspect Biol, 2010. **2**(11).

74. Lazcano, A. and J.L. Bada, *The 1953 Stanley L. Miller experiment: fifty years of prebiotic organic chemistry.* Orig Life Evol Biosph, 2003. **33**(3): p. 235-42.

75. Lecacheux, A., N. Biver, J. Crovisier, D. Bockelee-Morvan, et al., *Observations of water in comets with Odin.* Astronomy & Astrophysics, 2003. **402**(3): p. L55-L58.

76. Lincoln, T.A. and G.F. Joyce, *Self-Sustained Replication of an RNA Enzyme.* Science, 2009. **323**(5918): p. 1229-1232.

77. Lingwood, D. and K. Simons, *Lipid Rafts As a Membrane-Organizing Principle.* Science, 2010. **327**(5961): p. 46-50.

78. Lubeck, E. and L. Cai, *Single-cell systems biology by super-resolution imaging and combinatorial labeling.* Nature Methods, 2012. **9**(7): p. 743-U159.

79. Luisi, P.L., P. Walde, and T. Oberholzer, *Lipid vesicles as possible intermediates in the origin of life.* Current Opinion in Colloid & Interface Science, 1999. **4**(1): p. 33-39.

80. Markovitch, O. and D. Lancet, *Excess Mutual Catalysis Is Required for Effective Evolvability.* Artificial Life, 2012. **18**(3): p. 243-266.

81. Markovitch, O., D. Sorek, L.T. Lui, D. Lancet, and N. Krasnogor, *Is There an Optimal Level of Open-Endedness in Prebiotic Evolution?* Origins of Life and Evolution of Biospheres, 2012. **42**(5): p. 469-473.

82. Maurer, S.E., D.W. Deamer, J.M. Boncella, and P.A. Monnard, *Chemical Evolution of Amphiphiles: Glycerol Monoacyl Derivatives Stabilize Plausible Prebiotic Membranes.* Astrobiology, 2009. **9**(10): p. 979-987.

83. May, R.M., *Biological Populations with Nonoverlapping Generations - Stable Points, Stable Cycles, and Chaos.* Science, 1974. **186**(4164): p. 645-647.

84. McMullin, B., *John von Neumann and the evolutionary growth of complexity: Looking backwards, looking forwards ...* Artificial Life Vii, 2000: p. 467-476.

85. Mcshea, D.W., *Mechanisms of Large-Scale Evolutionary Trends.* Evolution, 1994. **48**(6): p. 1747-1763.

86. Miller, S.L., *A production of amino acids under possible primitive earth conditions.* Science, 1953. **117**(3046): p. 528-9.

87. Mills, D.R., F.R. Kramer, C. Dobkin, T. Nishihara, and S. Spiegelman, *Nucleotide-Sequence of Microvariant Rna - Another Small Replicating Molecule.* Proc Natl Acad Sci U S A, 1975. **72**(11): p. 4252-4256.

88. Mills, D.R., R.L. Peterson, and Spiegelm.S, *An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule.* Proc Natl Acad Sci U S A, 1967. **58**(1): p. 217-&.

89. Moran, P.A.P., *Random processes in genetics.* Mathematical Proceedings of the Cambridge Philosophical Society, 1958. **54**(01): p. 60-71.

90. Neef, A., A. Latorre, J. Pereto, F.J. Silva, M. Pignatelli, and A. Moya, *Genome Economization in the Endosymbiont of the Wood Roach Cryptocercus punctulatus Due to Drastic Loss of Amino Acid Synthesis Capabilities.* Genome Biology and Evolution, 2011. **3**: p. 1437-1448.

91. Noffke, N., D. Christian, D. Wacey, and R.M. Hazen, *Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the ca. 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia.* Astrobiology, 2013. **13**(12): p. 1103-1124.

92. Oparin, A.I., *Origin and evolution of metabolism.* Comp Biochem Physiol, 1962. **4**: p. 371-7.

93. Oparin, A.I., *Evolution of the concepts of the origin of life, 1924-1974.* Orig Life, 1976. **7**(1): p. 3-8.

94. Orgel, L.E., *Evolution of the genetic apparatus.* Journal of Molecular Biology, 1968. **38**(3): p. 381-93.

95. Orgel, L.E., *Molecular Replication.* Nature, 1992. **358**(6383): p. 203-209.

96. Orgel, L.E., *Prebiotic chemistry and the origin of the RNA world.* Critical Reviews in Biochemistry and Molecular Biology, 2004. **39**(2): p. 99-123.

97. Otto, S.P. and M.C. Whitlock, *The probability of fixation in populations of changing size.* Genetics, 1997. **146**(2): p. 723-33.

98. Pertea, M., *The human transcriptome: an unfinished story.* Genes, 2012. **3**(3): p. 344-360.

99. Pike, L.J., *The challenge of lipid rafts.* Journal of Lipid Research, 2009. **50**: p. S323-S328.

100. Pizzarello, S., S.K. Davidowski, G.P. Holland, and L.B. Williams, *Processing of meteoritic organic materials as a possible analog of early molecular evolution in planetary environments.* Proc Natl Acad Sci U S A, 2013. **110**(39): p. 15614-15619.

101. Pohorille, A. and D. Deamer, *Self-assembly and function of primitive cell membranes.* Research in Microbiology, 2009. **160**(7): p. 449-456.

102. Povolotskaya, I.S. and F.A. Kondrashov, *Sequence space and the ongoing expansion of the protein universe.* Nature, 2010. **465**(7300): p. 922-U7.

103. Powner, M.W., B. Gerland, and J.D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.* Nature, 2009. **459**(7244): p. 239-242.

104. Powner, M.W. and J.D. Sutherland, *Prebiotic chemistry: a new modus operandi.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2011. **366**(1580): p. 2870-2877.

105. Rasmussen, S., M. Bedau, L. Chen, D.W. Deamer, D.C. Krakauer, P.H. Norman, and S.F. Peter, *Protocells Bridging Nonliving and Living Matter*. 1st ed. 2009, London: MIT press. 684.

106. Rescigno, A., *Struggle for Life .I. 2 Species.* Bulletin of Mathematical Biophysics, 1967. **29**(2): p. 377-&.

107. Rescigno, A., *Struggle for Life .2. 3 Competitors.* Bulletin of Mathematical Biophysics, 1968. **30**(2): p. 291-&.

108. Ricardo, A., M.A. Carrigan, A.N. Olcott, and S.A. Benner, *Borate minerals stabilize ribose.* Science, 2004. **303**(5655): p. 196-196.

109. Ritson, D. and J.D. Sutherland, *Prebiotic synthesis of simple sugars by photoredox systems chemistry.* Nature Chemistry, 2012. **4**(11): p. 895-899.

110. Robertson, H.D., S. Altman, and J.D. Smith, *Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid presursor.* J Biol Chem, 1972. **247**(16): p. 5243-51.

111. Ruiz-Jarabo, C.M., A. Arias, C. Molina-Paris, C. Briones, E. Baranowski, C. Escarmis, and E. Domingo, *Duration and fitness dependence of quasispecies memory.* Journal of Molecular Biology, 2002. **315**(3): p. 285-296.

112. Ruiz-Mirazo, K., C. Briones, and A. de la Escosura, *Prebiotic Systems Chemistry: New Perspectives for the Origins of Life.* Chemical Reviews, 2014. **114**(1): p. 285-366.

113. Ruiz-Mirazo, K., J. Pereto, and A. Moreno, *A universal definition of life: Autonomy and open-ended evolution.* Origins of Life and Evolution of Biospheres, 2004. **34**(3): p. 323-346.

114. Saladino, R., J.R. Brucato, A. De Sio, G. Botta, E. Pace, and L. Gambicorti, *Photochemical Synthesis of Citric Acid Cycle Intermediates Based on Titanium Dioxide.* Astrobiology, 2011. **11**(8): p. 815-824.

115. Sanjuan, R., M.R. Nebot, N. Chirico, L.M. Mansky, and R. Belshaw, *Viral Mutation Rates.* Journal of Virology, 2010. **84**(19): p. 9733-9748.

116. Sauer, U., *Metabolic networks in motion: 13C-based flux analysis.* Mol Syst Biol, 2006. **2**: p. 62.

117. Schopf, J.W. and B.M. Packer, *Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia.* Science, 1987. **237**: p. 70-3.

118. Segre, D., D. Ben-Eli, D.W. Deamer, and D. Lancet, *The lipid world.* Origins of Life and Evolution of the Biosphere, 2001. **31**(1-2): p. 119-145.

119. Segre, D., D. Ben-Eli, and D. Lancet, *Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(8): p. 4112-4117.

120. Segre, D. and D. Lancet, *Composing life.* Embo Reports, 2000. **1**(3): p. 217-222.

121. Segre, D., B. Shenhav, R. Kafri, and D. Lancet, *The molecular roots of compositional inheritance.* Journal of Theoretical Biology, 2001. **213**(3): p. 481-491.

122. Shapiro, R., *Prebiotic Ribose Synthesis - a Critical Analysis.* Origins of Life and Evolution of the Biosphere, 1988. **18**(1-2): p. 71-85.

123. Shapiro, R., *Small molecule interactions were central to the origin of life.* Quarterly Review of Biology, 2006. **81**(2): p. 105-125.

124. Shenhav, B., A. Bar-Even, R. Kafri, and D. Lancet, *Polymer GARD: Computer simulation of covalent bond formation in reproducing molecular assemblies.* Origins of Life and Evolution of the Biosphere, 2005. **35**(2): p. 111-133.

125. Shenhav, B. and D. Lancet, *Prospects of a computational origin of life endeavor.* Origins of Life and Evolution of Biospheres, 2004. **34**(1-2): p. 181-194.

126. Shenhav, B., A. Oz, and D. Lancet, *Coevolution of compositional protocells and their environment.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2007. **362**(1486): p. 1813-1819.

127. Shenhav, B., D. Segre, and D. Lancet, *Mesobiotic emergence: Molecular and ensemble complexity in early evolution.* Advances in Complex Systems, 2003. **6**(1): p. 15-35.

128. Shenhav, B., A. Solomon, D. Lancet, and R. Kafri, *Early systems biology and prebiotic networks.* Transactions on Computational Systems Biology I, 2005: p. 14-27.

129. Sierra, S., M. Davila, P.R. Lowenstein, and E. Domingo, *Response of foot-and-mouth disease virus to increased mutagenesis: Influence of viral load and fitness in loss of infectivity.* Journal of Virology, 2000. **74**(18): p. 8316-8323.

130. Sole, R.V., S. Rasmussen, and M. Bedau, *Introduction. Artificial protocells.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2007. **362**(1486): p. 1725-1725.

131. Spiegelm.S, I. Haruna, I.B. Holland, Beaudrea.G, and D. Mills, *Synthesis of a Self-Propagating and Infectious Nucleic Acid with a Purified Enzyme.* Proc Natl Acad Sci U S A, 1965. **54**(3): p. 919-&.

132. Steinley, D., *K-means clustering: A half-century synthesis.* British Journal of Mathematical & Statistical Psychology, 2006. **59**: p. 1-34.

133. Strobeck, C., *N Species Competition.* Ecology, 1973. **54**(3): p. 650-654.

134. Summers, J. and S. Litwin, *Examining the theory of error catastrophe.* Journal of Virology, 2006. **80**(1): p. 20-26.

135. Szathmáry, E., M. Santos, and C. Fernando, *Evolutionary potential and requirements for minimal protocells.* Topics in Current Chemistry, 2005. **259**: p. 167-211.

136. Takeuchi, N., L. Salazar, A.M. Poole, and P. Hogeweg, *The evolution of strand preference in simulated RNA replicators with strand displacement: implications for the origin of transcription.* Biol Direct, 2008. **3**: p. 33.

137. Taylor, T.J., *From artificial evolution to artificial life*, in *Unpublished Ph.D. Thesis*1999, University of Edinburg: Edinburg.

138. Thompson, R.M., U. Brose, J.A. Dunne, R.O. Hall, et al., *Food webs: reconciling the structure and function of biodiversity.* Trends Ecol Evol, 2012. **27**(12): p. 689-697.

139. Trifonov, E.N., *Vocabulary of Definitions of Life Suggests a Definition.* Journal of Biomolecular Structure & Dynamics, 2011. **29**(2): p. 259-266.

140. Vaidya, N., M.L. Manapat, I.A. Chen, R. Xulvi-Brunet, E.J. Hayden, and N. Lehman, *Spontaneous network formation among cooperative RNA replicators.* Nature, 2012. **491**(7422): p. 72-77.

141. Vandermeer, J.H., *Interspecific Competition - New Approach to Classical Theory.* Science, 1975. **188**(4185): p. 253-255.

142. Varela, F.G., H.R. Maturana, and R. Uribe, *Autopoiesis: the organization of living systems, its characterization and a model.* Curr Mod Biol, 1974. **5**(4): p. 187-96.

143. Vasas, V., E. Szathmáry, and M. Santos, *Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(4): p. 1470-1475.

144. Vequi-Suplicy, C.C., K.A. Riske, R.L. Knorr, and R. Dimova, *Vesicles with charged domains.* Biochimica Et Biophysica Acta-Biomembranes, 2010. **1798**(7): p. 1338-1347.

145. Voet, D. and J. Voet, *Biochemistry*. 2nd ed. 1995, USA: John Wiley &ons.

146. Wilde, S.A., J.W. Valley, W.H. Peck, and C.M. Graham, *Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago.* Nature, 2001. **409**(6817): p. 175-178.

147. Wilke, C.O., *Quasispecies theory in the context of population genetics.* Bmc Evolutionary Biology, 2005. **5**.

148. Wills, Q.F., K.J. Livak, A.J. Tipping, T. Enver, A.J. Goldson, D.W. Sexton, and C. Holmes, *Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments.* Nat Biotechnol, 2013. **31**(8): p. 748-+.

149. Wurm, F.M., *CHO Quasispecies—Implications for Manufacturing Processes.* Processes, 2013. **1**(3): p. 296-311.

150. Zahnle, K., L. Schaefer, and B. Fegley, *Earth's earliest atmospheres.* Cold Spring Harb Perspect Biol, 2010. **2**(10): p. a004895.

8. PUBLICATIONS STEMMING FROM THIS THESIS

1. Markovitch, Inger, Shenhav and Lancet; Origins of Life and Evolution of Biospheres 40 (4-5), 484 (2010): Replication and Darwinian selection define life's origin.
2. Kafri, Markovitch and Lancet; Biology Direct 5:38 (2010): Spontaneous Symmetry Breaking in Early Molecular Networks.
   See also chapter 3.6 and [56].
3. Armgstrong, Markovitch, Zidovetzki and Lancet; Physical Biology 8, 066001 (2011): Replication of Simulated Prebiotic Amphphile Vesicles Controlled by Experimental Lipid Physiochemical Properties.
4. Markovitch and Lancet; Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems (ECAL11, 2011): Evolutionary Attributes of Simulated Prebiotic Metabolic Networks.
5. Markovitch and Lancet; Artificial Life 18, 3 (2012): Excess Mutual Catalysis Is Required for Effective Evolvability.
6. Markovitch, Sorek, Lui, Lancet and Krasnogor; Origins of Life and Evolution of Biospheres 42, 469 (2012): Is there an optimal level of open-endedness in prebiotic evolution?.
7. Markovitch and Lancet; Proceedings of the Twelfth European Conference on the Synthesis and Simulation of Living Systems (ECAL12, 2013): Prebiotic Evolution of Molecular Assemblies: From Molecules to Ecology.
8. Markovitch and Lancet; Journal of Theoretical Biology (2014): Multispecies population dynamics of prebiotic compositional assemblies.

9. DECLERATION

I declare that the thesis summarizes my independent research.

There were no collaborations related to the studies described in chapters 5.1, 5.2 and 5.4.

In the other chapters:

✓ Chapter 4.1.1: My role was in designing the motifs analyses.

✓ Chapter 5.3: My role was in comparing with the regular-GARD model.

✓ Chapter 5.5: My role was in conceiving, designing and closely supervising the research.

✓ Chapter 5.6: My role was in suggesting the question and building and simulating the new model.

10. FUNDING

This work was partly supported by EU-FP7 project MATCHIT and the Minerva Center for Life Under Extreme Planetary Conditions.

ווש, מתעופפת לה הסכין החדה,
המלבנים הלבנים מתעופפים,
תססס, קופץ לו השמן,
והנה הם עולים מהמעמקים,
הצ'יפסים הפריכים.

מוקדש, באהבה, לצ'יפסים בסן-מרטין. מי ייתן ותנצצו באור הניאון. לעד (עומר מרקוביץ', הצעת מחקר, 2010; Research proposal).

צהובה היא השמש,
בוהקת בשמיים ומעניקה חיים,
לפעמים משתקפת במים,
צהובה היא החמנייה, גדולה ועגלגלה,
אך לי שמור מקום בלב לצהובים אחרים,
קטנטנים ומרובעים.

מוקדש, באהבה, לשקדי מרק צהובים באשר הם. מי ייתן שתנצצו באורות ניאון באותו החן שבו צ'יפסים מצופי שומן נוצצים. לעד (עומר מרקוביץ', דו"ח ביניים, 2011; Interim report).

שלום ערן!, קוראים לי עומר ואני מסיים כעת את הדוקטורט שלי במכון וייצמן, בנושא ראשית החיים בכדור הארץ לפני כ4 מיליארד שנה. אני פונה אלייך בעקבות ביקורת שקראתי בעיתון הארץ על ספרך האחרון.

מידי שנה ושנה במהלך הדוקטורט, על כל תלמיד להגיש דו"ח המפרט את התקדמותו בשנה שחלפה. עקב ויכוחים שונים עם מנחה הדוקטורט שלי, פרופסור דורון לנצט, כחלק מהצעת המחקר שלי וכן דו"ח התקדמות, הקדשתי 2 שירים פרי עיטי לצ'פסים ולשקדי מרק וצירפתי אותם לדו"חות שהגשתי לוועדה.

כעת, כשעליי להגיש את תזת הדוקטורט שלי עד סוף שנת 2013, הייתי רוצה לבקש ממך לשקול לקרוא את שיריי ולהעביר עליהם ביקורת קצרה אותה אפרסם ביחד עם התיזה.
(כותרת התיזה היא: כימיה פרה-ביוטית כגישה לחקר ראשית החיים.)

אכבד כל החלטה שלך, ואני חושב שזה יהיה ממש מגניב אם תסכים לעשות כן (או לפחות להמליץ למי עליי לפנות).
תודה!. עומר מרקוביץ'.

"אל תאמרו לי שאני יודעת לבשל / המצרכים כבר נבראו / וכל שנותר לי הוא / לערבב ולחמם" (צאלה כץ)
"והנה הם עולים מהמעמקים, / הצ'יפסים הפריכים" (עומר מרקוביץ')

כבר ממבט ראשון קל לראות ששיריו של עומר מרקוביץ' מנסים להוות אלטרנטיבה, לאתגר תפיסה קיימת. אם התפיסה השלטת בנוגע למקור החיים הוא הורשה מבוססת-אלפבית של RNA, הרי שהיא מוטה לכיוון תרגום המידע הגלום בו לחומצות אמינו - כלומר לחלבונים. מרקוביץ' הופך את הקערה על-פיה, תהא זו קערת טוגנים או קערית מרק ובו שקדי מרק, למודל פואטי מבוסס פחמימות, תפוחי-אדמה וחיטה בהתאמה. "אך לי שמור מקום בלב לצהובים אחרים."

יכול אדם, אף כי אינו מוכרח לעשות כן, לשאול מה מחבר, יצירתית ומטבוליסטית, בין שני הרכיבים הפחממתיים; מה מדביק אותם יחדיו. והרי חלבון החיטה הוא הרי גלוטן, שמו קרוי על-שם דבק, וכך גם לגבי תפוחי-האדמה במידה פחותה, אז כיצד נדבק דבק לדבק? אני טוען, אם כי לא בלהט, שהטכניקה השירית של מרקוביץ' היא זו שמאפשרת את היחס בין הדבקים, יחס שהוא רפלקסיבי וטרנזיטיבי, אך לא בהכרח סימטרי. כל מרכיב מדביק את עצמו, ללא ספק, וגם אחד דבק בשני, אך אין בהכרח שצ'יפס הנוגעים בשקדי מרק יהיו לשקדי מרק הנוגעים בצ'יפס. ולפיכך, אין מדובר ביחס שקילות, ונדרש הדבק הפואטי, המאשאפ בלשון ימינו, שיאפשר את מודל הבריאה החלופי.

כאן עלינו להקדים ולשאול, אף כי אי אפשר, מפני שהטקסט שלעיל כבר נכתב, אז עלינו לאחר ולשאול, מי היה הראשון שהציע מודל הדבקה שכזה. אין צורך במחקר רב כדי להשיב שפבלו פיקאסו הוא האחד (וז'ורז' בראק הוא שניים), שהרי על שמם רשום המונח שדחק את הקלאסיציזם באמנות הצדה, לטובת המודרניזם, וזה שאיפשר למקוריות לנוח לצד היצירה הלא-מקורית, ואני מתכוון כמובן למונח: קולאז', שטבעו השניים בתחילת העשור השני של המאה ה-20. קולאז' פירושו - הדבק.

לפיכך, עד כה קשרנו בין המודל האמנותי האלטרנטיבי, שהניע את תחילת השירה המודרנית, לבין המודל הפרה-ביוטי המערכתי שהוביל לתחילת החיים, אך השאלה המהותית בשני המקרים היא שאלת גרעיניות המצרכים, או הצ'יפסים, או לחלופין, שאלת ההשתנות וההתפתחות. ביתר פירוט: האם המצרכים הגיעו בגודלם הטבעי, או שהם נוצרו במובן שאנו מכירים אותו מתוך תהליך הערבוב והחימום? האם צ'יפס הוא יותר כפיסי תפוחי אדמה, או יותר תוצר של טיגונם? האמנם ה RNA-הוא ראשון החומרים התורשתיים, או שמולקולות פשוטות יחסית קדמו לו בתפקיד זה?

מרקוביץ' רומז בשירתו לכיוון השני, לכיוון המצרכים הפשוטים, הבסיסיים, שהתהליך הביא אותם באופן שיטתי להקדים את ה .RNA-הבחירה בצ'יפס ובשקדי מרק אינה מקרית. מרקוביץ' מפנה זרקור לצ'יפסים של סן-מרטין. אותו חבל ארץ בצ'ילה, או שם של קפיטריה באוניברסיטה ישראלית, הוא למעשה שם של כל מקום. והחיבור לכל מקום הוא חיבורם של הצ'יפס ושקדי המרק. ולמעשה חיבור שכזה נוצר בכל מקום שמתרחשת עליו מלחמה. נודע סיפורה של סרייבו, בירת בוסניה (ויש שיאמרו והרצגובינה), שבה אזל מלאי תפוחי האדמה במהלך המצור במלחמה היוגוסלבית של שנות ה-90 של המאה העשרים. כדי להכין לילדים מלאי-החרדה מעט מזון נחמה, טיגנו להם צ'יפסים מבצק. ארנב

או ברווז, שואל ציור ניו-אייג' קלישאי, אך השאלה כאן היא האם "צ'יפס סרייבו" הוא צ'יפס או שקדי מרק. והרי אין סרייבו סרייבו בלבד, ואת מעדן הפסים המטוגן ניתן להשיג בדוכני פלאפל בפרוורי ערי שדה ישראליות.

ובעצם השאלה היא התשובה: כימיית המערכות, שדה שמבקש לחקור את היחסים שבין תגובות כימיות מרובות, ולתכלל את המידע כדי להבין כיצד הן מתפקדות יחד, היא הדבק, היא התהליך. ממש כמו שירה, שבה חומרי הגלם הרגשיים מוטחים על משטח השפה כדי ליצור שפה, אלפבית, יחידות חוזרות של צלילים או של בסיסים חנקניים. לפיכך שירתו של מרקוביץ' נוכחת ופועמת, כחלק אינהרנטי מן המחקר הכימי. התוצר הוא הנגלה, התהליך הוא הנחקר.

אני רואה ביצירתו של מרקוביץ' נדבך חשוב בקשר הקוולנטי שבין חקר ההברות לבין חקר הנוקלאוטידים, אך הייתי רוצה לסיים את הסקירה בחזון, נדבך נוסף של העניין, והוא פרויקט זינוטקסט של המשורר הקנדי כריסצ'ן בוק. בוק כתב שיר וקודד אותו כך שכל אות תתאים לשלשת DNA. ע"י הזרקת הגן X-P13 לבקטריה שידועה כעמידה לפיצוץ אטומי, נוצרה תגובה שבתורה נוצרה מולקולה חדשה, ובו שיר אחר שכתב בוק.... עד כה זינוטקסט עבד היטב בסימולציה, אבל נוצרו באגים בכתיבת השיר בנסיונות אמיתיים. ואולי זה מה שמשותף לשירת אמת ולמדע: הבאגים.

ערן הדס (Eran Hadas) / ספטמבר 2013