Title
Stochastic innovation: functional self-organization in simple systems

Author
Bradford, Justin Allen

Publication Date
2008-03-04

Peer reviewed|Thesis/dissertation

Stochastic innovation: functional self-organization in simple systems

by

Justin A. Bradford

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOPHYSICS

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Stochastic innovation: functional self-organization in simple systems

This thesis is dedicated to my family and my friends.

Acknowledgments

This dissertation would not exist without the support of many, remarkable people.

First, I want to thank Brendan and Sarah for their friendship, support, and encouragement. I hope they know how important their role is in the completion of this work and for my future endeavors. Second, I want to thank my mother and father for their limitless support and belief in me. They are fundamentally responsible for all the successes in my life.

To Ken, this dissertation could not exist without your guidance, patience, ideas, and direction. Thank you. And absolutely none of this would have been possible without Julie Ransom. She helped me solve so many problems, and whenever I lost my direction, she brought me back on course.

To Patsy Babbitt, for her enthusiasm for my work, contributions to its direction, and belief in my success. To Chris Voigt, for his support as my orals committee chair and his invaluable insight and ideas. To John Chodera and Yu Chen, for their encouragement, and providing me with role-models for success in graduate school. And to Vince Voelz and the entire Dill lab, for providing an environment to enjoy science and the day-to-day work underlying this dissertation.

And I want to thank Carol Morita, Colin Watanabe, Brad Nelson, Jennifer Lewis, and my brothers Corbin and Kevin for their boundless encouragement and utter determination to see me succeed. To Kelly and the Sham, for pushing me through to the finish, and, of course, to Skip and Oscar, for making me smile. And finally, to Liz Clarke, for her friendship and encouragement through some of the most difficult years.

Abstract

Stochastic innovation: functional self-organization in simple systems

by

Justin A. Bradford

Doctor of Philosophy in Biophysics

University of California, San Francisco

Professor Ken A. Dill, Chair

Our proposal is concerned with the organization of simple chemical catalysts into coordinated, complex catalytic systems. Theories regarding the origin of life must answer the question of how a pre-biotic, primordial soup of chemicals could organize into a structured system capable of supporting the origin of biological life. We are proposing a new conceptual principle to provide this organization.

Organization and complexity at a biological scale is driven by generally understood evolutionary principles. However, biological evolution requires a substantial infrastructure of chemical complexity: concrete information storage molecules capable of self-replication to provide a selective criteria. Information molecules describing a more stable, robust, or efficient method of replication become more common in succeeding generations. Furthermore, this process of development is intrinsically dependent on the environmental history it experiences. The information molecule becomes a reflection of the patterns of uncertainty and reliability inherent in the environment.

Other forms of addressing the problem of chemical organization have either taken the form of extremely simplified biological evolution ("polymer worlds") or innate chemical consequences of the form of self-organized criticality. We propose a novel, generalized concept: evolutionary organizing principles acting on simple chemical environments. These systems are simpler than, and proceed, any "genetic system" of evolutionary organization. Quite simply, we suggest a general evolutionary approach, devoid of the conventional biological infrastructure, and ask how simple chemistry and physics could accomplish this: a chemical evolution.

Our primary contribution is this idea and approaches to studying it. However, we

focus our research by proposing a single physically plausible model that serves as the basis of evolutionary action. While this may not be the model – or even one of many necessary models – underlying chemical evolution, we expect our work to provide a guide to further understanding this component of the origin of life.

Professor Ken A. Dill
Dissertation Committee Chair

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Summary

There are several examples of what might be called "stochastic innovation," whereby a biological, physical, or sociological system: (i) searches among viable options, then (ii) selects one or more of those options that is "best" by some metric, then (iii) locks in that selection for the future. In biology, the best-known example is Darwinian evolution[1], where variation is the term that describes the search step, and natural selection is the term that describes steps (ii) and (iii). Stochastic innovation occurs in learning and memory and neural development, where the correlated firings of neurons can lead to changes in synaptic strength[2–5], and to the development of the vascular system, where new blood vessels grow toward oxygen-deprived cells[6,7].

Stochastic innovation appears in other arenas, too. Human beings, businesses, and social organizations evolve through decision-making: they search among the options available to them, make self-serving choices, then remember and act on those decisions in the future. Social insects, like ants, search randomly for food, then lock in the discovery with chemical trails for the rest of the colony[8,9]. Computer models of artificial life and artificial societies show that stochastic innovation can meet apparent goals that were not programmed into them at earlier times[10,11]. The power of stochastic innovation is that it can lead to complex behaviors or organizational structures that are responsive to changes in the environment, even though such processes are unguided, unplanned, and stochastic.

The primary focus of this dissertation is whether stochastic innovation might also be achievable in chemistry and biochemistry. Can chemical and biochemical reactions be

chained together in complex and innovative ways, driven only by simple physicochemical search and selection processes? If so, it may be useful, not only as a tool in chemistry and biochemistry, but also for giving insights into the processes of chemical organization that may have occurred during pre-biotic evolution.

1.2 Motivation

While the search for a specific mechanism is important, this dissertation is motivated by more fundamental questions. "How did life begin?" is certainly a compelling topic, but it is ultimately only a sub-question of "How does a system of simple, non-intelligent components self-organize?"

1.2.1 The Origin of Life: A Rough Hypothesis

In the beginning, there were mere a handful of catalytic compounds[12–15]. Some simple organic and inorganic chemicals that shared four key traits: catalytic capability, an interdependence on substrate and products, the capability of association and complex formation, and interaction with some largely homogenous surfaces (such as minerals, clays, or lipid membranes).

From these starting conditions, a very primitive form of evolution began. The various fluctuations in the environment drove specific formation of compounds[16]. Sometimes, these compounds opened up novel catalytic pathways, and together they developed a meaningful metabolic ecology. From these humble beginnings, a simple, stable metabolic infrastructure emerged. The underpinnings were adaptive, a consequence of their environment, and they relentlessly built more elaborate structure[17]. There were three key consequences of this adaptive organization: RNA monomers formed and polymerization reactions for both RNA and amino acids formed[18].

Though random, some of these peptides and RNA polymers had simple structure and some of these had simple catalytic function[19–21]. In combination, they improved their local metabolic infrastructure: more RNA monomers, more amino acids, more robust catalysis, and novel catalytic pathways. These random peptides also provided useful structure and functional diversity for the catalytic properties of RNA[22]. Thus more, even random, peptides were useful, and a primitive RNA-based ribosome emerged. It generated more semi-structured peptides to facilitate RNA-enzyme function[23].

As this process occurred, a promising ecology of simple metabolic chemistry, supportive peptides, and catalytic RNA-enzymes were enclosed in some type of vesicle[12,24]. As the primitive ribosome became more refined, and the genetic code settled into place, modern biological evolution took hold. Increasingly sophisticated enzymes gradually displaced the earlier metabolic underpinnings and the last universal common ancestor (LUCA) had finally arrived[25].

1.2.2 Teleology

The natural world is responsible for many intricate, functional systems. It is not uncommon for observers to mistake a natural, functional system for one that required a pre-existing purpose or design for its origin. This is a teleological argument, and it conflates the existence of function by a system with the existence of purpose or design for that system's function. The canonical example is Paley's watchmaker analogy, from his "Natural Theology"[26]:

> This mechanism being observed (it requires indeed an examination of the instrument, and perhaps some previous knowledge of the subject, to perceive and understand it; but being once, as we have said, observed and understood), the inference, we think, is inevitable, that the watch must have had a maker: that there must have existed, at some time, and at some place or other, an artificer or artificers who formed it for the purpose which we find it actually to answer; who comprehended its construction, and designed its use.

However, Paley was bested by nearly two millenia is his watchmaker analogy. In the first century BCE, Cicero noted in his "De natura deorum"[27]:

> When you see a sundial or a water-clock, you see that it tells the time by design and not by chance. How then can you imagine that the universe as a whole is devoid of purpose and intelligence, when it embraces everything, including these artifacts themselves and their artificers?

But this argument, in all its variations, is based on a fallacy: while the things we make have a function to serve a purpose, not all things with functions were made for a purpose. To build an arch, you need scaffolding to support the intermediate structure. Once the keystone is in place, the scaffold is torn down. Nature makes stone arches, too. The current of a creek running through crevices in a pile of stone washes the "scaffold" away, leaving a standing arch of stone[28]. A creek may erode a pile of rock to leave an arch, but it was not the purpose of the creek to make an arch.

Purpose is in the eye of the beholder, and we see purpose in the useful "side-effects" of natural processes. The cell "sees" purpose in the metabolic chemistry it depends upon, but the enzymes in that cycle have no notion of that purpose. They simply do what they do. That natural arch "sees" purpose in the creek's erosion, but the creek's only concern is obeying gravity.

1.2.3 Self interest

Scientists often speak of the metaphorical "self-interest" of biological systems. Of course, the components of such systems have no notion of combined self-interest – they merely have a function which they perform. In the appropriate "context", such as a cell in an organism or an enzyme in cell, the component's function can contribute to its self-replication.

Likewise, there is no teleological purpose for a cell to be self-replicating, but rather, self-replicating systems are simply self-perpetuating. Self-replication is a self-reinforcing property. A form of positive feedback, the process fundamentally changes the character of the system. A self-reinforcing (or feedback-driven) process is the type of process that drives its "signal" above the environment's "noise"[29,30]. Cells do not have an explicit impetus "to go forth and replicate". Any given adaptation on a cell's processes – affecting its growth or replication – is not better or more purposeful if improving the ability to self-replicate. Simply those that do keep the "signal" from being lost in the "noise".

"Survival of the fittest" does appear to be a tautology, but the fittest (those that survive) share a fundamental property – an improved ability to grow and replicate – which ultimately does change the nature of the system. All natural mechanisms, from the perspective of its constituents, are tautological. Crystals are the things that crystallize; successful companies succeed. The question is not what, but why? Why does a crystal crystallize? Why do successful companies succeed? And why do the fittest survive?

However, self-replication, especially with the efficiency and precision of a biological system, is difficult[25]. The "context" necessary for a mere functional component to contribute to its self-replication is dauntingly complex. Even simpler self-replicating chemical systems, such as RNA polymers, must rely on a substantial pre-existing "context": such systems depend on an abundant and homogeneous supply of fairly complex molecules[31,32].

The resort to simpler replicators as infrastructure for more complex replicators

faces a recursive problem. With increasingly simpler "contextual" requirements for replication, the system would likely become less catalytically selective and efficient – thus increasingly less likely to independently provide the basis for more complex replicators[33]. It becomes increasingly probable that the "signal" of these simple replicators would be simply drowned out by the "noise" of side-reactions and other disruptions. There are jarring conceptual issues with the idea of a "DNA World" preceded by a "RNA World" preceded by some other replicator, and so on. The origin of life is probably not, exclusively, "replicators all the way down".

But what are our options? Given the extraordinarily short geological time-frame for the origin of the "LUCA population"[34,35] there must have been organizing processes at work. Biological organization has reasonable grounding on the known natural mechanisms of self-reinforcement through self-replication. But self-replication is not the only known natural mechanism of self-reinforcement. Clearly the process of crystallization is equivalent in the sense of self-reinforcement[36]. But more interestingly, is self-reproduction the only process of *adaptive* self-reinforcement at work in the natural world?

Returning to the notion of "self-interest" and, specifically, the non-intelligent functional components of a system, what is the "self-interest" of an enzyme? It only has one fundamental property: its function. So how could it act in its "self-interest"? Ultimately, it must act to increase its ability to function — it must act to increase its own productivity. If some functional component, an agent, required some specific resource for it to function, it would exist in one of three states:

1. no supply of the resource
2. numerous, diffuse, non-specific supplies of the resource
3. specific supplies (such as another agent) of the resource

The fundamental hypothesis behind this dissertation is simply: if an agent routinely depends on specific supplies (another agent) for its resource, it will develop long-term associations with that supply (that agent). And now that this agent has a more stable supply of its resources, it is now a more stable supplier of resources. With its more productive function, it is now a consistent supplier for some third agent. And with the incorporation of the third agent, it becomes the target for a fourth, and so on. In the end, a system organizes "around" – it builds structure "on" – specific functional relationships between agents.

Of course, some specific mechanism must exist for this process to occur. However,

the same is true for self-replication. The specific mechanism of biological replication arose through some other process, and that process was likely far less complex. It is likely that the mechanism underlying "functional association" will also be generally simpler and more probable. And if true, such mechanisms are not only applicable to chemistry, or even biochemistry, but it is likely such mechanisms are ubiquitous throughout the natural world.

1.2.4 A specific model

And while the broader questions are compelling, they all reduce to a specific question: how can a system of agents with no intelligence, with no inherit memory, and only the ability to statically function in reaction to the local environment, spontaneously self-organize into a system with these emergent properties? In other words, how can simple catalytic agents come to form a system with memory and adaptive behavior? How can simple molecules self-construct a metabolic system?

What chemical processes might have set the stage for Darwinian biological evolution? Even if the earliest stages of biology involved a "genetic world" for primitive transmission of genetic information from one generation to the next, it must have been preceded by an even earlier world of pre-biotic chemistry, involving, at the least, monomer synthesis, energy transduction in order to run that machinery, and the encoding of some set of chemical reactions that were worth propagating forward via the genetic mechanism.

This dissertation proposes a simple chemical mechanism, based on known physical principles and random processes, and testable by experiments, by which simple molecules could form increasingly complex forms of organization, possibly of the type that presaged the earliest biological evolution.

1.3 Overview of thesis

1.3.1 Related Work

The following chapter discusses existing work that is related and often foundational to this dissertation's core questions on the topic of organization.

1.3.2 Published Work

Our most compelling model was expressed as a specific interaction with an lattice surface. This work was published in the Proceedings of the National Academy of Science in June of 2007. This chapter is a slightly modified version of the paper and supplementary material published there.

1.3.3 Preliminary Work

A significant challenge of this work was to find a model that sufficiently embodied the basic questions. There were many, varied instances of precursor approaches that lead to the the core work presented. This initial work certainly could form the basis of future productive research.

1.3.4 Future Directions & Conclusions

The core model can be extended and generalized in a number of promising directions. Regardless, the core model encapsulates an example of a potentially fundamental mechanism of organization.

Chapter 2

Related Work

2.1 Similar Fields

Although this dissertation outlines a novel approach and focus on a critical subject, there are a number of fields of work related to our interest. Furthermore, this dissertation draws many ideas and concepts from these related works.

2.1.1 Game theory

Much theoretical work in economics and biology uses models based on iterated game theory, focusing on the issues of cooperation in groups of agents interacting over many successive rounds[37]. Related to this work, it is a commonly used model for studying self-organization of agents, such as the origin of cooperative behavior in the *Iterated Prisoner's Dilemma*[38]. However, a major shortcoming of this work, from our model's perspective, is the use of "intelligent" agents. Most game theoretic models employ agents which are capable of rational decision making, adapting their strategies to the current environment and in response to retained knowledge of their past behavior and that of other agents in the system.

The sub-field of evolutionary game theory improves this somewhat, as agents randomly choose pre-defined strategies and gradually move to historically more successful, neighboring strategies[39]. Still, these models frequently use strategies requiring rational agents employing previously learned information, such as the "tit-for-tat" behavior[38,40]. A noteworthy exception, however, are models using "zero-knowledge" agents. These agents

are simply successful or not and unaware as to why. They have a simple parameter which determines their behavior, and when selective pressure is applied, the system of agents eventually converges on the globally optimal, collective behavior[41].

Though interesting, this highlights a second limitation in this general class of research: it is limited to studying to the problem of "cheating" in cooperating groups. The only issue for the system to resolve is that of "convincing" all of the agents to work in the best interests of the group. In real systems, this "cheating" problem is likely only a sub-class of a broad range of potential inefficiencies and perturbations, most of which are external to the group of agents and inherently unavoidable. Although some models introduce noise or changing environments[39,42] the concept of a global optimum or limit behavior is relatively meaningless in an open-ended, evolving system. Our interest is how a subset of the agents can cooperate to be successful despite the action of other agents or the presence of external perturbations. And of specific interest is how these cooperative agent groups form and grow in response to continuous, and changing, perturbations.

2.1.2 Networks

A second field with major relevance is the study of networks. This area has received a great deal of attention in recent years, and a number of interesting properties have been observed. For example, a wide range of natural networks, such as gene regulation, protein interaction, neurons, and even computer networks have been found to share potentially significant structure at a "macroscopic" level[43–45]. They are generally all "small-world" (a short average path length between nodes) and "scale-free" (their node connectivities follow a power law distribution). This structure appears to provide stability and robustness for the system.

The origin of such properties is not as clear. The idea of "preferential attachment" during growth of the network has been proposed, where the probability of a new node being connected to an existing node is proportional to the number of edges already connected to the existing node[46]. This alone generates the "scale-free" and "small-world" properties of the network. Furthermore, a refinement of that idea, incorporating the additional concept of a node's fitness for being linked, helps to explain how such networks could radically change in structure over time. With merely "preferential attachment", old nodes will be more highly connected, but the addition of a new node with a higher fitness would allow it

to overcome its age disadvantage[45].

A significant criticism, however, is that this theory lacks any functional consideration. These networks do something, and the nature of their growth must take that into account. The connectivity of a newly added node should also contribute to the functioning of the network[47]. On the other hand, one can conjecture a functional analog of "preferential attachment" for many of the specific real networks. For example, a new page added to the World Wide Web is more likely to link to popular and well-linked existing sites. Also, the growth of a biochemical network likely involves a great deal of gene duplication and reuse of modular protein sub-units, so new proteins are more likely to interact with compounds that interact with many other proteins[46].

While specific, functional variations on the "prefential attachment" theme may be plausible, other research indicates that these variations have a major, and unique, impact on the resulting organization. The analysis of "motifs" in real networks revealed "microscopic" structure. A network's "motifs" are simply the small (3-4 node) subgraphs of which it is composed. Researchers found that the distribution of such motifs in real networks is distinct from that of random networks and, more importantly, random scale-free networks[48,49]. The common "motifs" even appear to vary between functional classes of networks. For example, information processing networks (such as those involving neurons or gene regulation) frequently have feed-forward loops, while ecological food webs have chains and bi-parallel motifs[48]. These results imply distinct, functionally constrained development, despite retaining similar high level properties. So, at the very least, it is clear that the generalized concept of preferential attachment is not sufficient to fully understand real networks. A better understanding of the specific mechanisms of their growth and development is necessary.

Another area of research, studying boolean networks, suggests the average node connectivity may reflect on a network's balance between stability and plasticity[50]. Overly "stable" systems would be unable to evolve in response to a new perturbation, but if excessively "plastic", the system would rapidly change, and lose its ability to function, when perturbed even slightly. The results from these networks show that a low average node connectivity results in simple, deterministic behavior, while a high connectivity gives chaotic behavior. In between, with a connectivity averaging 2-3 edges per node, a mixed, "edge of chaos" behavior provides a balance[51]. This is conjectured to be the point at which life and other complex systems exist. While this dissertation will not attempt to directly address

the problem of complexity, the nature of an agent group's attractor (stable, chaotic, or "complex") throughout its development might be an interesting property to consider.

2.1.3 System control theory: highly optimized tolerance

Another line of theoretical work, the idea "highly optimized tolerance" , proposes an alternate explanation for power laws and "scale-free" properties[52]. Furthermore, it is a counter to the significance of concepts such as the "edge of chaos" and "self-organized criticality" in biological systems[53]. By applying ideas from engineering and process control, the theory posits that the origin of complexity in evolved and designed systems is due to the accumulation of "control systems". With uncertainty in the system, controls (such as feedback loops) are necessary to compensate for variations, providing essential robustness[52,54].

Of particular interest to this proposal, a fundamental component of this theory is that functional (or domain specific) variations are critical to the result. The modular components of a biological system are highly "self-dissimilar", in that, for example, the mitochondria is radically different than the endoplasmic reticulum. While they may have some statistical similarities, it is unlikely that significant aspects of their structure and organization are a consequence of a development process completely independent of their function[54].

Concepts like "preferential attachment" and the "edge of chaos" rely on innate, unspecific, self-organizing principles which generate structure and complexity that is robust independently of the particular function of the network. Their resulting structure is "self-similiar", solving the problem of robustness with general structural concepts unrelated to function[55,56]. "Highly optimized tolerance" argues that structure and complexity is the consequence of the system evolving to provide robustness to the specific perturbations and uncertainty it experiences[52,57].

The end result, while retaining many of the "macroscopic" statistical properties, has important differences. A system evolving according to "optimized tolerance" becomes robust to the specific perturbations it has experienced, but not to potential perturbations is has yet to encounter. In fact, optimizing to minimize the impact of a particular uncertainty can make the system less robust to other perturbations[54]. The system becomes "robust-yet-fragile". The "edge of chaos" and "self-organized criticality" theories, lacking in functional

specifics, develop a system robust to general classes of uncertainty, rather than the specific ones it has encountered[55,56].

Our model shares much of the perspective implicit in the "highly optimized tolerance" theory of complexity. Our agents organize in response to specific fluctuations they encounter in the environment. They deal with the problem of robustness on a case-by-case basis, rather than as a general goal. Understanding the development and properties of such organization should provide a better understanding of real networks, something that ideas like "preferential attachment" and the "edge of chaos" alone cannot provide. Moreover, it may provide insight into the problem of increased fragility to other perturbations in exchange for robustness to some particular perturbation.

2.2 Chemical dynamics

An interesting consequence of refining the models of this dissertation, is that they come to resemble known, and studied, chemical dynamics approaches[58,59]. It does not appear that our model reduces to any specific, categorized chemical dynamical system, but a better understanding of its relation to established fields is useful.

The intent is to describe a system which forms structure around functionally coupled components. Furthermore, the environment (external fluctuation) has a strong effect on the availability and sustainability of the resulting structure (through functional properties of the components). As a chemical reaction model, one would expect this system to be unique and fairly complex. However, this approach does obviously build upon the ideas derived from the study of existing simple systems and general models.

2.2.1 BZ Reaction / Bulk Chemistry

The obvious starting point is the BZ reaction[29,59,60]. Capable of generating temporal oscillations, the system could potentially be modified and extended to generate more complex, environmentally dependent reaction processes. These potentially oscillating systems have fixed-point attractor states which bifurcate in certain parameter ranges. In the fixed point regime, these systems settle into simple steady states where reactant concentrations no longer vary with time. With the appropriate parameters, these fixed points bifurcate to give rise to a closed loop limit cycle. The system can no longer settle into a

steady state, and instead, continuously loops through the reaction phase space, forming an oscillating reaction.

It is possible to describe a basic version of organization with a few simple bulk chemistry reactions. However, adaptation would require reaction rates to be dependent on the specific state of the environment. For example:

$$\begin{aligned} A + 1 \leftrightarrow A1 &\rightarrow A2 \rightarrow A + 2 \\ B + 2 \leftrightarrow B2 &\rightarrow B3 \rightarrow B + 3 \\ B + A2 \leftrightarrow A2B &\rightarrow C + 3 \end{aligned}$$

A binds 1 to form the $A1$ intermediate; this intermediate is converted to the $A2$ intermediate, which then dissociates into A and 2. A similar process occurs for B, converting 2 to 3. However, B can also bind to the $A2$ intermediate, resulting in a new complex C. This sufficiently describes our desired system: organization (formation of C) is dependent on function (1 drives formation of $A2$, which is an input for B) and is shielded by the environment (free or solitary 2). Ultimately, this is simply a chemical reaction with competitive inhibition.

This basic process could be expanded with positive and negative feedback loops. For example, the complex C could be autocatalytic, accelerating the rate of the $B + A2 \rightarrow C$ reaction. Or the complex could somehow sequester free 2 or otherwise inhibit the free 2 inhibition of complex formation. Ultimately, this is likely not the most productive approach to pursue. It is difficult to imagine how this process alone could be a significant mechanism for complex, multi-component organization.

2.2.2 Reaction-diffusion systems

Complex organization will likely be most pronounced in a non-continuum setting, and reaction-diffusion model approaches are the most promising. Here the components of the system can vary in time and space, potentially providing the infrastructure for more elaborate self-organization.

Percolation

There are a set of basic model classes based on the idea of a connected network of nodes. Each node has some state, which changes as a consequence of the states of neighboring/adjacent nodes. Also, there may be some non-local contribution to a node state change (such as a global field affecting the network). This set can generally be considered a "percolating" system[58].

In percolation, some node "activity" propagates to neighboring nodes[61]. A simple example is a "forest fire" model. Nodes are either empty or contain a tree. Some spark on a tree node causes it to set fire (become active). The fire can then spread to neighboring nodes containing trees. The activity (fire) percolates through the system, following certain pathways (adjacent trees).

A system can exhibit directed percolation, where activity propagation is asymmetric in some dimensions. Frequently, this is simply the consequence of a time dimension – the forest fire cannot propagate backwards through time. However, it can be a more complex form of directed propagation. With the forest fire model, one can imagine an "easterly wind", where the fire is more likely to spread to trees to the left than to the right. Also, there are dynamical percolation classes, where nodes have some memory of their state. The fire is less likely to return to previously "burned" nodes.

"Voter model" classes are similar. In these systems, a node simply adopts the state of a neighboring node[62]. This class in particular has well-defined "adsorbing" (or static) states, where all nodes have the same state and the system no longer has any dynamic properties. They can also develop "interfaces". Large continuous sections of this system might all be of the same state, and the dynamic interactions only occur at nodes where these homogenous sections meet.

The role of global fields has been explored with simple Ising models of magnetism. In an Ising model, nodes have two possible states (or "spins"): up or down. In addition to local, neighboring interactions, which cause spin-flips or spin-exchanges, a global "magnetic field" might bias the probability of a node changing its spin in a certain direction.

Multi-component models

While useful, these percolation classes have some limitations and are ill-suited for describing a complicated, multi-component chemical reaction process. An alternative frame-

work is to consider diffusing components or particles (components undergoing a random walk). These components might also interact. The simplest example is a single-component annihilating random walk (ARW)[63]. A number of A particles randomly diffuse. Adjacent A particles react to form some inert particle, which does not contribute to the dynamics of the system. Effectively, adjacent A particles annihilate each other.

This can be expanded with branching: with some probability, adjacent A particles might generate a third A, rather than annihilating[63]. The particular number of adjacent particles necessary to branch or annihilate (and the specific number of particles that are created or annihilated) can be adjusted. Further, this can be expanded to multi-component systems, where some combination of adjacent A and B particles annihilate or branch, and so on. At certain dimensionalities and certain branching/annihilating rules, these processes can be reduced to simpler single-component random walks or even simple percolation systems.

With these multi-component models, it is sometimes possible to generate "blockades". Potentially reactive particles are separated by some non-reactive particles. This is generally a phenomenon seen at low dimensionality, as with an increasing number of dimensions it becomes possible for reactive particles to circumvent blockades.

2.2.3 Surfaces & lattices

One of the major characteristics of our model is that agent interactions are mediated by some coupled, functionally related, resource. Simply, an agent A is likely to associate with an agent B because A creates 2 which B uses. For our model, an obvious desire is for B agents to become spatially localized with 2 resources, which are likely to be near A agents. One potential solution is to assume adsorption to some catalytic surface. The presence of 2 resources facilitates a B agents adsorption to the surface.

There are example systems, though largely in reverse, of such lattice models. For example, there are interacting monomer-monomer systems on catalytic surfaces[64,65]. An A can bind to a free lattice site. Adjacent As on the lattice have a repulsive interaction, and vice versa for B binding. In these models, adjacent ABs react to annihilate, freeing the lattice positions. Like the Voter Model mentioned previously, this approach can generate reactive interfaces separating regions of similar (non-reactive) components[58].

Diffusive or frozen

One consideration with models involving adsorption to a surface is movement on the lattice. In "diffusive" models, the adsorbed particles can move, while in "frozen" models, they cannot[58]. In our model, we assume that agents are recruited to the surface by resources, which are likely to be near their generating agent. As our ultimate interest is in the interaction between agents, a "diffusive" model is probably the best approach. However, we would expect that adsorbed agents and resources have greatly reduced mobility, relative to those not associated with the surface.

2.2.4 Aggregation / Deposition

Another class of reaction models to consider are those involving aggregation and deposition. For the former, the models involve the aggregation of particles into blocks, which can then move as one, or split into fragments, and so on. For the latter, particles can deposit onto some surface forming layers. The key aspects of these groups are the conditions/rules upon which aggregations split or deposited particles dissociate. In the specific case of deposition, particles (or groups of particles) can generally only dissociate from the topmost layer and only from "terrace" edges (particles having an empty adjacent lattice node)[58].

However, it is not clear how our mechanism would be described in such a model. A key component of our idea is the notion of a diverse range of functional components capable of forming novel structures as a result. The aggregation/deposition approach would seem to be incompatible with these requirements, but it still might provide a useful source of ideas as we continue to explore the possible implementations of our system.

2.3 Our model in relationship to this related work

A general outline of a potential model: An agent A on the surface generates a constant supply of resource 2. The resources slowly diffuse on the lattice. An agent B is recruited to the surface by one or more of the resources. The agent diffuses on the lattice and is likely to encounter to A agent, upon which the two react to form some complex agent C. However, in the presence of an abundant external supply of resource 2, there will be adsorbed resources distributed across the surface, rather than simply in the vicinity of the

agent As. In this case, the B agents will spend a larger fraction of their time exploring local regions of the surface devoid of A agents.

Although the model above is a sufficient starting point, there are many additional variations and additions to consider. First, we will likely want to consider the disassociation of the C complex back into its component agents (or possibly into other novel agents). Second, we might consider a process by which associated agents are removed from the lattice due to abundant unassociated free resources. In this case, agents could "get stuck" in a very local, unproductive region on the lattice and need a "free" period to recover and explore spatially and find new pockets of associable agents for complex formation. This could give rise to a proper "Brownian ratchet" mechanism, with some additional adjustments.

Also, there is the possibility of unassociable "pairs" of agents an agent which generates a resource that a second agent uses, but the two agents are unable to form a complex. Or we could make the rate of complex association or disassociation dependent on their recent productivity (ie. rate of resource conversion). An unproductive complex might be more likely to dissociate into its component agents, for example.

The recruitment process for adsorption/desorption could involve additional complexities. An agent could have repulsive interactions with some resources. Similarly, some agents could attract or repulse each other directly. A potential consequence of this course might be the introduction of homogenous regions and reactive interfaces. In such a case, our notion of organization might be the formation of spatially specialized regions, rather than complex formation, which grow in complexity/organization by interactions at their interfaces with other homogenous region "complexes".

Chapter 3

Stochastic innovation of chemical reaction networks

The material in this chapter was published in the *Proceedings of the National Academy of Sciences* as:

Stochastic innovation as a mechanism by which catalysts might self-assemble into chemical reaction networks

Justin A. Bradford[1] **and Ken A. Dill**[2]
[1]Graduate Group in Biophysics
[2]Department of Pharmaceutical Chemistry
University of California
San Francisco, CA 94143
Email: justin@maxwell.ucsf.edu, dill@maxwell.ucsf.edu

Abstract

We develop a computer model for how two different chemical catalysts in solution, A and B, could be driven to form AB complexes, based on the concentration gradients of a substrate or product that they share in common. If A's product is B's substrate, B will be attracted to A, mediated by a common resource that is not otherwise plentiful in the environment. By this simple physicochemical mechanism, chemical reactions could spontaneously associate to become chained together in solution. According to the model, such

catalyst self-association processes may resemble other processes of "stochastic innovation," such as Darwinian evolution in biology, that involve a search among options, a selection among those options, and then a lock-in of that selection. Like Darwinian processes, this simple chemical process exhibits cooperation, competition, innovation, and a preference for consistency. This model may be useful for understanding organizational processes in prebiotic chemistry and for developing new kinds of self-organization in chemically reacting systems.

chemical evolution, self-organization, abiogenesis, catalytic chains

3.1 Introduction

Here, we propose a simple model. Our goal is not to explain some existing body of data, because we know of none that pertains. Rather, our goal here is to propose a type of organizing principle that has not been explored before, as far as we know, but that is based on well established physicochemical principles and that can be tested by experiments. Our initial motivation for this work was to understand some puzzles of prebiotic chemistry, where, it could be argued, the field is just as limited by a lack of specific testable models at the moment as it is by a lack of experiments.

3.1.1 Model of Agents and Resources

We focus here on catalysts, such as enzymes or simple surfaces. We call a catalyst an "agent." An agent converts a substrate to a product; we label agents alphabetically (see Fig. 3.1). We call a substrate or product a "resource"; we label resources numerically. We assume that agents are MichaelisMenten catalysts; i.e., they bind to their substrate before converting the substrate to a product. In our model, resources may be supplied by the external environment. Such environmental resources may vary with time in a random or controlled way by external forces, but, for simplicity, we assume they are uniformly distributed in space.

Fig. 3.1 shows an example. Agent A converts a substrate 1 to a product 2. Agent B converts substrate 2 to product 3. Key components of our model are the common resources, which are substrates or products that serve in common among different types of agents. For example, in Fig. 3.1, resource 2 is a common resource because it is both a product of A and a reactant for B. Figs. 3.1 and 3.3 also show that if agents A and B

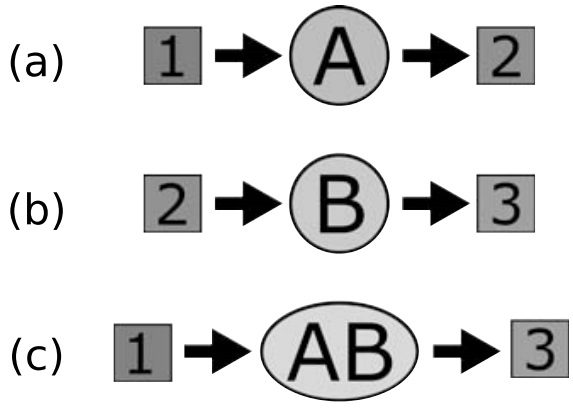


Figure 3.1: Agents (lettered circles) and resources (numbered squares). (a) Agent of type A converts substrate 1 to product 2. (b) Agent of type B converts substrate 2 to product 3. (c) When agents A and B are complexed together, two reactions are chained together, converting substrate 1s to product 3s.

come together by some process, then the AB complex is a "machine" that converts 1s to 3s, mediated by the intermediary resource 2s.

Agent B (Fig. 3.1) may take up a substrate molecule 2 from either of two sources: either the 2 was produced as the output from a nearby A agent, or the 2 was supplied externally from the environment, if 2s are available from external sources. Because we assume that Bs are MichaelisMenten catalysts, a B will bind to its substrate, a 2 in this case. Bs will concentrate around 2s simply because Bs flow down their chemical potential gradients, in the same way that solutes in chromatographic mobile phases will seek out and bind to stationary-phase surfaces for which they have affinity.

3.1.2 Principles of Attraction and Shielding

There are two possibilities for each B agent (Fig. 3.2): attraction or shielding.

1. Attraction. As attract Bs through the following indirect mechanism: As produce 2s; 2s are localized near the As; those 2s attract Bs, concentrating the Bs around the As, thus leading to more AB complexation than would have occurred without the intermediary 2s (see Fig. 3.2a). This enhancement happens when: (i) A agents are present, (ii)

1s (the substrates for As) are plentiful, and (iii) 2s are depleted in the environment (i.e., available only at small or zero concentrations). AB complexation introduces into the system an "innovation," i.e., an ability to produce 3s from 1s, an ability that does not simply and directly result from the presence of A or B alone (see Fig. 3.1c). Significant chaining together of agents is an emergent property of our system. In short, AB complexes are driven to form through mutual indirect attraction, mediated by 2s, the common resource, but this attraction occurs only when 2s are depleted from the environment.

2. Shielding. In contrast, when environmental 2s are plentiful and available in all directions, Bs will not selectively migrate toward As (see Fig. 3.2b). We call this "shielding." (A more biological example of shielding is chemotaxis. A bacterium will swim toward a point source of food, except if food is uniformly distributed everywhere in space; then the bacterium would not migrate preferentially toward any one single point source. Chemotaxis, of course, is a complex process, but it illustrates how a favored direction of motion can result from simple physicochemical forces that change when environmental resources vary.) In the present model, attraction and shielding are simple consequences of concentration gradients.

3.1.3 Details of the Model

Here is our model for how stochastic innovation might arise in a system of chemical catalysts. We assume that catalyst agents can adsorb to a surface. The prebiotic origin of life may have involved surfaces, such as minerals or clays, on which reactions took place[12,24,66].

Our model involves two compartments. First, there is a surface lattice where all of the reactions take place. Second, that surface is in direct contact with a bulk solution just above it, which serves as a chemical potential "bath," a source of agents and resources for the surface. The surface simply provides a mechanism for trapping the 2s and Bs, slowing their escape from the As, providing the basis for the AB complexation enhancement. There are N_A molecules of each agent type in the simulation, and E_N (E_1, E_2, ...) represent the concentrations of that resource in the bulk solution. Fig. 3.3 shows the steps: (i) 1s attract As from the bulk onto the surface lattice, (ii) As produce 2s, (iii) which then attract Bs, (iv) leading to a machine in which As and Bs are clustered on the surface to produce 3s from 1s. Agents and resources diffuse rapidly throughout the bath, so they can be regarded as being in equilibrium within the bath over the time scale of the processes that happen

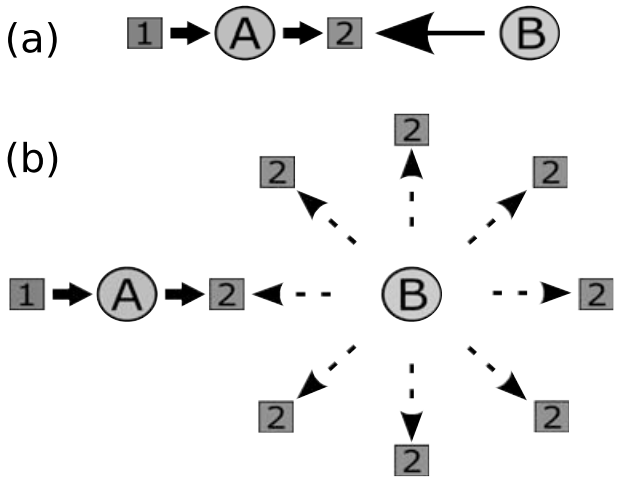


Figure 3.2: Attraction and shielding. (a) Attraction. Bs are attracted to 2s, which are produced by As, hence Bs are attracted to As. (b) Shielding. When 2s are plentiful in the environment, Bs are attracted to them in all directions, hence have no special net tendency to associate with As.

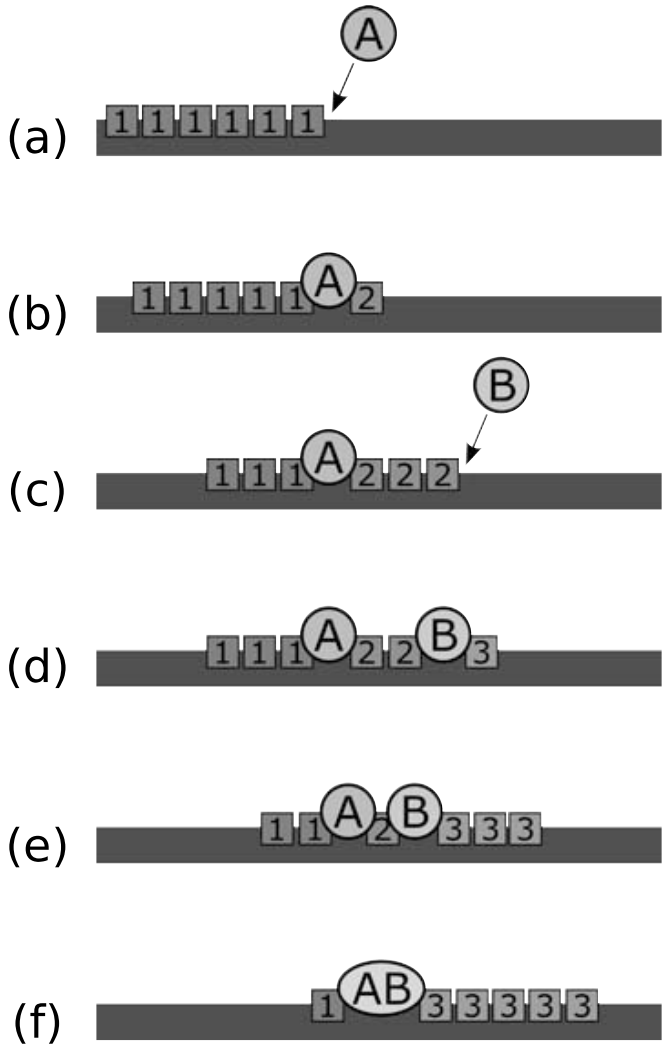


Figure 3.3: Attraction on the lattice. (a) Agent A leaves the bulk and binds to the surface lattice in a region where resource 1s are concentrated. (b) Agent A converts 1s to 2s. (c) Agent B leaves the bulk and binds to the surface lattice where 2s are concentrated, which tends to be near the As that produced them. (d) Agent B produces 3s. (e and f) Agent B associates with A (e), forming a complex, which is now (f) a machine that converts 1s to 3s.

on the surface. The lattice has $L_x \times L_y$ lattice sites; it just serves to coarse-grain the spatial localization. Each lattice site is large enough to be occupied by multiple molecules of different types at the same time.

We believe this is the minimal model that captures how catalysts might self-associate in solution. We use the Gillespie algorithm[67] for the simulation, and during any given time interval, any or all of the following processes may occur (see Fig. 3.5).

1. A resource molecule or an agent molecule may drop down from the bulk and associate with the surface lattice, with rate coefficient k_{on}. To keep the model simple, all four types of molecule bind the lattice with the same rate constant.
2. Any resource or agent molecule on the surface may detach from the surface and be released into the bulk, with rate coefficient k_{off}.
3. Because of the MichaelisMenten binding property of each agent, an agent will attach more rapidly to a site where substrate is concentrated, in proportion to the concentration of its substrate at that site, with a rate coefficient k_{coop}.
4. An agent molecule A can convert any 1s to 2s on its lattice site with a rate coefficient k_{12}.
5. An agent molecule B can convert any 2s to 3s on its lattice site with coefficient k_{23}.
6. Agents and resources can diffuse laterally on the surface lattice with coefficient $D_{lattice}$.
7. An agent A and agent B on the same lattice site can associate with each other, with rate coefficient k_{form}, forming a new species of agent, the AB complex.
8. An AB complex may dissociate, either in the bulk or on the surface, with rate coefficient k_{decay}. Supporting information provide further details and typical values of the parameters (Table 3.1), as well as a pseudocode implementation (see section 3.7). (Source code is available on request.)

3.2 Results

3.2.1 Exploring Attraction and Shielding

Our computer simulations show that if an agent A produces 2s at a rate that is faster than the 2s diffuse away, and if the A is bound to the surface, then 2s will be concentrated near the As on the surface. Agents of type B, which use 2s as substrates, explore space stochastically but will be attracted to the 2s, on average, thus binding to

Parameter	Units	Values	Description
N_A	molecules	10	Number of each agent type
E_N	molecules	10^4	Environment resources available in bulk
k_{on}	time^{-1}	0.005	Basic rate of association with the lattice
k_{off}	time^{-1}	0.001	Dissociation from the lattice
k_{coop}	$\text{time}^{-1} \cdot \text{molecules}^{-1}$	1	Resource-mediated association with lattice
$D_{lattice}$	$\text{length} \cdot \text{time}^{-1}$	0.001	Spatial diffusion on the lattice
k_{x-y}	$\text{time}^{-1} \cdot \text{molecules}^{-1}$	0.1	Agent catalysis of resource conversion
k_{form}	$\text{time}^{-1} \cdot \text{molecules}^{-1}$	0.1	Formation of complex with lattice bound, adjacent agents
k_{decay}	time^{-1}	0.001	Dissociation of agent complexes
$L_x L_y$	length $\cdot$ length	10^4	Area of lattice surface

Table 3.1: Parameters of the lattice model. The units for each parameter, the standard value used in simulations, and the role of the parameter. Molecules are a count of agents or resources. Time is in arbitrary simulation units. Length in in units of a lattice site.

surface lattice sites near the most productive As. If there is a mutual affinity of As for Bs, AB complexes will form on the surface.

Fig. 3.6 shows: (i) attraction, the situation in which Bs migrate to As to form complexes, driven by depletion of the common resource, 2; (ii) shielding, the situation in which common resource 2 is plentiful in the environment, so Bs are not selectively attracted to As; and (iii) a control simulation showing that when As produce no common resource 2s, there is essentially no formation of AB complexes, even though there is some intrinsic affinity between the As and Bs. In this model, AB complex formation is a nonequilibrium process; it happens only in the presence of a gradient of 2s concentrated around the As. The complex formation process can result from either highly productive As or the environmental depletion of 2s from the system.

The rate of complex formation: increases with the productivity of the agents and decreases with the degree of environmental shielding. AB complexation is largest when As are productive (requiring that bulk 1s are plentiful). AB complexation is reduced by shielding (when environmental 2s are plentiful) (Fig. 3.4).

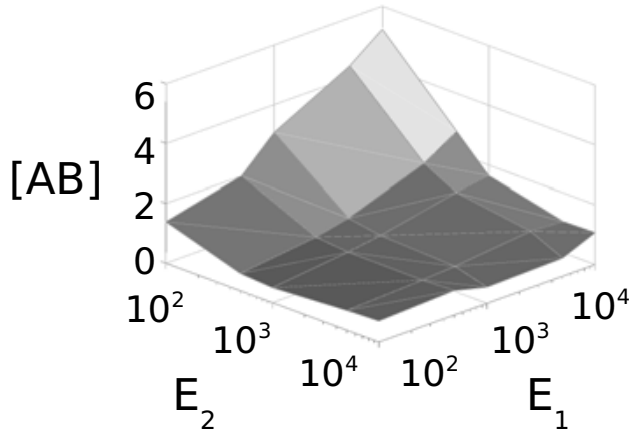


Figure 3.4: Steady-state numbers of AB complexes formed, mediated by resource 2, vs. the supply of resource 1 (E_1) and the supply of resource 2 from the environment (E_2). This shows that: (a) more 2s produced by As lead to more complex formation and (b) that more shielding due to 2s in the environment leads to reduced complex formation.

3.2.2 Some Properties of the Model: Cooperation, Competition, Consistency, and Innovation

Our agents cooperate with each other. When the environment provides no substrate for catalyst B, B will migrate toward A via its substrates. The model shows a driving force for innovation: when the common resource is depleted, the system evolves an ability to create 3s directly from 1s, thus creating two chemical reactions chained together where there were only two isolated reactions before.

Our model also exhibits competition. Fig. 3.7 shows a version of the model in which As convert 1s to 2s and Bs convert 2s to 3s, as before, but now there is an additional agent, labeled A^*. A^* is a superior version of A. A^* converts 1s to 2s faster than As catalyze the same conversion. Fig. 3.7 b and c show that complex formation increases with As productivity. For comparison, Fig. 3.7b just shows independent experiments of a single A interacting with B: curve $[A^*B]$ shows what happens when the A^* produces 2s rapidly, and curve $[AB]$ shows what happens when the A produces 2s more slowly. In contrast, Fig. 3.7c shows A and A^* together in the same solution, competing for Bs. A^* outcompetes

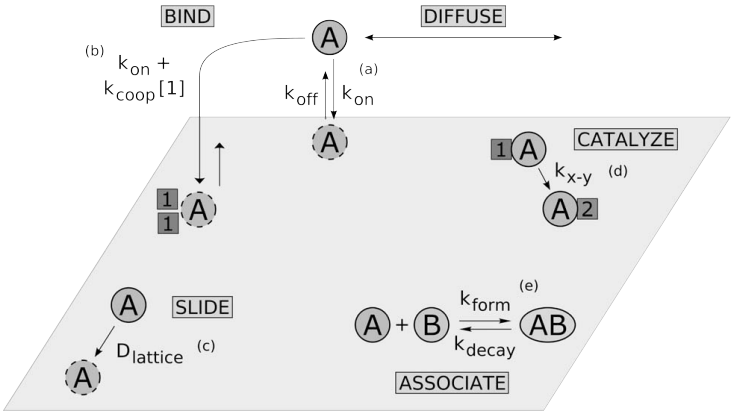


Figure 3.5: Processes in the lattice model. (a) Molecules exchange between the bulk solution and bound to a random lattice site. (b) Agents have an additional binding rate at lattice regions with their input resource. (c) Bound molecules move about the surface. (d) Agents convert input resources to output resources at their site. (e) Two agents at a site can form an agent complex.

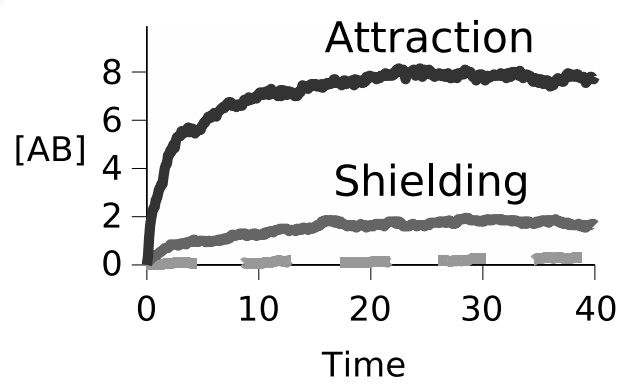


Figure 3.6: Numbers of AB complexes formed vs. time. Dark line: under attraction conditions, no environmental supply of resource 2 ($E_1 = 10^4$, $E_2 = 0$). Light line: under shielding conditions, with resource 2 provided by the environment ($E_1 = 10^4$, $E_2 = 10^3$). Dashed line: control experiment; no 2s are available because the environment has none, and As are unproductive because of the absence of 1s ($E_1 = 0$, $E_2 = 0$). (Time is in units of 1,000 simulation units for this and following plots.)

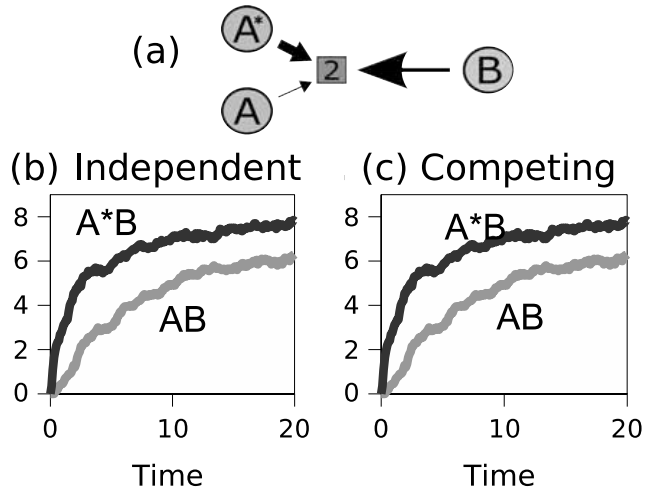


Figure 3.7: Competition. (a) Agent B can associate with either agent A or agent A^*, a superior producer of resource 2. (b) Complex formation increases as the productivity of As increase. (c) Agent A competes with a superagent A^* (for A^*, $k_{x-y} = 0.1$ and for A, $k_{x-y} = 0.001$). Competition enhances the difference between A^*B and AB complex formation.

A to associate with the Bs. Comparing Fig. 3.7c with Fig. 3.7b shows a nonadditivity because of the competition for the same parameters: When A and A^* compete in the same solution, the "rich get richer." Within our simple chemical model system, this competition resembles Darwinian selection, except that our metric of "success" is complex formation, whereas the metric of success in biological systems is survival.

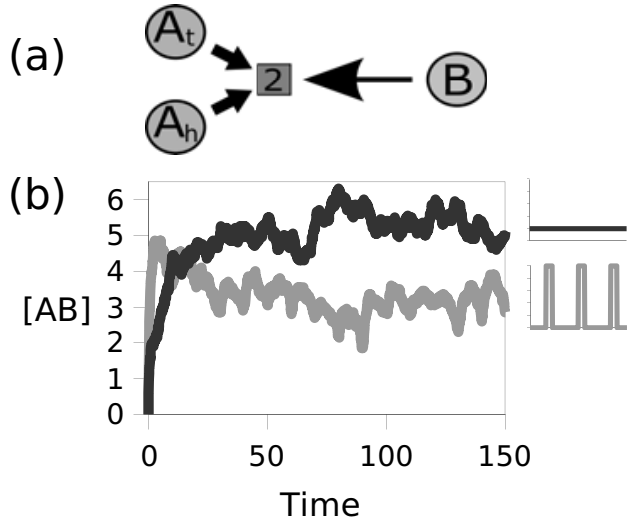


Figure 3.8: Consistency. (a) Agent B can associate with either agent A_t (tortoise), which produces resources at a constant rate (E_t is constant vs. time), or agent A_h (hare), which produces resources in larger, occasional bursts (E_h pulses vs. time). The time-averaged productions of resources is the same for both agents. (b) Agent A_h, producing resources for 2,000 of every 10,000 simulation time units, forms less complex with agent B than agent A_t.

Our model shows that consistency has value, exhibiting "tortoise and hare" behavior (Fig. 3.8). One type of agent, A_t, the tortoise, is a slow consistent producer of 2s. The other type of agent, A_h, the hare, is highly productive, but only in short bursts. Even though the time-averaged productivity in converting 1s to 2s is identical here for these two types of agents, the tortoise wins. The tortoises form complexes with Bs at a faster rate than the hares form complexes with Bs. Thus, sustained consistency is more effective for complex formation than high-activity-burst behavior.

3.2.3 Other Functional Hierarchies

This model indicates how more complex chains of catalytic activity could be formed. Fig. 3.9a shows a linear series of catalytic agents that become chained together to convert 1s to 2s to 3s to 4s, etc. Fig. 3.9b shows the time dynamics of formation, when the environment is simply providing 1s. It shows that the final endstate machine, a concatenation of catalysts A, B, C, D, E, F, G, and H, grows monotonically populated, and that no intermediate smaller machine is ever substantially populated during the time course of development. Thus, multiple catalysts can be driven together, potentially into a variety of topological arrangements, including metabolic chains, networks, and cycles.

These results bear on an idea that has been called "irreducible complexity." It has been argued that complex biological and prebiotic chemical systems could not have arisen by simple physicochemical processes, because there would have been no selective advantage for each of the putative incremental changes along the way[68]. In that view, what good is half an eye? An organism would not be served by anything less than a full eye, so intermediate structures would not have imparted enough value to survive natural selection. In that view, "irreducible" refers to a system that would fail to function if any one component is removed, and irreducible complexity refers to the idea that such systems require design and could not be developed by stochastic innovation. The counterargument, seen in computer simulations, for example[10], has been that stochastic innovation works differently: evolution does not "know" the final end-goal in advance, but finds it through a random search in indirect, incremental steps.

Fig. 3.9 shows our model version of an irreducible system (a chain of reactions that convert 1s ultimately into 9s; it would fail to do so if any step is removed). This system ultimately converts 1s into 9s through a multistep process in which no step is dispensable. Yet, this machine arises in our model from simple physicochemical processes that have not involved any sort of "design" in advance to achieve this particular goal of producing 9s. In the model, this chaining together of chemical reactions is driven by blind physicochemical forces.

3.3 Discussion

3.3.1 Evolutionary "Gaps"

In evolution, there are sometimes evolutionary gaps in the fossil record, situations in which lifeforms X and Y or features X and Y are known but where there is no evidence of the steps in between. Fig. 3.9b shows how such gaps arise in our simple machine. In short, the intermediate states are unstable. The steps are downhill. One evolutionary step leads to the next, quickly followed by the next, and so on, without pausing. In the evolutionary metaphor, half an eye never appears as a stable state because such a state is quickly driven by even stronger evolutionary forces to form a complete eye, maybe for a different purpose than the half-eye. Such two-state transitions are also common in protein folding, for example, where the denatured state is followed in time by a partly structured state that is immediately followed by an even more structured state, etc., until the molecule becomes fully folded into the native structure. At the earliest stages of folding, the protein does not know that it is headed toward the native state; it is just seeking a situation that is marginally better than its previous state.

The alternative is to have no evolutionary gap. Fig. 3.9 ce shows how that alternative situation arises in our simple model, for a different set of parameters. In these cases, intermediate states are stable and populated during the evolution of the full molecular machine. The system can stall at intermediate points. In both situations (stable intermediates or two-state behavior), this simple model of chemical association resembles behaviors observed in Darwinian biological systems: Evolutionary gaps sometimes appear, and sometimes they don't.

3.3.2 Implications for Prebiotic Chemistry

There have been two general models for prebiotic evolution: genetics-first (GF) or metabolism-first (MF)[31,32,69]. In GF, some genetic machinery, or capability for self-reproduction, is presumed to arise at an early stage; for example, using RNA molecules[19,70]. Biology then evolves from that point. Once genetic machinery exists, there are many plausible models for the later stages, based on hypercycles[71,72], and/or based on the many powerful RNA and protein evolution experiments that have been performed[20,73,74]. The challenge in accepting GF as a first step, however, is that it is complex and requires the

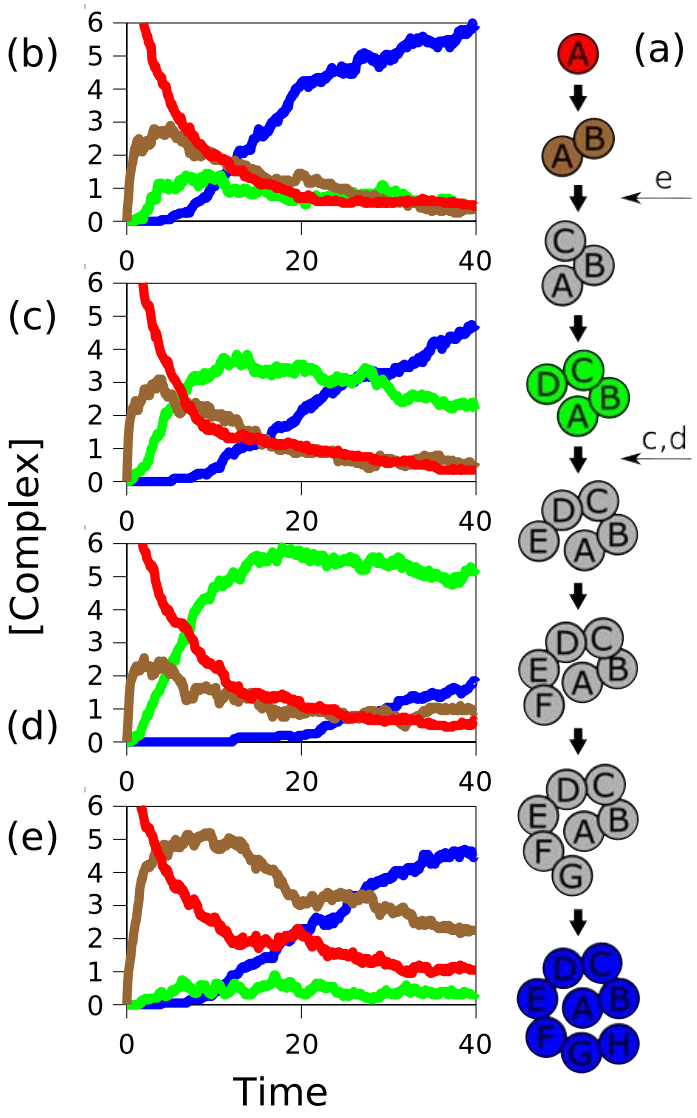


Figure 3.9: Formation of multiagent complexes. (a) Eight agents assemble sequentially to form a catalytic chain, using only the input resource for agent A. (b) When there is no environmental shielding, the full chain assembles rapidly, with no stable intermediates. (c) The presence of environment resources ($E_5 = 100$) shields the step between agents D and E, resulting in longer-lived minor intermediates. (d) A 10-fold stronger environmental supply ($E_5 = 1,000$) results in the stable dominance of the A-D intermediate chain over the complete chain. (e) The presence of environment resources ($E_3 = 100$) at an earlier stage, between B-C, leads to an initially stronger, but ultimately minor, intermediate chain.

joint emergence of catalysis, compartmentation, and heritability, all at the same time[32]. de Duve has noted[31] that GF "is accepted much less for its likelihood than for the lack of an alternative."

In the MF model[31,66,75], chemical reactions become chained together evolutionarily before the appearance of genetic machinery. Although the requirements for MF are, in theory, more elementary than for GF, a key question about MF is how catalysts might become organized on their own, in the absence of a genetic system[76,77].

We believe the present model of stochastic innovation based on attraction and shielding among chemical catalysts provides a plausible mechanism by which simple metabolic chains and cycles of reactions could have come together, perhaps at least long enough for a genetic system to then emerge. Of course, an important virtue of ultimately having a genetic system is that it provides much longer term "memory" for the "lock-in" step than does nongenetic propagation, where memory is merely provided by a ratio of off-rates to resource fluctuation times.

A key distinction between stochastic innovation, explored here, and design-based innovation, in which a complex system is engineered and constructed by a designer, is that stochastic innovation involves no implicit "goals" and no guidance toward a particular purpose. The Darwinian paradigm shows how increasing complexity and order can arise from processes that do not involve guidance through intelligence or design. Darwinian evolution is a process of elimination ("evolution-away-from"), rather than a process of design ("evolution-toward"). Stochastic innovation achieves evolution-away-from by search, selection, and lock-in. The present model has the three features of stochastic innovation.

1. Search. The B agents diffuse randomly through space and find 2s, based on a mutual binding affinity.
2. Selection. The B agents associate with A agents, if the common resource 2s are not present in the environment, leading to AB complex formation.
3. Lock-In. If the off-rate for dissociation of AB complexes is slower than the average frequency of resource depletion disasters of the common resource in the environment, then the complexes will be stable beyond the time scale of a single resource disaster.

Here is a possible experimental test of our model. Two enzymes, A and B, would be selected, based on having a common resource (e.g., 2). A's product is known to be B's substrate. The As would be covalently linked to a surface, such as a chromatography sta-

tionary phase. A pair of fluorescent probes would be attached: a donor probe is attached to each A molecule and an acceptor probe to each B molecule. This would allow for monitoring the concentration of AB complexes, through fluorescence quenching, for example. Two experiments would be performed. In one experiment, B molecules in a mobile phase would flow across the stationary phase in the absence of added 2s, but in the presence of 1s, the substrate for As, and then the amount of AB complex formation, would be measured. The other experiment would be identical, except that a high concentration of 2s would also be present in the mobile phase. The present model suggests that, if the relative on-rates, off-rates, and diffusion coefficients are roughly as given in Table 3.1, AB complexes should be more concentrated in the first experiment.

3.3.3 Related Modeling Efforts

The present model differs from others that have been used to explore chemical self-organization. Some agent-based models[10,11,78] involve computer-based rules rather than the laws of physical chemistry that are of interest here. Mass-action models of early evolution have been developed[79,80], but they also are not focused on the microscopic chemical mechanisms. Models of hypercycles[71,72] pertain to molecular systems that already have a genetic system, whereas our interest here is in molecular systems having no genetic system. Recently, there have been interesting studies of complexity and fragility in biological systems[54,57], but those treatments presume some preexisting mechanism of stochastic innovation.

Because the work of Turing in the early 1950s[81], much mathematical modeling in chemistry, chemical engineering, and biology has explored pattern formation[29]. Pattern formation is often modeled by using lattice models such as the present one, or nonequilibrium coupled spatiotemporal differential equations, sometimes with added stochastic noise terms. We believe that our model, too, could be readily cast in the form of such a coupled set of "reaction-diffusion" equations.

However, the present work differs from those treatments in certain respects. First, ours addresses not just how molecules can move and organize in space, but how chemical reactions can move and organize in space and in so doing become chained together into more complex reactions. Second, pattern formation usually involves some particular nonlinear component term in the model's mathematics, often arising from saturable or cooperative

binding, for example[60]. Our model, too, involves saturable binding, but the most important nonlinear component of our model is what might be called "reflexive catalysis." That is, at the same time that our agents catalyze reactions among the resources, our common resources also perform a sort of catalysis in the process of agentagent association. In addition, our resources can be regarded as "ephemeral catalysts," catalyzing complex formation only under certain nonequilibrium conditions and losing their catalytic power at equilibrium or when resources are plentiful.

3.3.4 Second Law of Thermodynamics

This model describes a spontaneous process of organization, and hence might lead to the suspicion that it violates the laws of thermodynamics. It does not. Because our agents are assumed to be Michaelis-Menten binders, it means that binding is spontaneous, i.e. ΔG is negative. We have also assumed that our catalytic conversions have reasonable yields. For example, *A*s convert 1s to 2s at a nonzero rate and in nonzero product concentrations. This, too, is standard physical chemistry, and violates no laws. What about other situations: can stochastic innovation also produce catalytic cycles in which there are steps that are uphill in free energy? Yes. Then our model would simply require a coupling to some other downhill step, say the conversion of ATP to ADP. Our explicit focus here has been on how downhill processes could lead catalysts to associate. Uphill processes could also lead to association, but would require consideration of the coupling to energy supplies. As a metaphor, you can describe a factory processes and products, without the distraction of also describing its wall plugs and its energy usage. Such energy balances cannot violate physical laws, but it is not necessary to treat it explicitly here.

3.4 Conclusion

A well known process in chemistry is the binding and association of molecules, driven by thermodynamic forces. Here, we consider whether catalyst molecules might be driven to associate with each other, through typical binding forces, but based on their molecular functions. Functional driving forces are well-known in biology, through the principles of evolution, but are not yet much studied in chemistry. We propose a model for how different Michaelis-Menten enzymes or catalysts might tend to associate, driven by the production or depletion of common resources. The agents do not associate if the common resource is

plentiful. We call this the shielding principle. In this way, agents organize adaptively, and complexity can form from simpler systems. In our model, "function dictates structure," a reversal of the paradigm in which "structure dictates function."

Our model of stochastic innovation for chemical catalysts has three components: (i) a method for sampling a space of possibilities (Bs can search for As or not, depending on the presence of 2s), (ii) selection of a "favorable" outcome (Bs become more productive of 3s when their substrate 2s are depleted, if they associate with As), and (iii) a form of memory that locks in this outcome (the off-rates of Bs from As is slower than the time constant for fluctuations of 2s in the environment. Our process resembles other stochastic innovation processes, such as Darwinian evolution, in that agents compete, cooperate, value consistency, and innovate. The model is experimentally testable. It is a possible model for how metabolic reactions might have become chained together, at least primitively, before the formation of a genetic system. And, it might lead to ways to self-assemble chemical reactions in solution.

3.5 Materials and Methods

Fig. 3.5 shows that B agents can associate with A agents through two routes: (i) slow random migration on the surface, after Bs bind to the lattice, or (ii) rapid diffusion in the bulk, then binding to the lattice. The latter is more targeted, because Bs bind preferentially (through parameter k_{coop}) where 2s are located, potentially near the As. Hence, the intrinsic rate of AB complexation depends on k_{on}, k_{off}, and $D_{lattice}$.

Intermediary 2s can affect AB complexation through two mechanisms. First, AB complexation is promoted if the number of lattice-bound 1s is high and if As are productive (k_{12} is large). The number of 1s on the surface depends on 1s in the bulk (E_1) and the adsorption equilibrium constant, k_{on}/k_{off}. The number of surface As is then determined by k_{coop} and dependent on the concentration of lattice-bound 1s. These bound molecules produce bound 2s at the rate k_{12}. However, if catalytic rates are much slower than k_{off}, resources disappear from the surface before they can become concentrated enough to attract agents. Second, AB complexation is slowed by the environmental supply of 2s (E_2), which adsorb to the surface randomly and serve as decoys that attract Bs to nonproductive sites (where there are no As).

In order that 2s remain near As long enough to mediate AB complexation, lateral

diffusion on the surface is set to be relatively slow, so dissociation of an AB complex on the surface is typically followed by reassociation. In contrast, dissociation of an AB complex in the bulk solution seldom leads to reassociation, because diffusion in the bulk is set to be fast. Therefore, most of the permanent dissociation of AB complexes occurs in the bulk solution. Also, the catalytic rates used in our simulations are large or comparable to the lateral diffusion rates.

3.6 Acknowledgments

We thank Patsy Babbitt, Don Hilvert, Tack Kuntz, Chris Voigt, John Chodera, and Yigal Nochomovitz for helpful discussions. This work was supported by a National Science Foundation predoctoral fellowship (to J.A.B.) and National Institutes of Health Grant GM34993 and the Sandler Foundation (to K.A.D.).

3.7 Pseudo-code implementation of lattice model

```
count is array of (position, molecule)
rate is array of (position, process, molecule)
do loop
  # accumulate a total rate for all processes (in the bulk
  #   and all lattice sites)
  total_rate = 0
  # reset the bulk counts of resources according to the
  #  current "universe" state; these are generally constant
  #  or pulsed functions of E_i
  for each i in resources
    count(bulk, i) = universe(E_i, time)
  end

  # AB complex in the bulk decays
  rate(bulk, decay_AB) = k_decay * count(bulk, AB)

  # increment the total rate with the decay rate in the bulk
  total_rate += rate(bulk, decay_AB)
  for each s in lattice sites
    for each i in molecules
      rate(s, diffusion, i) = D_lattice * count(s, i)
      rate(s, bind, i) = k_on * count(bulk, i) / area
      rate(s, unbind, i) = k_off * count(s, i)
```

```
    end

    # increment the binding rates of agents with the input
    #   resource mediated cooperative binding
    rate(s, bind, A) += k_coop * count(s, 1) * count(bulk, A) / area
    rate(s, bind, B) += k_coop * count(s, 2) * count(bulk, B) / area
    rate(s, bind, AB) += k_coop * count(s, 1) * count(bulk, AB) / area
    # the resource conversion reactions
    rate(s, reaction, A) = k_1_2 * count(s, A) * count(s, 1)
    rate(s, reaction, B) = k_2_3 * count(s, B) * count(s, 2)
    rate(s, reaction, AB) = k_1_3 * count(s, AB) * count(s, 1)
    # formation and decay of the AB complex
    rate(s, form, AB) = k_form * count(s, A) * count(s, B)
    rate(s, decay, AB) = k_decay * count(s, AB)
    # increment the total rate with the summed rate of all
    #   processes at this lattice site
    total_rate += sum(rate(s, *))
  end
  # Gillespie algorithm to choose when and what event happens next

  # the time increment for this event
  time += (1 / total-rate) * log(1.0 / random(0, 1))

  # what the event is.
  # position: bulk or specific lattice site
  # process: what happens at that position (binding, unbinding, etc)
  # molecule: what agent or resource binds, unbinds, etc

  choose (position, process, molecule)
      with probability = rate(position, process, molecule) /
total_rate
  # now perform the chosen event:

  if process = diffusion:
    count(position, molecule) -= 1
    count(random_adjacent_position(position), molecule) += 1
  end

  if process = bind:
    count(bulk, molecule) -= 1
    count(position, molecule) += 1
  end

  if process = unbind:
```

```
    count(position, molecule) -= 1
    count(bulk, molecule) += 1
  end

  if process = reaction:
    if molecule = A:
      count(position, 1) -= 1
      count(position, 2) += 1
    end

    if molecule = B:
      count(position, 2) -= 1
      count(position, 3) += 1
    end

    if molecule = AB:
      count(position, 1) -= 1
      count(position, 3) += 1
    end
  end

  if process = form:
    count(position, A) -= 1
    count(position, B) -= 1
    count(position, AB) += 1
  end

  if process = decay:
    count(position, AB) -= 1
    count(position, A) += 1
    count(position, B) += 1
  end
end
```

Chapter 4

Preliminary Work

A large portion of the research underlying this dissertation was the exploration of potential model frameworks. Initial results lead to subsequent refinement and re-conceptualization of our basic models, ultimately resulting in the work presented in the previous chapter. Nonetheless, these precursor models did provide a variety of interesting results, and could certainly form the basis of future, productive research. For this reason, we briefly describe the core concepts and key results from our preliminary work.

4.1 "Scorecard" model

Based on the idea of molecules adhering to a surface, and the specific combination of molecules on the surface convey catalytic properties (Fig. 4.1). The "scorecards" represent potential catalytic actions, and a number of agents representing available surfaces. The surface can partially match scorecards to produce a limited catalysis. The reactant and product of the agent are also the molecules which bind to form the catalytic surface.

4.1.1 Competition & Cooperation

We use a variety of scorecards that require the same resource they output. For example, an agent full of 1s produces 1s, and an agent full of 2s produces 2s, and so on. Each resource is generated by the background at the same, low probability, and we see a "competition" between resources for control of the available agents.

We extend this system by adding "cooperation" between a subset of resources. A pair of cooperating scorecards uses one resource as input and produces a second resource,

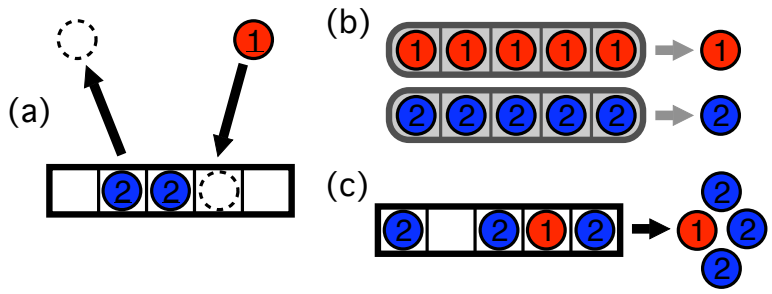


Figure 4.1: Introduction to resources, agents, and scorecards. (a) Resources (numbered circles) exchange between empty spaces on agents (rectangles) and the bulk solution. (b) Scorecards (grayed capsules) define possible catalytic functions. (c) An agent produces new resources based on the similarity of its resource composition to the available scorecards.

with the two scorecards in the pair exchanging the role of input and output for the two resources. This provides the cooperating resources an advantage in competition against the simple "selfish" behavior (producing oneself).

To build a system with strongly modal "states", we introduced two scorecards, which consumed the other's resources to generate resources of its own input type. With this feedback process, we were looking for a system that would quickly move from an intermediate, mixed condition, to one or another strongly defined "states" of activity.

4.1.2 Cycles

Using a looped set of scorecard reactions, we attempt to demonstrate a stable periodic behavior. The first scorecard would use 1s as input, and produce 2s at a very high rate. The second would use the 2s to produce 3s, and so on, until resource N, which activated a scorecard reaction to produce 1s, completing the loop.

To improve the cyclical quality of the previous system, we also studied a more elaborate state change construction. For each resource, we had two scorecards. With a poor match (say, less than half of an agent's bound resources matched the scorecard), the agent would produce the current resource and consume the previous resource (relative to its input) in the cycle. With a good match, the agent would produce the next resource and consume itself. Resources would "pull themselves up" to establish dominance, and then once dominant, resources would "push themselves down" to pass to the next state.

4.1.3 Robustness

How long do reactions survive when the input resources are unreliable (eg, random on-off states of background production)? We compare simple, non-recursive reactions, selfish recursive reactions, and cooperative two-step recursive reactions. We show a special case of this system in Figure 4.2, where there is an initial supply of resources, but no further resource contribution from the environment. We that recursive feedback ("self" and "coop" behavior) performs much more strongly than the non-reinforcing reaction ("none" behavior). Further, we see that cooperative behavior generally out-lives the purely selfish reactions.

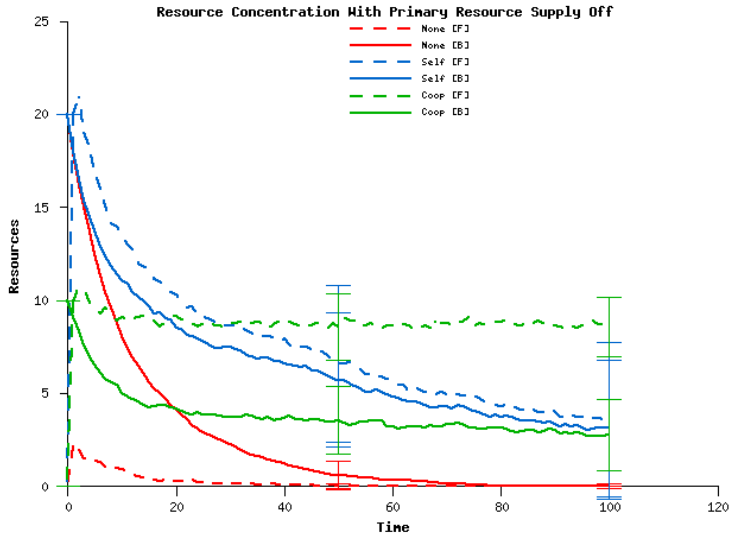


Figure 4.2: Decay of resources with non-reinforcing ("none"), selfish, and cooperative behavior. Solid lines indicate the population of agent-bound resources, and dotted lines show the population of free, unbound resources. While the "none" agent simply decays exponentially, the reinforcing behaviors are more complex. Selfish is initially far more productive, but it is eventually overtaken by the cooperative behavior.

4.2 "Scorecards" with specialized agents (surfaces)

A limitation of the previous model was homogeneity of surfaces, preventing two or more distinct catalytic pathways from strongly establishing on the availabe surface. To

open this possibility, we introduced distinct surface types; different surfaces preferentially bind different resources (Fig. 4.3).

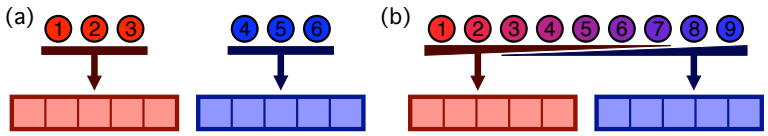


Figure 4.3: Specialized agents interacting with different resources. For example, (a) different agents might bind distinct sets of resources, or (b) agents might vary in affinity across some spectrum of resources.

4.3 Spatial model

Following on the idea of surface specialization, we experimented with spatial degrees of freedom. The intent here was to again open the possibility of multiple, specialized catalytic pathways within a system, but to avoid introducing highly specific parameterization, such as the various surface properties in the previous model.

A major departure from the previous models is to replace scorecard based agents with simple catalytic agents – an agent has an input and output resource, and performs the conversion whenever an input resource is available. Organization in the model now results from relative spatial arrangement of the agents. However, in addition to random diffusion, agent movement is affected by adjacent input resources. Converting an adjacent resource causes a slight displacement of the agent in that direction (Fig. 4.4).

4.4 Analytic "pools" approximation

We attempted to generalize the idea in the previous model into an abstract analytical model. Each type of agent would produce resources, which would necessarily be in close proximity to the producing agent. Thus, we introduced a notion of resource "pools". An agent produced resources into its local pool, which gradually "decayed" to the general pool (Fig. 4.5).

Assuming an agent is physically drawn to its reactant resources, there would be two competing "forces" of attraction to specific local pools and shielding from local pools by the general pool. Based on this action, we assume agents could come to associate spatially

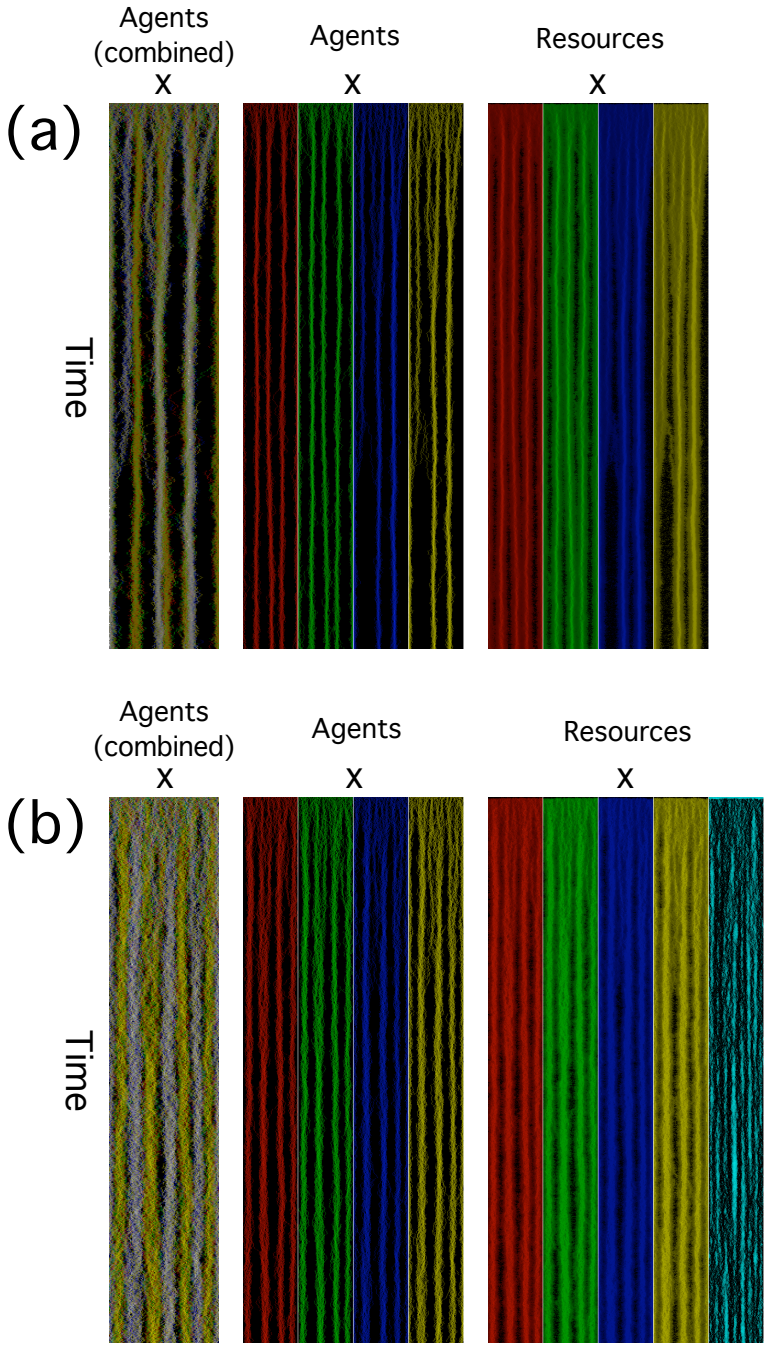


Figure 4.4: Spatial organize of cooperating agents. (a) Pairs of cooperating agents (red/green and blue/yellow) quickly co-localize spatially due to their shared resources. (b) If agent function is also limited by a common, scare "food" resource (cyan in resources), we find that distinct pairs of cooperating agents spread out evenly across the available space.

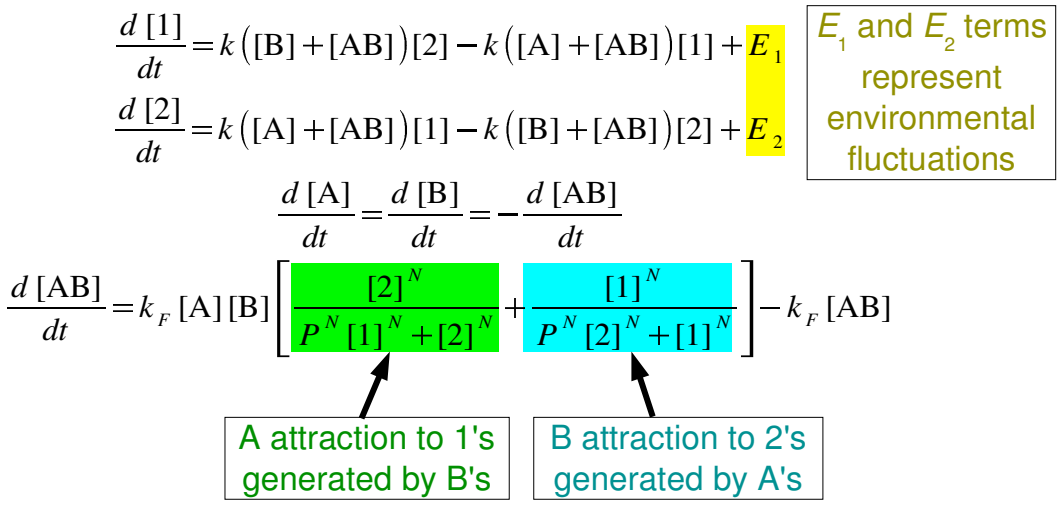


Figure 4.5: Equations for a simple cooperative pair of agents. Change in resource concentration depends on the function of the two agents (A & B) and the agent complex (AB), as well as a term representing environmental fluctuations. Formation of the complex depends on two sigmoidal terms, representing the availability of input resources through the paired agent versus those available from the general solution. These sigmoidal terms represent the attraction of an agent to its pair due to the pair's production of its input resource.

with other agents. The basic system leads to environmental dependence (Fig. 4.6) and opens the possibility for cooperative systems to mitigate environmental instability (Fig. 4.7).

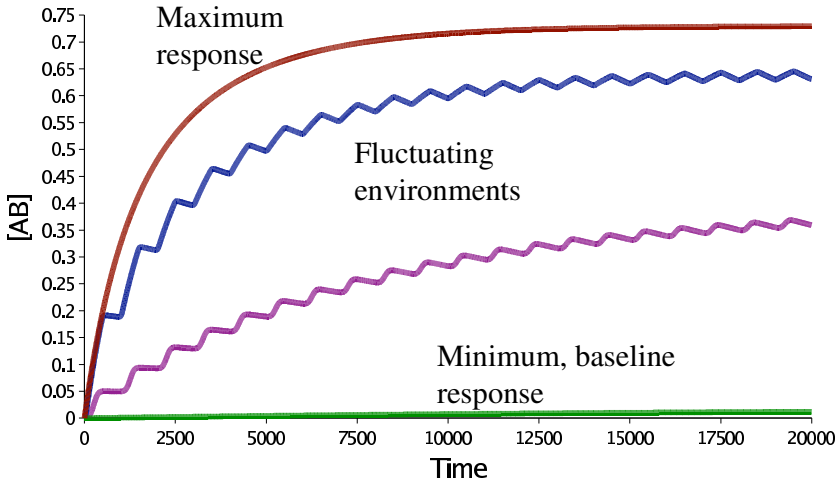


Figure 4.6: Response to different environments. We vary the availability of resource 1 in a cooperating pair with no resource 2s available from the environment. When 1s are highly abundant, we approach a maximum concentration of complex. When no 1s are available, we see a very small concentration of complex, due to random interactions. We also observe the response when 1s are provided by the environment sinusoidally, with different amplitudes.

4.5 Explicit lattice

Although the analytical pool model captured many interesting properties, it was helpful to pursue a more explicit model of how these agents were physically attracted and shielded. The resulting models were based on the idea of a surface upon which agents and resources are bound. As with the 1D model, agent movement is biased by adjacent resources. Converting an adjacent resource can displace the agent, moving it into the resource's lattice site. This simple, local behavior results in the migration of agents towards spatial regions with high resource concentrations (Fig. 4.8).

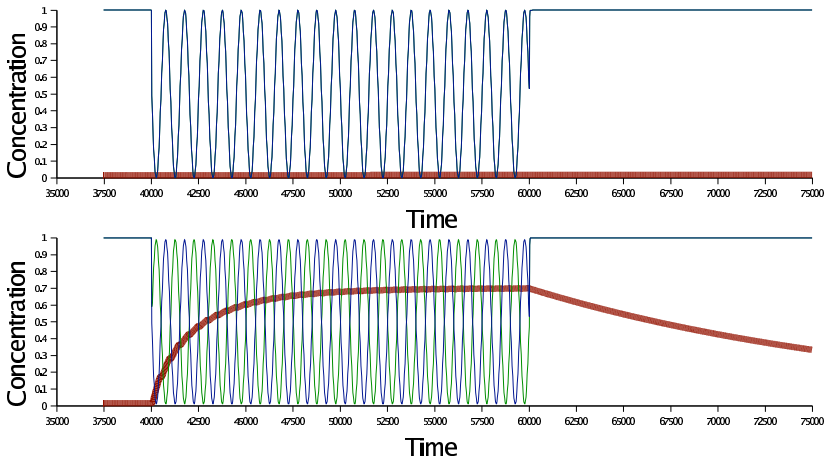


Figure 4.7: Cooperating pair response to in-phase and out-of-phase perturbation. In a simple cooperating pair, the availability of both resource 1 and 2 varies sinusoidally with time. When the availability of resources is in-phase, very little complex forms, similar to the baseline response. And when the resources are provided out-of-phase, complex concentration approaches that of maximal activity.

4.6 Current Model

Finally, we considered a new variation on lattice interactions. Agents and resources can exchange between the lattice and an adjacent bulk solution. Unbound agent interaction with bound reactants leads to cooperative lattice binding of the agent. This effectively results in agents preferentially binding to the surface at regions of high reactant resource concentration. This general model is the basis of the work presented in the preceding chapter.

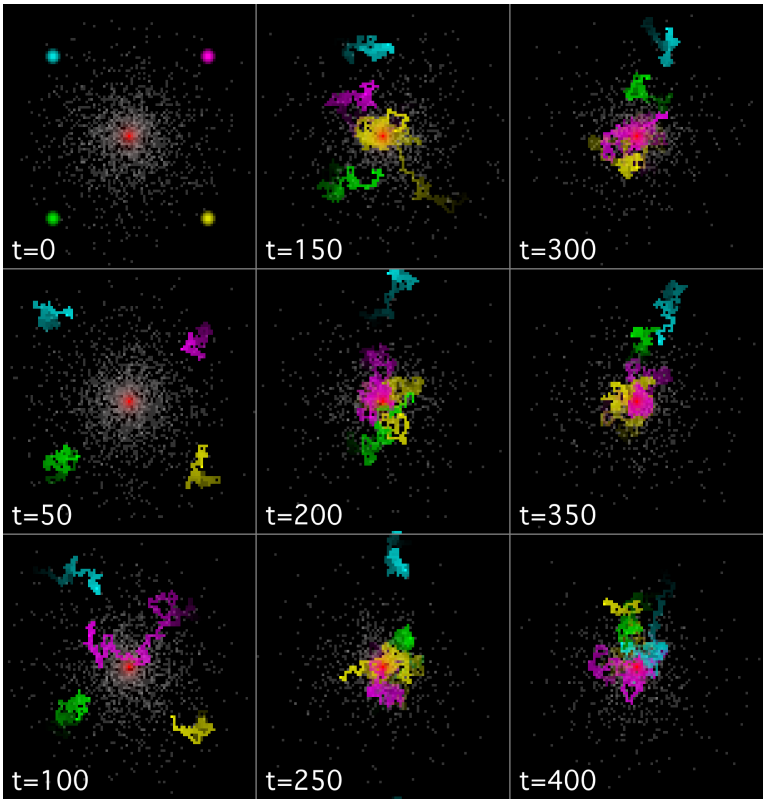


Figure 4.8: Movement of agents on lattice surface over time. Fixed at the center (red dot) is an agent A. Its product, shown by white dots (brightness indicates concentration), is also the reactant for B agents. Each B leaves a color that fades with time at its current lattice site, indicating diffusional trails of nearby Bs towards A.

Chapter 5

Future Directions

5.1 Simple Extensions

For the sake of simplicity, this research has focused on a very simple model of catalytic behavior. Specifically, catalysis is limited to a basic transformation, or map, from some discrete input to some discrete output. Furthermore, association of agents into structures assumes the resulting complexes merely join individual functions. This model can be easily extended with slight more complex notions of catalysis and complex behavior.

5.1.1 Synthesis, decomposition, and displacement reactions

Although it is possible abstractly express a variety of interesting reactions with the current model, it would be useful to consider a more flexible definition of catalysis. To expand the model, it is worth considering agents which explicitly take more than one substrate or produce more than one product. In addition to our current "transformation" (or isomerization) ($A \rightarrow B$) reaction, adding synthesis ($C + D \rightarrow E$), decomposition ($F \rightarrow G + H$), and displacement ($I + J \rightarrow K + L$) should offer a reasonably expressive range of reaction forms[82].

5.1.2 Modulators

It is very common for simple catalytic processes to respond to modulators, eg. various modes of promotion or inhibition[83,84]. Even in the absence of intelligence, agents are capable of reacting to the presence of resources independent of their input and output

resources. Clearly, potential experimental systems with relevance to this dissertation are capable of responding to "promoters" (or cofactors) and "inhibitors". Our model could be extended to consider variations in the rate of function for an agent based on the local concentration of resources independent of the set of resources actually used and produced.

5.1.3 Novel catalysis

Upon agent association, this model assumes that the combined complex only sequentially joins its constituent functions. While this might be the most common case, there are examples of association leading to novel catalytic behavior through exaptation[85,86]. Our model could be extended to consider the effect of novel catalytic ability upon formation of a novel complex. Capability of this nature would likely increase the adaptability of the model by expanding the dimensionality of fitness space, making isolated functional regions more accessible[87].

5.2 Going further: Generalizing the model

5.2.1 Probability surface

An interesting generalization of this model is to forgo explicit, discrete models of resource and agent flow, in exchange for 2-D gaussian functions to represent population of the lattice. Using a probability surface should provide a very close approximation of the original model's behavior, while being far more computationally efficient. This would allow for simulations with much larger surface areas and far greater diversity in agent and resource types.

5.2.2 Matrix evolution

The next step in generalization requires a fundamental abstraction of the lattice space. A complex matrix representation of this system, and corresponding operator reproducing the "attraction and shielding" behavior, could reproduce the critical behavior of the current model. In short, this system would appear to be expressible in terms of linear algebra, and if so, would offer the opportunity for much deeper mathematical analysis. While it is not yet clear how to construct a complete translation, this is certainly a promising approach for future work.

5.2.3 Non-ribosomal peptide synthesis

Experimental demonstration of the principles of our theoretical models are inevitably required. And while beyond the scope of this dissertation, it is worthwhile to note specific "real world" examples that might assist future experimental work. With that in mind, there is one such system that is particularly compelling: non-ribsomal peptide synthesis (NRPS)[88,89].

This brief synopsis of NRPS is included with the hope that it will provide insight on future experimental work related to this dissertation. It is:

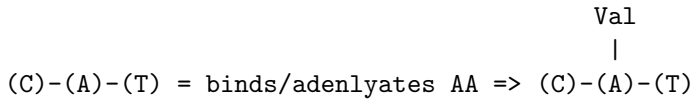
a) abnormal from the view of canonical cellular processes, in that it is most commonly used to express peptide structures not possible with the ribosomal approach; and
b) a very common biochemical process that is remarkably analogous to the general theoretical mechanism of this dissertation.

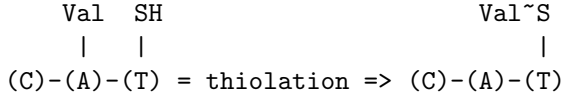
The NRPS mechanism uses linear sequences of "modules", each binding a specific amino acid and adding it to a peptide chain. This is the synthetic route for many important biomolecules, often with antibiotic/antiviral properties, such as bactracin, gramicidin and surfactin[90]. These complexes are some of the largest proteins in the cell. This mechanism also has some similarities to polyketide synthesis (PKS), and there are example systems using a mix of NRPS and PKS pathways, such as the biosynthesis of epothilone[91].

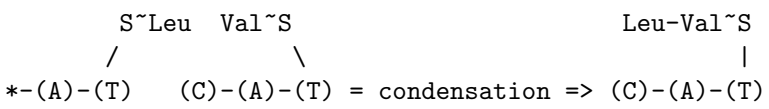
A single "module" has three key domains[89,92]:

(C)ondensation: catalyzes bonding of the module's amino acid and the amino-acid/peptide provided by a preceeding module.
(A)denlyation: binds a specific type of amino-acid (AA) and adenylates it. The promisicuity of the various (A) domains does vary.
(T)hiolation: accepts the amino acid from the A domain, and is primed for ligation with an "upstream" amino acid on its condesation domain.

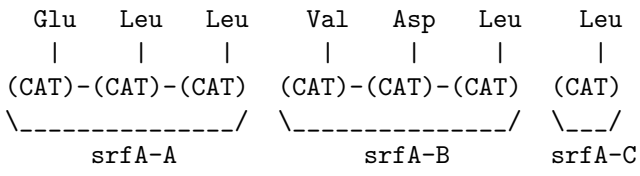
Optionally, there can be other amino acid modifying domains following the thiolation (T) domain, such as epimerization or methylation functions[89,92]. However, a basic step follows this process:







Also, several modules are often combined into a single protein (a "chain"), and most synthesis pathways involve a sequence of several chains. For example, surfactin synthesis in B. subtilis[92,93]:

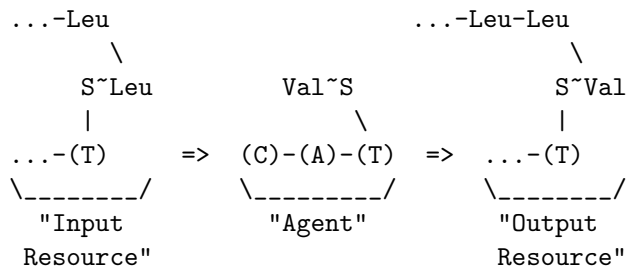


The various chains involved in a particular synthesis are usually encoded in a single operon[94]. Also, the "synthethic order" is almost always preserved in the gene sequence order in the operon, but experiments disrupting this expression "structure" did not affect the synthetic function[90]. Perhaps these details have something to do with the evolution of the system, which suggests the linearity is possibly related to organization of the nature of our model.

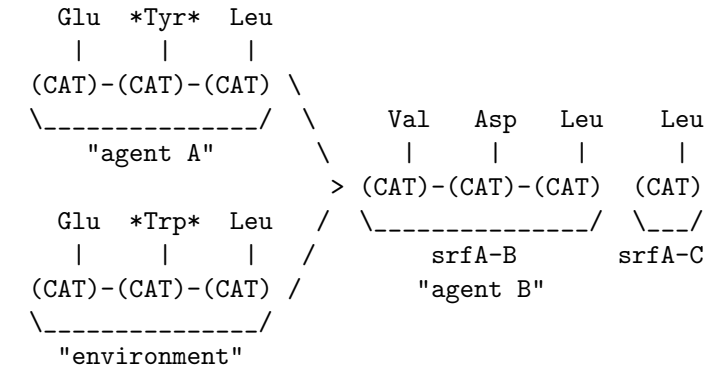
There appears to be specificity between chains, but there is also some degree of "mix and match", too, in that synthesis does not always follow the "canonical" sequences of chains[95]. However, the specificity between chains (or, analogously, the preference for a particular input) appears to involve two mechanisms:

1. Non-covalent interactions between chains, eg. the N- and C- terminus of sequential chains bind specifically[95], and
2. Condensation domains have some preferences for the "upstream" amino acid (side-chain size, charge, polarity, chirality, etc)[92,96].

As for this dissertation, this system seems like a good candidate for experimental tests. To clarify, one would use modules as both agents and resources, eg:



Using the surfactin pathway as a model system, one could engineer variations on the srfA-A module to require a new, unique amino acid. Due to the biomedical interest in NRPS products, and the relative ease of working with such a modular system, there has been work done on modifying and engineering new chains[92,93].



So the Tyr/Trp concentrations effectively control the "resource production of agent A" and the "environmental shielding", respectively. With this, some simple, potential experiments:

Microscopy. Apply fluorescent pairs on the *agent A* and *agent B* chains. With increasing Tyr concentration, fluorescence transfer should increase ("activation"). With increasing Trp, transfer should decrease ("shielding").

Kinetics. The two chains need to find each other spatially and properly orient for the surfactin synthesis to occur. With a constant supply of Tyr, we assume that the two chains will remain in close proximity and proper orientation. In the absence of Tyr,

the two chains will diffuse apart. An assay for surfactin – with sufficient time-resolution – could measure the reaction rate of the multi-component assembly. Potentially some form of pulse-flow experiments might prove useful: starting "cold", there would be a lag time before surfactin is produced. Our model would predict an increase in the reaction rate, approaching some maximum. After reaching this maximum, suspend the Tyr input for various short periods of time. If NRPS is an example of our model, the reaction rate should recover with shorter lag times as a function of the time of the Tyr input was suspended.

Chapter 6

Conclusion

6.1 A general theory of organization

This dissertation is concerned with the origin of a peculiar class of systems. These "composite systems" are complex organizations of multiple components performing a specific function. Our specific focus is in the physical principles that could drive this process of organization, and their ability to explain the origin of organized, group behavior from simple components.

Biological evolution is certainly capable of producing these composite systems, but it would appear our model can achieve the same with far fewer "infrastructural" assumptions. That is, our mechanism is not simply a restatement of biological evolution. Although they share a number of similarities, our model is not just an equivalent, but simplified, version. Our hypothesis is a natural organizing process independent – perhaps parallel, perhaps a precursor, perhaps even more fundamental – to the "self-replication" notions underlying conventional thinking on complex organization.

And this same disconnect is perhaps the basic motivation for the various extensions on the neo-Darwinian mechanism: Kauffman and self-organization[97], Margulis and symbiotic theory[98], Kirschner and facilitated variation[7]. Self-replication is not the fundamental unit of organization; rather, it is a particularly powerful and productive instance of a more fundamental process of organization.

The question is how the notion of the "attraction & shielding" principle applies here. This dissertation has largely assumed association is a consequence of physical proximity, but that has perhaps limited our intuition of the model. In biology, attraction and

shielding uses the terms of phenotype and fitness landscapes: it is the selective pressure – of pre-existing function – on the underlying genetics that guides organization[99,100].

A functional agent on the lattice creates a "probability well" on the lattice that dependent agents are more likely to explore with random physical movement. A primitive mitochondria creates a similar "probability well" on the fitness landscape that primitive eukaryotes are likely to explore with random genetic variation. It is possible that a mechanism for improving "spatial co-localization in a chemical model" is ultimately the same thing as a mechanism for improving "fitness landscape gradients in a biological model".

If so, this suggests that we are near a more general theory of organization. The functional relationship of components is the basic principle. The mechanisms of association vary in detail, but the important question is to understand how the function of a pre-existing system comes to define the structural properties of its subsequent system.

6.2 Summary of findings

For better or worse, we have likely raised more questions than we have answered. Our fundamental idea is that specifically entangled, functional relationships are the "crossroads" of organization. This is certainly not the only mechanism of organization at work in the world, but of those known, it is one of few that is truly responsive to the nature of the environment – it is one of few that are deeply adaptive to their world.

In a sense, this model defies the notion that "structure defines function" by adding the equally important idea that "function defines structure". This is not a thermodynamic system: it is not crystal formation. The system does not develop solely in response to structural variation. Rather, functional relationships define new structure, which define new functions, which define new structure, and so on. It is a constant exchange, in a way far deeper than the boolean (yes/no, survive/die) response that function plays in a purely neo-Darwinian model.

We have demonstrated the potential of this mechanism for meaningful, even profound, organization is systems as simple as trivial catalytic chemistry. Our chemical model is specific and based on a plausible physical basis. It demonstrates the ability for basic cooperation, competition, and hierarchical development. We expect there to be more diverse chemical mechanisms with the same basic qualities. Our approach is simply an initial example.

We also expect that these basic mechanisms are not bound to chemical systems. The basic premise – systems organize around specific functional relationships – is not deeply tied to the specifics of this model. The analogies are clear, whether in economics, sociology, or science. Entangled function defines the available paths of adaptive change.

Bibliography

[1] Mayr E. *What Evolution Is*. Basic Books, New York, 2001.

[2] Hebb DO. *The Organization of Behavior*. Wiley, New York, 1949.

[3] Kelso SR, Ganong AH, and Brown TH. Hebbian synapses in hippocampus. *Proceedings of the National Academy of Sciences*, 83:5326–5330, 1986.

[4] Goda Y and Davis GW. Mechanisms of synapse assembly and disassembly. *Neuron*, 40:243–264, 2003.

[5] Alberts B, Johnson A, Lewis J, M Raff, Roberts K, and Walter P. *Molecular Biology of the Cell*, page 1238. Garland Science, New York, 2002.

[6] Ferrara N, Gerber HP, and LeCouter J. The biology of vegf and its receptors. *Nature Medicine*, 9:669–676, 2003.

[7] Kirschner MW and Gerhardt JC. *The Plausibility of Life*, pages 143–176. Yale Press, New Haven, CT, 2005.

[8] Goss S, Aron S, Deneubourg JL, and Pasteels JM. Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, 76:579–581, 1989.

[9] Detrain C, Natan C, and Deneubourg JL. The influence of the physical environment on the self-organised foraging patterns of ants. *Naturwissenschaften*, 88:171–174, 2001.

[10] Lenski RE, Ofria C, Pennock RT, and Adami C. The evolutionary origin of complex features. *Nature*, 423:139–144, 2003.

[11] Epstein JM and Axtell R. *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, Cambridge, MA, 1996.

[12] Hanczyc MM, Fujikawa SM, and Szostak JW. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science*, 302:618–622, 2003.

[13] Miller SL. A production of amino acids under possible primitive earth conditions. *Science*, 117:528–529, 1953.

[14] Miller SL and Urey HC. Organic compound synthesis on the primitive earth. *Science*, 130:245–251, 1959.

[15] Kvenvolden K, Lawless J, Pering K, Peterson E, Flores J, Ponnamperuma C, Kaplan IR, and Moore C. Evidence for extraterrestrial amino-acids and hydrocarbons in the murchison meteorite. *Nature*, 228:923–926, 1970.

[16] Kochavi E, Bar-Nun A, and Fleminger G. Substrate-directed formation of small biocatalysts under prebiotic conditions. *Journal of Molecular Evolution*, 45:342–351, 1997.

[17] Hartman H. Speculations on the origin and evolution of metabolism. *Journal of Molecular Evolution*, 4:359–370, 1975.

[18] Rode BM. Peptides and the origin of life. *Peptides*, 20:773–786, 1999.

[19] Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, and Cech TR. Self-splicing rna: autoexcision and autocyclization of the ribosomal rna intervening sequence of tetrahymena. *Cell*, 31:147–157, 1982.

[20] Bartel DP and Szostak JW. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261:1411–1418, 1993.

[21] Yao S, Ghosh I, Zutshi R, and Chmielewski J. Selective amplification by auto- and cross-catalysis in a replicating peptide system. *Nature*, 396:447–450, 1998.

[22] Breaker RR and Joyce GF. Self-incorporation of coenzymes by ribozymes. *Journal of Molecular Evolution*, 40:551–558, 1995.

[23] Wolf YI and Koonin EV. On the origin of the translation system and the genetic code in the rna world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct*, 2:14, 2007.

[24] Segre D, Ben-Eli D, Deamer DW, and Lancet D. The lipid world. *Origins of Life and Evolution of the Biosphere*, 31:119–145, 2001.

[25] Woese CR. On the evolution of cells. *Proceedings of the National Academy of Sciences*, 99:8742–8747, 2002.

[26] Paley W and Paxton J. *Natural theology*, page 6. Gould and Lincoln, 1860.

[27] Cicero MT. *The Nature of the Gods (De Natura Deorum)*, page 159. Penguin Classics, 1972.

[28] Miser HD, Trimble KW, and Paige S. The rainbow bridge, utah. *Geographical Review*, 13:518–531, 1923.

[29] Murray JD. *Mathematical Biology*. Springer, New York, 2002.

[30] Garnier S, Gautrais J, and Theraulaz G. The biological principles of swarm intelligence. *Swarm Intelligence*, 1:3–31, 2007.

[31] de Duve C. Selection by differential molecular survival: a possible mechanism of early chemical evolution. *Proceedings of the National Academy of Sciences*, 84:8253–8256, 1987.

[32] Smith E and Morowitz HJ. Universality in intermediary metabolism. *Proceedings of the National Academy of Sciences*, 101:13168–13173, 2004.

[33] Doudna JA and Lorsch JR. Ribozyme catalysis: not different, just worse. *Nature Structural & Molecular Biology*, 12:395–402, 2005.

[34] Oberbeck VR and Fogleman G. Estimates of the maximum time required to originate life. *Origins of Life and Evolution of the Biosphere*, 19:549–560, 1989.

[35] Lazcano A and Miller SL. How long did it take for life to begin and evolve to cyanobacteria? *Journal of Molecular Evolution*, 39:546–554, 1994.

[36] Hartman P and Perdok WG. On the relations between structure and morphology of crystals. *Acta Crystallographica*, 8:49–52, 1955.

[37] Shubik M. Game theory, behavior, and the paradox of the prisoner's dilemma: Three solutions. *Journal of Conflict Resolution*, 14:181–193, 1970.

[38] Axelrod R and Dion D. The further evolution of cooperation. *Science*, 242:1385–1390, 1988.

[39] Lomborg B. Game theory vs. multiple agents: The iterated prisoner's dilemma. *Artificial Social Systems*, 4:69–93, 1994.

[40] Sandholm TW and Crites RH. Multiagent reinforcement learning in the iterated prisoner's dilemma. *BioSystems*, 37:147–166, 1996.

[41] Huck S, Normann HT, and Oechssler J. Zero-knowledge cooperation in dilemma games. *Journal of Theoretical Biology*, 220:47–54, 2003.

[42] Akiyama E and Kaneko K. Dynamical systems game theory and dynamics of games. *Physica D*, 147:221–258, 2000.

[43] Watts DJ and Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[44] Jeong H, Tombor B, Albert R, Oltvai ZN, and Barabási AL. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[45] Albert R and Barabási AL. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[46] Barabási AL and Albert R. Emergence of scaling in random networks. *Science*, 286: 509–512, 1999.

[47] Vázquez A, Dobrin R, Sergi D, Eckmann JP, Oltvai ZN, and Barabási AL. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proceedings of the National Academy of Sciences*, 101: 17940–17945, 2004.

[48] Milo R, Shen-Orr S, Itzkovitz S, Kastan N, Chklovskii D, and Alon U. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[49] Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EZ, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, and Young

RA. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298: 799–804, 2002.

[50] Shmulevich I, Kauffman SA, and Aldana M. Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proceedings of the National Academy of Sciences*, 102: 13439–13444, 2005.

[51] Fox JJ and Hill CC. From topology to dynamics in biochemical networks. *Chaos*, 11: 809–815, 2001.

[52] Carlson J and Doyle J. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60:1412–1427, 1999.

[53] Bak P, Tang C, and Wiesenfeld K. Self-organized criticality. *Physical Review A*, 38: 364–374, 1988.

[54] Csete ME and Doyle JC. Reverse engineering of biological complexity. *Science*, 295: 1664–1669, 2002.

[55] Solé RV, Manrubia SC, Benton M, Kauffman S, and Bak P. Criticality and scaling in evolutionary ecology. *Trends in Ecology & Evolution*, 14:156–160, 1999.

[56] Bak P and Paczuski M. Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences*, 92:6689–6696, 1995.

[57] Carlson JM and Doyle J. Complexity and robustness. *Proceedings of the National Academy of Sciences*, 99:2538–2545, 2002.

[58] Ódor G. Universality classes in nonequilibrium lattice systems. *Reviews of Modern Physics*, 76:663, 2004.

[59] Sagues F and Epstein I. Nonlinear chemical dynamics. *Dalton Transactions*, pages 1201–1217, 2003.

[60] Strogatz SH. *Nonlinear Dynamics and Chaos*. AddisonWesley, Reading, MA, 1994.

[61] Essam JW. Percolation theory. *Reports on Progress in Physics*, 43:834–912, 1980.

[62] Suchecki K, Eguíluz VM, and San Miguel M. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Physical Review E*, 72:036132–036139, 2005.

[63] Cardy J and Täuber UC. Theory of branching and annihilating random walks. *Physical Review Letters*, 77:4780–4783, 1996.

[64] Ziff RM, Gulari E, and Barshad Y. Kinetic phase transitions in an irreversible surface-reaction model. *Physical Review Letters*, 56:2553–2556, 1986.

[65] Voigt CA and Ziff RM. Dynamic behavior of the monomermonomer surface reaction model with adsorbate interactions. *Journal of Chemical Physics*, 107:7397–7401, 1997.

[66] Wächtershäuser G. Before enzymes and templates: theory of surface metabolism. *Microbiological Reviews*, 52:452–484, 1988.

[67] Gillespie DT. General method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

[68] Behe M. *Darwins Black Box.* Free Press, New York, 1996.

[69] Orgel LE. The origin of life–a review of facts and speculations. *Trends in Biochemical Sciences*, 23:491–495, 1998.

[70] Joyce GF and Orgel LE. *The RNA World*, pages 49–77. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1999.

[71] Eigen M and Schuster P. Stages of emerging life–five principles of early organization. *Journal of Molecular Evolution*, 19:47–61, 1982.

[72] Eigen M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

[73] Wright MC and Joyce GF. Continuous in vitro evolution of catalytic function. *Science*, 276:614–617, 1997.

[74] Keefe AD and Szostak JW. Functional proteins from a random-sequence library. *Nature*, 410:715–718, 2001.

[75] Kauffman SA. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119: 1–24, 1986.

[76] de Duve C. *Blueprint for a Cell*, pages 123–170. Neil Patterson, Burlington, NC, 1991.

[77] Orgel LE. Self-organizing biochemical cycles. *Proceedings of the National Academy of Sciences*, 97:12503–12507, 2000.

[78] Troisi A, Wong V, and Ratner MA. An agent-based approach for modeling molecular self-organization. *Proceedings of the National Academy of Sciences*, 102:255–260, 2005.

[79] Jain S and Krishna S. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proceedings of the National Academy of Sciences*, 98:543–547, 2001.

[80] Pfeiffer T, Soyer OS, and Bonhoeffer S. The evolution of connectivity in metabolic networks. *PLoS Biology*, 3:1269–1275, 2005.

[81] Turing AM. A physical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London*, 237:3772, 1952.

[82] Zumdahl SS and Zumdahl SZ. *Chemistry*, chapter 4. Houghton Mifflin, 2000.

[83] Tinoco I, Sauer K, and Wang JC. *Physical Chemistry*, chapter 8. Prentice Hall, 1995.

[84] Stryer L. *Biochemistry*. W.H. Freeman & Company, 1995.

[85] Gould SJ and Vrba ES. Exaptation; a missing term in the science of form. *Paleobiology*, 8:4–15, 1982.

[86] Brosius J. Rnas from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*, 238:115–134, 1999.

[87] Gordon R. Evolution escapes rugged fitness landscapes by gene or genome doubling: the blessing of higher dimensionality. *Computers & Chemistry*, 18:325–331, 1994.

[88] Melius P. Non-random non-ribosomal assembly of amino acids in proteins and proteinoids. *BioSystems*, 11:125–132, 1979.

[89] Challis GL and Naismith JH. Structural aspects of non-ribosomal peptide biosynthesis. *Current Opinion in Structural Biology*, 14:748–756, 2004.

[90] Moffitt MC and Neilan BA. The expansion of mechanistic and organismic diversity associated with non-ribosomal peptides. *FEMS Microbiology Letters*, 191:159–167, 2000.

[91] Cane DE, Walsh CT, and Khosla C. Harnessing the biosynthetic code: Combinations, permutations, and mutations. *Science*, 282:63–68, 1998.

[92] Finking R and Marahiel MA. Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology*, 58:453–488, 2004.

[93] Eppelmann K, Stachelhaus T, and Marathiel MA. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*, 41:9718–9726, 2002.

[94] Marahiel MA. Multidomain enzymes involved in peptide synthesis. *Federation of European Biochemical Societies*, 307:40–43, 1992.

[95] Hahn M and Stachelhaus T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proceedings of the National Academy of Sciences*, 101:15585–15590, 2004.

[96] Belshaw PJ, Walsh CT, and Stachelhaus T. Aminoacyl-coas as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, 284:486–489, 1999.

[97] Kauffman SA. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[98] Margulis L. *Origin of Eucaryotic Cells*. Yale University Press, New Haven, CT, 1970.

[99] Watson RA, Reil T, and Pollack JB. Mutualism, parasitism, and evolutionary adaptation. *Proceedings of Artificial Life*, 7:170–178, 2000.

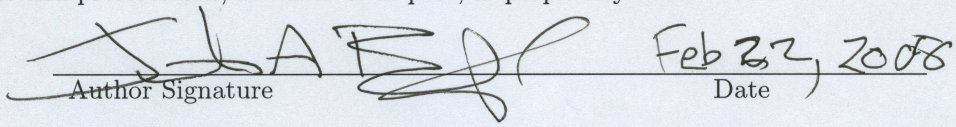
[100] Watson RA and Pollack JB. A computational model of symbiotic composition in evolutionary transitions. *BioSystems*, 69:187–209, 2003.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature Date