Sequence selection by dynamical symmetry breaking in an autocatalytic binary polymer model

Fellermann, Harold; Tanaka, Shinpei; Rasmussen, Steen

Published in:
Physical Review E

DOI:
10.1103/PhysRevE.96.062407

Publication date:
2017

Document version
Final published version

Citation for published version (APA):
https://doi.org/10.1103/PhysRevE.96.062407

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
Sequence selection by dynamical symmetry breaking in an autocatalytic binary polymer model

Harold Fellermann*
Interdisciplinary Computing and Complex Biosystem Research Group, School of Computing, Newcastle University, I Science Square, Newcastle upon Tyne, NE4 5TG, United Kingdom

Shinpei Tanaka
Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima 739-8521, Japan

Steen Rasmussen
Center for Fundamental Living Technology (FLinT) Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark and Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, New Mexico 87501, USA

(Received 15 August 2017; published 13 December 2017)

Template-directed replication of nucleic acids is at the essence of all living beings and a major milestone for any origin of life scenario. We present an idealized model of prebiotic sequence replication, where binary polymers act as templates for their autocatalytic replication, thereby serving as each other’s reactants and products in an intertwined molecular ecology. Our model demonstrates how autocatalysis alters the qualitative and quantitative system dynamics in counterintuitive ways. Most notably, numerical simulations reveal a very strong intrinsic selection mechanism that favors the appearance of a few population structures with highly ordered and repetitive sequence patterns when starting from a pool of monomers. We demonstrate both analytically and through simulation how this “selection of the dullest” is caused by continued symmetry breaking through random fluctuations in the transient dynamics that are amplified by autocatalysis and eventually propagate to the population level. The impact of these observations on related prebiotic mathematical models is discussed.

DOI: 10.1103/PhysRevE.96.062407

I. INTRODUCTION

The ability of nucleic acids to serve as templates for their own replication via the well-known Watson-Crick pairing is at the heart of all life known today. Consequently, the onset of such chemical sequence information and its replication are crucial cornerstones in most theories about the origin of life [1,2].

While the appearance of first molecular replicators is still subject to ongoing debate [3,4], we suspect [5,6] that catalysis combined with replication of the catalytic molecules is the critical molecular invention. Catalysis enables access to energy and resources while replication both preserves and enables combinatorial exploration of the catalyst. Competition over common and scarce resources would cause a selection pressure on the catalyst that directs the random exploration process of copy error mutations. The resulting evolutionary search would eventually select replicator species with increasingly advantageous catalytic properties [7]—thus exemplifying Darwinian chemical evolution.

Several simple models have been developed to elucidate the prebiotic emergence and evolution of nucleic acid replicators. Most notably, random catalytic network models have been introduced to predict the emergence of autocatalytic cycles in pools of arbitrary molecules with random cross-catalytic activity [8–10]. More closely related to nucleic acids are catalytic polymer models, where the catalyzed polymer ligation reaction is modeled as concatenation of random strings [11–19]. Most studies of polymer models concern themselves with the appearance of catalytic cycles when the catalytic activity of polymers is assigned either randomly or dependent on the sequence around the ligation site. Common conclusion of these studies is that random catalytic activity among a set of molecules is sufficient to promote the appearance of reflexively autocatalytic sets, i.e., groups of molecules that are able to jointly replicate themselves through cross catalysis, see, e.g., Refs. [9,13,20].

In this study, we take the appearance of catalysis and replication for granted and raise the following critical questions: Do catalysis, replication, competition, cooperation, and selection indeed allow for the free exploration of the chemical sequence space or are some prebiotic evolutionary fates more common than others? If the latter, then what is driving the evolution in such chemical networks? Ultimately, we want to better understand how prebiotic chemistry with reversible reactions can give rise to the emergence of history, i.e., can lead to a series of path dependent events, where the outcome of a past selection event impacts what becomes accessible to evolution in the future.

In order to focus on these questions, we take a radical stance and study a polymer model in which autocatalysis is assumed to work perfectly for all molecular species. There are several reasons to do so. First, if we believe chemical evolution to have evolved more and more effective self-replicators, then it should be just as informative to understand the limiting case of perfect autocatalysis as it is to understand the limiting case of completely random chemistry. Second, by demanding perfect autocatalysis, we are able to derive closed analytic expressions for all involved molecular species. These allow us to scrutinize the reaction system with unprecedented analytic rigor, and we are able to align our simulation results with the analytic derivations.

*harold.fellermann@newcastle.ac.uk
We have previously reported on this model of self-replicating binary polymers where exact autocatalytic replication introduces a selection bias toward few specific replicator motifs and population structures [21]. Here we present a full analysis of this system and several prebiotically motivated model variants.

In the remainder of the article, we give a detailed exposition of our simple binary polymer model with exact autocatalysis. We derive analytical results for the two important limiting cases of purely random as well as purely autocatalytic replication. We proceed by presenting simulation results of the prebiotically interesting case, where autocatalysis is dominant but augmented with relatively rare random ligation. For this scenario we reveal a strong inherent selection pressure toward replication strands and to transform them into their activated single-strand configuration.

Since $\mathcal{A}^*$ is enumerable, we can take it as basis for an infinite-dimensional state space: Each state vector $x$ denotes a population within which every species $k \in \mathcal{A}^*$ spans one dimension and its associated coordinate $x_k \in \mathbb{R}^+$ specifies the concentration of the respective species. Standard dynamical systems analysis is complicated by the infinite dimensionality of this state space. Previous studies circumvented this difficulty by introducing a maximal strand length [15,17] or focusing on very small monomer pools [18,25]. Here we present a mathematical formalism that is able to deal with the infinite dimensionality for the most part analytically.

Assuming mass action kinetics, we can derive dimensionless reaction kinetic equations for the molar concentrations $x_k$ of species $k \in \mathcal{A}$ as

$$\frac{dx_k}{dt} = \left( \sum_{j,k \in \mathcal{A}^*} x_j - \sum_{i,j \neq k} x_k \right) + \alpha \left( \sum_{i,j \neq k} x_i x_j - \sum_{k \in \mathcal{A}^*} x_j x_k \right) + \beta \left( \sum_{i,j \neq k} x_i x_j x_k - \sum_{j,k \in \mathcal{A}^*} x_i x_j x_k \right) = f_k,$$

where we have chosen a time scale $t = c_0^{-1}$, and introduced the nondimensional random ligation rate $\alpha = c_1/c_0$, and the nondimensional autocatalytic ligation rate $\beta = c_2/c_0$. The above summations consider every pair of species $i$ and $j$ which satisfy $i,j = k$ and $k,j = i$ or $j,k = i$. Formally, the sum operators are defined by means of the indicator functions $1_{i,j = k} = 1$ if $i,j = k$ else 0 as follows:

$$\sum_{i,j \neq k} := \sum_{i,j \in \mathcal{A}^*} 1_{i,j \neq k},$$

$$\sum_{j,k = i \atop k,j = i} := \sum_{i,j \in \mathcal{A}^*} (1_{j,k = i} + 1_{k,j = i}).$$

Note that the second definition may involve an infinite sum and is only defined if the right-hand side is defined. For finite-size systems, which we consider in this article, the sum is always defined and allows us to write down the reaction equation system in a compact closed form without introducing any arbitrary truncation.

Equation (4) allows us to perform standard stationarity and stability analysis for the two limiting cases of purely random ligation as well as purely autocatalytic ligation.
A. Purely random and purely autocatalytic limit cases

In the absence of autocatalysis ($\beta = 0$), the system approaches a stationary and stable equilibrium distribution, where the concentration of each polymer decays exponentially with its length:

$$
(x_{\text{exp}}^*)_k = \frac{1}{\alpha} e^{-b[k]},
$$

which can be readily confirmed by inserting (7) into (4) for $\beta = 0$. Here $b$ is a constant determined by the boundary condition. Importantly, there is no selection for specific sequence patterns.

We will now show that this situation changes radically in the limit of purely autocatalytic ligation ($\alpha = 0$). First, since decomposition is now the only reaction that introduces novel species, any stationary state can only consist of initially present strands or their substrands. For any strand that is populated in the stationary state, all its substrands must be populated as well. This can be proven by contradiction: Assume there exists a steady state $x^*$ with $x^*_m > 0$ but $x^*_k = 0$ for some $k, j = m$. From Eq. (4) it follows:

$$
\frac{dx_k}{dt} \bigg|_{x^*} = \sum_{i \neq j = k \neq j} \frac{d(x_i x_j x_k)}{dx_i} + \beta \left( \sum_{i \neq j = k \neq j} \frac{d(x_i x_j x_k)}{dx_i} - \sum_{j \neq k = i \neq j} \frac{d(x_i x_j x_k)}{dx_i} \right) = 0
$$

since $f_k(x^*) > 0$, $x^*$ cannot have been a steady state. Thus, the assumption is false and species $k$ must be populated as well.

We call any strand that is not a substrand of any other strand in the population a “chief” and the set of its substrands its “clan.” An example of such a “chief-clan” structure is shown in Fig. 1. We denote the set of all chiefs of $x$ by $A^\dagger$ and the set of all clan members by $A^\dagger_k$. Using this nomenclature, the stationary chiefs of the purely autocatalytic case are those substrands of initially present strands that survived decomposition and are not substrands of other surviving species.

Interestingly, under our assumption of exact replication (3), stationary species concentrations no longer decay exponentially with their length. Instead, any proper clan member is maintained at a constant concentration given by the replication rate constant $\beta$:

$$
x^*_k = \begin{cases} 
\beta^{-1/2}, & k \in A^\dagger - A^\dagger_k, \\
0, & k \not\in A^\dagger_k 
\end{cases}
$$

This can again be readily confirmed by inserting (9) into (4) with $\alpha = 0$.

Concentrations of chief species $m \in A^\dagger$ are unconstrained in the steady state, apart from what is determined by the boundary condition. This is so because the steady state condition for chiefs

$$
\frac{dx_m}{dt} \bigg|_{x^*} = -\sum_{i, j = m} x^*_i x^*_j x^*_m + \beta \sum_{i, j = m} x^*_i x^*_j x^*_m
$$

$$
= x^*_m \left( \beta \sum_{i, j = m} -1 - \sum_{i, j = m} 1 \right) = 0
$$

is always fulfilled, independent of $x_m^*$.

To understand how stationary solutions respond to fluctuations, we derive the entries of the functional matrix as follows:

$$
\frac{\partial f_k}{\partial x_i} \bigg|_{x^*} = \sum_{i \neq j = k, j \neq i} \frac{\partial x_k}{\partial x_i} - \sum_{i \neq j = k} \frac{\partial x_k}{\partial x_i} + \beta \left( \sum_{i \neq j = k} \frac{\partial(x_i x_j x_k)}{\partial x_i} - \sum_{j \neq k = i \neq j} \frac{\partial(x_i x_j x_k)}{\partial x_i} \right) = \left( \sum_{j \neq k = i \neq j} 1 - \sum_{j \neq k = i \neq j} 1 - \delta_{ij} \sum_{j \neq k = i \neq j} 1 \right),
$$

where $\delta_{ij} = 1$ if $i = j$, else 0 is the Kronecker symbol, and chief concentrations are assumed to be $\beta^{-1/2}$ for simplicity.

The equation simply expresses a detailed balance relation: Increasing the amount of any clan species results in a flux to any species that uses the increased strand as prefix or suffix, while simultaneously introducing a flux away from all species needed as reactants for the formation of those longer species. Again, increasing the level of chief species ($l = m$) does not introduce any flux in the population structure. Notably,

the stability of a stationary population is entirely determined by the topology of the reaction graph, not by the reaction rates.

To fully determine the stability of stationary solutions requires solving for the eigenvalue spectrum of the functional matrix. The analytical treatment is complicated by the fact that the dimension of our system is formally infinite. We therefore resolve to numerically probe the eigenvalue spectrum of hundreds of different trial populations (see Table I). We find
TABLE I. Stability of clans with different number of chiefs.

<table>
<thead>
<tr>
<th></th>
<th>$P_{ss}$</th>
<th>$P_{st}$</th>
<th>$P_{sc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.40</td>
<td>0.60</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.20</td>
<td>0.80</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>0.07</td>
<td>0.93</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>0.02</td>
<td>0.98</td>
<td>0.00</td>
</tr>
<tr>
<td>14</td>
<td>0.01</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>18</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Clans with two chiefs

<table>
<thead>
<tr>
<th></th>
<th>$P_{ss}$</th>
<th>$P_{st}$</th>
<th>$P_{sc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.72</td>
<td>0.28</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.21</td>
<td>0.79</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.04</td>
<td>0.95</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td>12</td>
<td>0.00</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>14</td>
<td>0.00</td>
<td>0.95</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Clans with four chiefs

<table>
<thead>
<tr>
<th></th>
<th>$P_{ss}$</th>
<th>$P_{st}$</th>
<th>$P_{sc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.82</td>
<td>0.18</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.19</td>
<td>0.81</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td>12</td>
<td>0.00</td>
<td>0.99</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: $|m|$, the length of the chief(s); $P_{ss}$, the probability to find stable spirals; $P_{st}$, the probability to find stable nodes; $P_{sc}$, the probability to find unstable stationary points. The number of sampling was $10^4$. Only combinations of chiefs that had the same number of “0” and “1” were sampled.

that the vast majority of stationary chief-clan structures have indeed no positive eigenvalues. Moreover, we find that the number of zero eigenvalues either equals the number of chiefs in the population or is one greater than that. Up to two of these zero eigenvalues result from mass conservation. The remaining ones express the fact that chief-clan structures with competing chiefs are connected by lines of linearly stable solutions. For example, in a population with competing chiefs 01 and 10, all populations that satisfy the constant solution $x_0 = x_1 = x^*$ are linearly stable, independent of how material is distributed among the chiefs.

We note that linear stability analysis only detects local stability and does not discriminate the relative stability of one solution over another. Moreover, our analysis does not consider boundary conditions, which might render some chief-clan structures less likely than others if not outright impossible.

How many potential equilibrium solutions are there? We can roughly estimate the potential stable chief-clan structures as a function of the longest population member as follows: there are $2^{|m|}$ different chiefs of length $|m|$. If any of these chiefs can be either absent or present in the population, then there are $2^{2^{|m|}}$ different potential steady-state populations. Despite the constraints of Eq. (9), there are superexponentially many stable potential equilibrium populations. Even for strands as short as 10, this amounts to more than $10^{300}$ possible stationary chief-clan structures. One should bear in mind, however, that this estimation ignores the boundary condition imposed by the initial (and conserved) material amount.

So far our analysis has been based on molar concentrations. If material is sparse, however, then it is more accurate to adopt the view that molecular species are present in discrete amounts. Under this view, the continuous equations (1)–(3) still describe the expected tendencies of the dynamic process but with the difference that species can become extinct and disappear entirely rather than being diluted to arbitrary small (but nonvanishing) concentrations. Under pure autocatalysis, this is especially relevant for chief species, as they are not constrained to potentially high-equilibrium amounts. If chief species disappear as the result of random fluctuations, then they become permanently extinct.

B. Combined random and autocatalytic ligation

We now study systems where both ligation reactions act in combination. In particular, we analyze how the system transitions from the exponential equilibrium solution to stable chief-clan structures when varying the ratio of the random and autocatalytic ligation rates.

To answer this question, we sample the stochastic process defined through (1)–(3) via exact stochastic simulation using a Gillespie algorithm. In order to tackle the dimensionality of the problem, the algorithm generates for the current state $x(t)$ a list of potential follow-up states on the fly and evaluates transition rates $r_i$ according to Eq. (4). For any finite-size population, this transition list will also be finite.

Simulations are started from a monomer pool with $x_0(0) = x_{00}(0) = 2100$, $\beta = 10^{-3}$, and $\alpha$ varying from $10^{-3}$ to $10^{-5}$. Under sufficiently rare random ligation, we expect these settings to stabilize chief-clan structures up to length 10, and under only random ligation, an exponential distribution up to length eight. With the given $\beta$, in a population distributed according to (9), random and autocatalytic ligation would be equally likely for a random ligation rate constant of $\alpha = \beta^{1/2} = 10^{-2}$.

After $t = 1000$ time units, populations are compared to both equilibrium solutions (7) and (9) using the cosine similarities

$$s_{ss} = \frac{1}{\alpha^{|A^*|}} \sum_{k \in A^*} x_k e^{-\beta k^{1/2}}$$

$$s_{sc} = \frac{1}{\alpha^{|A^*|}} \sum_{k \in A} x_k e^{-\beta k^{1/2}}$$

where $A_0 = A^* \setminus \partial A^*$, $s_{ss}$ is the distance toward the corresponding exponential equilibrium, and $s_{sc}$ measures the distance to the smallest enclosing chief-clan structure, i.e., the smallest chief-clan structure that has a nonzero species count for any species that is present in the population. Figure 2 shows the transition from stable chief-clan structures into random distributions. With increasing random ligation rate...
FIG. 2. Cosine similarity between simulated population samples and ideal exponential distributions (red) and between the samples and their smallest enclosing chief-clan structure (blue), the latter for different threshold values. Species with occupancies below the threshold are not considered for the distance calculation.

α, observed populations cease to resemble their original chief-clan structure. Visual inspection of the results revealed that these deviations are mainly caused by few random ligation products that are not numerous enough to establish their clan. When filtering out such outliers, we found that the remaining core populations resemble more closely their corresponding chief-clan structure, essentially until α reaches $10^{-2}$. At the same time, populations approach equilibrium distributions, because autocatalysis becomes too slow to generate chief-clan structures for the novel species created by random ligation. For high-enough α, deviation from the exponential equilibrium distribution is again mainly caused by rare long strands, whose expectation number from Eq. (7) is well below a single molecule.

These results indicate that chief-clan structures cannot develop above a critical random ligation rate less than $\beta^{1/2}$. Approaching this critical threshold, chief-clan structures become succeedingly obscured by random ligation products in low copy number.

C. Rare random ligation under strong autocatalysis

With the above insights we now concern ourselves with the prebiotically interesting parameter regime where comparatively rare random ligations occasionally introduce novel species into populations with strong autocatalytic replication. When starting from a pool of monomers, random ligation will create longer species that get amplified via autocatalytic replication and generate their clans via decomposition. We expect this process to generate chief-clan structures satisfying Eq. (9) of a size that is just able to accommodate the overall amount of material.

Our question now becomes the following: Does autocatalytic replication with rare random ligation generate all of the overwhelmingly many potential chief-clan structures or is there some reason for some chief-clan structure to be more prominent than others?

Using the results of 1000 simulation runs (simulated until $t = 100$), we perform single-linkage hierarchical clustering based on cosine distance,

$$d(x, x') = 1 - \frac{\sum_{k \in A^*} x_k x'_k}{|x||x'|}.$$  (14)

We find that about 77% of the simulation results concentrate in six clusters with characteristic population structures, shown in Fig. 3. This is a very strong selection pressure, given that all our kinetic equations are sequence agnostic—especially under the light of the results of the previous discussion.
Interestingly, the selected populations have very few chiefs composed of short repeating sequence motifs such as 01 in cluster of Fig. 3(a) (bootlace), 0011 in clusters of Fig. 3(b) and 3(f) (pinecone1), and 00001111 and 11110000 in clusters of Fig. 3(c) and 3(d) (pinecone2). In about 2% of the simulation runs, we even observe complete demixing of the monomers into homopolymers composed exclusively of 0’s and 1’s, respectively [cluster in Fig. 3(e), two-towers].

We also observe that some clusters exhibit an “inverted” population structure where chiefs are more populated than other clan members [Fig. 3(c), 3(d) and 3(f)]. This is due to the boundary condition where the total material amount is too low to maintain a longer chief but too high for the chief concentration to be comparable to the equilibrium value. As a consequence, the highly populated chief will more likely create—via random ligation—species outside its clan structure, which either die out or cause the whole population to transition into another structure. Indeed, we have observed in simulation that these metastable structures exhibit stronger random fluctuations and potentially transition into completely distinct chief-clan structures [21].

We emphasize that the stochastic simulation results are rather unexpected if not even in apparent contradiction to the results of the linear stability analysis derived above. If linear stability theory allows for almost any chief-clan structure to be (at least meta-)stable, then why does that stochastic simulation exhibit such a strong bias toward selecting very few and very regular population structures? What constitutes this “survival of the dullest”?

### III. MECHANISMS OF STRUCTURE SELECTION

We now propose two possible explanations for the symmetry breakdown in this replicator ecology: On the one hand, we identify a selection bias in the transient process of creating equilibrated, noninverted chief-clan structures. On the other hand, we identify a selection bias in the steady-state dynamics of an established equilibrated population. Both selection biases steer the system toward highly repetitive sequence patterns, and we hypothesize that both of them are responsible for the strong selection observed in stochastic simulation.

#### A. Selection in locally equilibrated populations

We first discuss the equilibrium scenario. Consider an equilibrated population structure that satisfies (9) and comprises two chiefs \(a\) and \(b\). The clans of these chiefs always overlap (because monomers take part in all clans) and the overlap forms a pool of common resources for the two clan structures to compete over. As in the end of the previous section, we assume random ligations to be sufficiently rare and use equilibrated chief-clan structures as initial conditions.

It is illustrative to first investigate a minimal system of four species, where two chiefs 01 and 10 compete over the shared resources 0 and 1. We know from the above that mass conservation of the two monomers fixes this system on a two-dimensional (2D) plane, which features a continuous line of marginally stable steady states due to condition (9). The line connects the two boundary states, formed by populations with only one surviving chief species.

In a stochastic treatment of this system, the line of connected equilibria translates into a drift term that vanishes for equilibrium solutions. Unlike in the deterministic treatment, random fluctuations can now redistribute material along the line of marginal steady states with virtually no deviation from equilibrium. The two boundary states are absorbing boundaries, as the walk terminates if one of the competing chiefs becomes extinct (and cannot be repopulated in the absence of random ligations).

Interestingly, diffusion of the random process also becomes position dependent: Randomly converting a single chief molecule from one species to the other requires the abundance of both chief species, one entering as reactant to be degraded and the other as catalyst to enhance its own ligation. Diffusion is therefore the highest when the product of the two chiefs is maximized and goes toward zero when approaching the boundaries, where either of the two species becomes sparse.

In the Appendix, we develop a one-dimensional random walk, constrained to the line of connected steady states, that approximates the two-dimensional case. For this approximation, we are able to analytically derive a Fokker-Planck equation that features zero drift and position-dependent diffusion:

\[
\frac{\partial}{\partial t}P(x_{01},t) \approx \frac{\partial^2}{\partial x_{01}^2}[D(x_{01})P(x_{01},t)],
\]

where

\[
D(x) = \frac{x_01(x_01-1)}{x_01+x_{01}}.
\]

The parabolic diffusion term has a remarkable effect on the first hitting time, i.e., the typical time of chief coexistence. Because diffusion is higher in the center of the manifold (where both chiefs are populated about evenly) the system is more prone to leave this state and move toward states of lower diffusion. This diffusion induced effective drift drives the system toward regions where one chief outnumbers the other. The average time for the system to reach one of the absorbing states is therefore orders of magnitude shorter than for a random walk process with comparable, but constant, diffusion. Figure 13 of the Appendix compares the cumulative probability functions of the hitting time distributions for the original two-dimensional system, our one-dimensional approximation, and a random walk process with constant diffusion.

Perhaps contrary to expectations raised by the previous linear stability results, we now see that there is a natural tendency for minority species to be driven into extinction. This effective drift is ultimately induced by position-dependent diffusion along the line of competing chief populations.

Note that this effective drift can only be observed among chiefs that can absorb all the material of their competitor: Otherwise, both chiefs can survive. For example, two chiefs 010 and 101 can coexist because neither can absorb all the material of the other. Yet, deviations in the chief-clan structure from the constant solution can sometimes absorb a chief with different material than its competitor, as we discuss now.

Figure 4 shows how a short chief, 00, can be driven into extinction as the result of competition with longer chiefs: Simulations are initialized with equilibrated bootlace structures of varying length, plus 00, and measure the probability of survival of the shorter chief after 1000 time units as a function of its initial concentration for varying lengths of the longer chief. In this case, chiefs in a bootlace structure, such as
00, cannot absorb all the material of 00, but the short chief can nevertheless be absorbed by fluctuations of the chief-clan structure, which in turn deviates from the exact solution. The simulation shows the clear tendency of extinction of shorter clans in competition.

While the above reveals the source of competition and extinction in our model, and indicates that simulations should generate populations with few chiefs, it does not explain why the dynamics select those particular chiefs that feature simple repetitive motifs.

An answer to this question lies in the way perturbations are distributed among competing clans, as shown schematically in Fig. 5: When species in the overlap of competing clans are perturbed through random decomposition of longer strands, detailed balance causes an opposing flow from the perturbed species toward longer strands. This outflow can be partitioned into three components: material flowing into species belonging exclusively to clan A, clan B, or their overlap C:

\[
\frac{\partial f_i^{A \cup B \cup C}}{\partial x_l} = \frac{\partial f_i}{\partial x_l},
\]

where we have introduced the “partial” flow \( \frac{\partial f_i^A}{\partial x_l} \) from \( l \) into the species in \( A \):

\[
\left. \frac{\partial f_i^A}{\partial x_l} \right|_{x^*} = -\left( \sum_{l_{j=A} \atop j \neq A} 1 + \sum_{l_j \in A} 2 \right) \approx - \sum_{l_j \in A} 1.
\]

Thus, the flow into \( A \) (\( B \)) scales with the number of strands in \( A \) (\( B \)) that have the perturbed strand as their prefix or suffix. Since regular chiefs can form longer strands with more repetitive sequences than irregular chiefs, we expect regular chiefs to outperform irregular ones in this competition.

**Summary.** In equilibrium, strands with high population numbers outcompete low-populated strings in competition for resources and this makes it difficult for randomly created chiefs to survive. In some cases this process can be described by a Fokker-Planck equation, which expresses a position-dependent diffusion. Further, long clans can absorb material faster than short clans because they offer more reaction pathways to absorb fluctuations (due to an increased number of prefix and suffix matches). With a given amount of material, regular chiefs can form longer clans than irregular ones.

**B. Selection far from equilibrium**

We now discuss the transient dynamics of a typical process of structure generation that starts from a pool of monomers,
where—as before—random ligation is small enough for the system to locally equilibrate after each random ligation event. Thus, after spontaneous formation of an intermediate chief through random ligation, the new chief will accumulate material via autocatalysis in order to satisfy the equilibrium condition (9). The result is an inverted population structure where chief species accumulate the majority of the material. Therefore, the next ligation is more likely to happen among two chiefs than among any other species in the population—thus creating a strand with repetitive sequence patterns [21]. Irregular chiefs, on the other hand, require additional substrands to be formed from now-less-abundant clan members. The process repeats until chief concentrations become lower than the equilibrium solution, at which point there is no surplus material to grow longer strands.

The above verbal argument can be refined into a quantitative estimation of selection likelihoods: Assuming that populations grow by irreversible random ligation and subsequent equilibration, the likelihood to obtain one population from a previous one via a single ligation event is proportional to the equilibrium abundance of the reactants, which is largely determined by the equilibrium condition (9). Assuming that the surplus material is evenly distributed among chief species, we can calculate the probability to obtain any final population from the initial condition by multiplying the conditional probabilities of ligations along each ligation pathway that can lead to the final population in question. However, because of the double exponential scaling of potential chief-clan structures as a function of chief length, the reaction system leads to a combinatorial explosion that quickly exceeds computing resources. Justified by Eq. (17) and the results of Fig. 4, we introduce the additional assumption that a new chief can only form in a population if it is at least as long as any existing chief. With this additional assumption, the combinatorial problem can be tackled.

We systematically perform these likelihood estimations for different initial monomer amounts and estimate the 90% most likely ligation pathways. The resulting population structures are classified according to their sequence motifs and compared to the results of stochastic simulations for selected pool sizes. The computed likelihoods—shown in Fig. 6—are generally in good agreement with the simulations.

**Summary.** During the transient, equilibration moves most material into chiefs. If there is a single chief, then it will most likely react with itself to create the next chief (scenario 01 + 01 → 0101). The key here are inverted population structures [21] where most material is in the chief species. To make irregular sequences during the transient, distinct building blocks are needed. If one chief has absorbed the material, then there is less material for the other chief, thus reducing the propensity of forming the irregular chief (scenario 10 + 0 → 100). Again most material is in the chief. To make a particular irregular chief during the transient, the two substrings have to ligate in correct order (scenario 10 + 0 → 100 vs. 0 + 10 → 010), so the symmetry of the reaction network plays as only one of two possible reactions produce a nonrepetitive sequence. Finally, regular chiefs require the build-up of fewer intermediate species, i.e., 01010101 requires only 01 and 0101, whereas 01110001 would require 01, 11, 00, 01, 0111, and 0001 or a similar number of other intermediates.

![Graph](image)

**FIG. 6.** Estimated likelihood to form populations of strands with certain motifs as a function of the monomer pool size with balanced monomer compositions. Note the exponential scale of the x axis. Overlaid bar charts show results of full stochastic simulations for various system sizes. The table lists population frequencies (sim) and their associated likelihood estimates (est) for the pool sizes probed in simulation. See text for details.

### C. Distinguishing the alternative selection mechanisms

We have presented two alternative driving mechanisms that might explain the selection of few chiefs with short repetitive motifs. Which of these mechanisms is responsible for the results presented in Figs. 2 and 3?

In order to discriminate the two selection mechanisms, we perform numerical simulations and measure—over the course of time—the number of species and chiefs, an activity measure $A$ that measures the surplus of material in chief species through the definition

$$A(x) = \sum_{x_i \in A} \left( \frac{x_i}{x^\alpha} \right)^2 \left( \frac{x_i}{x} - 1 \right)^2,$$

as well as the cosine similarity $s_{ccs}$ of the population with the smallest enclosing chief-clan structure.

Figure 7 shows representative results for the formation of a bootlace structure under the parameters of Fig. 2 with $\alpha = 10^{-7}$. The stepwise elongation of the bootlace structure is reflected by a steady increase in species count but a nongrowing number of either one or two chiefs. While the structure grows, the activity generally decreases, until it levels off around $t = 30$, where the structure is fully established. Each formation of new chiefs in the structure leads to a temporary decrease in the otherwise maximal cosine similarity to the corresponding chief-clan structure. This decrease in similarity is caused by the temporarily high population count of the previous chief, which now violates the constant solution.
Similarity is regained with the autocatalytic flow of matter into the newly formed chief and is accompanied by a decrease in activity. These observations indicate that the selection of the bootlace structure occurs from the onset and is caused by the far-from-equilibrium selection mechanism that is active during the transient period. Yet, once the structure is fully developed around $t = 30$, the equilibrium selection mechanism stabilizes the developed population and prevents the occurrence of new chiefs. As a result, fluctuations in species and chief numbers, activity, and similarity are short lived.

Visual inspection of several dozens simulation runs confirms this general behavior independent of the nature of the selected structure. The lower the random ligation rate $\alpha$, the more prominent the far-from-equilibrium selection. For the parameters chosen (cf. Fig. 2), clear chief-clan structures form and remain stable for $\alpha$ between $10^{-7}$ and $10^{-6}$. For $\alpha = 10^{-5}$, the dynamics still yield a transient selection, but the selected chief-clan structure is no longer maintained in equilibrium, as random ligations become too common. Finally, for $\alpha = 10^{-4}$ and above, stable chief-clan structures can no longer form (at least without applying a threshold for the structure identification).

IV. MODEL VARIANTS

We now present several variations of our original reaction system to demonstrate that the reported selection pressure is structurally robust and can be observed in similar replicator systems that more closely resemble the chemistry of nucleic acid replicators.

A. Increased alphabet size

Alphabet size has a significant impact on the selected population structures. Here we show populations produced from a pool of four types of monomers ($A = \{0,1,2,3\}$). Although introducing new monomer types makes the system more complex quickly, the selection of chief-clan structures does happen as shown in Fig. 8. The six largest clusters contain about 25% of all the populations produced. They all feature permutations of 0123 as motifs. They are thus a counterpart of the bootlace structure with motif 01. As in the case of a
Figure 9 shows the largest and second-largest clusters found in simulation. The seven clusters shown in Fig. 10 occupy about 70% of all the clusters produced. This value is slightly smaller than the case in Fig. 3, but is still significant fraction, indicating a strong dynamical selection. On the other hand, the occupation of each cluster becomes different from the case in Fig. 3. The bootlace structure, for example, is still a biggest cluster, but it occupies only about 11% in comparison with 48% shown in Fig. 3. Pinecone1 [Fig. 10(b) and 10(c)] and two-towers [Fig. 10(d)] increase their occupancy. Three new clusters [Figs. 10(e), 10(f) and 10(g)] appear as replacing the bootlace’s occupancy. One also notices that all the structures shown are symmetric about the center axis, reflecting the complementarity.

D. Existence of inert food species

Several previous studies assume the existence of food species [17,18] that cannot be produced by autocatalysis and thus have to be present in the system in the beginning or have to be fed continuously into the system. In our basic model, monomers are the only food species. We now analyze a variant where any species up to a certain length stabilities of the structures. If we start simulations from a pool of monomers, then monomers must spontaneously ligate to become longer than the population structure, which likely connects to the different stabilities of the structures.

C. Sequence complementarity

Sequence complementary hybridization is a characteristic feature of template-directed nucleic acid replication. We introduce sequence complementarity into our model by defining a complementarity operator \( r : \mathbb{A}^* \to \mathbb{A}^* \) recursively as follows:

\[
\begin{align*}
\bar{0} &= 1, \\
\bar{1} &= 0, \\
\bar{l.m} &= \bar{l}.\bar{m},
\end{align*}
\]

This allows us to replace the original template directed ligation reaction (3) by its sequence-complementary counterpart:

\[
l + m + \bar{l}.\bar{m} \xrightarrow{r} l.m + \bar{l}.\bar{m}.
\]

Figure 10 shows a result of hierarchical clustering from 1000 runs. The seven clusters shown in Fig. 10 occupy about 70% of all the clusters produced. This value is slightly smaller than the case in Fig. 3, but is still significant fraction, indicating a strong dynamical selection. On the other hand, the occupation of each cluster becomes different from the case in Fig. 3. The bootlace structure, for example, is still a biggest cluster, but it occupies only about 11% in comparison with 48% shown in Fig. 3. Pinecone1 [Fig. 10(b) and 10(c)] and two-towers [Fig. 10(d)] increase their occupancy. Three new clusters [Figs. 10(e), 10(f) and 10(g)] appear as replacing the bootlace’s occupancy. One also notices that all the structures shown are symmetric about the center axis, reflecting the complementarity.

B. Mutations

Evolution arises from variation and selection. In our basic model, variation is entirely due to random ligation. It is interesting to see how the reported sequence selection in our model is affected when potential mutation opens an additional channel for variation. To this end, point mutations are introduced as a reaction:

\[
l' + m' + l.m \xrightarrow{c_{1}[l.m]} l'.m' + l.m,
\]

where \( l'.m' \) is one bit different from \( l.m \) and \( c_3/c_2 = 10^{-4} \).

Figure 9 shows the largest and second-largest clusters found in simulation. The largest cluster [Fig. 9(a)], occupying 34% of all populations, is based on the bootlace structure [Fig. 3(a)], whereas the second-largest cluster [Fig. 9(b)], occupying 10%, is based on the pinecone1 structure [Fig. 3(b)], judging from the tetramer and pentamer sequence motifs. In Fig. 9(a), despite the fairly large mutation rate, perturbations arising from point mutations only produce short additional chiefs, and the bootlace structure remains unperturbed in its longer sequences. On the other hand, the distortion in Fig. 9(b) is significant. Actually, all the species shorter than trimers appear. This indicates that the robustness against mutations depends on the binary alphabet, symmetry is broken at the level of dimers and propagates from there through the population structure. From the seventh cluster onwards [Fig. 8(b)], structures lose order and become more irregular.
FIG. 10. The structures produced by sequence complementary autocatalysis. The parameters used are as follows: $\alpha = 10^{-10}$, $\beta = 10^{-7}$ with $2 \times 10^5$ of 0 and 1. These seven clusters occupy about 70% of all the populations produced.

V. DISCUSSION

We have presented a simple model of binary self-replicating polymers, in which exact autocatalysis gives rise to the formation of characteristic chief-clan structures, where the sequence information of the longest replicators completely determines the population distribution of the molecular ecosystem. When starting from a pool of monomers, only a few, highly regular, sequence patterns are selected by the dynamics.

Our derivations attribute the selection for repetitive sequences in part to reaction symmetries and in part to a high transient concentration of material in the long sequences. Due to the reaction conditions during the transient, most material is concentrated in the long sequences, the chiefs. In the transient formation of the longer sequences (mainly from the abundant chiefs) repetitive sequences require fewer reaction pathways and are also less susceptible to potential side reactions due to their symmetries. Transient selection is sustained in steady state, since repetitive sequences are also able to absorb random fluctuations faster than nonrepetitive sequences, thereby out-

FIG. 11. A stability diagram of the bootlace structures with their chief length $|m|$ against the maximum length of food species, $l_{food}$. Crosses and circles represent where they are stable and unstable, respectively.

$\langle x \rangle$ being the averaged distribution of species, depending on their length $|k|$ when $l_{food} = 8$. Crosses represent a case starting from a prebuilt bootlace structure, whereas circles represent a case starting from a pool of monomers. Solid lines represent exponential functions. Dashed line is the constant solution, $\beta^{-1/2}$.

$\langle x \rangle$ to initiate autocatalytic reactions. It is unlikely then, that autocatalytic species appear spontaneously unless the concentration of monomers is extremely high. Circles in Fig. 12 show the distribution when the simulation is started with the same amount of monomers as above but without a preconstructed bootlace structure (crosses in Fig. 12). Only dimers appear in this case. Therefore, it is unlikely, if not impossible, for this system to achieve spontaneously autocatalytic species and the resulting distribution maintained by autocatalysis.

This barrier between autocatalytic and nonautocatalytic states depends strongly on the size of food species. The longer the food species, the higher the barrier. This observation is important, for example, to a real chemistry or origin-of-life scenarios. The available monomer concentration determines how long the food species is tolerated to achieve an autocatalytically maintained state.
report the emergence of "concentration landscapes" in states is selected by noise induced drift, as encountered in the to commit to one or the other or because a smaller set of stable equilibria, then selection is likely to occur, either because If the kinetics of the polymer chemistry allows for multiple environment), then any polymer chemistry is expected to ap-} 

arrant catalysis over random ligation, and relatively constant our work applies (sufficiently high molecule numbers, dom-} 

ositive replication rates [18], random cross catalysis [15], size that many of our general results also hold for similar binary polymers, up to a length where material becomes sparse and 

accommodate for the additional material by generating larger of molecular species in steady state. Instead, the system will 

but not by the total material concentration. Increasing the total population numbers are only determined by reaction constants 

populations on the system level, and, second, steady-state 

in the transient, which ultimately determines the fate of entire mole- 

ules that we report here stem from symmetries in the reaction 

random catalyzing polymer networks, which we can now identify as equivalents to chief-clan structures.

Notably, this dynamical selection is intrinsic to our model and does not require an a priori fitness function. This is important for the study of, for example, prebiotic evolution and origins of life [26–28], as well as protocell research [6,29,30], since it implies that the selection occurs by means of both environmental as well as internal factors. This aspect was not elucidated by previous studies where selection was either induced by means of an external fitness function or by assigning randomly attributed catalysis rates.

While we reported some of the dynamics of this system in an earlier communication [21], we here deliver exhaustive theoretical derivations for the observed selection mechanisms. Whereas our previous report covered only one particular parametrization, we here derive that repetitive sequences are selected as soon as the flux through autocatalytic reactions outweighs the flux through random ligation reactions. The stronger the difference in those fluxes, the more prominent become the selected patterns. Moreover, analysis of several model variants, that were not previously reported, indicates that the observed selection of repetitive sequences is structurally robust and expected to occur in many similar chemistries.

It is commonly understood that random fluctuations are mostly significant in systems with small molecular copy numbers and that, in the limit of large numbers, results of stochastic simulations typically approach expectation values that can be obtained from equivalent deterministic models. The situation is radically different in the system presented in this study: First, autocatalytic replication amplifies small random fluctuations in the transient, which ultimately determines the fate of entire populations on the system level, and, second, steady-state population numbers are only determined by reaction constants but not by the total material concentration. Increasing the total monomer concentration will not increase the concentration of molecular species in steady state. Instead, the system will accommodate for the additional material by generating larger polymers, up to a length where material becomes sparse and thus remains susceptible to copy number fluctuations.

Although our exact findings have been derived by analyzing a specific reaction system (exact autocatalysis), we hypothesize that many of our general results also hold for similar binary polymer models, for example, those employing sequence-dependent replication rates [18], random cross catalysis [15], material flow.

Namely, if studied under conditions where the analysis of our work applies (sufficiently high molecule numbers, dominant catalysis over random ligation, and relatively constant environment), then any polymer chemistry is expected to approach some equilibrium distribution. The phase-space profile of these equilibrium states will dictate which populations are stable and how matter is distributed among their members. If the kinetics of the polymer chemistry allows for multiple equilibria, then selection is likely to occur, either because these are isolated stable states and an initial population has to commit to one or the other or because a smaller set of stable states is selected by noise induced drift, as encountered in the chemistry studied in this article. For example, Bagley et al. [15] report the emergence of “concentration landscapes” in randomly catalyzing polymer networks, which we can now identify as equivalents to chief-clan structures.

As long as selected populations introduce novel species that provide resources and catalysts for higher units of selection (i.e., longer polymers of a net-catalyzing set), the chemistry meets in principle the conditions for continued selection dynamics during a transient that starts from monomers or short oligomers. In the model system discussed in this article, continued selection is particularly prominent, since every newly created species is an auto-catalyst. Continued selection is expected to stall once the chemistry does not introduce novel resources or novel catalysts.

In contrast, we expect that the particular repetitive sequence motifs that we report here stem from symmetries in the reaction graph of exact autocatalysis and are likely not prominent in random catalytic chemistries.

The investigation of related chemistries employing se- 

quence dependent and randomly assigned catalysis is left for future research.

APPENDIX: FOKKER-PLANCK DERIVATION

We study the minimal competing system with four species, where two competing chiefs 01 and 10 compete over resources 0 and 1. Mass conservation maintains the dynamics on a two-dimensional manifold, essentially \( x_{01} \) and \( x_{10} \) (\( x_0 = x_1 \) can then be inferred).

For brevity, we introduce the notation \( X = 01, Y = 10 \), with concentrations \( x = x_{01}, y = x_{10}, z = x + y \), and indiscriminately refer to monomers by \( A \) with concentration \( a = x_0 = x_1 \).

Motivated by the observation that there is a line of steady-state solutions (any combination of \( x \) and \( y \) that fulfills \( a = \beta^{1/2} \)) that are all of marginal stability (zero eigenvalues), we introduce a simplification that constrains the system to maintain equilibrium by only distributing material among chiefs.

To achieve this, we construct a constrained stochastic process, where no two degradations nor two liggations can follow each other without an intermediate event of the other type. Each net degradation-ligation event thus maintains the initial distance from the equilibrium line.

Technically, we introduce a stochastic net process with two Markovian degradation reactions and two delayed time ligation reactions whose rates \( \beta \) are functions of the time \( t - t' \) since the last degradation event. With

\[
\int_{t'}^{\infty} \beta(t - t') dt = \beta, \tag{A1}
\]

we ensure that the original ligation rate is maintained in the long term. Informally, the constrained process can be written as

\[
X \xrightarrow{1} 2A \xrightarrow{\beta(t-t')} X, \tag{A2}
\]

\[
X \xrightarrow{1} 2A \xrightarrow{\beta(t-t')} Y, \tag{A3}
\]

\[
Y \xrightarrow{1} 2A \xrightarrow{\beta(t-t')} X, \tag{A4}
\]

\[
Y \xrightarrow{1} 2A \xrightarrow{\beta(t-t')} Y. \tag{A5}
\]
samples for each of distribution functions have been obtained from 1000 numerical samples for each of purple, the original 2D process defined by Eqs. (1) and (3); green, our 1D approximation of equation (A7); and blue, a random walk with constant diffusion for reference.

We now let

$$\beta(t \rightarrow t) \rightarrow \beta_0$$

approach a Dirac distribution, so that ligation follows degradation immediately.

Degradation events occur as previously with rate constant 1, but only a certain fraction [e.g., $y/(x + y)$ for Eq. (A3)] lead to an actual change in the system state. The net conversion rates are therefore the degradation rate (having been set to one per unit time in the dimensionless system) times the probability to lead to a system change. The net process therefore reads:

$$X \xrightarrow{(z-x)/z} Y, \quad (A7)$$
$$Y \xrightarrow{x/z} X. \quad (A8)$$

Note that the total propensity of this net process, $(z-x)/z + x/z = z$, equals the total propensity of degradations in the original process.

To compare the quality of this approximation, we sampled 1000 instantiations of the original 2D process and of our 1D simplification and compare the cumulative probability distributions of first hitting times for walks that start at $x(0) = \frac{x}{z}$ and reach one of the absorbing boundaries 0 or $z$.

Results are shown in Fig. 13.

The results reveal that the 1D approximation generates hitting time distributions with the same general shape than the original process but overestimates typical hitting times—that is, the unconstrained process typically terminates faster than the constrained process. The approximation is made more accurate for late hitting times and approaches the original hitting time distribution.

Figure 13 also reveals that both the original and the constrained process complete orders of magnitude faster than a random walk where $X$ transforms into $Y$ and vice versa with constant rate $z$.

We now derive the master and Fokker-Planck equations for the stochastic process (A7) and (A8). Since mass is conserved in this system, this corresponds to a one-dimensional random walk, here expressed for the species $X = 01$:

$$\ldots \Rightarrow x = 1 + \frac{(z-1)(z-x+1)/z}{(x-1)(z-x)/z} x \xrightarrow{x(z-x)/z} x + 1 = \ldots \quad (A9)$$

The stochastic process has the associated master equation

$$\frac{\partial}{\partial t} P(x,t) = -2\frac{x(z-x)}{z} P(x,t) + \frac{(x-1)(z-x+1)}{z} P(x-1,t) + \frac{(x+1)(z-x-1)}{z} P(x+1,t), \quad (A10)$$
$$= -2\frac{x(z-x)}{z} P(x,t) + \frac{x(z-x)}{z} P(x-1,t) + \frac{x}{z} P(x-1,t) - \frac{z-x}{z} P(x-1,t) - \frac{1}{z} P(x-1,t)$$
$$+ \frac{x(z-x)}{z} P(x+1,t) - \frac{x}{z} P(x+1,t) + \frac{z-x}{z} P(x+1,t) - \frac{1}{z} P(x+1,t), \quad (A11)$$
$$= \frac{x(z-x)}{z} [P(x+1,t) - 2P(x,t) + P(x-1,t)]$$
$$+ \frac{z-2x}{z} [P(x+1,t) - P(x-1,t)] - \frac{1}{z} P(x+1,t) - \frac{1}{z} P(x-1,t), \quad (A12)$$
$$= \frac{x(z-x)-1}{z} [P(x+1,t) - 2P(x,t) + P(x-1,t)] + \frac{z-2x}{z} [P(x+1,t) - P(x-1,t)] - \frac{2}{z} P(x,t). \quad (A13)$$

Approximating

$$P(x+1,t) - P(x-1,t) \approx 2 \frac{\partial}{\partial x} P(x,t), \quad (A14)$$
$$P(x+1,t) - 2P(x,t) + P(x-1,t) \approx \frac{\partial^2}{\partial x^2} P(x,t), \quad (A15)$$

we derive a Fokker-Planck equation with zero drift and nonconstant diffusion:

$$\frac{\partial}{\partial t} P(x,t) \approx \frac{x(z-x)-1}{z} \frac{\partial^2}{\partial x^2} P(x,t) + \frac{2}{z} \frac{z-2x}{z} \frac{\partial}{\partial x} P(x,t) - \frac{2}{z} P(x,t) \quad (A16)$$
\[ z - 2x \frac{\partial}{\partial x} P(x,t) + \frac{x(z-x) - 1}{z} \frac{\partial^2}{\partial x^2} P(x,t) + \frac{z - 2x}{z} \frac{\partial}{\partial x} P(x,t) - \frac{2}{x} P(x,t), \quad (A17) \]

\[ = \frac{\partial}{\partial x} \left[ \frac{x(z-x) - 1}{z} \frac{\partial}{\partial x} P(x,t) \right] + \frac{\partial}{\partial x} \left[ \frac{z - 2x}{z} P(x,t) \right], \quad (A19) \]

\[ = \frac{\partial^2}{\partial x^2} \left[ \frac{x(z-x) - 1}{z} P(x,t) \right], \quad (A20) \]

\[ = \frac{\partial^2}{\partial x^2} \left[ D(x) P(x,t) \right], \quad (A22) \]

with nonconstant diffusion \( D(x) = \frac{\xi(x)}{2} \). Note that diffusion is maximal for \( x = z/2 \), i.e., when both chiefs are equally populated, and becomes negative for \( x = 0 \) and \( x = z \), which reflects the fact that these are absorbing boundary conditions.