Bivology: A Computational Framework for Mapping the Chemistry-to-Life Transition

Quantitative Predictions for the Eigen Error Threshold, Autocatalytic Coupling, Spatial Dynamics, and Proto-Darwinian Selection at the RNA Error Floor

Jim Lauletta
Independent Researcher
March 2026
minimallife2@gmail.com · jimlauletta.com · creationthruchemistry.com

---

We present Bivology (virtual biology), a six-step computational simulation framework designed to map the parameter space governing the transition from prebiotic chemistry to proto-Darwinian replication systems. Using a digital chemistry engine implementing the Eigen error threshold, autocatalytic coupling, spatial diffusion dynamics, vesicle encapsulation, and template-directed heredity, we identify three quantitative thresholds necessary and sufficient for the emergence of heritable replication from random chemistry.

The simulations confirm: (1) the Eigen threshold behaves as a sharp phase transition (Sarle bimodality coefficient $\beta > 0.555$ *across all tested error rates); (2) a minimum autocatalytic coupling of* $1.5\times$ *is required to survive replication at the RNA error floor (*$\varepsilon = 0.07$*); (3) ensemble statistics at the RNA floor yield* $P(survival) = 0.62$*, characterizing the transition as probabilistic rather than deterministic; (4) spatial rescue via molecular diffusion fails at biologically plausible rates, identifying pore channel geometry as a critical constraint; (5) vesicle encapsulation alone cannot bootstrap functional chemistry from random sequences, confirming Eigen's catch-22; and (6) a minimum seed fraction of 3% functional sequences at encapsulation produces reliable population takeover via template-directed replication (effective error rate* $\varepsilon_T = 0.025$*), with a cliff transition between collapse and full heredity emergence.*

These results generate six independently testable predictions for physical microfluidic reactor experiments targeting the chemistry-to-life transition. The framework is proposed as a computational companion to staged experimental origin-of-life programs.

Keywords: *origin of life; Eigen error threshold; autocatalytic sets; quasispecies; protocell; template replication; computational astrobiology*

---

1. Introduction

The transition from prebiotic chemistry to self-replicating, evolvable molecular systems is among the most fundamental unsolved problems in science. Despite substantial theoretical and experimental progress, the parameter space governing this transition has not been systematically mapped. Computational approaches offer a complementary path: by implementing simplified but physically grounded models of replication, selection, and encapsulation, simulations can identify thresholds, phase transitions, and minimum requirements that are difficult or impossible to probe experimentally.

The central theoretical constraint was identified by Eigen in 1971 [Eigen, 1971]. For molecular evolution to begin, replicating sequences must copy themselves with sufficient fidelity that useful sequences are preserved across generations. Above a critical error rate — the *error catastrophe* threshold — genetic information dissolves into noise faster than selection can act. This threshold imposes a fundamental tension: non-enzymatic chemistry produces error rates of 7–26% per base [Eigen, 1971], while minimum viable genomes require hundreds of bases, implying stable sequence lengths far shorter than functionally meaningful.

A parallel body of work on autocatalytic sets, initiated by Kauffman [Kauffman, 1971] and formalized through the RAF (Reflexively Autocatalytic Food-generated) framework [Hordijk and Steel, 2004], demonstrates that self-sustaining chemical reaction networks emerge at modest levels of catalytic activity. Recent computational work has extended these models to spatially structured environments [Hordijk et al., 2018] and encapsulated protocell populations. However, the quantitative coupling between autocatalytic organization, error-threshold dynamics, and encapsulation — and the minimum conditions required for all three to operate simultaneously — has not been characterized in an integrated framework.

Here we present Bivology (*virtual biology*), a six-step simulation series that integrates these processes sequentially. Each step addresses a specific scientific question, produces quantitative outputs, and generates a testable prediction for physical experimental systems. The framework is designed as a computational companion to staged laboratory programs targeting the chemistry-to-life transition, providing baseline parameter maps and statistical distributions against which experimental results can be compared.

2. Model and Methods

2.1. Digital Chemistry Engine

The core simulation operates on sequences of length $L = 12$ drawn from a four-symbol alphabet $\{A, B, C, D\}$, representing a minimal analog of RNA bases. A target sequence $T =$ (ABCABCABCABC) encodes the autocatalytic loop motif A→B→C, repeated across the sequence. Fitness is defined as:

$$f(s) = \begin{cases} f_{\text{base}}(s) \cdot b & \text{if } s \text{ contains complete loop} \\ f_{\text{base}}(s) + f_{\text{loop}}(s) \cdot (b-1) \cdot 0.3 & \text{otherwise} \end{cases} \tag{1}$$

where $f_{\text{base}}(s) = \text{match\_fraction}(s, T)$, $f_{\text{loop}}(s)$ is the fraction of complete A→B→C triplets in s, and b is the autocatalytic boost parameter — the key experimental variable representing the replication advantage conferred by a functional reaction loop.

Replication proceeds by fitness-proportionate selection with per-position error rate ε. Template-directed replication (Step 6) uses a complementary alphabet $\{A \leftrightarrow C,\ B \leftrightarrow D\}$ with reduced error rate $\varepsilon_T = \varepsilon \times 0.35$, modeling primitive base-pairing without enzymatic assistance. All simulations were implemented in Python 3 using NumPy and SciPy.

2.2. Simulation Architecture

The six simulation steps build progressively on this engine:

Step 1. Threshold characterization. Well-mixed population, $N = 400$, $G = 200$ generations, error rates $\varepsilon \in \{0.01, 0.05, 0.10, 0.15, 0.20, 0.26\}$.

Step 2. Coupling sweep. 72-condition grid, $b \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0\}\times$, error rates $\varepsilon \in \{0.03–0.26\}$, $N = 400$, $G = 150$.

Step 3. Ensemble statistics. 60 replicates per condition at critical boost $b = 1.5\times$, boundary error rates $\varepsilon \in \{0.05–0.16\}$.

Step 4. Spatial dynamics. 14×14 cell grid, 15 sequences/cell, diffusion rate 10%/generation, hot patch $b = 2.5\times$, background $b = 1.45\times$.

Step 5. Protocell encapsulation. Vesicle population, $N_{\text{max}} = 120$, division threshold 36 sequences, lysis threshold 4 sequences, lipid economy model, $G = 150$.

Step 6. Heredity sweep. Seed fractions $\{0\%, 1\%, 3\%, 5\%, 10\%, 20\%\}$, 8 replicates each, template replication for functional sequences, $G = 200$.

2.3. Measurement

The primary measurement is the *quasispecies fraction*: the proportion of population sequences within Hamming distance 2 of target T. Secondary measurements include Shannon entropy (information content), Sarle's bimodality coefficient [Sarle, 1990]:

$$\beta = \frac{\hat{\mu}_3^2 + 1}{\hat{\mu}_4 + \frac{3(n-1)^2}{(n-2)(n-3)}} \tag{2}$$

where $\hat{\mu}_3$ and $\hat{\mu}_4$ are the sample skewness and excess kurtosis. Values $\beta > 0.555$ indicate bimodal distributions, signaling a phase transition rather than gradual degradation. Additional measurements include survival probability (quasispecies fraction > 0.10), coefficient of variation (vesicle diversity), and template replication fraction (Step 6).

3. Results

Table 1: Summary of six-step simulation results and reactor predictions.

#	Simulation	Key Finding	Reactor Prediction
1	Error threshold phase transition	Sharp bimodal collapse at $\varepsilon \approx 0.125$; $\beta > 0.555$	Threshold confirmed as first-order-like transition
2	Autocatalytic coupling sweep	Minimum viable coupling $= 1.5\times$ at RNA floor ($\varepsilon = 0.07$)	ΔpH > 1.2; Fe-S rate enhancement $> 1.5\times$
3	Ensemble statistics (60 reps)	P(survival) $= 0.62$ at RNA floor; $\beta > 0.555$	∼6 reactor runs sufficient to detect signal
4	Spatial dynamics (14×14 grid)	Rescue failed: diffusion too slow vs. collapse timescale	Pore channel geometry must enable fast mixing
5	Protocell encapsulation	Bootstrap failure: encapsulation cannot discover function	Stage I chemistry must precede Stage III encapsulation
6	Heredity + seed fraction sweep	Min viable seed = 3%; cliff transition; $\varepsilon_T = 0.025$	$> 3\%$ functional events per encapsulation required

3.1. The Eigen Threshold Is a Sharp Phase Transition (Step 1)

At error rates below the threshold the population maintains a coherent quasispecies. Above it, information collapses to noise. The transition is discontinuous: at $\varepsilon = 0.10$, quasispecies fraction $= 0.301$; at $\varepsilon = 0.15$, collapse to 0.011. The bimodality coefficient exceeds 0.555

at all tested error rates, confirming a phase-transition character. The in-silico threshold appears near $\varepsilon \approx 0.125$, consistent with theoretical expectations for $L = 12$ sequences over a four-symbol alphabet.

3.2. Minimum Autocatalytic Coupling (Step 2)

Sweeping 72 parameter combinations maps the full threshold surface. Without autocatalytic coupling ($b = 1.0\times$), the threshold sits at $\varepsilon \approx 0.065$ — just *below* the RNA error floor of 0.07. A modest 1.5× replication advantage shifts it to $\varepsilon \approx 0.106$, above the floor.

Table 2: Eigen threshold as a function of autocatalytic coupling at $N = 400$, $L = 12$.

Boost	**Threshold** ε	**vs. RNA floor (0.07)**	**Outcome**
1.0×	≈ 0.065	Below	Collapse
1.5×	≈ 0.106	Above	Viable
2.0×	≈ 0.112	Above	Viable
3.0×	≈ 0.148	Above	Viable
5.0×	≈ 0.178	Above	Viable

The 1.5× minimum coupling is physically interpretable: a pH gradient $\Delta\text{pH} > 1.2$ across a mineral membrane, or an iron-sulfur catalytic rate enhancement $> 1.5\times$ relative to background, is sufficient to cross the fidelity barrier [Lane and Martin, 2012, Russell and Martin, 2004]. Neither value is extreme; both are plausible under alkaline hydrothermal vent conditions.

3.3. Statistical Structure at the RNA Floor (Step 3)

Running 60 independent replicates at $b = 1.5\times$ and $\varepsilon = 0.07$ reveals $P(\text{survival}) = 0.62$, with bimodality coefficient $\beta = 0.601$. Replicates cluster into survived ($f_{QS} > 0.10$) and collapsed ($f_{QS} < 0.10$) groups with few intermediate outcomes. This is the hallmark of a genuine phase transition.

The 62% survival probability directly informs experimental design. Approximately six independent reactor experiments are sufficient to detect this signal at 80% statistical power ($\alpha = 0.05$). The simulation provides the comparison baseline: quasispecies fraction > 0.36 (mean $+2\sigma$ of the collapse distribution) constitutes a statistically significant positive reactor result.

3.4. Spatial Dynamics and the Diffusion Constraint (Step 4)

A 14×14 grid models the hydrothermal vent as a spatially heterogeneous environment. A central hot patch ($b = 2.5\times$) is surrounded by background cells ($b = 1.45\times$, below the solo-survival threshold). Diffusion rate was 10%/generation.

Spatial rescue did not occur at any tested patch size (radius 1–4 cells). This null result identifies a minimum diffusion rate requirement separate from the coupling threshold: the diffusion timescale must be shorter than the error catastrophe timescale. For the physical reactor, this constrains channel cross-sectional dimensions to the range where oligonucleotide diffusion times are short relative to the characteristic sequence collapse timescale.

3.5. Encapsulation and the Bootstrap Problem (Step 5)

Introducing protocell dynamics — vesicle growth, division at 36 sequences, lysis below 4 sequences, and a lipid economy — tests whether encapsulation can rescue threshold-adjacent chemistry from fully random initial sequences. The population collapsed to near-zero quasispecies fraction across all 150 generations. This computationally confirms Eigen's catch-22: accurate copying requires information, but accumulating information requires accurate copying.

Encapsulation is a *propagation* mechanism, not a discovery mechanism. The null result is physically correct: the RNA world hypothesis proposes that functional sequences arise in solution and are subsequently encapsulated, not that they originate within protocells [Szostak et al., 2001].

3.6. Heredity Emergence via Seeded Encapsulation (Step 6)

Seeding a fraction f_s of vesicles with a single functional sequence, combined with template-directed replication ($\varepsilon_T = 0.025$), produces a cliff transition:

Seed fraction	**Replicates**	**P(survival)**	**Mean QS (last 20 gen)**
0%	8/8	0%	0.001 ± 0.002
1%	8/8	0%	0.000 ± 0.000
3%	8/8	**100%**	$\mathbf{0.830 \pm 0.017}$
5%	8/8	100%	0.828 ± 0.018
10%	8/8	100%	0.856 ± 0.012
20%	8/8	100%	0.840 ± 0.035

At seed fractions $\leq 1\%$, every replicate collapses. At $\geq 3\%$, every replicate achieves population-level takeover (mean $f_{QS} = 0.83$). There are no intermediate outcomes. Above the threshold, template replication at $\varepsilon_T = 0.025$ drives functional sequences to dominance via a double advantage: higher fitness *and* lower effective error rate. Heredity — stable heritable information encoded in sequence structure — has emerged.

The 3% seed fraction translates to a physical prediction: one in approximately 33 encapsulation events must capture a functional sequence for reliable heredity establishment.

4. The Three-Requirement Conclusion

The six simulations converge on a minimal characterization of what is necessary and sufficient for heritable replication to emerge from random chemistry:

Requirement 1 — Autocatalytic coupling $\geq 1.5\times$.
The geochemical environment must deliver at least $1.5\times$ replication advantage to functional loop sequences relative to random chemistry. In physical terms: ΔpH > 1.2 across mineral membrane, or Fe-S catalytic rate enhancement $> 1.5\times$. Below this coupling, error catastrophe is unavoidable regardless of sequence length, population size, or time available.

Requirement 2 — At least one functional sequence arising in solution.
Encapsulation cannot discover function from scratch. A functional sequence must arise by non-enzymatic chemistry before encapsulation. This need not be common — only sufficient to ensure $\geq 3\%$ of encapsulation events capture a functional sequence.

Requirement 3 — Encapsulation before dilution.
Once functional sequences exist in solution, lipid encapsulation drives population takeover via two coupled selection mechanisms: vesicle-level (functional vesicles grow and divide; non-functional lyse) and sequence-level (template copying provides lower error rate and higher fitness simultaneously).

Each requirement is independently measurable in a physical reactor. None requires extraordinary chemistry. Each is a threshold, not a wall.

5. Discussion

5.1. Relationship to Prior Computational Work

The RAF framework of Hordijk and Steel [Hordijk and Steel, 2004] provides the formal mathematical foundation for autocatalytic set theory, demonstrating that such sets form readily at modest catalytic activity and can be sustained within encapsulated compartments [Hordijk et al., 2018]. The present framework extends this work in two directions: it integrates the Eigen error threshold explicitly as a controllable parameter (rather than assuming a fixed error rate), and it characterizes the *statistical structure* of the transition via ensemble methods, producing probability distributions rather than single-trajectory results.

The spatial null result (Step 4) is consistent with the general finding that spatial structure alone does not rescue threshold-adjacent chemistry without fast diffusion. The geometry constraint derived here — diffusion timescale shorter than the error catastrophe timescale — provides a quantitative design criterion not previously stated in this form.

The heredity emergence result (Step 6) is broadly consistent with the Szostak protocell model [Szostak et al., 2001], in which vesicles containing replicating sequences grow at the expense of empty vesicles. The minimum seed fraction of 3% is consistent with the expectation from branching process theory that stochastic loss becomes negligible when functional sequences are present in more than one copy per vesicle, here demonstrated in the full dynamic model.

5.2. Limitations

Model abstraction. The digital chemistry alphabet and simplified fitness function abstract away specific chemical details of RNA replication. Quantitative thresholds (1.5× coupling, 3% seed fraction, $P = 0.62$) are model-dependent and should be treated as order-of-magnitude estimates. The key contribution is the *structure* of results — phase transitions, cliff transitions, minimum requirements — rather than specific numerical values.

Parameter sensitivity. The sensitivity of reported thresholds to sequence length L, alphabet size, and fitness function shape has not been fully characterized. Results expected to be robust include the bimodal character of the phase transition, the existence of a minimum seed fraction, and the failure of spatial rescue at slow diffusion. Results expected to be model-dependent include the specific values $\varepsilon^* \approx 0.125$, $b_{\min} = 1.5\times$, and $f_{s,\min} = 3\%$. Sensitivity analysis is ongoing.

Population size and sequence length. Computational constraints required $N = 200$–400 and $L = 12$, substantially smaller than biologically realistic systems. Ensemble replication partially compensates, but finite-size effects are acknowledged.

Template mechanism. The template fidelity factor of 0.35 is a plausible but uncalibrated estimate. Real non-enzymatic template copying involves complex thermodynamic and kinetic factors including template-monomer binding affinity and per-base misincorporation rates.

5.3. Implications for Experimental Design

The simulation results suggest four specific design constraints for physical microfluidic reactor experiments:

1. Geochemical conditions must deliver $\geq 1.5\times$ catalytic rate enhancement for functional sequence motifs, measurable via comparative polymerization rate assays with and without mineral surface catalysis.

2. Channel geometry must ensure diffusion timescales shorter than the error catastrophe timescale ($\sim$10–50 generation equivalents), constraining channel cross-sections to approximately 1–10 μm for typical oligonucleotide diffusion coefficients.

3. Polymer selection assays (Stage II) must confirm quasispecies dynamics in solution before encapsulation; the simulation predicts this signal in $\sim$62% of reactor runs at minimum coupling conditions.

4. Encapsulation efficiency must achieve $\geq 3\%$ functional sequences per encapsulation event, verifiable by single-vesicle fluorescence assays.

6. Conclusions

The Bivology simulation framework maps the parameter space of the chemistry-to-life transition across six sequential steps. The principal findings are:

- The Eigen error threshold is a sharp phase transition ($\beta > 0.555$), not a gradual degradation.

- Minimum autocatalytic coupling = $1.5\times$ to survive the RNA error floor, corresponding to achievable hydrothermal vent geochemical conditions.

- $P(\text{survival}) = 0.62$ at minimum coupling and RNA error floor; $\sim$6 independent experiments sufficient to detect the signal.

- Spatial rescue fails at biologically plausible diffusion rates, identifying pore channel geometry as a critical reactor design parameter.

- Encapsulation cannot bootstrap functional chemistry from random sequences, confirming the necessity of ordered experimental staging.

- Seeded encapsulation with template replication produces a cliff transition at 3% seed fraction, above which population-level heredity emerges reliably.

Three requirements are necessary and sufficient: autocatalytic coupling $\geq 1.5\times$; at least one functional sequence arising in solution; and encapsulation before dilution. All three are independently measurable in physical experiments. The simulation framework provides quantitative baseline distributions for comparison.

Acknowledgments

This research was conducted independently, without institutional funding. All simulations were implemented in Python 3 using NumPy and SciPy. Simulation code is available from the author on request.

References

Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971. doi: 10.1007/BF00623322.

Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227:451–461, 2004. doi: 10.1016/j.jtbi.2003.11.020.

Wim Hordijk, Jonathan Naylor, Natalio Krasnogor, and Harold Fellermann. Population dynamics of autocatalytic sets in a compartmentalized spatial world. *Life*, 8(3):33, 2018. doi: 10.3390/life8030033.

Stuart A. Kauffman. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics*, 1:71–96, 1971. doi: 10.1080/01969727108545830.

Nick Lane and William F. Martin. The origin of membrane bioenergetics. *Cell*, 151:1406–1416, 2012. doi: 10.1016/j.cell.2012.11.050.

Michael J. Russell and William Martin. The rocky roots of the acetyl-CoA pathway. *Trends in Biochemical Sciences*, 29:358–363, 2004. doi: 10.1016/j.tibs.2004.05.007.

Warren S. Sarle. The BIMODAL procedure. In *SUGI 15 Proceedings*, pages 190–203. SAS Institute, 1990.

Jack W. Szostak, David P. Bartel, and Pier Luigi Luisi. Synthesizing life. *Nature*, 409: 387–390, 2001. doi: 10.1038/35053176.