Supporting Information — Lack of evolvability in self-sustaining autocatalytic networks

Vera Vasas, Eörs Szathmáry, and Mauro Santos

A Replication-mutation equilibrium frequencies

Table S1. Rank order distribution (from highest to lowest equilibrium frequencies) of compositional assemblies with size $N_{\min} = 3$ resulting from the Eigen equation.

Assembly #	Rank	Molecular counts									
94	1	{0	0	1	0	0	0	2	0	0	0}
20	2	{0	0	0	0	0	0	3	0	0	0}
116	3	{0	0	2	0	0	0	1	0	0	0}
103	4	{0	0	1	0	1	0	1	0	0	0}
98	5	{0	0	1	0	0	1	1	0	0	0}
109	6	{0	0	1	1	0	0	1	0	0	0}
45	7	{0	0	0	0	1	0	2	0	0	0}
30	8	{0	0	0	0	0	1	2	0	0	0}
66	9	{0	0	0	1	0	0	2	0	0	0}
92	10	{0	0	1	0	0	0	1	0	1	0}
93	11	{0	0	1	0	0	0	1	1	0	0}
197	12	{1	0	1	0	0	0	1	0	0	0}
91	13	{0	0	1	0	0	0	1	0	0	1}
152	14	{0	1	1	0	0	0	1	0	0	0}
18	15	{0	0	0	0	0	0	2	0	1	0}
19	16	{0	0	0	0	0	0	2	1	0	0}
70	17	{0	0	0	1	0	1	1	0	0	0}
175	18	{1	0	0	0	0	0	2	0	0	0}
17	19	{0	0	0	0	0	0	2	0	0	1}
130	20	{0	1	0	0	0	0	2	0	0	0}
28	21	{0	0	0	0	0	1	1	0	1	0}
110	22	{0	0	1	1	0	1	0	0	0	0}
96	23	{0	0	1	0	0	1	0	0	1	0}
49	24	{0	0	0	0	1	1	1	0	0	0}

Assembly #	Rank	Molecular counts									
190	25	{1	0	0	1	0	0	1	0	0	0}
75	26	{0	0	0	1	1	0	1	0	0	0}
65	27	{0	0	0	1	0	0	1	1	0	0}
119	28	{0	0	2	1	0	0	0	0	0	0}
68	29	{0	0	0	1	0	1	0	0	1	0}
82	30	{0	0	0	2	0	1	0	0	0	0}
117	31	{0	0	2	0	0	1	0	0	0	0}
118	32	{0	0	2	0	1	0	0	0	0	0}
81	33	{0	0	0	2	0	0	1	0	0	0}
64	34	{0	0	0	1	0	0	1	0	1	0}
120	35	{0	0	3	0	0	0	0	0	0	0}
34	36	{0	0	0	0	0	2	1	0	0	0}
29	37	{0	0	0	0	0	1	1	1	0	0}
71	38	{0	0	0	1	0	2	0	0	0	0}
44	39	{0	0	0	0	1	0	1	1	0	0}
115	40	{0	0	2	0	0	0	0	1	0	0}
43	41	{0	0	0	0	1	0	1	0	1	0}
108	42	{0	0	1	1	0	0	0	1	0	0}
114	43	{0	0	2	0	0	0	0	0	1	0}
200	44	{1	0	1	1	0	0	0	0	0	0}
32	45	{0	0	0	0	0	2	0	0	1	0}
54	46	{0	0	0	0	2	0	1	0	0	0}
112	47	{0	0	1	2	0	0	0	0	0	0}
113	48	{0	0	2	0	0	0	0	0	0	1}
107	49	{0	0	1	1	0	0	0	0	1	0}
201	50	{1	0	2	0	0	0	0	0	0	0}
184	51	{1	0	0	0	1	0	1	0	0	0}
69	52	{0	0	0	1	0	1	0	1	0	0}
42	53	{0	0	0	0	1	0	1	0	0	1}
27	54	{0	0	0	0	0	1	1	0	0	1}
156	55	{0	1	2	0	0	0	0	0	0	0}
179	56	{1	0	0	0	0	1	1	0	0	0}
111	57	{0	0	1	1	1	0	0	0	0	0}
104	58	{0	0	1	0	1	1	0	0	0	0}
139	59	{0	1	0	0	1	0	1	0	0	0}
99	60	{0	0	1	0	0	2	0	0	0	0}
63	61	{0	0	0	1	0	0	1	0	0	1}
145	62	{0	1	0	1	0	0	1	0	0	0}
134	63	{0	1	0	0	0	1	1	0	0	0}
97	64	{0	0	1	0	0	1	0	1	0	0}
15	65	{0	0	0	0	0	0	1	1	1	0}
76	66	{0	0	0	1	1	1	0	0	0	0}

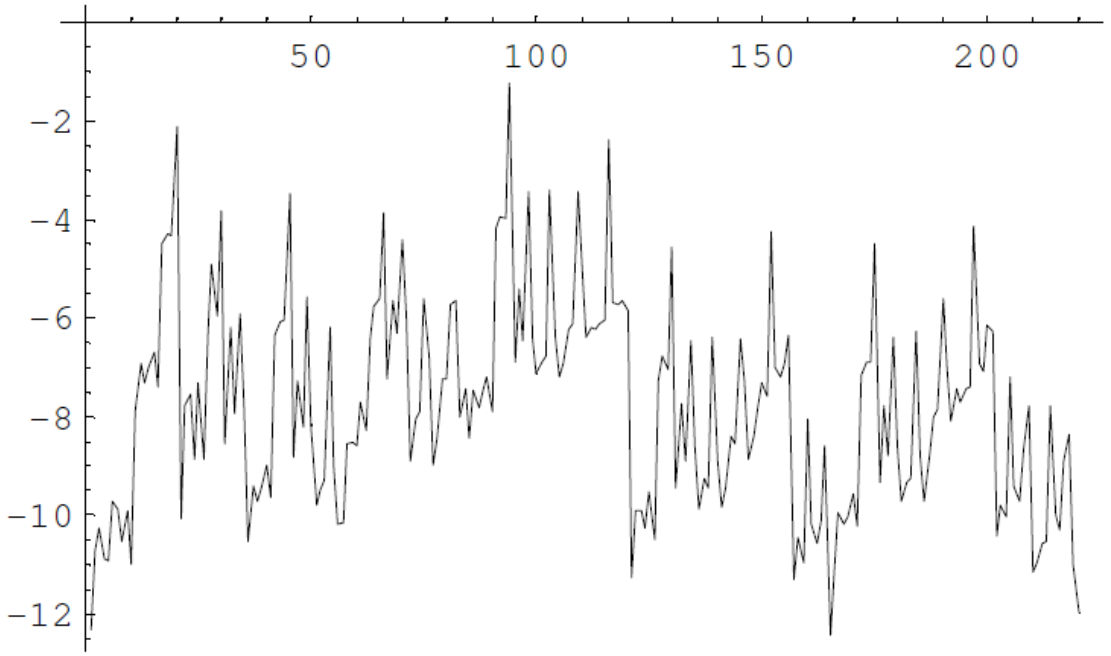
Assembly #	Rank	Molecular counts									
102	67	{0	0	1	0	1	0	0	1	0	0}
128	68	{0	1	0	0	0	0	1	0	1	0}
174	69	{1	0	0	0	0	0	1	1	0	0}
101	70	{0	0	1	0	1	0	0	0	1	0}
173	71	{1	0	0	0	0	0	1	0	1	0}
95	72	{0	0	1	0	0	1	0	0	0	1}
198	73	{1	0	1	0	0	1	0	0	0	0}
106	74	{0	0	1	1	0	0	0	0	0	1}
12	75	{0	0	0	0	0	0	1	0	1	1}
155	76	{0	1	1	1	0	0	0	0	0	0}
191	77	{1	0	0	1	0	1	0	0	0	0}
153	78	{0	1	1	0	0	1	0	0	0	0}
14	79	{0	0	0	0	0	0	1	1	0	1}
129	80	{0	1	0	0	0	0	1	1	0	0}
199	81	{1	0	1	0	1	0	0	0	0	0}
100	82	{0	0	1	0	1	0	0	0	0	1}
172	83	{1	0	0	0	0	0	1	0	0	1}
105	84	{0	0	1	0	2	0	0	0	0	0}
89	85	{0	0	1	0	0	0	0	1	1	0}
154	86	{0	1	1	0	1	0	0	0	0	0}
205	87	{1	1	0	0	0	0	1	0	0	0}
80	88	{0	0	0	2	0	0	0	1	0	0}
67	89	{0	0	0	1	0	1	0	0	0	1}
79	90	{0	0	0	2	0	0	0	0	1	0}
146	91	{0	1	0	1	0	1	0	0	0	0}
47	92	{0	0	0	0	1	1	0	0	1	0}
127	93	{0	1	0	0	0	0	1	0	0	1}
25	94	{0	0	0	0	0	1	0	1	1	0}
13	95	{0	0	0	0	0	0	1	0	2	0}
150	96	{0	1	1	0	0	0	0	0	1	0}
16	97	{0	0	0	0	0	0	1	2	0	0}
196	98	{1	0	1	0	0	0	0	1	0	0}
195	99	{1	0	1	0	0	0	0	0	1	0}
193	100	{1	0	0	2	0	0	0	0	0	0}
84	101	{0	0	0	3	0	0	0	0	0	0}
86	102	{0	0	1	0	0	0	0	0	1	1}
88	103	{0	0	1	0	0	0	0	1	0	1}
23	104	{0	0	0	0	0	1	0	0	2	0}
151	105	{0	1	1	0	0	0	0	1	0	0}
61	106	{0	0	0	1	0	0	0	1	1	0}
194	107	{1	0	1	0	0	0	0	0	0	1}
132	108	{0	1	0	0	0	1	0	0	1	0}

Assembly #	Rank	Molecular counts									
177	109	{1	0	0	0	0	1	0	0	1	0}
35	110	{0	0	0	0	0	3	0	0	0	0}
209	111	{1	1	1	0	0	0	0	0	0	0}
214	112	{2	0	0	0	0	0	1	0	0	0}
22	113	{0	0	0	0	0	1	0	0	1	1}
149	114	{0	1	1	0	0	0	0	0	0	1}
87	115	{0	0	1	0	0	0	0	0	2	0}
189	116	{1	0	0	1	0	0	0	1	0	0}
90	117	{0	0	1	0	0	0	0	2	0	0}
74	118	{0	0	0	1	1	0	0	1	0	0}
11	119	{0	0	0	0	0	0	1	0	0	2}
33	120	{0	0	0	0	0	2	0	1	0	0}
83	121	{0	0	0	2	1	0	0	0	0	0}
188	122	{1	0	0	1	0	0	0	0	1	0}
160	123	{0	2	0	0	0	0	1	0	0	0}
73	124	{0	0	0	1	1	0	0	0	1	0}
192	125	{1	0	0	1	1	0	0	0	0	0}
50	126	{0	0	0	0	1	2	0	0	0	0}
48	127	{0	0	0	0	1	1	0	1	0	0}
62	128	{0	0	0	1	0	0	0	2	0	0}
218	129	{2	0	1	0	0	0	0	0	0	0}
148	130	{0	1	0	2	0	0	0	0	0	0}
143	131	{0	1	0	1	0	0	0	0	1	0}
85	132	{0	0	1	0	0	0	0	0	0	2}
78	133	{0	0	0	2	0	0	0	0	0	1}
59	134	{0	0	0	1	0	0	0	0	2	0}
144	135	{0	1	0	1	0	0	0	1	0	0}
58	136	{0	0	0	1	0	0	0	0	1	1}
31	137	{0	0	0	0	0	2	0	0	0	1}
60	138	{0	0	0	1	0	0	0	1	0	1}
180	139	{1	0	0	0	0	2	0	0	0	0}
164	140	{0	2	1	0	0	0	0	0	0	0}
135	141	{0	1	0	0	0	2	0	0	0	0}
208	142	{1	1	0	1	0	0	0	0	0	0}
187	143	{1	0	0	1	0	0	0	0	0	1}
178	144	{1	0	0	0	0	1	0	1	0	0}
185	145	{1	0	0	0	1	1	0	0	0	0}
46	146	{0	0	0	0	1	1	0	0	0	1}
24	147	{0	0	0	0	0	1	0	1	0	1}
147	148	{0	1	0	1	1	0	0	0	0	0}
26	149	{0	0	0	0	0	1	0	2	0	0}
140	150	{0	1	0	0	1	1	0	0	0	0}

Assembly #	Rank	Molecular counts									
217	151	{2	0	0	1	0	0	0	0	0	0}
133	152	{0	1	0	0	0	1	0	1	0	0}
72	153	{0	0	0	1	1	0	0	0	0	1}
55	154	{0	0	0	0	2	1	0	0	0	0}
40	155	{0	0	0	0	1	0	0	1	1	0}
77	156	{0	0	0	1	2	0	0	0	0	0}
137	157	{0	1	0	0	1	0	0	0	1	0}
183	158	{1	0	0	0	1	0	0	1	0	0}
53	159	{0	0	0	0	2	0	0	1	0	0}
176	160	{1	0	0	0	0	1	0	0	0	1}
182	161	{1	0	0	0	1	0	0	0	1	0}
206	162	{1	1	0	0	0	1	0	0	0	0}
37	163	{0	0	0	0	1	0	0	0	1	1}
39	164	{0	0	0	0	1	0	0	1	0	1}
138	165	{0	1	0	0	1	0	0	1	0	0}
131	166	{0	1	0	0	0	1	0	0	0	1}
142	167	{0	1	0	1	0	0	0	0	0	1}
52	168	{0	0	0	0	2	0	0	0	1	0}
125	169	{0	1	0	0	0	0	0	1	1	0}
170	170	{1	0	0	0	0	0	0	1	1	0}
41	171	{0	0	0	0	1	0	0	2	0	0}
181	172	{1	0	0	0	1	0	0	0	0	1}
38	173	{0	0	0	0	1	0	0	0	2	0}
186	174	{1	0	0	0	2	0	0	0	0	0}
6	175	{0	0	0	0	0	0	0	1	1	1}
207	176	{1	1	0	0	1	0	0	0	0	0}
203	177	{1	1	0	0	0	0	0	0	1	0}
51	178	{0	0	0	0	2	0	0	0	0	1}
141	179	{0	1	0	0	2	0	0	0	0	0}
136	180	{0	1	0	0	1	0	0	0	0	1}
7	181	{0	0	0	0	0	0	0	1	2	0}
122	182	{0	1	0	0	0	0	0	0	1	1}
123	183	{0	1	0	0	0	0	0	0	2	0}
9	184	{0	0	0	0	0	0	0	2	1	0}
167	185	{1	0	0	0	0	0	0	0	1	1}
215	186	{2	0	0	0	0	1	0	0	0	0}
169	187	{1	0	0	0	0	0	0	1	0	1}
204	188	{1	1	0	0	0	0	0	1	0	0}
21	189	{0	0	0	0	0	1	0	0	0	2}
163	190	{0	2	0	1	0	0	0	0	0	0}
57	191	{0	0	0	1	0	0	0	0	0	2}
56	192	{0	0	0	0	3	0	0	0	0	0}

Assembly #	Rank	Molecular counts									
168	193	{1	0	0	0	0	0	0	0	2	0}
161	194	{0	2	0	0	0	1	0	0	0	0}
171	195	{1	0	0	0	0	0	0	2	0	0}
124	196	{0	1	0	0	0	0	0	1	0	1}
3	197	{0	0	0	0	0	0	0	0	2	1}
216	198	{2	0	0	0	1	0	0	0	0	0}
202	199	{1	1	0	0	0	0	0	0	0	1}
158	200	{0	2	0	0	0	0	0	0	1	0}
126	201	{0	1	0	0	0	0	0	2	0	0}
36	202	{0	0	0	0	1	0	0	0	0	2}
8	203	{0	0	0	0	0	0	0	2	0	1}
213	204	{2	0	0	0	0	0	0	1	0	0}
212	205	{2	0	0	0	0	0	0	0	1	0}
162	206	{0	2	0	0	1	0	0	0	0	0}
2	207	{0	0	0	0	0	0	0	0	1	2}
4	208	{0	0	0	0	0	0	0	0	3	0}
5	209	{0	0	0	0	0	0	0	1	0	2}
211	210	{2	0	0	0	0	0	0	0	0	1}
159	211	{0	2	0	0	0	0	0	1	0	0}
219	212	{2	1	0	0	0	0	0	0	0	0}
10	213	{0	0	0	0	0	0	0	3	0	0}
166	214	{1	0	0	0	0	0	0	0	0	2}
210	215	{1	2	0	0	0	0	0	0	0	0}
121	216	{0	1	0	0	0	0	0	0	0	2}
157	217	{0	2	0	0	0	0	0	0	0	1}
220	218	{3	0	0	0	0	0	0	0	0	0}
1	219	{0	0	0	0	0	0	0	0	0	3}
165	220	{0	3	0	0	0	0	0	0	0	0}

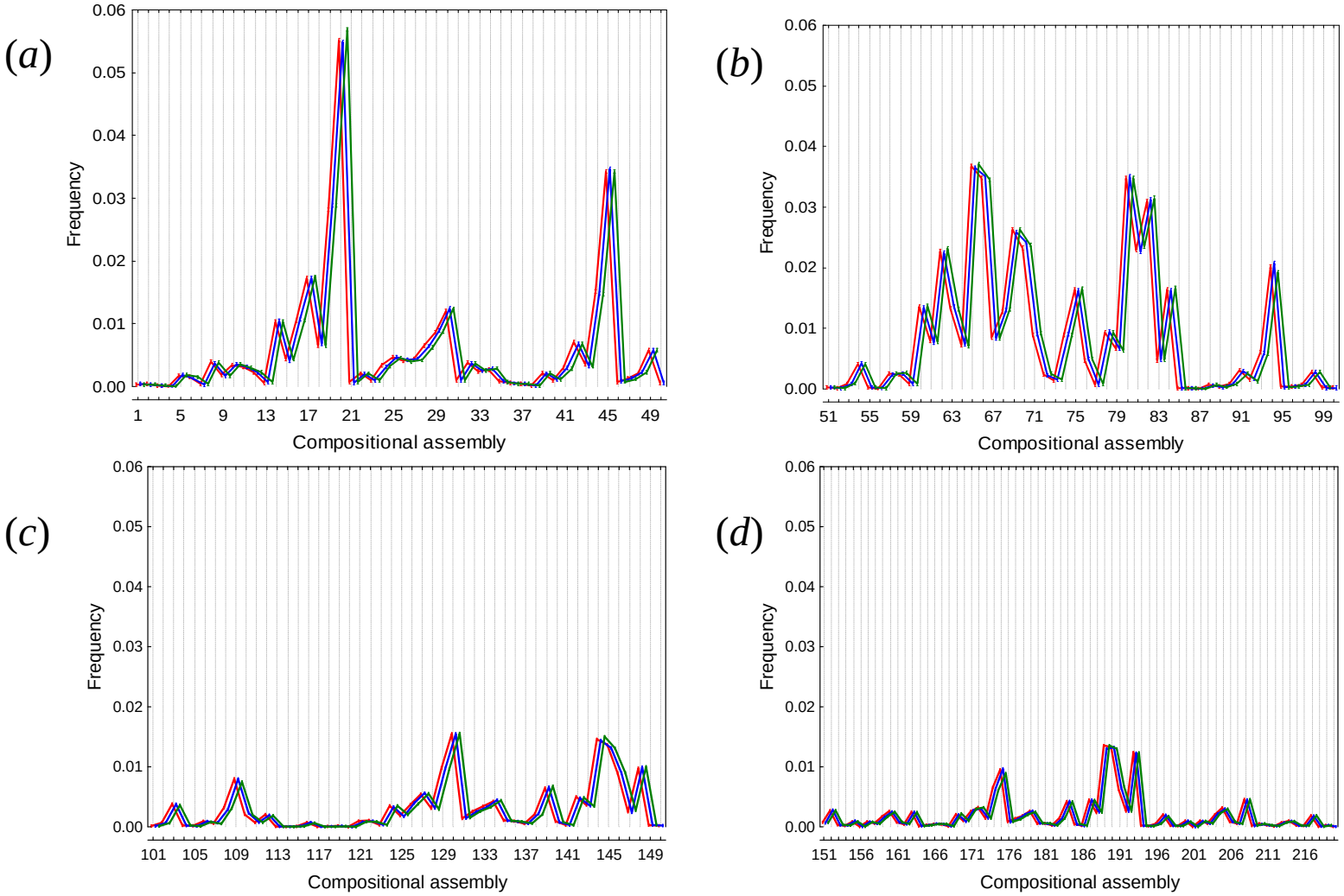
Figure S1. Log of the stationary distribution for the $\Omega = 220$ compositional assemblies in Table S1.



B Population dynamics of compositional assemblies within Eigen's framework

We statistically investigated the simulated behavior of the GARD system modeled within Eigen's framework after some slight modifications of the MATLAB m-file originally used by Segré *et al* (14). An initial 10-long vector assembly of size $N_{\min} = 3$ was randomly seeded. Discrete stochastic changes in the assembly were dictated by Equation [1] and at each time step the count of molecules of each species (Δn_i) was calculated from a Poisson distribution with average $F_i\left(\boldsymbol{\eta}^{\mathbf{10}}\right)\Delta t$, where Δt defines the time scale of the process (set to 0.05 sec). After reaching size $2N_{\min}$ the assembly was randomly divided and a daughter assembly of size $N_{\min} = 3$ was kept. The process continues during ~1,000 growth and split steps before sampling the final daughter assembly, which represents a single data point. The replication-mutation equilibrium distribution frequency of all possible assemblies was estimated from 100,000 independent data points adding up to one run (i.e., the ratio of sample size to number of assemblies of size $N_{\min} = 3$ is ~455). For the background situation without imposing selection in the GARD kinetic model compositional assembly $\boldsymbol{\eta}_{20}^{\mathbf{10}}$ ranked now first according to its replication-mutation equilibrium frequency. To test whether or not the system can significantly depart from the asymptotic steady-state solution dictated by Equation [1], simulations were also run imposing selection for two different target compositions. We have thus simulated three different situations: (*i*) GARD dynamics following Equation [1] (background case); (*ii*) GARD dynamics using as a target a compositional assembly at low replication-mutation equilibrium frequency; and (*iii*) using as a target a compositional assembly at intermediate replication-mutation equilibrium frequency. For each set of conditions six independent runs were performed, and the data set was analyzed by means of two-way analysis of variance (ANOVA) using the estimated replication-mutation equilibrium frequencies of each compositional assembly (transformed as $\arcsin\sqrt{q}$) as the dependent variable, and compositional assembly ($\boldsymbol{\eta}_k^{\mathbf{10}}$, $k = 1, 2, \cdots 220$) and simulated situation as fixed effects.

Figure S2. The plotted frequencies $(a)-(d)$ were obtained from computer simulations of the GARD model mimicking the analytical scenario for the exact solution of Eigen's equations. The number of possible compositional assemblies is $\Omega = 220$. The profiles of background (red line) and selection frequency distributions (green line: compositional target $\boldsymbol{\eta}_{94}^{10}$; blue line: compositional target $\boldsymbol{\eta}_{124}^{10}$) were tested by two-way ANOVAs (frequencies transformed as $\arcsin\sqrt{q}$), which detected statistically significant differences only in the red *vs.* green comparison: group-by-assembly interaction $F_{219,2200} = 0.93,\ P = 0.744$ with $\boldsymbol{\eta}_{124}^{10}$ as the target; $F_{219,2200} = 1.43,\ P < 0.001$ with $\boldsymbol{\eta}_{94}^{10}$ as the target.



C Hidden Compartmentalization in the β_{ij} Matrix

The compartmentalization of the β_{ij} matrix was evaluated based on the criterion

$$Crit = \left(\frac{\sum_{i=1}^{N_G} \sum_{i=j}^{N_G} \beta_{ij}}{{N_G}^2} * \sum_{i=1}^{N_G} \sum_{j=1}^{N_G} d_{ij} + \sum_{i=1}^{N_G} \sum_{j=1}^{N_G} \beta_{ij} s_{ij} \right) - \left(\sum_{i=1}^{N_G} \sum_{j=1}^{N_G} \beta_{ij} d_{ij} + \frac{\sum_{i=1}^{N_G} \sum_{i=j}^{N_G} \beta_{ij}}{{N_G}^2} * \sum_{i=1}^{N_G} \sum_{j=1}^{N_G} s_{ij} \right),$$

where $d_{ij} = 1$ if the nodes *i* and *j* belong to different compartments, otherwise $d_{ij} = 0$; and $s_{ij} = 1$ if the nodes *i* and *j* belong to the same compartment, otherwise $s_{ij} = 0$. The equation is modified from Krause et al. (2003. *Nature* 426:282-285); existing links within compartments and missing links between compartments increase, while missing links within compartments and existing links between compartments decrease the criterion. A simple hill-climbing algorithm was implemented in R (http://www.r-project.org/) for finding the compartment distribution where *Crit* reaches its maximum.

The quasi-compartmentalization was revealed (a) by systematically deleting links whose removal led to the greatest increase of compartmentalization, and (b) by deleting weak links below certain cut-off values. Due to computational limits, only the second method was employed in the case of the larger β_{ij} matrix. Cut-offs were set to be the *m*th power of 10, ranging from the $m = -6$ to $m = 2$ for $N_G = 10$ and from $m = -9$ to $m = 4$ for $N_G = 100$. For both calculations, the networks corresponding to the highest compartmentalization value were selected. Both methods lead to the quasi-compartmentalization shown in Fig. 2. Similarity of composomes and compartments is calculated as the percentage of molecules within a composome that belong to each compartment.

D Convergence to the late-time stationary distribution

Parameters in Equation [1] were $\Delta t = 0.05$, $k_i = 10^{-2}\,\text{sec}^{-1}$, $k_{-i} = 10^{-5}\,\text{sec}^{-1}$; and the $N_G \times N_G$ matrix for the catalytic enhancement factors β_{ij} were sampled from a lognormal distribution with parameters $\mu = -4$ and $\sigma = 4$. Let an initial random vector grow from $N_{\min} = 40$ to $2N_{\min}$ before random splitting and continue the process for 2,000 time steps (delay time from $N_{\min}$ to $2N_{\min}$ was 40). From this point on follow the time-dependent change in the concentration of each molecular species for an additional number of time steps (4,000; 10,000; 16,000). Obtain the covariance matrix of molecular concentrations ($\left[\mathrm{C}_1^{4,000}\right]$; $\left[\mathrm{C}_1^{10,000}\right]$; $\left[\mathrm{C}_1^{16,000}\right]$). Now repeat the procedure (with different initial random vectors) 1,000 times for each number of time steps and compare the covariance matrices (e.g., $\left[\mathrm{C}_1^{4,000}\right]$ *vs.* $\left[\mathrm{C}_i^{4,000}\right]$; $i = 2, 3, \cdots, 1001$) with a Mantel test by randomly permuting 10,000 times the rows and columns of one of the matrices. Mantel test indicates whether or not the covariance pattern of concentrations is similar between the two runs, and the results are shown in Fig. S3. A good approximation to the late-time stationary distribution seems to be reached after 10,000 – 16,000 additional time-steps; that is, a consistent estimate of the covariance matrix for the time-dependent change in the concentration of each molecular species is obtained.

Figure S3. Results of the Mantel tests. Average correlation for 4,000 time-steps was $r_{(t=4,000)} = 0.0659$, with an average probability $P = 0.0717$; for 10,000 time-steps $r_{(t=10,000)} = 0.4533,\ P = 0.0018$; and for 16,000 time steps $r_{(t=16,000)} = 0.4566,\ P = 0.0012$.

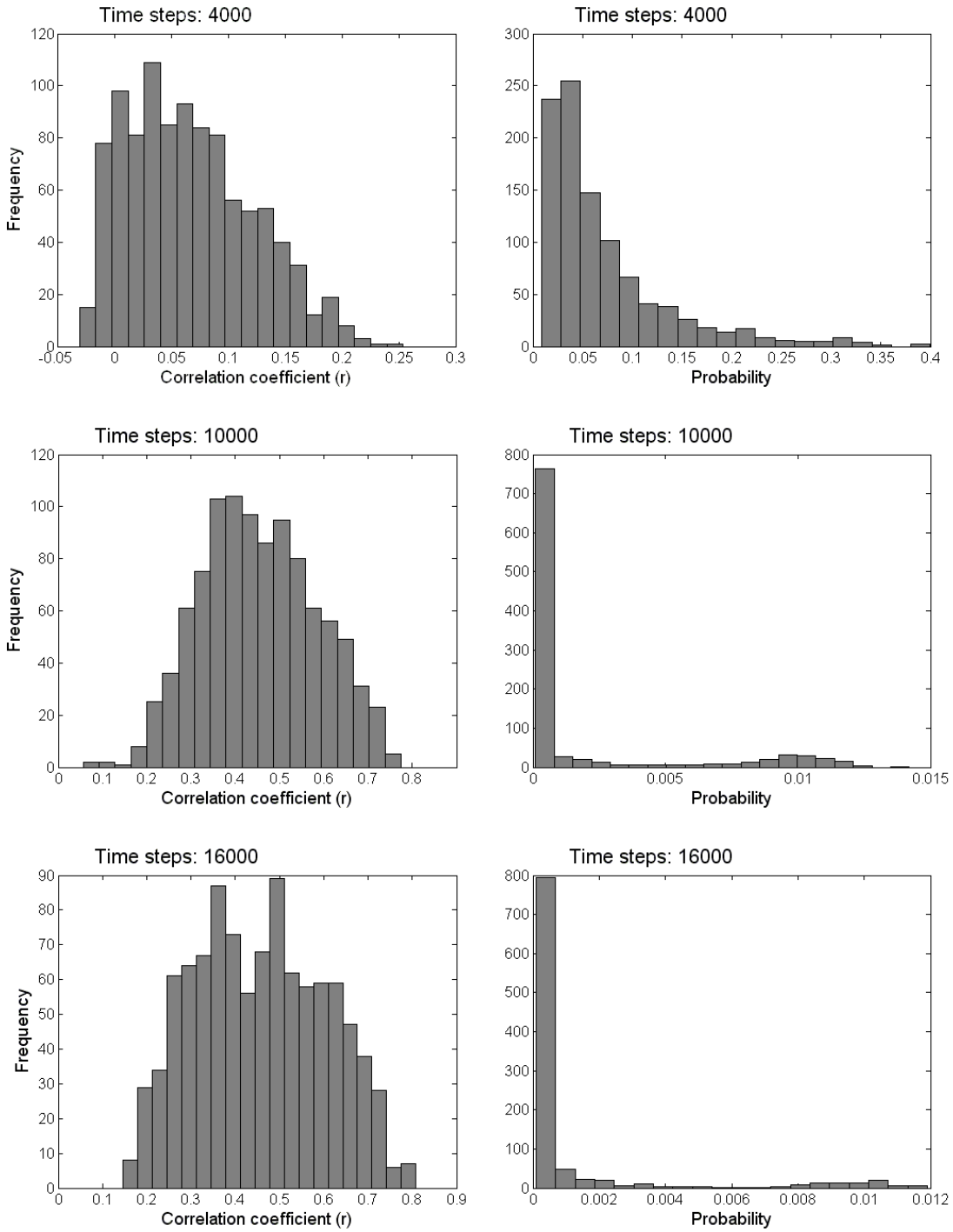


Figure S4. Principal component analysis of the covariance matrix of time-dependent molecular concentrations after 16,000 time steps (~400 generations). Eigenvalues are plotted relative to the total variance. It is clear that there are only a few PCs whose eigenvalues are relatively large. The first 5 PCs explain 96.6% of the total variance.