

1 **Metabolically Coupled Replicator Systems: Overview of an RNA-World model concept of**
2 **prebiotic evolution on mineral surfaces**

3
4
5 **Tamás Czárán**^{1,*}

6 ¹MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, H-1117 Pázmány
7 Péter sétány 1/c, Budapest, Hungary
8 Email: czaran@caesar.elte.hu

9
10 **Balázs Könnnyű**²

11 ²Eötvös Lorand University, Department of Plant Systematics, Ecology and Theoretical Biology, H-
12 1117 Pázmány Péter sétány 1/c, Budapest, Hungary
13 Email: konnyu@caesar.elte.hu

14
15 **Eörs Szathmáry**^{1,2,3}

16 ¹MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, H-1117 Pázmány
17 Péter sétány 1/c, Budapest, Hungary

18 ²Eötvös Lorand University, Department of Plant Systematics, Ecology and Theoretical Biology, H-
19 1117 Pázmány Péter sétány 1/c, Budapest, Hungary

20 ³Parmenides Institute for the Conceptual Foundations of Science, Kirchplatz 1, D-82049
21 Munich/Pullach, Germany
22 E-mail: szathmary.eors@gmail.com

23

24 *Corresponding author

25 **Abstract**

26 Metabolically Coupled Replicator Systems (MCRS) are a family of models implementing a simple,
27 physico-chemically and ecologically feasible scenario for the first steps of chemical evolution
28 towards life. The hypothetical starting point of the scenario is a large population of RNA(-like)
29 macromolecules produced abiotically on a suitable spot of prebiotic Earth, attached to a mineral
30 surface capable of binding both the macromolecules and the monomers they are made of. Evolution
31 sets in as soon as any one of the RNA molecules become *autocatalytic* by engaging in template
32 directed self-replication from activated monomers, so that its population size starts increasing
33 exponentially. Competition for the finite external supply of monomers ignites selection favouring
34 RNA molecules with catalytic activity helping self-replication by any possible means. The most
35 straightforward way of providing such catalytic help is to become a *replicase ribozyme* offering a
36 new self-copying mechanism, even if it is only marginally more efficient than the one available
37 before. An additional way is through increasing monomer supply by contributing to monomer
38 synthesis from external resources, i.e., by evolving *metabolic enzyme activity*. *Retroevolution* may
39 build up an increasingly autotrophic, cooperating community of metabolic ribozymes running an
40 increasingly complicated and ever more efficient metabolism.

41 Maintaining such a cooperating community of metabolic replicators raises two serious
42 ecological problems: one is keeping the system *coexistent* in spite of the different replicabilities of
43 the cooperating replicators; the other is *constraining parasitism*, i.e., keeping “cheaters” in check.
44 Surface-bound MCRS provide an automatic solution to both problems: the coexistence of
45 cooperating replicators and their parasite resistance are the consequences of assuming the local
46 nature of metabolic interactions. In this review we present an overview of results published in
47 previous articles, showing that these effects are, indeed, robust in different MCRS implementations,
48 by considering different environmental setups and realistic chemical details in a few different
49 models. We argue that the MCRS model framework naturally offers a suitable starting point for the
50 future modelling of membrane evolution and extending the theory to cover the emergence of the
51 first protocell in a self-consistent manner. The coevolution of metabolic, genetic and membrane
52 functions is hypothesized to follow the *progressive sequestration* scenario, the conceptual blueprint
53 for the earliest steps of protocell evolution.

54
55 **Keywords:** early molecular community, stability, coexistence, spatially explicit model, cellular
56 automata

57 **Graphical Abstract**

58 **1. Introduction**

59
60
61 The problem of the origin of life is a scientific question, but one with a strong historical dimension.
62 The historical aspect raises at least two difficulties which seem impossible to overcome. First, those
63 who venture into the field of prebiotic evolution should be prepared to accept the fact that very
64 likely noone will ever be able to factually verify or falsify claims on any hypothetical series of
65 events that would have produced the first living organism, simply because no fossil proof of any
66 kind can be hoped for from the enormous distance of over 3 billion years ago to support such
67 hypotheses. Second, we have no clue on what alternative histories of prebiotic chemical evolution
68 could have existed on the prebiotic Earth, since the chemical universalities of all recent organisms
69 suggest that the actual history is unique, although it is well possible that different attempts had been
70 made by radically different prebiotic chemical systems, and the one that successfully launched life
71 as we know it today had won the competition between those possible candidate systems at a very
72 early phase.

73
74 Therefore, studying the process of prebiotic evolution is largely restricted to the domain of

75 the *possible*, not the *actual*: we may search for scenarios that are feasible from a physical-chemical
76 point of view and are reconcilable with the chemical organization of recent forms of life. Even
77 though we cannot tell with certainty what actually happened, we may have reasonably strong
78 scientific arguments to decide what could have happened and what not (Eschenmoser, 2007).
79 *Systems chemistry* (von Kiedrowski et al. 2010) offers a wide range of theoretical and experimental
80 methods for constructing and testing possible evolutionary scenarios of prebiotic evolution, from
81 the very beginning to the emergence of the first living cell. We attempt to sketch such a scenario in
82 this paper, one that we believe is both feasible and open to further improvements through the
83 inclusion of more detailed and more realistic physical and chemical mechanisms.

84 Molecular interactions had shaped the chemical evolution of prebiotic macromolecular
85 structures with a selective power almost as efficient as that of competitive and sexual interactions
86 driving the evolution of living creatures today. In view of all that we know – or suspect – about the
87 earliest phases of the origins of life these molecular interactions may have played the key roles in
88 the transformation of matter from inanimate to animate.

89 Most students of the origin of life agree that chemical evolution must have started with the
90 formation of small organic molecules (like formaldehyde, hydrogen-cyanide etc.) through abiotic
91 (geo)chemical processes. Lacking sufficiently accurate information of the climatic and geochemical
92 environment on Earth four billion years ago, the study of even this initial step of the wake of life is
93 largely speculative and often controversial with respect to the actual details (Martin and Russel
94 2003; Miller, 1953; Miyakawa et al. 2002; Monnard et al. 2003; Powner et al. 2009;
95 Wächtershäuser, 1990), so much so that the hypothetical initial sets of prebiotic organic compounds
96 show a large variety across the literature. Whatever their chemical identities were, those small
97 organic molecules must have reacted with each other to produce macromolecules which later
98 formed macromolecular complexes or communities by hypothesized self-assembly processes or
99 coexistence mechanisms of different kinds (Chen and Walde, 2010; Cleaves et al. 2012; Deamer
100 and Weber, 2010; Ehrenfreund and Cami, 2010; Ferris, 2006; Garay, 2011; Johnson et al. 2008;
101 Miller, 1953; Miyakawa et al. 2002; Orgell, 2004; Powner et al. 2009; Rushdi and Simoneit, 2001).
102 The mechanisms of self-assembly and macromolecular community formation are often theoretically
103 problematic, either because the assumptions of the underlying (toy) models are too schematic or
104 because they are physically or chemically unrealistic (Morowitz et al. 2000; Pross, 2004;
105 Szathmáry, 2006; Segré et al. 2001). The actual chemical and evolutionary details of the many
106 different scenarios are usually implicit, so it is often difficult to see how the envisioned
107 macromolecular complex or community could be a self-sustaining and self-regulated unit of life or
108 of evolution (Gánti, 1987; Rasmussen et al. 2009).

109 Historically the first prebiotic replicator community model was Eigen's *hypercycle* (Eigen
110 and Schuster, 1979). It was conceived to solve the chicken-and-egg problem of reliable replication:
111 one would need a long replicase in the first place that would be able to accurately copy itself. This
112 poses the question how the first such long replicase could have emerged and persisted without the
113 copying accuracy required? Lacking an efficient replicase the copying process is hampered by
114 frequent errors, leading to an *error catastrophe* (Eigen and Schuster, 1979) for sequences longer
115 than the *error threshold*. The hypercycle was thought to solve the problem by splitting the long
116 sequence into short ones which are not prone to the error catastrophe. To avoid competition among
117 the fragments they are organized in a structure such that they help each other's replication in a
118 cyclical topology. The cumulated size of the members in a hypercycle may exceed the size limit set
119 by the error threshold for single sequences. Theoretical considerations have proven that the simple
120 hypercycle cannot be evolutionarily stable (Boerlijst, 2000; Boerlijst and Hogeweg, 1991; Bresch et
121 al. 1980; Kim and Jeong, 2005). Two types of mutants both may ruin the cooperation of the
122 replicators in a hypercycle: *selfish parasites* (replicators helping themselves but not the downstream
123 neighbour in the cycle) cut the cyclical flow of benefits, whereas *shortcut parasites* (helping
124 another member of the hypercycle instead of the downstream neighbour) exclude some members

125 from the circular flow of benefits, repeated shortcut mutations ultimately reducing the hypercycle to
126 a single member (Fig.1).

127

128 **1.1. The Metabolic Replicator paradigm**

129 Our approach, like the hypercycle, is one of the “evolution of coexistent replicator
130 communities” scenarios, aimed at explaining the dynamical stability and the evolvability of a
131 hypothetical community of macromolecules provided by an initially random RNA World (Gilbert,
132 1986; Joyce, 2002). The *Metabolically Coupled Replicator System (MCRS)*, we believe, is the most
133 feasible candidate suggested so far for a prebiotic chemical supersystem that may have evolved into
134 the first cellular form of life, the *protocell*. We shall explain below why we think so.

135

136 The central proposal of the RNA-World scenario is that the first evolvable entities on
137 prebiotic Earth may have been RNA (or RNA-like – (Eschenmoser, 2007; Hall, 2004; Robertson
138 and Joyce, 2010)) macromolecules, and the first protocell is the evolutionary product of an RNA(-
139 like) macromolecular community, the members of which were connected by a specific set of
140 mutually advantageous interactions. This suggestion is appealing for a number of very important
141 reasons. RNA inherently embodies the first two of the three indispensable *infrabiological*
142 (Szathmáry et al, 2005) components of living systems (*metabolism*: catalytically channelled
143 reaction network producing compounds necessary for reproduction; *genetics*: hereditary
144 information transmission through template replication; and *membrane*: partial separation of the
145 biological entity from the outside world). First, RNA has been proven to possess a wide range of
146 enzymatic activities (Chen et al. 2007; Landweber et al. 1998; Lilley, 2003) that are absolutely
147 necessary for driving even a very primitive metabolism. Second, RNA is inherently modular, i.e., it
148 is composed of a few chemical modules (nucleotides) in a linear arrangement, so that the sequence
149 of the modules may carry information. Sequences may be of a virtually unrestricted variety, and the
150 unambiguous complementation of the modules allows for the template replication of the sequences
151 (Szathmáry, 2006), i.e., for genetic information transmission through generations of RNA
152 molecules. It is this inherent dual (metabolic and genetic) role of RNA which earned the name
153 “*Metabolically Coupled Replicator System*” to our prebiotic evolutionary scenario. The third
154 infrabiological component (membrane) may be the product of subsequent evolution within the
155 metabolic RNA community, or its function might have been initially supplied by specific
156 environmental conditions, as we will show later (Branciamore et al. 2009).

157 Of course there are large gaps in our knowledge with respect to prebiotic chemistries
158 capable of delivering activated modules for the replication of early RNA-World molecules. Yet, we
159 have no other choice at present but assuming that the RNA-World was initially absolutely
160 “heterotrophic”, that is, the first RNA(-like) macromolecules were randomly assembled from
161 activated modules which in turn were the products of so far largely unknown geochemical
162 processes. “Black smokers” (hot and high pressure volcanic vents thousands of meters below sea
163 level in the ocean-beds) seem to be reasonably good candidates for having supplied the modules
164 (Deamer and Weber, 2010; LaRowe and Regnier, 2008; Orgel, 2004), but we are still far from even
165 an established hypothesis on this topic. Fortunately, there is much more known about the
166 possibilities of non-template-directed RNA synthesis from activated monomers (nucleotides).
167 Experimental results suggest that mineral (clay) surfaces – like that of montmorillonite – can
168 catalyse spontaneous bond formation between activated nucleotides (Ferris, 2006, Ferris et al.
169 1999), resulting in RNA molecules of different lengths and random nucleotide sequences.

170 Once a sufficiently diverse random set of RNA molecules is available, the stage is set for the
171 evolution of a sustainable, cooperative RNA-World scenery to play out (Copley et al. 2007;
172 Manrubia and Briones, 2007). An essential criterion of this to happen is that the RNA molecules
173 originally produced by spontaneous bond formation become *template replicated*. In fact this is the
174 single most crucial condition for evolution to set in and select for RNA assemblies somewhat more

175 efficient in replication than others. It seems very reasonable to assume that some random RNA
176 molecules – or an assembly of a few different ones – generated on the mineral surface might have
177 had a weak RNA-replicase activity. This would have been sufficient to ignite the selection process
178 for a gradual increase of replicase activity (Attwater et al. 2010; Johnson et al. 2001). The snag with
179 this straightforward reasoning is empirical: even though RNA replicase ribozymes are actively
180 searched for in many laboratories (Johnson et al. 2001), the ones discovered so far are incapable of
181 replicating themselves because they are longer than the longest template they can handle (Attwater
182 et al. 2010; Wochner et al. 2011). In spite of the lack of a real breakthrough in this respect so far,
183 this is one of the most promising directions of experimental research on prebiotic evolution: we
184 seem to be quite close to having a proper RNA replicase ribozyme at hand (Attwater et al. 2010;
185 Wochner et al. 2011).

186 There is another, comparably important criterion to be met for the evolution of the RNA-
187 World towards the first living cell to proceed, and that one is *ecological* in nature. Assume that we
188 have a few different self-replicating RNA replicase ribozymes and a constant supply of activated
189 monomers from a geochemical source. The initial excess of resources (monomers) allows all the
190 replicases to reproduce and establish their own populations. These exponentially increasing
191 replicase populations will inevitably exhaust the constant monomer supply sooner or later,
192 ultimately reducing the concentration of available monomers in the environment to a break-even
193 level at which even the *fastest replicating* (i.e., fittest) replicase population stops growing. At the
194 break-even resource level the populations of all other replicases are already *decreasing*, and they
195 will continue doing so until they go extinct. This is the result of *competition* among the different
196 replicase species for monomers, leading to the survival of the fittest, i.e., the victory of the most
197 efficient replicase. According to the Gause principle of competition (Meszéna et al. 2006) the
198 maximum number of coexistent replicator populations is equal to the number of different resources
199 they exploit. If the replicators do not discriminate with respect to the different monomers that they
200 use for their own replication, then there can be only a single winner. If the different monomer types
201 (A, U, C, G for RNA) count as different resources, i.e., if different replicators use the monomers
202 differentially, then the number of potentially coexistent replicators is equal to the number of
203 monomer pairs (two in this case: A+U and C+G, (Szilágyi et al. 2013)). Since the differential use of
204 monomers by the replicators – i.e., a marked difference in their A+U and C+G demand - would
205 represent a severe constraint on their function (replicase activity), it seems reasonable to assume
206 that the monomer pool constitutes a single resource, which implies a single winner. That is, for
207 more than a single replicator species to coexist, some mechanism is needed that circumvents the
208 problem of competitive exclusion, because further evolutionary improvements of the victorious
209 replicase depend on its *cooperation* with RNA molecules helping its own population growth.

210 The straightforward ally could be a ribozyme catalysing a reaction that produces monomers
211 from another geochemically supplied resource, i.e, a simple *metabolic enzyme*. The adoption of a
212 single-step metabolism would be driven by the selective pressure on the replicase to exploit new
213 resources present in the environment from which extra supplies of activated monomers can be
214 produced. Ribozymes from the random-sequence replicator population of the RNA-World may be
215 selected for the useful metabolic function and copied by the replicase, which in turn benefits from
216 the increased monomer supply. This mutualistic interaction (cooperation) of the replicase and the
217 metabolic ribozyme allows for a shift of the system towards *autotrophy* through the *construction of*
218 *a new niche* that, thanks to their cooperation, becomes available for both the replicase and the
219 metabolic ribozyme. The new niche is the potential to exploit the new compound – a resource thus
220 far useless in replication – that the metabolic replicator is able to convert to activated monomers.
221 The repeated inclusion of new metabolic ribozymes into the evolving RNA replicator community
222 implies increasing autotrophy and metabolic efficiency of the reaction network through the process
223 of *retroevolution of metabolism* ((Horowitz, 1945), Fig.2.). Our Metabolically Coupled Replicator
224 System (MCRS) has been developed for studying the dynamical properties and the evolutionary

225 potential of such a community of cooperating ribozymes. The main questions to answer with the
226 models are:

227

228 • Can a metabolically coupled set of replicators be coexistent in spite of the inevitable
229 competitive interaction between the different replicator types? If so, under what
230 environmental conditions does coexistence occur?

231 • How many metabolic replicators can be coexistent in MCRS?

232 • Can the MCRS resist the invasion of parasitic replicators which use the monomers and the
233 service of the replicase for their reproduction but do not contribute to monomer production
234 or replication at all?

235 • Can metabolically active replicators develop from a random replicator set?

236 • Is there any further evolutionary potential in MCRS through the acquisition and the
237 development of new replicator functions?

238

239 **2. General assumptions and the mean-field version of MCRS**

240 Below we detail the basic assumptions of the MCRS model family, first specifying the
241 mean-field version in which no spatial structure of the replicator community is considered. After
242 showing that the mean-field model is not viable, we turn to the assumptions related to the spatial
243 structure of the surface-bound RNA World, and specify the details of the spatially explicit core
244 version of the MCRS scenario.

245

246 **Assumption 1.** *The chemical identity of early replicators.* The MCRS framework does not make
247 explicit assumptions with respect to the chemical identity of prebiotic replicators, but
248 straightforward general principles constrain the possibilities to modular (and, consequently, digital)
249 structures capable of unlimited heredity (Szathmáry, 2006). These constraints practically exclude
250 the majority of known chemical entities from among the plausible molecule types, except for
251 variants of recent nucleic acids and proteins (Eschenmoser, 2007; Hall, 2004; Robertson and Joyce,
252 2010; Nielsen, 2009). Most researcher of the origin of life today agree that RNA, or RNA-like
253 molecules are by far the most likely entities responsible for booting up life on Earth 3-4 billion
254 years ago (Chen et al. 2007; Gilbert, 1986; Joyce, 2002; Robertson and Joyce, 2010). The MCRS is
255 built on the RNA world scenario allowing for some chemical variations but maintaining the
256 postulates of a modular, template-replicated macromolecule as the basic chemical entity of prebiotic
257 evolution.

258 **Assumption 2.** *Error-free replication.* As explained earlier, one of the most difficult “missing links”
259 in the MCRS scenario is that of RNA replication. The sequence of a relatively simple, yet
260 sufficiently accurate RNA-dependent RNA polymerase ribozyme has not been discovered so far.
261 Evolving such a replicase ribozyme is one of the biggest challenges for recent *in vitro* RNA
262 evolution experiments (Attwater et al. 2010; Johnson et al. 2001; Rohatgi et al. 1996; Wochner et al.
263 2011). Lacking an efficient RNA replicase ribozyme we need to assume for the time being that the
264 template replication of RNA molecules was nevertheless possible at the time of the wake of life.
265 The most straightforward solution would be to suppose that there was a – so far undiscovered –
266 replicase ribozyme present in the RNA world after all, which seems not to be unrealistic given the
267 promising experimental results lately. A minor difficulty arises from the omission of the fact that
268 any template replication is prone to mismatch errors (*mutations*) resulting in copies slightly
269 different from the template. In fact this is the *error catastrophe* problem that the coexistence models
270 of prebiotic evolution (i.e., the hypercycle, (Eigen and Schuster 1979), the stochastic corrector
271 model (Szathmáry and Demeter, 1987), parabolic growth models (Szathmáry and Gladkih, 1989)
272 and MCRS (Czárán and Szathmáry, 2000; Károlyi et al. 2002)) are meant to solve in the first place,
273 but it is essentially circumvented by the assumption that the genetic information to be transmitted is
274 split into short sequences. Therefore MCRS makes the simplifying assumption that RNA replication

275 is error-free on the ecological time scale for which the coexistence of metabolic replicators is
276 investigated.

277 **Assumption 3. Double-stranded RNA.** Another difficulty related to the problem of experimental
278 RNA replication is that even if the complementary strand can be formed, the copy cannot be
279 separated from the template without imposing chemical conditions on the system that are very far
280 from any reasonable assumption of prebiotic environmental conditions (Szathmary, 2006; Patzke
281 and von Kiedrowski, 2007). For lack of empirical knowledge on this issue we are again forced to
282 assume that strand separation does occur somehow due to a mechanism so far unknown. As an
283 initial simplifying assumption we assume that the sister strands of replicating RNA molecules are
284 identical – an assumption that will be relaxed later (see Section 4).

285 **Assumption 4. Enzymatic activity of replicators.** Many different RNA molecules are known to take
286 part in several vital biochemical processes of recent cells as catalysts (*ribozymes*, (Cech, 2009)).
287 Early prebiotic RNA world systems must have relied mostly on the catalytic potential of ribozymes,
288 because translation and thus more efficient protein enzymes are later achievements of evolution.
289 The broad catalytic potential of RNA molecules was justified in different independent experimental
290 studies (Bartel and Unrau, 1999; Chen et al. 2007; Landweber et al. 1998, Lilley, 2003).

291 **Assumption 5. Metabolism.** The key assumption of the MCRS is that each member of a set of
292 different replicator types (i.e., replicator macromolecules of different nucleotide sequences)
293 catalyses a single reaction in a hypothetical metabolic reaction network in which their own building
294 blocks (*monomers*) are produced. Therefore monomers for replication are self-supplied only in the
295 presence of a complete set of metabolic replicators (Fig.3.A.); any one of them missing halts
296 monomer production altogether. Notice that we do not yet assume any explicit topology and
297 stoichiometry for the metabolic reaction network here, even though it might be of substantial effect
298 on the actual dynamics of the metabolic replicator system.

299

300 Based on these assumptions the mean-field version of the MCRS (Czaran and Szathmary,
301 2000) model can be set up, in which the change of the frequencies (concentrations) of the metabolic
302 replicators (f_i) are given as

303

$$304 \quad \frac{df_i}{dt} = f_i(k_i \cdot M - \varphi(f)) \quad , \quad \text{Eq. 1}$$

305

306 where k_i is the replicator-specific growth rate, $\varphi(f)$ is the outflow function which keeps the total
307 concentration of replicators constant within the system, without altering their relative frequencies.
308 M is the efficiency of metabolism, the network of chemical reactions in which each of the
309 individual reactions is specifically catalysed by one of the metabolic replicators. Metabolic
310 efficiency is calculated as the geometric mean of the replicator frequencies (concentrations) within
311 the system:

312

$$313 \quad M = \left(\prod_{i=1}^n f_i \right)^{\frac{1}{n}} \quad , \quad \text{Eq. 2}$$

314

315 where n is the number of essential metabolic replicator types in the system. The metabolic function

316 M is the same for all the replicator types, because it represents the concentration of the product of
317 metabolism, i.e., the supply of monomers, which is the single common resource of self-reproduction
318 for all the replicators present in the system. Therefore the only parameter that determines the growth
319 rate of replicator i is its replication rate k_i in Eq.1. Consequence: the replicator of highest k_i
320 competitively excludes all the other ones and the metabolic community collapses in the mean-field
321 version of the MCRS model (Czárán and Szathmáry, 2000). In fact the system is exterminated
322 already by the exclusion of the *first* essential metabolic replicator type, because $f_i = 0$ for any i
323 implies $M = 0$ in Eq.2. Note that Eigen and Schuster (Eigen and Schuster, 1979) had considered and
324 outright rejected a model of similar dynamics, precisely because it is not coexistent in a well-mixed
325 system.

326

327 **3. The spatial version of the Metabolically Coupled Replicator System – the Metabolic Replicator** 328 **Model (MRM)**

329 The disappointing conclusion of the mean-field model turns to its exact opposite with the
330 assumption that the MCRS is bound to a mineral surface, so that the interactions of the replicators
331 (metabolic cooperation and competition for monomers) become locally context dependent.

332 Experimental data of very different sorts provide strong indirect support for the idea:
333 mineral underwater surfaces (rocks of pyrite, clay minerals like montmorillonite, etc.) can be
334 catalysts for nucleotide binding (Ferris, 2006, Ferris et al. 1999); they might be responsible for the
335 homochirality of biomolecules (Hazen et al. 2001, Joshi et al. 2011); they are supposed to have
336 assisted membrane production and thus the formation of the first proto-cells (Hanczyc et al. 2007);
337 and they may have protected replicators from the harmful effects of UV radiation (Biondi et al.
338 2007).

339

340 **3.1. Space-related assumptions of the spatially explicit MCRS model**

341 **Assumption 6.** *Replicators are bound to mineral surfaces.* The most probable arena for prebiotic
342 replicator evolution may have been on mineral surfaces which can bind RNA molecules reversibly
343 through divalent cations (Franchi et al. 2003). Detachment and re-attachment of parts of the
344 macromolecules result in their caterpillar-like movement on the surface, which is in turn responsible
345 for their limited rate of spatial mixing – a feature that later will be shown to be of crucial
346 importance for their population dynamics. The two-dimensional arena is a lattice of binding sites,
347 each site harbouring a single replicator at a time. Replicator movement is represented by swapping
348 the contents of neighbouring sites. We specify the details of replicator movement in Section 5.

349 **Assumption 7.** *Initial replicator diversity generated by spontaneous polymerisation.* Surface-
350 catalysed RNA polymerisation results in a diverse pool of oligo- and polynucleotides of different
351 lengths and random nucleotide sequences (Copley et al. 2007; Garay, 2011, Ma et al. 2007;
352 Manrubia and Briones, 2007). This random community of replicators is then selected for useful
353 metabolic functions contributing to monomer production.

354 **Assumption 8.** *Local metabolic interactions on the surface.* The limited mobility of replicators on
355 the mineral surface makes their metabolic and competitive interactions local. Local metabolic
356 interactions mean that the metabolite molecule produced by a ribozyme replicator needs to be
357 delivered to the ribozyme catalysing the next reaction of metabolism before the metabolite decays
358 or desorbs from the surface. This requires that the corresponding metabolic replicators be
359 sufficiently close to each other in space. As an implicit proxy to this criterion we assume that all the
360 metabolically essential replicators need to be present within a certain area called the *metabolic*
361 *neighbourhood* (Fig.3.C.) around a replicator so that it has a sufficient local monomer supply for its
362 replication. This corresponds to the local application of Eq. 2 within each metabolic neighbourhood
363 instead of the whole replicator community.

364 **Assumption 9.** *Surface diffusion of metabolites.* The detailed chemical nature of precursors,
 365 intermediary metabolites and monomers is disregarded in the MCRS, just like the topology of the
 366 metabolic reaction network itself. What we implicitly consider are a few general features of small
 367 molecules in relation to their movement on and detachment from the mineral surface. We assume
 368 that small molecules move on the surface faster than macromolecules do, and they can desorb from
 369 the surface with a probability higher than replicators. Both of these assumptions reflect that small
 370 molecules (e.g. monomers) are certainly less attached to the surface than macromolecules.

371 **Assumption 10.** *Local competition for monomers.* Like metabolic interactions, competition is also
 372 local in the spatially explicit MCRS model: replicators within the *replication neighbourhood* (cf.
 373 Fig.3.C.) of an empty site compete for the possibility to put a copy of themselves onto the focal
 374 empty site. The chance of replicator I_i to win depends on its replication parameter k_i and its local
 375 monomer supply M_i .

376 3.2. The stochastic cellular automaton implementations of MCRS

377 *Basic model setup.* The computer implementation of the Metabolically Coupled Replicator System
 378 scenario is a series of stochastic cellular automaton (SCA) models: the Metabolic Replicator Model
 379 (MRM) family. A set of n different, metabolically active ribozyme replicators are assumed to
 380 compete for the monomers which they produce themselves in cooperation, through catalysing the
 381 reactions of a simple metabolism (Fig.3.A.). Each replicator occupies a site of the SCA lattice
 382 representing the mineral surface on which all the interactions take place. The opposite margins of
 383 the lattice are merged forming a toroidal structure to avoid edge effects. The number of possible
 384 states for a site is $n + 1$, including the “empty” state and the n different occupied states. The lattice
 385 size we used throughout the simulations was 300×300 , which is sufficiently large to avoid strong
 386 periodic effects but is still manageable in terms of computer resources. One generation (from t to $t +$
 387 1) consists of elementary updates equal in number with the number of sites in the lattice (90.000).
 388 The updating algorithm is random: the state of each site is updated once per time unit on average, in
 389 a random order (*asynchronous updating rule*).

391 *Update processes: replication and decay.* Empty and occupied sites are updated by separate
 392 algorithms. Occupied sites turn to the “empty” state (*replicator decay*) with the constant replicator
 393 decay probability p_d . “Empty” sites can become occupied by a copy of one of the replicators from
 394 within the *replication neighbourhood*; the replicators there compete for the focal empty site. The
 395 chance of a replicator to win the competition and put a copy of itself to the empty site depends on
 396 its replication parameter and the local monomer supply within the *metabolic neighbourhood* of the
 397 focal replicator (Fig.1.C.). The size of the metabolic neighbourhood is considered proportional to
 398 the average distance that a small molecule (metabolite or monomer) can cover by surface diffusion
 399 before it either desorbs from the surface or is consumed in a replication process. The individual
 400 “claim” C_f of the replicator f for occupying the empty site depends on its monomer supply M_f and
 401 its specific replication rate k_f as

$$402 \quad C_f = k_f \cdot M_f \quad , \quad (Eq. 3)$$

405 and

$$406 \quad M_f = n \sqrt{\prod_{i=1}^n x_i(f)} \quad , \quad (Eq. 4)$$

407 where $x_i(f)$ is the number of type replicator i within the metabolic neighbourhood of the focal

408 replicator f , and i runs through all replicator types needed to catalyse the metabolic reactions ($i = 1,$
 409 \dots, n). Thus, the local monomer supply of the focal replicator f depends on the presence of *all*
 410 metabolic replicators within its own metabolic neighbourhood – with any one of the n metabolic
 411 replicator types missing the corresponding $x_i(f) = 0$ and thus also $M_f = 0$. This in turn implies no
 412 local monomer production and therefore no chance of replication for the focal replicator f . Each
 413 replicator within the replication neighbourhood of an empty site has a chance to occupy the empty
 414 site with a copy of itself:

$$415 \quad p_f = \frac{C_f}{C_e + \sum_m C_m} \quad , \quad (\text{Eq. 5})$$

416 where m runs through all replicators within the replication neighbourhood of the focal replicator f ,
 417 and C_e is a constant representing the claim of the empty site for remaining empty. Obviously, the
 418 probability that the empty site remains empty is

$$419 \quad p_e = \frac{C_e}{C_e + \sum_m C_m} \quad . \quad (\text{Eq. 6})$$

420 Note that in the basic MRM there is no specialised replicase replicator in the system. It is implicitly
 421 supposed here that the replicase “service” is supplied either by the mineral surface itself, or by a
 422 very rudimentary replicase ribozyme which is present in excess on the surface. This implicit
 423 assumption will be relaxed later by the explicit inclusion of a replicase ribozyme.

424 *Replicator diffusion.* The movement of replicators on the mineral surface is implemented using the
 425 Toffoli-Margolus algorithm: randomly chosen 2x2 blocks of sites are rotated by 90° left or right
 426 with equal (0.5) probability (Toffoli and Margolus, 1987). The intensity of replicator diffusion is
 427 scaled by the average number D of diffusion steps per site per generation. Note that even $D = 0$
 428 represents some minimum mixing of replicators on the surface, due to the fact that each newborn
 429 copy is placed into a site different from – adjacent to – the one occupied by the parent (template).

430
 431 *Structured (porous) habitat.* The basic model was also modified to account for the dynamical effects
 432 of an *ab ovo* compartmentalised, i.e., structured, habitat. Many of the possible minerals on which
 433 prebiotic replicator evolution might have taken place are in fact of a porous structure (Fig.4.). The
 434 pores, which are connected by capillary channels, represent compartments relatively separated from
 435 other pores. In this spatially structured version of the MRM the pores take the role of interaction
 436 (metabolic and replication) neighbourhoods: each pore is considered as an open stirred-tank reactor
 437 connected by the in- and outflow of small molecules and, occasionally, of macromolecular
 438 replicators as well. Each pore can support a certain number of replicators (*pore capacity*), and the
 439 concentrations of small molecules (“resources” and “monomers”) are explicitly followed. A given
 440 fraction of streaming small molecules can dock within the pore reducing pore capacity. The
 441 metabolic efficiency (M) of a pore is calculated based on its replicator and monomer contents,
 442 taking the monomer threshold for replication into account. Since the structured model is more
 443 explicit in terms of chemistry, and somewhat different in terms of spatial structure compared to the
 444 basic model, it is of very high importance to evaluate its predictions against those of the basic
 445 MRM.

446 447 **4. Spatially explicit simulations**

448 The most striking result of the spatially explicit models of the MCRS scenario is that all
449 implementations are very robustly coexistent within a broad range of their space-related parameters.
450 This suggests that local interactions among a set of metabolically essential ribozyme replicators are
451 sufficient to maintain their cooperation and to neutralise, or at least to reduce, the competitive
452 effects which drive the mean-field system (cf. Section 2) to extinction. A typical run of the non-
453 structured simulation yielded the time series on Fig. 5, with 4 metabolic replicators of substantially
454 different replication parameters k_i .

455

456 **4. 1. The ecology of the spatial models**

457 The space-related parameters of the basic MRM which are relevant for the coexistence of the
458 metabolic replicator community are: 1) metabolic neighbourhood size, 2) replication neighbourhood
459 size and 3) the diffusion parameter of the replicators. Fig.6. summarizes the results of a series of
460 simulations scanning through the space of these three parameters.

461 What explains the fundamental difference in the dynamics of the mean-field model and the
462 spatial models? The answer is that in the spatial model the local range of metabolic cooperation
463 gives an indirect advantage to *rare* replicator types through local metabolism, because their
464 metabolic neighbourhoods are easily complemented by the more common types, therefore they
465 have a better chance for replication than the common types, most of which lack the presence of the
466 rare type within their metabolic neighbourhood. The larger the difference in frequency between two
467 replicator types the larger the advantage of rarity.

468 Eq.3. implies that the fitness of a replicator (I_f) consists of two components – a *direct* and an
469 *indirect* one. k_f , the specific and constant replication parameter is the direct fitness component: low
470 k values provide few opportunities for replication, thus the density of the corresponding replicator
471 type in the community is low (the replicator is rare). The advantage of the rare type comes from the
472 indirect fitness component (M_f), and it acts through the better local monomer supply of the rare
473 types on average. The complete metabolic replicator community is coexistent when the fitnesses of
474 all the replicator types are equal: $C_i = C_j$ for any (i, j) . That is, replicators of low direct fitness
475 compensate for their handicap by a higher indirect fitness. Since the indirect fitness component
476 increases with rarity, the negative feedback of replicator population density on fitness regulates the
477 community to coexistence in a broad range of the parameter space.

478 Obviously, the spatial parameters of the model (metabolic neighbourhood size, replication
479 neighbourhood size and replicator mobility) affect coexistence through their effects on the indirect
480 components of replicator fitnesses. Let us consider these in turn.

481 *Metabolic neighbourhood:* The size of the metabolic neighbourhood is a proxy to the distance that
482 metabolites and monomers travel by surface diffusion before disappearing either by desorption or
483 by reaction (cf. Assumption 8). Very small metabolic neighbourhoods mean a very localised
484 metabolism, which translates to a strong advantage of rarity: low frequency metabolic replicators
485 with very small metabolic neighbourhoods have better chances for replication than common ones,
486 because their indirect fitness component is very high. Increasing the metabolic neighbourhood shifts
487 the system towards the mean-field approximation; in the limit case of the metabolic neighbourhood
488 being equal to lattice size (300x300) we arrive at the mean-field model which we know to go
489 extinct (cf. Section 2).

490 *Replication neighbourhood:* The replication neighbourhood of an empty site corresponds to the
491 distance to which the „offspring” of a replicator can be placed from its parent. Common sense
492 suggests that this should not be large, because it is difficult to imagine the mechanism which could
493 put the copy far from the template in spite of the relatively strong adherence of both to the surface.
494 A long-distance movement by the copy would require its detachment from, and then its distant re-
495 attachment to the surface – a very unlikely series of events indeed. Even so, increasing the size of
496 the replication neighbourhood has an obvious mixing effect: it decreases the probability that the

497 offspring remains close to the parent, i.e., the chance of aggregated pattern formation decreases.
498 Note, however, that increasing the replication neighbourhood *does not* shift the system towards the
499 mean-field case, because the metabolic advantage of rarity remains the same.

500 *Replicator mobility (diffusion)*: Faster replicator movement means better mixing, too. It is
501 obviously advantageous for the coexistence of the metabolic replicator community, especially if the
502 metabolic neighbourhood is small. Less mixing would lead to the aggregation of conspecific
503 replicators, which would drastically decrease the chance of metabolic complementation on the
504 spatial scale of local metabolism (i.e., at the scale set by metabolic neighbourhood size). Replicator
505 mobility (diffusive mixing) increases the overall fitness of the community by increasing the number
506 of complete metabolic neighbourhoods, and thus the indirect fitness of *all* replicators in the system.

507 The combined effects of these three spatial parameters on the stationary states of the MRM system
508 are shown on Fig.6. (Könnyű and Czárán, 2013). The best conditions for the coexistence of
509 metabolically coupled replicator communities are at relatively small metabolic neighbourhood sizes
510 and intensive replicator mixing (the latter condition seen at large replication neighbourhoods and/or
511 high replicator mobility).

512 The parameters of the spatially structured „pore-model” analogous to metabolic
513 neighbourhood size and replicator mobility in MRM are pore size (i.e., the maximum number of
514 replicators fitting into a pore) and replicator migration, respectively. Fig.7. shows that within the
515 coexistent section of the space of these two parameters the trend in the pore-model is the same as in
516 MRM: larger pore size decreases, whereas more replicator mobility increases the fitness (and the
517 mean density of the replicator community). Since the pore model is more explicit in terms of
518 chemical detail (i.e., it considers the constant input of a „resource compound” which can be
519 converted to monomers by the replicator community of a pore, provided it is metabolically
520 complete), the convergence of the results of the two models is encouraging.

521

522 The original problem which MRM (and the hypercycle model) intended to solve is the
523 maintenance of genetic information surpassing the error threshold and sufficient to code for a
524 machinery complicated enough to be capable of its own reproduction (Eigen and Schuster, 1979;
525 Kun et al. 2005; Maynard-Smith, 1979; Niesert, 1987; Niesert et al. 1981; Takeuchi and Hogeweg,
526 2007). Considering this problem as the central one, there is another parameter of the MRM of
527 crucial importance: the number of different replicator types that the model can keep coexistent, i.e.,
528 the maximum attainable system (genome) size (n). One simple constraint is trivial: system size
529 cannot exceed the maximum number of replicators fitting into the metabolic neighbourhood, or else
530 a complete local metabolism is impossible, so larger metabolic neighbourhoods should be able to
531 harbour larger systems. However, increasing metabolic neighbourhood size decreases the advantage
532 of rarity at the same time; therefore we expect the largest possible viable systems to be maintained
533 at intermediate metabolic neighbourhood sizes. The actual attainable system size is also limited by
534 the level of spatial mixing – the more intensive it is, the larger the biggest sustainable system should
535 be. We have tested maximum viable system size as the function of space-related model parameters
536 both in MRM and in the pore-model (Fig. 8.), and the results confirm these expectations: high
537 mixing and intermediate metabolic neighbourhood sizes allow for the coexistence of over 10
538 replicators. Towards the limit of infinite diffusion the Metabolic Replicator Model approaches
539 Wilson’s trait group mechanism of coexistence (Maynard-Smith and Szathmáry, 1995; Szathmáry,
540 1992; Wilson, 1975).

541

542 **4.2. Parasites, complementary strands and facultative cooperators in MRM**

543 The metabolic replicator system, like any cooperative community, is exposed to “cheaters”, i.e.,
544 individuals taking advantage of cooperation by others, but not investing into cooperation
545 themselves. Such free-riders enjoy the fitness advantage of reduced resource investment compared

546 to cooperators, and they spread in the community until cooperation breaks down altogether. This is
547 what happens in parasite-infected cooperative communities without a proper reward/punishment
548 scheme in effect. Intentional rewarding or punishment is, of course, out of question in
549 macromolecular communities. The cooperating members of a hypercyclically coupled replicator
550 community have no means of feeding back the damage from parasitism to the parasite itself. This is
551 why the naked hypercycle is doomed to collapse upon the emergence of mutant replicators acting as
552 selfish or shortcut-parasites (cf. Introduction).

553 The only conceivable parasite of the MCRS is one that uses up monomers for its replication
554 but does not contribute to monomer production (Fig.3.B.). Any mutant failing to contribute to the
555 common good is a parasite; therefore we expect a whole range of different parasites – a parasitic
556 quasispecies - to emerge in any metabolic replicator system. Neglecting the slight differences in
557 their dynamically relevant parameters we lump the members of the parasitic quasispecies into a
558 single replicator category. The metabolic cooperation mechanism of the surface-bound MCRS
559 provides an “automatic” delivery of efficient punishments to such selfish parasites: wherever they
560 pop up and start spreading, monomer production is impaired, which in turn locally stops the
561 replication of all replicators including the parasite. Since extinctions occur only where parasites
562 prevail, local extinction decimates the parasite more than cooperators. This mechanism is
563 sufficiently powerful to keep the parasitic quasispecies in check even if it has the highest replication
564 parameter in the community (Fig.9.A.). High parasite replicability is a feasible assumption, since no
565 enzymatic function constrains the secondary structure of a parasite: it will be selected for fast
566 reproduction, i.e, it should be short and loosely folded.

567
568 The most relevant parameter of the MRM with respect to its parasite resistance is not the
569 replication rate of the parasite, but replicator mobility. At limited replicator mobility even a very
570 fast reproducing (of large k_p) parasite will attain a low steady state frequency in the MCRS, but at
571 high mobilities the parasite can destroy cooperation (Branciamore et al. 2009; Czárán and
572 Szathmáry, 2000). Since very high mobilities are not reasonable to assume for surface-bound
573 macromolecules, this extreme case does not constrain the feasibility of the model. The feasible
574 space-related parameter range of a viable MRM exposed to parasitic mutants is therefore small to
575 moderate metabolic neighbourhood sizes and small to moderate replicator mobilities. We did not
576 explicitly tackle the dependence of diffusibility and desorption rate on replicator length in this
577 model, but we may safely assume that short parasitic replicators move faster and desorb easier from
578 the surface than longer, metabolically active replicators do. These two effects of sequence
579 shortening are assumed to quench each other: one is advantageous and the other is detrimental for
580 parasite persistence.

581
582 Note that the parasitic quasispecies can be rather heterogeneous with respect to length and
583 replication parameter, but for the cooperating replicator community the only dynamically relevant
584 feature of a parasite is its being a cheater (i.e., that it does not contribute to monomer production).
585 The replicability of the parasite is quite irrelevant for the cooperators, as they can repress them
586 through the metabolic “punishment” mechanism anyway. The difference in the replication rates of
587 different parasites plays a role only in inter-parasite competition: the fastest replicating type of the
588 parasitic quasispecies excludes all the other types (Könnnyű and Czárán, 2013), in perfect
589 accordance with the Gause principle ((Meszéna et al. 2006), Fig.9.B.). Thus the outcome of a
590 typical MRM + parasitic quasispecies simulation is the coexistence of all cooperators and the fastest
591 parasite, with the latter attaining a low and steady equilibrium frequency in the community.

592 Two special modifications of the MCRS model deserve mention here, because their
593 dynamical consequences are somewhat similar to that of introducing parasites. The first such
594 modification is relaxing the template/copy identity postulate (Assumption 3). RNA template and
595 copy strands are not identical but complementary in their nucleotide sequences, and possibly very

596 different in secondary structure (except for palindromes, (Boza et al. 2014; Ivica et al. 2013)). The
597 complementary strand (i.e., the “gene”) of a metabolically active ribozyme is, in all probability,
598 functionally inactive, which makes the copy of the ribozyme similar to a parasite from an ecological
599 point of view: it consumes monomers, but it does not contribute to producing them. The difference
600 is that the copy is the offspring of a ribozyme, and the copy of a copy is a functional ribozyme
601 again, which is not the case with a real parasite. Assuming that the functional (ribozyme) forms
602 have lower replicability than the complementary strands (Ivica et al. 2013; Könnyű and Czárán,
603 2014) because of their – presumably more compact – secondary structure, the system behaves like
604 the MRM + parasite quasispecies model, except that the “gene” copies do not exclude each other:
605 all the complementary strands coexist with the metabolically active ribozyme forms. The MCRS
606 proved to be viable in this pheno/geno version as well (Fig.10.A.), which is not a big surprise given
607 the robust parasite resistance of the original MRM (Könnyű and Czárán, 2014).

608 The other special modification of the model is the inclusion of *facultative metabolic*
609 *cooperators*: replicators which increase the efficiency of metabolism, but are not essential for
610 monomer production. The metabolic benefit provided by a facultative cooperator may come, for
611 example, from its acting as a co-factor of another, essential metabolic ribozyme. Such facultative
612 cooperators are very similar to parasites in their dynamical properties, except that their negative
613 effect of diluting the local assembly of essential ribozymes is counteracted by their positive effect
614 on metabolic efficiency. Of course they also coexist with the original MRM, and exclude other
615 parasites which do not help metabolism (Könnyű and Czárán unpub.) (Fig.10.B.).

616

617 **5. Adaptive evolution in MRM**

618 The stable coexistence of the core of MRM (the metabolically essential replicator set) with a non-
619 functional parasitic replicator is the most important feature of the surface-bound MCRS from the
620 viewpoint of its evolvability. The benefit of the presence of a parasitic replicator lies in its pre-
621 adaptive value: it remains persistent in the functioning MCRS without causing much damage, and it
622 can freely mutate to obtain new functions potentially increasing the fitness of the community of
623 cooperating replicators. The most straightforward adaptive enhancements of the system may
624 advance through improvements of the existing metabolic ribozymes, simply by selection towards
625 better, or more specialised, enzymatic activities.

626

627 **5.1 Adaptations improving metabolic efficiency**

628 Better catalyst may drive better metabolism, and local replicator communities fed by more efficient
629 metabolism will obviously displace others from the surface by competition. However, adaptations
630 towards better metabolic functions are necessarily traded off with replicability: more efficient
631 ribozyme structures tend to be more compact, therefore they are also more difficult to unfold and
632 replicate. On the other hand, less efficient enzymes can be more versatile in terms of possibly
633 catalysing more than a single reaction of metabolism, if two (or more) different, but energetically
634 similar foldings of the macromolecule are possible, and each has some catalytic activity with
635 respect to a metabolic reaction different from those of the other foldings. Such “promiscuous”
636 catalytic activities have been reported both for protein (Khersonsky and Tawfik, 2010; O’Brien and
637 Herschlag, 1999) and for RNA (Ancel and Fontana, 2000; Schultes and Bartel, 2000) enzymes. The
638 different catalytic effects of promiscuous ribozymes are also in a trade-off relation one with the
639 other, for at least two different reasons. First, spending time in one of the foldings means that the
640 other folding is inactive, i.e., the ribozyme works in time-sharing mode. Second, the fact that the
641 molecule is able to trans-fold to other secondary structures implies that none of its secondary
642 structures is very stable (compact): a handicap with respect to its catalytic activity in each folding.
643 These trade-off constraints are reflected in the following supplementary assumption applied in the
644 next modification of MRM:

645

646 **Assumption 11. Trade-offs in replicator features.** We assume two-way trade-offs among the
 647 features of metabolic ribozymes with potentially two different catalytic activities. The first trade-off
 648 is between the two enzymatic activities, the second one is between the enzymatic activities and the
 649 replicability of the metabolic replicator molecule. These trade-off constraints restrict the available
 650 combinations of the three features below the trade-off surface shown on Fig.11 (Könnyű and
 651 Czárán, 2011). The parameters b and g of the trade-off function scale the strength of the trade-off
 652 between the two enzymatic activities and the enzymatic activities and replicability, respectively.

653
 654 Replacing the numbers of type i ribozymes $x_i(f)$ with the *total type i activity* of the different
 655 replicators k ($\sum_{k=1} E_{i,k}(f)$) within the metabolic neighbourhood of the focal replicator f leads to the
 656 substitution of Eq.4 with

$$657 \quad M_f = n \sqrt[n]{\prod_{i=1}^n \sum_{k=1} E_{i,k}(f)} \quad , \quad \text{Eq. 7}$$

658 in which each metabolic replicator k within the given metabolic neighbourhood counts with *only*
 659 *one* of its enzymatic activities i , which is drawn at random with weights of choice proportional to
 660 the actual activities $E_{i,k}$ of replicator k . If the focal replicator f is the one copied from among the
 661 candidates in the replication neighbourhood, then the copy is either identical with its template or –
 662 with a small probability – it is a mutant with its catalytic activities and replicability constrained by
 663 the trade-off surface, but otherwise chosen at random.

664 Depending on the shape of the trade-off function of catalytic activities (parameter b), and on
 665 replicator mobility D , the simulations reveal two different outcomes: the system ends up either in a
 666 dominantly “specialist” replicator community consisting of single-activity metabolic ribozymes, or
 667 in the dominance of “generalists”, i.e., catalytically less efficient but bifunctional replicators
 668 (Fig.12.). The criteria for ribozyme specialization proved to be hard trade-off (low values of b)
 669 between the two catalytic activities, and moderate replicator mobility – both criteria falling in the
 670 most feasible zone of the parameter space. Hard trade-off means that the sum of the two enzyme
 671 activities of “generalists” (bifunctional ribozymes) is less than that of any of the two “specialists”.
 672 Of course more mixing (larger D) is beneficial for specialization, because it prevents the
 673 aggregation of identical templates and copies, which prevents metabolic complementation of
 674 specialists. Note that parasites (replicators with both of their catalytic activities next to zero, but
 675 with very high replicability) are kept at very low frequencies in this model implementation, just like
 676 in all previous ones, provided that replicator mobility is not extremely high (Könnyű and Czárán,
 677 2011).

678 5.2. The evolutionary acquisition of new functions by the metabolic replicator community

680 Besides improving the catalytic activities of existing members of the metabolic replicator
 681 community as explained above, even more innovative adaptations might come from adopting new
 682 functions by mutants of either a core replicator or the parasite of the system. The mutant may be a
 683 new metabolic ribozyme possibly opening a new, more efficient chemical route to monomer
 684 production, but it may obtain other functions increasing the fitness of the cooperating replicator
 685 community in radically different ways. Such adaptations may, for example, accelerate the
 686 replication process itself through improving the replicase ribozyme, thus increasing the fitness of
 687 each replicator in the community; or they might contribute to the completion of the system with the
 688 third essential infrabiological (Szathmáry et al, 2005) component of life: the membrane envelope.

689
 690 *Replicase evolution:* For evolutionary adaptation to take place within the MCRS, template
 691 replication has to work one way or another. This requires the replicase function to be available from
 692 the outset, either as a service of the environment (some mineral surfaces are shown to have a basic
 693 catalytic activity helping spontaneous RNA template replication – (Ferris, 2006)) or in the form of a

694 simple replicase ribozyme of the initial random replicator population capable of copying itself and
695 other RNA replicators (Fig.13.A. (Könnyű et al. 2008)). Assuming that the metabolic replicator
696 community has already domesticated a parasite for the replicase function, that replicase can mutate
697 to become worse or better in its role. Allowing for mutations both in the negative and the positive
698 direction of replicase activity we studied the evolutionary dynamics of the core MRM + replicase
699 system, based on the following supplementary assumption:

700

701 **Assumption 12.** *Beneficial and deleterious mutations of the replicase.* Parasites can mutate to
702 obtain increasing template replicase activity (beneficial for MCRS) or replicase inhibitory effects
703 (deleterious for MCRS). Both these types of mutation are traded off with the replicability of the
704 parasite: the stronger the functional effect of the mutation (in any direction), the smaller the direct
705 fitness component (k) of the replicator.

706

707 Simulations reveal that inhibitory parasites disappear from the system, because 1) they kill off
708 metabolic replicators locally more efficiently, thus committing suicide faster, or 2) they evolve to
709 higher replicability to the expense of their inhibitory effect. On the other hand, beneficial mutants
710 spread and achieve substantial frequencies (Fig.13.B.) in the replicator community, provided that
711 the trade-off relation of replicase activity and replicability is not very rigid, and replicator mobility
712 is not too high. Moreover, the overall density of the replicator community also increases as higher
713 replicase activity builds up, indicating the evolutionary benefit of improving an aspecific replicase
714 function to the whole system.

715

716 6. Perspectives of the MCRS approach

717

718 First a few comments on the relationship of the systems surveyed here to Gánti's various
719 suggestions of chemical supersystems are in order. The 'classis' 1971 Hungarian edition of the
720 Principle of Life (Gánti, 1971) coined the term 'chemoton', but then it referred to the system
721 doublet made of an autocatalytic metabolic cycle and a template replicator only. In this sense we
722 have also dealt with similarly organized infrabiological systems (Szathmáry et al, 2005). But there
723 is a crucial difference: the idea of hereditary catalytic effects by the templates entered Gánti's
724 thinking only towards the end of the seventies only (Gánti, 1978), and then strictly within the fully-
725 fledged metabolism-boundary-genetic material tripartite systems that is nowadays being referred to
726 as the chemoton We think, however, that catalytic reaction channelling must have been an
727 indispensable feature of any chemical supersystem maintaining even a minimal metabolism and
728 capable of self-reproduction (Deamer and Weber, 2010; Meléndez-Hevia et al. 2008; Orgel, 2000,
729 2004; Pross, 2004; Szathmáry et al, 2005). Therefore, the most important agents of an early
730 metabolism-replicatorsystem must have been the catalysts which, through the metabolism they
731 drive, can produce their own building blocks, using externally supplied raw materials. The early
732 RNA-World hypothesis provides an excellent starting point for a feasible scenario of the origin of
733 life, because RNA is involved in two of the three infrabiological functions of life: a wide spectrum
734 of catalytic activities for driving practically any *metabolism*, and a large variability of template-
735 complementary module (nucleotide) sequences to attain *unlimited heredity* and self-reproduction
736 (Chen et al. 2007; Gilbert, 1986; Joyce, 2002; Landweber et al. 1998).

737

738 Evolution requires reproduction, i.e., self-copying of sufficient accuracy. Even though we
739 cannot yet pinpoint the agent which could have been able to copy itself at the earliest stages of
740 chemical evolution, RNA, or RNA-like macromolecules are the primary suspects for this role as
741 well. Recent laboratory experiments are very promising, with their results getting ever closer to the
742 discovery of a self-replicating ribozyme, i.e., an RNA replicase that copies diverse RNA molecules
743 of at least its own size with a sufficiently low mutation rate. Once we have that, the stage is ready
744 for the evolution of an increasingly complex metabolism supplying monomers for the replicase

744 population, through the sequential adoption of other replicators playing the roles of metabolic
745 enzymes in a metabolic reaction network that becomes increasingly autotrophic (*retroevolution*,
746 (Szathmáry, 2007)). This process is driven by the ecological pressure towards occupying (or
747 constructing) new niches: the inclusion of new compounds supplied by the environment for
748 metabolism as old external resources become exhausted, one after the other, by the exponentially
749 increasing replicator population.

750 The other ecological constraint on the dynamics of the evolving replicator community is the
751 avoidance of competitive exclusion of any of the ribozymes playing a vital role in maintaining
752 metabolism and replication. The metabolically coupled replicator system (MCRS) model was
753 developed to demonstrate that this is possible, if the system is bound to a mineral surface, thereby
754 increasing the viscosity (i.e., limiting the spatial mixing) of the interacting replicators. The MCRS
755 model resists parasitic replicators in the sense that, under physico-chemically reasonable
756 assumptions, parasites cannot kill the system, even though they remain persistent at low
757 frequencies. Deleterious mutants of either the metabolic cooperators or the parasites are doomed to
758 extinction, because by hindering obligatory cooperation they decimate neighbouring cooperators
759 and thus cut their own monomer supply – in effect, they behave as suicide bombers, and go extinct.

760 The MCRS model framework may be improved in two main directions: in depth, by
761 explicitly considering important chemical details that have not been addressed so far; and in
762 extension, by broadening the approach to include completely new directions of MCRS evolution.
763 Some in-depth variants of MRM are being studied already: besides the “pheno-geno” version
764 considering replication to produce complementary strands ((Könnyű and Czárán, 2013), cf. Section
765 4.2), simple explicit metabolic reaction topologies with explicit metabolite and monomer
766 production and diffusion have been shown to work (Kőrössi et al, *unpub.*). These simulation studies
767 also show the limits of the surface-bound MRM. The size and the topology of the metabolic
768 network, just like the number of possibly coexistent replicators, are constrained mainly by the same
769 spatial factors that make the system work: the local nature of interactions, and the limited range of
770 the surface diffusion of replicators and metabolites. The limits set by these spatial constraints can be
771 pushed further out only by extending the MRM approach to new mechanisms of selection. The key
772 to such modifications is the unavoidable, but rarely fatal, presence of parasitic replicators in the
773 cooperating replicator community.

774 Neutral mutants of persistent parasites are free to random-walk across the sequence space
775 and may find functions beneficial for the cooperating replicator community. Such converted
776 parasites can be adopted by the system and might radically increase its fitness, by opening new,
777 efficient metabolic routes, improving replication (cf. Section 5.2), or producing membranogenic
778 molecules and trans-membrane channels.

779 Membrane production is the critical step towards the occurrence of the first protocell,
780 allowing for a new organizational level to occur, and new mechanisms for its evolution. Acquiring
781 the ability of membrane synthesis could provide the replicator community with *individuality*, of
782 profound evolutionary consequences. Autonomous membrane production could be achieved
783 through some mutant parasites evolving to ribozymes catalysing the production of membranogenic
784 (amphipathic) molecules from other metabolites, and the spontaneous insertion of their product into
785 the expanding membrane (Fig.14.). Encapsulating a replicase-aided MCRS into self-supplied
786 membrane compartments would establish a more effective, new level of selection for further
787 evolution of the system – it would be the organizational level of the *protocell*. The stoichiometric
788 coupling of membrane production to metabolism ensures the synchrony of doubling metabolite
789 content and membrane surface, which warrants the possibility of protocell fissions maintaining the
790 original volume/surface ratio through indefinitely many generations. Once in place, the membrane
791 capsule can adopt selective permeability functions or even active pumping of resource compounds
792 into the protocell, by evolving specific membrane-bound ribozymes (Khvorova et al. 1999).
793 Through such adaptations the protocell could achieve independence from the mineral substrate and

794 enter a new evolutionary regime: that of the internal reorganization of genetic, metabolic and
795 transport functions, towards the cellular state as we know it in recent organisms. The simultaneous
796 (co-)evolution of the genetic, the metabolic and the membrane subsystems could have occurred
797 through the progressive sequestration scenario (Szathmáry, 2007), with metabolism becoming more
798 complex, membrane channels more selective and genetic material organized in chromosomes.
799 Modelling the early phases of protocell evolution along these lines is the intended direction of our
800 future extensions to MRM; Fig.14. is a caricature of the idea, the model implementation of which is
801 a task for the future.

802

803

804 **Competing interests**

805 The author(s) declare that they have no competing interests

806

807 **Authors' contributions**

808 TC, BK and ESz designed, analysed and interpreted the introduced studies. All authors contributed
809 to writing the manuscript and approved the final version.

810

811 **Acknowledgements**

812 TC and BK acknowledge financial support from the Hungarian Research Foundation (OTKA Grant
813 No. K100806). E. Sz. was supported by European Research Council under the European
814 Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement No.
815 [294332] and partly by EU COST action CM1304 "Emergence and Evolution of Complex
816 Chemical Systems."

817

818 **References**

- 819
- 820 AnceL, W.L., Fontana, W., 2000. Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool. (Mol.*
821 *Dev. Evol.)* 288, 242–283.
- 822 Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A., Holliger, P., 2010. Ice as a protocellular
823 medium for RNA replication. *Nature Commun.* 1, 76.
- 824 Bartel, D.P., Unrau, P.S., 1999. Constructing an RNA world. *Trends. Genet.* 15, M9–M13.
- 825 Biondi, E., Branciamore, S., Maurel, M.C., Gallori, E., 2007. Montmorillonite protection of an UV-
826 irradiated harpine robozyme: evolution of the RNA world in a mineral environment. *BMC*
827 *Evol. Biol.* 7, Suppl 2.
- 828 Boerlijst, C., Hogeweg, P., 1991, Spiral wave structure in pre-biotic evolution: hypercycles stable
829 against parasites. *Phisyca D.* 48, 17–28.
- 830 Boerlijst, C., 2000. Spirals and spots: novel evolutionary phenomena through spatial self-
831 structuring. In Dieckmann U., Law R., Metz, J.A.J. (Eds.), *The Geometry of Ecological*
832 *Interactions*, IIASA and Cambridge University Press, Cambridge pp. 171–182.
- 833 Boza, G., Szilágyi, A., Kun, A., Santos, M., Szathmáry, E., 2014. Evolution of the division of labor
834 between genes and enzymes. *PLoS Comput. Biol.* 10, e1003936,
835 doi:10.1371/journal.pcbi.1003936.
- 836 Branciamore, S., Gallori, E., Szathmáry, E., Czárán, T. 2009. The origin of life: chemical evolution
837 of metabolic sytem in a mineral honeycomb? *J. Mol. Biol.* 69, 458-469.
- 838 Bresch, C., Niesert, U., Harnasch, D. 1980. Hypercycles, parasites and packages. *J. Theor. Biol.* 85,
839 399–405.
- 840 O'Brien, P.J., Herschlag, D., 1999. Catalytic promiscuity and the evolution of new enzymatic
841 activities. *Chem. Biol.* 6, R91–R105.
- 842 Cech, T.R., 2009. Crawling out of the RNA World. *Cell.* 136, 599–602.
- 843 Chen, I.A., Walde, P., 2010. From self-assembled vesicles to protocells. *Cold Spring Harb. Perspect.*
844 *Biol.* 2, a002170.
- 845 Chen, X., Li, N., Ellington, A.D., 2007. Ribozyme catalysis of metabolism in the RNA world.
846 *Chem. Biodiv.* 4, 633–655.
- 847 Cleaves, H.J., Ii, H., Scott, M.A., Hill, F.C., Leszczynski, J., Sahai, N., Hazen, R., 2012. Mineral-
848 organic interfacial processes: potential roles in the origins of life. *Chem. Soc. Rev.* 41, 5502–
849 5525.
- 850 Copley, S.D., Smith, E., Morowitz, H.J., 2007. The origin of the RNA world: Co-evolution of genes
851 and metabolism. *Bioorg. Chem.* 35, 430–443.
- 852 Czárán, T., Szathmáry, E., 2000. Coexistence of replicators in prebiotic evolution. In Dieckmann U.,
853 Law R., Metz, J.A.J. (Eds.), *The Geometry of Ecological Interactions*, IIASA and Cambridge
854 University Press, Cambridge pp. 116–134.
- 855 Deamer, D.W., Weber A.L., 2010. Bioenergetics and life's origins. *Cold Spring Harb. Perspect.*
856 *Biol.* 2, a004929.
- 857 Ehrenfreund, P., Cami, J., 2010. Cosmic carbon chemistry: from the interstellar medium to the early
858 Earth. *Cold Spring Harb. Perspect. Biol.* 2, a002097.
- 859 Eigen, M., Schuster, P. 1979. *The hypercycle*, first ed. Springer-Verlag, Berlin.
- 860 Eschenmoser, A., 2007. The search for the chemistry of life's origin. *Tetrahedron.* 63, 12821-12844.
- 861 Ferris, J.P., 2006. Montmorillonite-catalysed formation of RNA oligomers: the possible role of
862 catalysis in the origins of life. *Philos. T. Roy. Soc. B.* 36, 1777–1786.
- 863 Ferris, J.P., Hill, R.A., Liu, R., Orgel, L.E., 1996. Synthesis of long prebiotic oligomers on mineral
864 surfaces. *Nature.* 381, 59–61.
- 865 Franchi, M., Ferris, J.P., Gallori, E., 2003. Cations as mediators of the adsorption of nuclec acids on
866 clay surfaces in prebiotic environments. *Origins Life Evol. B.* 33, 1–16.
- 867 Garay, J., 2011. Active centrum hypothesis: the origin of chiral homogenity and the RNA-World.

868 BioSystems. 103, 1–12.

869 Gánti, T., 1971. Az élet princípium (The principles of life), first ed. Gondolat, Budapest (In
870 Hungarian).

871 Gánti, T., 1978. Az élet princípium (The principles of life), second (revised) ed. Gondolat, Budapest
872 (In Hungarian).

873 Gánti, T., 1987. The principle of life, first ed. OMIKK, Budapest.

874 Gilbert, W., 1986. Origin of life: the RNA world. *Nature*. 319, 618.

875 Hall, N., 2004. The quest for the chemical roots of life. *Chem. Commun.* 11, 1247–1252.

876 Hanczyc, M.M., Mansy, S.S., Szostak, J.W., 2007. Mineral surface brane assembly. *Origins Life
877 Evol. B.* 37, 67–82.

878 Hazen, R.M., Filley, T.R., Goodfriend, G.A., 2001. Selective adsorption of L- and D-amino acids on
879 calcite: implications for biochemical homochirality. *P. Natl. Acad. Sci. USA.* 98, 5487–5490.

880 Horowitz, N.H. 1945. On the evolution of biochemical syntheses. *P. Natl. Acad. Sci. USA.* 31,
881 153–157.

882 Ivica, N.A., Obermayer, B., Campbell, G.W., Rajamani, S., Gerland, U., Chen, I.A., 2013. The
883 paradox of dual roles in the RNA world: resolving the conflict between stable folding and
884 templating ability. *J. Mol. Evol.* 77, 55–63.

885 Johnson, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., Bartel, D.P., 2001. RNA-catalyzed
886 RNA polymerization: accurate and general RNA templated primer extension. *Science.* 292,
887 1319–1325.

888 Johnson, A.P., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Lazcano, A., Bada, J.L., 2008. The Miller
889 volcanic spark discharge experiment. *Science.* 322, 404.

890 Joshi, P.C., Aldersley, M.F., Ferris, J.P., 2011. Homochiral selectivity in RNA synthesis:
891 montmorillonite-catalyzed quaternary reactions of D, L-purine with D, L- pyrimidine
892 nucleotides. *Origins Life Evol. B.* 41, 213–236.

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

894 Károlyi, Gy., Scheuring, I., Czárán, T., 2002. Metabolic network dynamics in open chaotic flow.
895 *Chaos.* 12, 460–469.

896 Khersonsky, O., Tawfik, D.S. 2010. Enzyme promiscuity: a mechanistic and evolutionary
897 perspective. *Annu. Rev. Biochem.* 79, 471–505.

898 Khvorova, A., Kwak, Y.G., Tamkun, M., Majerfeld, I., Yarus, M., 1999. RNAs that bind and change
899 the permeability of phospholipid membranes. *P. Natl. Acad. Sci. USA.* 96, 10649–10654.

900 von Kiedrowski, G., Sijbren, O., Piet, H., 2010. Welcome Home, Systems Chemists! *J. Sys. Chem.*
901 1, 1.

902 Kim, P.-J., Jeong, H., 2005. Spatio-temporal dynamics in the origin of genetic information. *Physica
903 D.* 203, 88–99.

904 Könnyű, B., Czárán, T., Szathmáry, E., 2008. Prebiotic replicase evolution in a surface-bound
905 metabolic system: parasites as a source of adaptive evolution. *BMC Evol. Biol.* 8, 267.

906 Könnyű, B., Czárán, T., 2011. The evolution of enzyme specificity in the metabolic replicator model
907 of prebiotic evolution. *PloS One.* 6, e20931.

908 Könnyű, B., Czárán, T., 2013. Spatial aspects of prebiotic replicator coexistence and community
909 stability in a surface-bound RNA world model. *BMC Evol. Biol.* 13, 204.

910 Könnyű, B., Czárán, T., 2014. Phenotype/genotype sequence complementarity and prebiotic
911 replicator coexistence in the metabolically coupled replicator system. *BMC evol. Biol.* 14,
912 234.

913 Kun, Á., Santos, M., Szathmáry, E., 2005. Real ribozymes suggest a relaxed error threshold. *Nature
914 Genetics.* 37, 1008–1011.

915 Landweber, L.F., Simon, P.J., Wagner, T.A., 1998. Ribozyme engineering and early evolution.
916 *BioScience.* 48, 94–103.

917 LaRowe, D.E., Regnier, P. 2008. Thermodynamic potential for the abiotic synthesis of adenine,

- 918 cytosine, guanine, thymine, uracil, ribose and deoxyribose in hydrothermal systems. *Origins*
919 *Life Evol. B.* 38, 383–397.
- 920 Lilley, D.M.J., 2003. The origins of RNA catalysis in ribozyme. *Trends. Biochem. Sci.* 28, 495–501.
- 921 Nielsen, P.E., 2009. Peptid nucleid acids as prebiotic and abiotic genetic material, In Rasmussen,
922 S., Bedau, M.A., Chen, L., Deamer, D.W., Krakauer, D.C., Packard, N.H., Stadler P.F. (Eds.),
923 *Protocells: Bridging non-living and living matter.* MTI Press, Cambridge, pp. 337–346
- 924 Niesert, U., Harnasch, D., Bresch, C., 1981. Origin of life between Scylla and Charybdis. *J. Mol.*
925 *Evol.* 17, 348–353.
- 926 Niesert, U., 1987. How many genes to start with? A computer simulation about the origin of life.
927 *Origins Life Evol. B.* 17, 155–169, .
- 928 Ma, W., Yu, C., Zhang, W., Hu, J., 2007. Nucleotide synthetase ribozymes may have emerged first
929 in the RNA world. *RNA.* 13, 2012–2019.
- 930 Manrubia, S.C., Briones, C. 2007. Modular evolution and increase of functional complexity in
931 replicating RNA molecules. *RNA.* 13, 97–107.
- 932 Martin, W., Russel, M.J., 2003. On the origins of cells: a hypothesis for the evolutionary transitions
933 from abiotic geochemistry to chemoautotrophic prokaryotes and from prokaryotes to
934 nucleated cells. *Philos. T. Roy. Soc. B.* 358, 59–85.
- 935 Maynard-Smith, J., 1979. Hypercycles and the origin of life. *Nature.* 280, 445–446.
- 936 Maynard-Smith, J., Szathmáry, E., 1995. *The major transitions in evolution.* Oxford University
937 Press, Oxford.
- 938 Meléndez-Hevia, E., Montero-Gómez, N., Montero, F., 2008. From prebiotic chemistry to cellular
939 metabolism—the chemical evolution of metabolism before Darwinian natural selection. *J.*
940 *Theor. Biol.* 252, 505–519.
- 941 Meszéna, G., Gyllenberg, M., Pásztor, E., Metz, J.A.J., 2006. Competitive exclusion and limiting
942 similarity: an unfied theory. *Theor. Popul. Biol.* 69, 68–87.
- 943 Miller, S.L., 1953. A production of amino acids under possible Earth conditions. *Science.* 117, 528–
944 529.
- 945 Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H.J., Miller, S.L., 2002. Prebiotic synthesis
946 from CO at-mospheres: implications for the origins of life. *P. Natl. Acad. Sci. USA.* 99,
947 14628–14631.
- 948 Monnard, P.-A., Kanavarioti, A., and Deamer, D.W., 2003. Eutectic phase polymerization of
949 activated ribonucleotide mixtures yields quasi-equi-molar incorporation of purine and
950 pyrimidine nucleobases. *J. Am. Chem. Soc.* 125, 13734–13740.
- 951 Morowitz, H.J., Kostelnik, J.D., Yang, J., Cody, G.D., 2000. From the cover: the origin of
952 intermediary metabolism. *P. Natl. Acad. Sci. USA.* 97, 7704–7708.
- 953 Orgel, L.E., 2000. Self-organizing biochemical cycles. *P. Natl. Acad. Sci. USA.* 97, 12503–12507.
- 954 Orgel, L.E., 2004. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol.*
955 *39,* 99–123.
- 956 Parsons, I., Lee, M.R., Smith, J.V., 1998. Biochemical evolution II: origin of life in tubular
957 microstructures on weathered feldspar surfaces. *P. Natl. Acad. Sci. USA.* 95, 15173–15176.
- 958 Patzke, V., von Kiedrowski, G., 2007. Self replicating systems. *Arkivoc.* 8, 293–310.
- 959 Powner, M.W., Gerland, B., Sutherland, J.D., 2009. Synthesis of activated pyrimidine
960 ribonucleotides in prebiotically plausible conditions. *Nature.* 459, 239–242.
- 961 Pross, A., 2004. Causation and the origin of life. Metabolism or replication first? *Origins Life Evol.*
962 *B.* 34, 307–321.
- 963 Rasmussen, S., Bedau, M.A., McCaskill, J.S., Packard, N.H., 2009. A roadmap to protocells. In
964 Rasmussen, S., Bedau, M.A., Chen, L., Deamer, D.W., Krakauer, D.C., Packard, N.H., Stadler
965 P.F. (Eds.), *Protocells: Bridging non-living and living matter.* MTI Press, Cambridge, pp. 71–
966 100.
- 967 Robertson, M.P., Joyce, G.F., 2010. The origins of the RNA world. *Cold Spring Harb. Perspect.*

968 Biol. 2, 1–22.

969 Rohatgi, R., Bartel, D.P., Szostak, J.W., 1996. Nonenzymatic, template-directed ligation of
970 oligoribonucleotides is highly regioselective for formation of 3' - 5' phosphodiester bond. J.
971 Am. Chem. Soc. 118, 3340–3344.

972 Rushdi, A.I., Simoneit, B.R., 2001. Lipid formation by aqueous Fischer-Tropsch-type synthesis
973 over a temperature range of 100 to 400 degrees C. *Origins Life Evol. B.* 31, 103–118.

974 Scheuring, I., Czárán, T., Szabó, P., Károlyi, G., Toroczkai, Z., 2003. Spatial models of prebiotic
975 evolution: soup before pizza? *Origins Life Evol. B.* 33, 319–355.

976 Schultes, E.A., Bartel, D.P., 2000 One sequence, two ribozymes: implication for the emergence of
977 new ribozyme fold. *Science.* 289, 448–452.

978 Segré, D., Ben-Eli, D., Deamer, D.W., Lancet, D., 2001. The lipid world. *Origins Life Evol. B.* 31,
979 119–145.

980 Szathmáry, E., 1992. Viral sex, levels of selection, and the origin of life. *J. Theor. Biol.* 159, 99-109.

981 Szathmáry, E., 2006. The origin of replicators and reproducers. *Philos. T. Roy. Soc. B.* 361, 1761–
982 1776.

983 Szathmáry, E., 2007. Coevolution of metabolic networks and membranes: the scenario of
984 progressive sequestration. *Philos. Trans. R. Soc. London B.* 362, 1781–1787.

985 Szathmáry, E., Demeter, L. 1987. Group selection of early replicators and the origin of life. *J.*
986 *Theor. Biol.* 128, 463–486.

987 Szathmáry, E., Gladkih, I., 1989. Sub-exponential growth and coexistence of non-enzymatically
988 replicating templates. *J. Theor. Biol.* 138, 55–58.

989 Szathmáry, E., Santos, M., Fernando, C., 2005. Evolutionary potential and requirements for
990 minimal protocells. *Top. Curr. Chem.* 259, 167–211.

991 Szilágyi, A., Zachar, I., Szathmáry, E., 2013. Gause's principle and the effect of resource
992 partitioning on the dynamical coexistence of replicating templates. *PLoS Comput. Biol.* 9,
993 e1003193.

994 Takeuchi, N., Hogeweg, P., 2007. Error-threshold exists in fitness landscapes with lethal mutants.
995 *BMC Evol. Biol.* 7, 15–27.

996 Toffoli, T., Margolus, N., 1987. Cellular automata machines: a new environment for modelling.
997 MIT Press, Cambridge.

998 Wächtershäuser, G., 1990. Evolution of the first metabolic cycles. *P. Natl. Acad. Sci. USA.* 87, 200–
999 204.

1000 Wilson, D.S., 1975. A Theory of group selection. *P. Natl. Acad. Sci. USA.*, 72, 143–146.

1001 Wochner, A., Attwater, J., Coulson, A., Holliger, P., 2011. Ribozyme-catalyzed transcription of an
1002 active ribozyme. *Science.* 332, 209–212.

1003

1004

1005 **Figure legends**

1006 **Graphical abstract: Metabolically coupled replicator system.** The metabolic replicator system
1007 with four autocatalytic metabolic replicators (I_i , $i = 1, \dots, 4$ within the circular arrows). M is the
1008 metabolic reaction network supported by the metabolic replicators as enzymes (solid lines) and
1009 producing monomers for their replication (dashed lines).

1010 **Figure 1. Parasites of the hypercycle.** P_1 : selfish parasite; P_2 : short-cut parasite. Based on
1011 (Scheuring et al. 2003).

1012 **Figure 2. A schematic representation of the retroevolution of metabolism.** The evolution of
1013 metabolically active replicators k (pV_k , $k = n, m, q, \dots$) catalysing an increasingly complex network
1014 (here: chain) of metabolic reactions (solid arrows and coloured folded structures) to produce
1015 monomer V . Reactions are included in the metabolic network sequentially as monomers (V) and
1016 then monomer precursors (A and B) are depleted from the environment (right diagram) by the
1017 increasing replicator population. n , m and q are stoichiometric constants.
1018

1019 **Figure 3. The MCRS concept and neighbourhood definitions of the spatially explicit MCRS**
1020 **model. Panel A:** The metabolic replicator system with four autocatalytic metabolic replicators (I_i , i
1021 $= 1, \dots, 4$ within the circular arrows). M is the metabolic reaction network supported by the
1022 metabolic replicators as enzymes (solid lines) producing monomers for their own replication
1023 (dashed lines). **Panel B:** The relation of metabolic (I_i , where $i = 1, \dots, 3$) and parasitic (P) replicators
1024 to metabolism. Parasites consume monomers produced by the metabolic network but do not
1025 contribute to metabolism by catalytic activity. **Panel C:** Neighbourhood definitions on the non-
1026 structured surface of the spatially explicit model. X is an empty site of the cellular automaton lattice,
1027 I_i ($i = 1, \dots, 4$) are the metabolic replicators. Dark grey sites are the replication neighbourhood of the
1028 empty site (*von Neumann neighbourhood* in this case) and light grey sites constitute the metabolic
1029 neighbourhood of replicator I_1 (*3x3 Moore neighbourhood* in this case). (From (Könnnyü and
1030 Czárán, 2013))

1031 **Figure 4. Structure of a real mineral surface and the model. Panel A** SEM image of a resin cast
1032 of an etch-pit network near the surface of a weathered Shap alkali feldspar (scale bar 20 μm). The
1033 cast was made by impregnating the feldspar with Araldite resin under vacuum, curing, and
1034 dissolving away the feldspar in concentrated HF. The surface of the feldspar is off the bottom of the
1035 micrograph, and the image is of a pile of two-dimensional networks that have fallen over to lie on
1036 top of each other. Because the resin is flexible, parts of the networks are curved. The original etch-
1037 pits were developed on edge dislocations very nearly parallel to b (*horizontal*) and c (*vertical*) in the
1038 perthite contact plane close to 601 of the monoclinic feldspar (SEM picture and caption from Fig.2.
1039 of (Parsons et al. 1998)). **Panel B** $I_{1..4}$ are the metabolic replicators; M is metabolism. Solid arrows
1040 represent the flux of resources (raw materials) outside the pores chemically transformed inside the
1041 pores in nucleotides by the catalytic activity of replicators (ribozyme). White arrows mean the
1042 catalytic effect of metabolic replicators helping metabolism. P represents a parasitic replicator that
1043 uses the monomers supplied by metabolism, but it does not help producing them. (From
1044 (BranCIamore et al. 2009))

1045 **Figure 5.: A typical run of MRM.** Parameters: system size (number of metabolic replicators) $n =$
1046 4 ; replicator mobility $D = 4$, size of metabolic neighbourhood: 5×5 (Moore); size of replication
1047 neighbourhood: von Neumann; replication parameters of replicators: $k_1 = 3$ (blue), $k_2 = 5$ (red), $k_3 =$
1048 7 (green), $k_4 = 9$ (orange),
1049

1050 **Figure 6. Coexistence of metabolic replicators as a function of replicator diffusion (D),**
1051 **metabolic (h) and replication (r) neighbourhood size.** The panels of the figure differ in the
1052 number of diffusion steps per generation: **Panel A:** $D = 0$, **Panel B:** $D = 1$, **Panel C:** $D = 4$ and
1053 **Panel D:** $D = 100$. x - and y -axes are the sizes of metabolic neighbourhoods (h) and replication
1054 neighbourhoods (r), respectively (N : von Neumann neighbourhood; 3: 3×3 , 5: 5×5 , 7: 7×7 , 25: 25×25
1055 and 37: 37×37 Moore neighbourhoods). The grayscale shades correspond to average replicator
1056 densities (% occupied) on the whole grid at the end of the simulations (i.e., at $t = 1.000$). The
1057 numbers within the cells of the tables indicate coexistent/extinct replicate simulations out of five
1058 repetitions with the same parameter set and different pseudo-random number sequences. Based on
1059 (Könnyű and Czárán, 2013).

1060 **Figure 7. The effect of migration and pore size on total replicator density in the pore-model.**
1061 Fixed parameters: resource input ($r = 2$) and system size ($n = 5$). From (Branciamore et al. 2009).
1062

1063 **Figure 8. The maximum number of coexisting metabolic replicators as the function of**
1064 **replicator diffusion (D), metabolic (h) and replication (r) neighbourhood size.** The panels of the
1065 figure differ in the number of diffusion steps per generation: **Panel A:** $D = 0$, **Panel B:** $D = 4$ and
1066 **Panel C:** $D = 100$ x - and y -axes are the sizes of metabolic neighbourhoods (h) and replication
1067 neighbourhoods (r) respectively (N : von Neumann neighbourhood; 3: 3×3 , 5: 5×5 , 7: 7×7 , 25: 25×25
1068 and 37: 37×37 Moore neighbourhoods). The number within a cell of the panel shows the maximum
1069 attainable system size (n_{max}) for the corresponding parameter set. Other parameters: $p_d = 0.2$, $C_e =$
1070 2.0 , $k_i = 3.0 + 2.0i$ ($i = 0, \dots, n_{max}$). From (Könnyű and Czárán, 2013). **Panel D:** Relationship
1071 between system size and minimal pore size necessary for coexistence in the pore model. The
1072 migration parameter was $d = 0.8$. From (Branciamore et al. 2009).

1073 **Figure 9.: MRM and parasite(s).** Panel A: Parameters: system size (number of replicators): 3 +
1074 parasite (black); D : 4, size of metabolic neighbourhood: 3×3 (Moore); size of replication
1075 neighbourhood: von Neumann; replication parameters of metabolic replicators: $k_1 = 3$ (blue), $k_2 = 5$
1076 (red), $k_3 = 7$ (green), and parasite $k_p = 9$ (black). **Panel B:** Parameters: system size (number of
1077 replicators): 4 metabolic and 4 parasite replicators; D : 4, size of metabolic neighbourhood: 5×5
1078 (Moore); size of replication neighbourhood: von Neumann; replication parameters of replicators:
1079 $k_{1m} = 3.0$ (blue), $k_{1p} = 4.0$ (light grey), $k_{2m} = 5.0$ (red), $k_{2p} = 6.0$ (middle grey), $k_{3m} = 7.0$ (green), k_{3p}
1080 $= 8.0$ (dark grey), $k_{4m} = 9.0$ (orange) and $k_{4p} = 10.0$ (black); subscripts m and p denote metabolic
1081 and parasite replicator types, respectively.
1082

1083 **Figure 10. Typical runs of specially modified MRM. Panel A:** The pheno/geno version of MRM.
1084 Parameters: system size (number of replicators): 4 phenotype and 4 genotype replicators; D : 4, size
1085 of metabolic neighbourhood: 3×3 (Moore); size of replication neighbourhood: 37×37 (Moore);
1086 replication parameters of replicators: $k_{1p} = 3.0$ (blue), $k_{1g} = 4.0$ (blue), $k_{2p} = 5.0$ (red), $k_{2g} = 6.0$ (red)
1087 , $k_{3p} = 7.0$ (green), $k_{3g} = 8.0$ (green), $k_{4p} = 9.0$ (orange) and $k_{4g} = 10.0$ (orange); subscripts p and g
1088 denote phenotype-forms (solid lines) and genotype-forms (dashed lines) of replicator types,
1089 respectively. **Panel B:** MRM with a facultative metabolic cooperater. Parameters: system size
1090 (number of replicators): 3 essential metabolic and facultative metabolic replicators; D : 4, size of
1091 metabolic neighbourhood: 3×3 (Moore); size of replication neighbourhood: von Neumann;
1092 replication parameters of replicators: $k_1 = 3$ (blue), $k_2 = 5$ (red), $k_3 = 7$ (green), and the facultative
1093 cooperater: $k_p = 9$ (orange).

1094 **Figure 11. The $E_1 - E_2 - k$ trade-off surface.** The trade-off function constrains the phenotypes of
1095 emerging mutant replicators to below the surface given by

1096
$$k(E_1, E_2) = \left[E_{\max}^g - \left[E_1^b + E_2^b \right]^{\frac{1}{b^g}} \right]^{\frac{1}{g}} \cdot \frac{k_{\max} - k_{\min}}{E_{\max}} + k_{\min} .$$

1097 Fixed parameters: $k_{\min} = 2.0$, $k_{\max} = 4.0$, $E_{\max} = 10.0$.

1098 **Panel A:** convex function representing strong trade-off both between the two enzyme activities
 1099 E_1/E_2 and between enzyme activities and replication rate, E/k . **Panel B:** a function with convex
 1100 (strong) E_1/E_2 trade-off and concave (weak) E/k trade-off . **Panel C:** concave (weak) E_1/E_2 and
 1101 convex (strong) E/k trade-off. **Panel D:** both the E_1/E_2 and the E/k trade-offs are concave (weak).
 1102 From (Könnyű and Czárán, 2011).

1103
 1104 **Figure 12. Frequencies of replicator types. Panel A:** The steady-state frequencies of specialist and
 1105 generalist replicators as a function of b (the strength of the trade-off between enzymatic activities),
 1106 at $D = 0$; **Panel B:** the same, at $D = 5$. Other parameters: $p_m = 0.01$ (mutation rate), $g = 1.0$ (the
 1107 strength of the trade-off enzymatic activities and replication rate), $E_{\max} = 10$ (maximal enzymatic
 1108 activities) and $k_{\max} = 2.5$ (maximal replication rate) at the 150.000th generation. Note that the
 1109 frequency of parasitic replicators is less than 1% everywhere in this parameter setting, so we have
 1110 not plotted it here. Based on (Könnyű and Czárán, 2011).

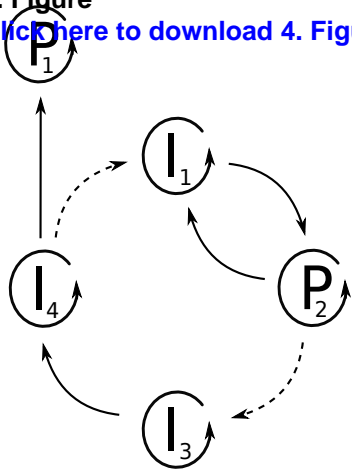
1111
 1112 **Figure 13. The benefit of evolving a sequence-specific replicase replicator. Panel A:** Metabolic
 1113 system with a parasite evolved into a replicase (R). Dashed-dotted lines represent sequence-
 1114 specific replicase activity. Other arrows and letters are the same as in Figure 1. **Panel B:** The effect
 1115 of an evolving replicase replicator on the dynamics of the metabolic replicator community.
 1116 Replication parameters: $k_1 = 2$ (blue), $k_2 = 4$ (red), $k_3 = 6$ (green), and parasite/replicase $k_p = 8$
 1117 (orange). Black line: replicase activity (scale on the second y axis). Based on Könnyű et al. 2008.

1118
 1119 **Figure 14. The Metabolically Coupled Repricator System enclosed in a self-produced**
 1120 **membrane vesicle (“protocell”).** The metabolic replicator set ($I_{1..4}$) with a replicase (R), a lipid
 1121 synthetase (L) and a membrane channel forming replicator (T) added. M produces membranogenic
 1122 molecules (*black triangles*) which are transformed to membrane molecules (black rectangles) by the
 1123 lipid synthetase (L) replicator. New lipid molecules are inserted into the membrane spontaneously.
 1124 Transporter replicators (*grey rectangle with a T*) insert themselves into the membrane to form trans-
 1125 membrane channels which selectively let small metabolic precursor molecules (*black stars*) enter
 1126 the vesicle.

1127

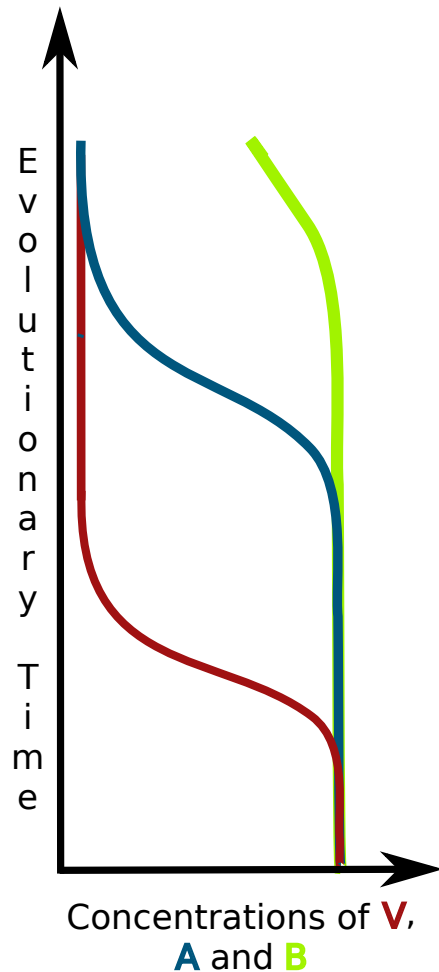
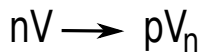
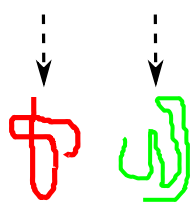
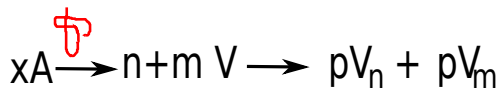
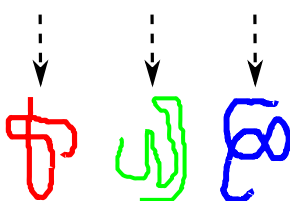
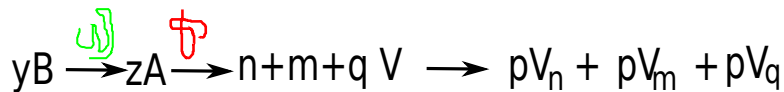
4. Figure

[Click here to download 4. Figure](#)



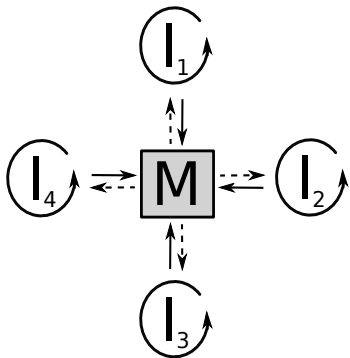
4. Figure

[Click here to download 4. Figure: Fig02.pdf](#)

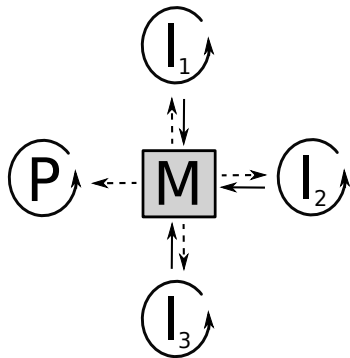


4. Figure

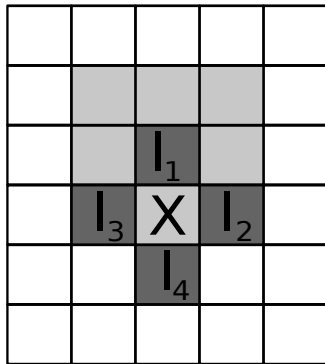
[Click here to download 4. Figure: Fig03.pdf](#)



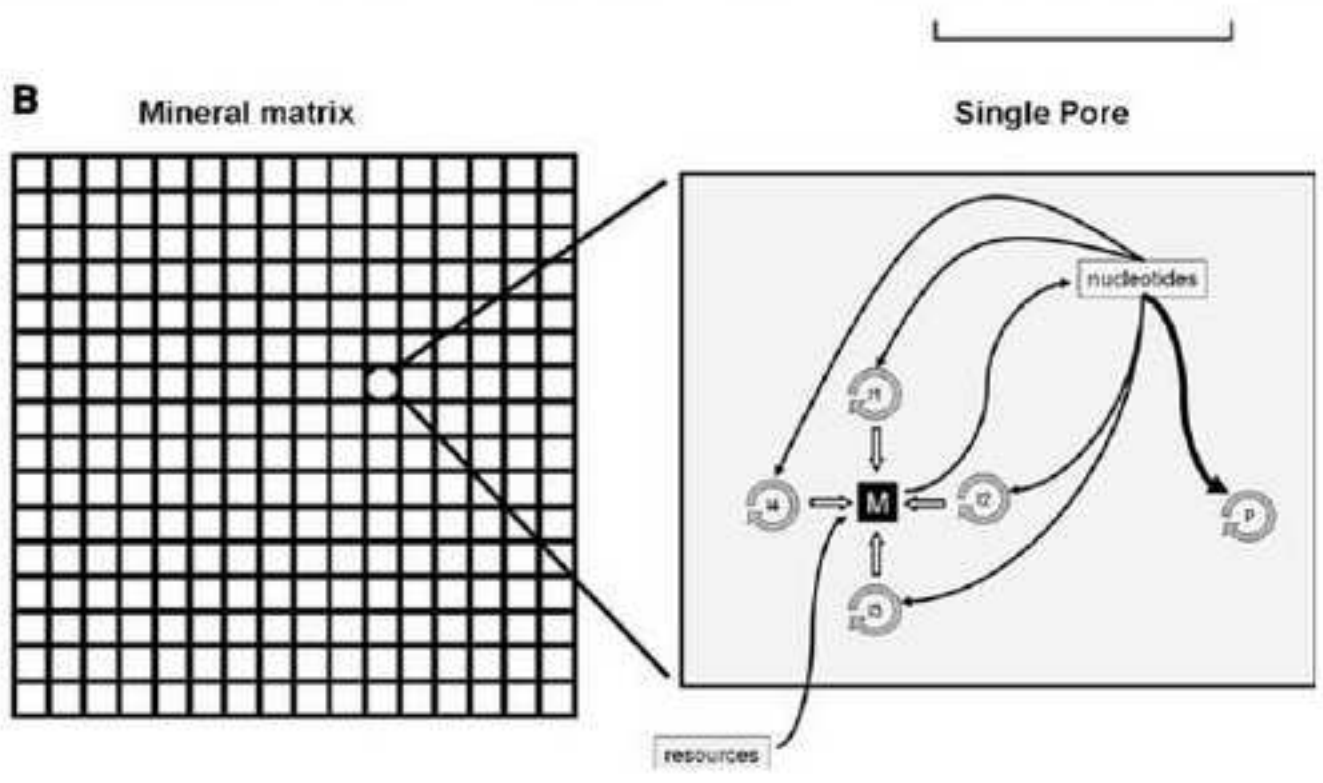
B



C

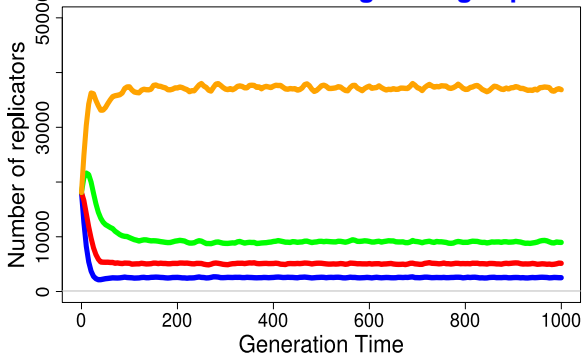


4. Figure
[Click here to download high resolution image](#)



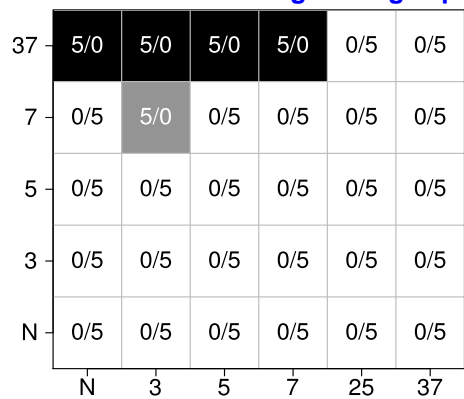
4. Figure

[Click here to download 4. Figure: Fig05.pdf](#)

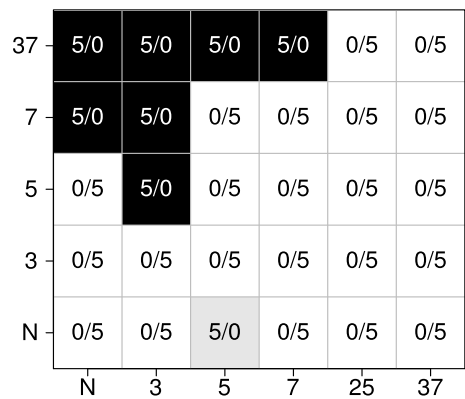


4. Figure
[Click here to download 4. Figure: Fig06.pdf](#)

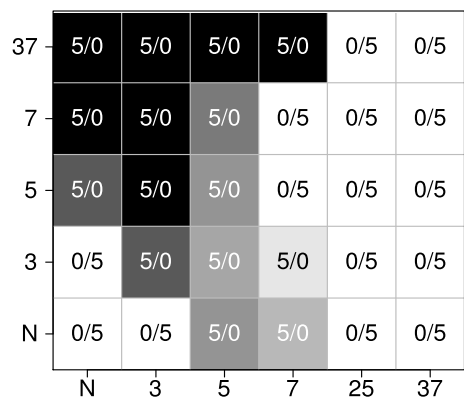
Size of replication neighbourhood (r)



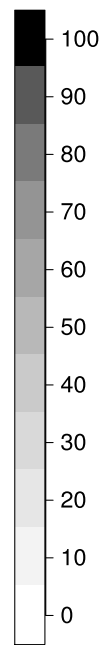
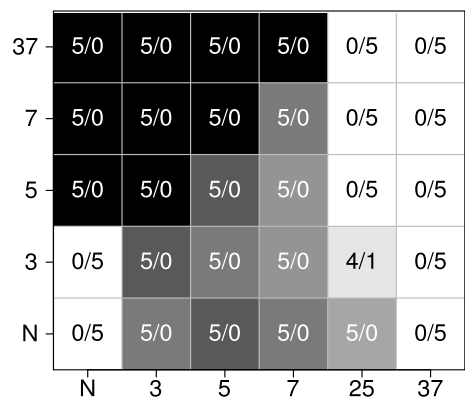
B



C



D



Size of metabolic neighbourhood (h)

4. Figure
[Click here to download high resolution image](#)

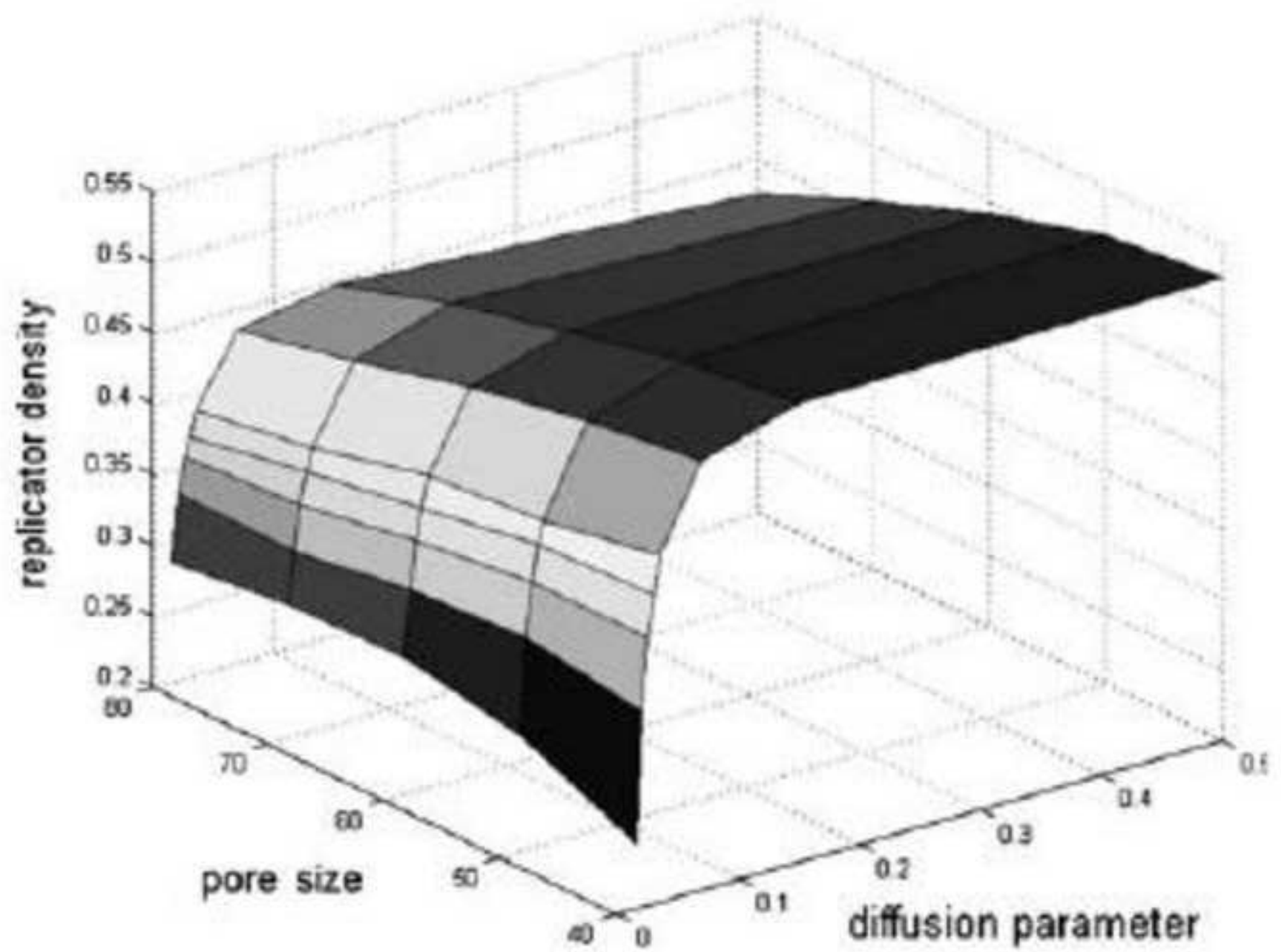
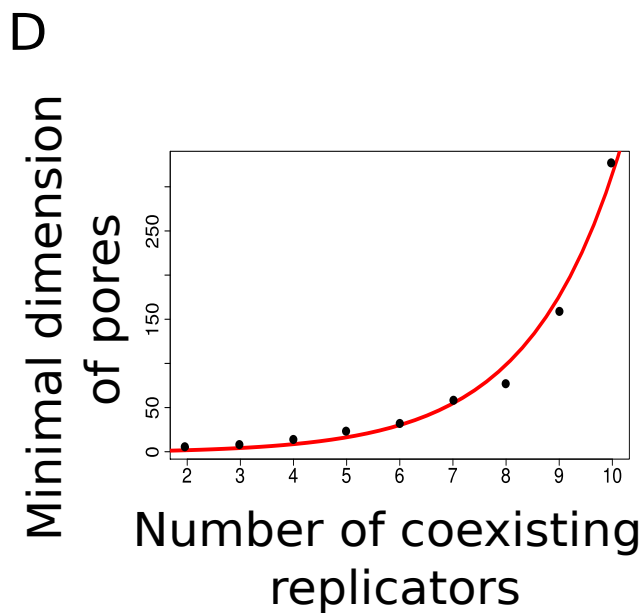
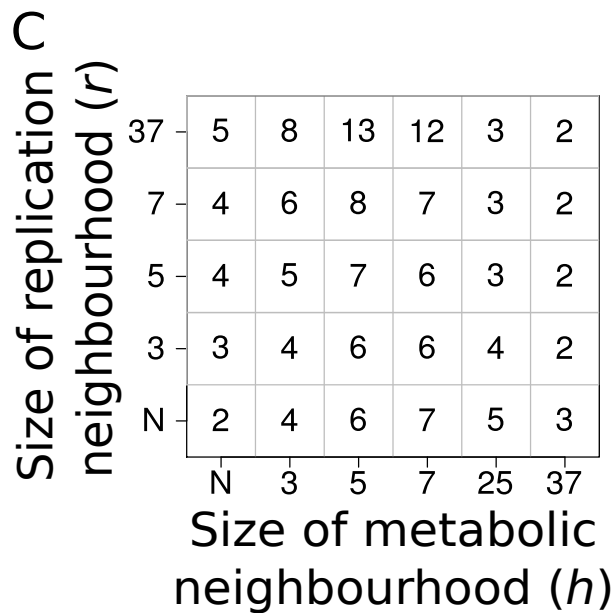
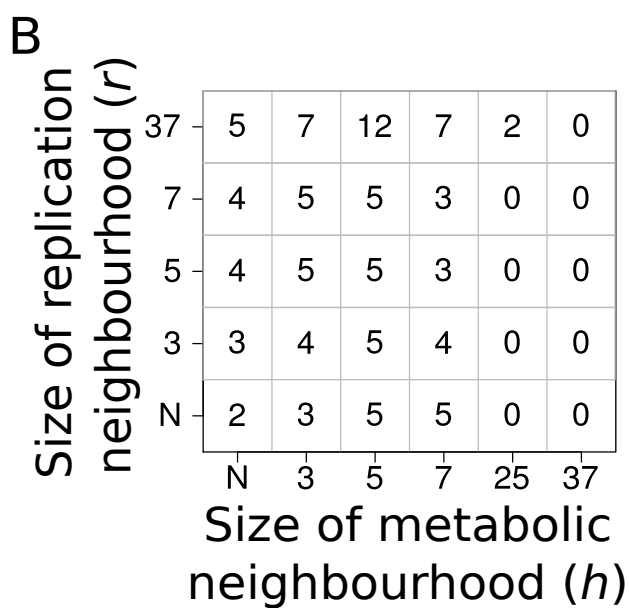
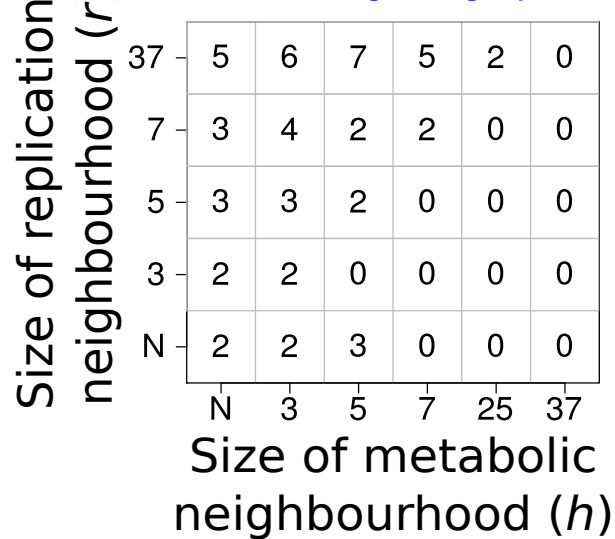
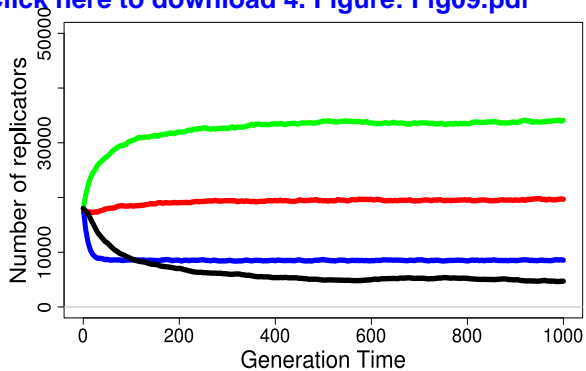


Figure 4
[Click here to download 4. Figure: Fig08.pdf](#)

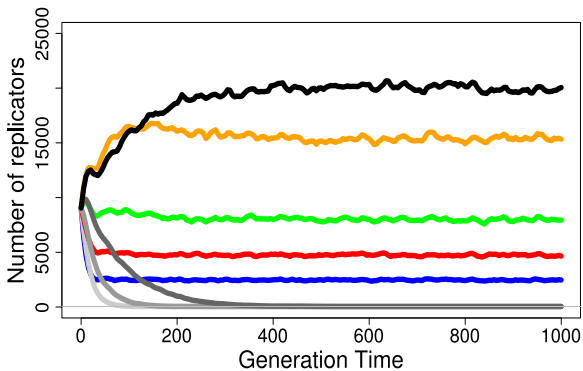


4. Figure

[Click here to download 4. Figure: Fig09.pdf](#)

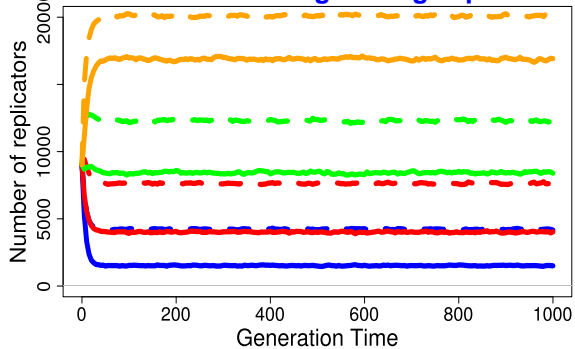


B

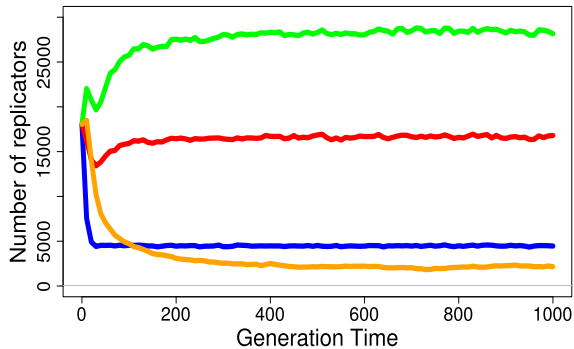


4A Figure

[Click here to download 4. Figure: Fig10.pdf](#)



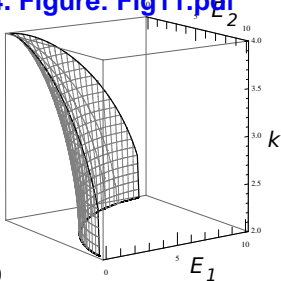
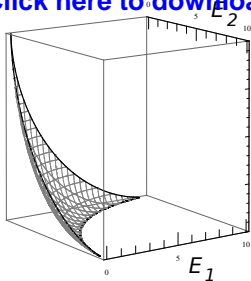
B



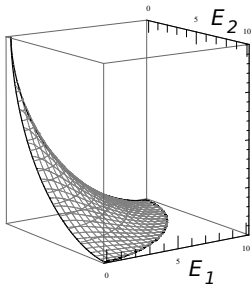
A. Figure

[Click here to download 4. Figure: Fig11.pdf](#)

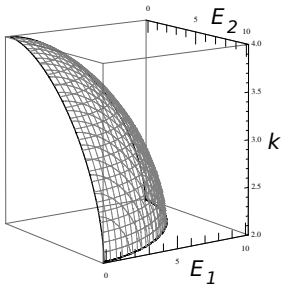
B



C

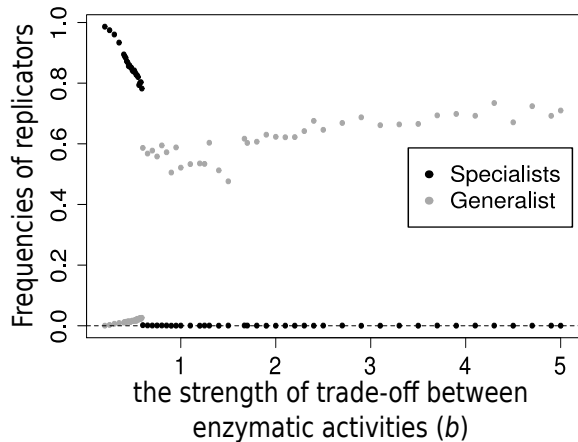


D

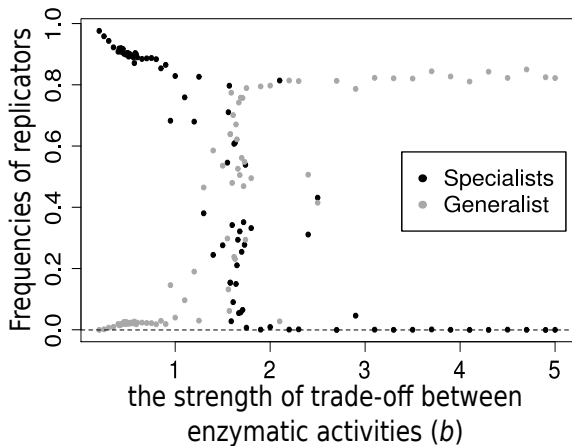


4 Figure

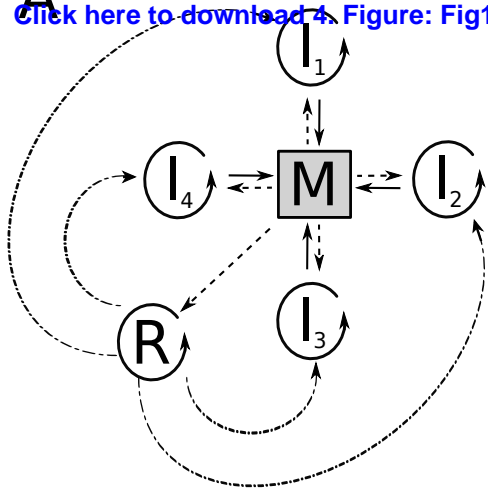
[Click here to download 4. Figure: Fig12.pdf](#)



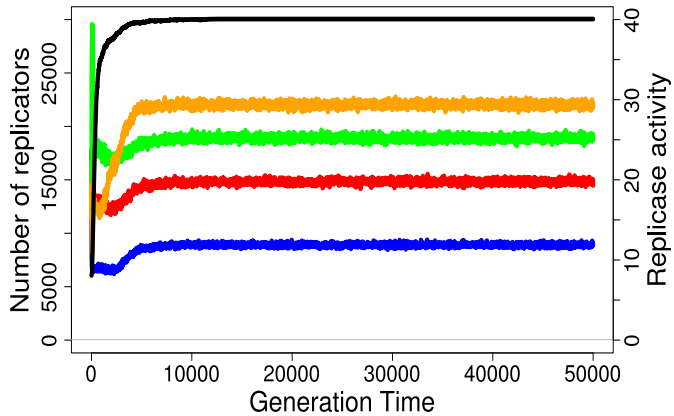
B



A Figure
[Click here to download 4. Figure: Fig13.pdf](#)



B



4. Figure

[Click here to download 4. Figure: Fig14.pdf](#)

