

## Research



**Cite this article:** Sinai S, Olejarz J, Neagu IA, Nowak MA. 2018 Primordial sex facilitates the emergence of evolution. *J. R. Soc. Interface* **15**: 20180003.

<http://dx.doi.org/10.1098/rsif.2018.0003>

Received: 2 January 2018

Accepted: 5 February 2018

### Subject Category:

Life Sciences – Mathematics interface

### Subject Areas:

astrobiology, biomathematics, evolution

### Keywords:

multiplicity reactivation, origin of life, complexity, evolution, protocells, sex

### Authors for correspondence:

Sam Sinai

e-mail: [samsinai@g.harvard.edu](mailto:samsinai@g.harvard.edu)

Martin A. Nowak

e-mail: [martin\\_nowak@harvard.edu](mailto:martin_nowak@harvard.edu)

†These authors contributed equally to this study.

# Primordial sex facilitates the emergence of evolution

Sam Sinai<sup>1,2,†</sup>, Jason Olejarz<sup>1,†</sup>, Iulia A. Neagu<sup>1,3</sup> and Martin A. Nowak<sup>1,2,4</sup>

<sup>1</sup>Program for Evolutionary Dynamics, <sup>2</sup>Department of Organismic and Evolutionary Biology, <sup>3</sup>Department of Physics, and <sup>4</sup>Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

ORCID iD: [SS, 0000-0003-0301-9201](https://orcid.org/0000-0003-0301-9201); [JO, 0000-0003-3571-7709](https://orcid.org/0000-0003-3571-7709); [IAN, 0000-0002-0757-0853](https://orcid.org/0000-0002-0757-0853); [MAN, 0000-0001-5489-0908](https://orcid.org/0000-0001-5489-0908)

Compartments are ubiquitous throughout biology, and they have very likely played a crucial role at the origin of life. Here we assume that a protocell, which is a compartment enclosing functional components, requires  $N$  such components in order to be evolvable. We calculate the timescale in which a minimal evolvable protocell is produced. We show that when protocells fuse and share information, the timescales polynomially in  $N$ . By contrast, in the absence of fusion, the worst-case scenario is exponential in  $N$ . We discuss the implications of this result for the origin of life and other biological processes.

## 1. Introduction

A defining characteristic of living organisms is their ability to replicate and evolve [1]. A major objective of research on the origin of life is, therefore, to find plausible chemical systems that are capable of self-replication. The ‘RNA world hypothesis’ is a leading framework encompassing theories about the role of RNA in the origin of life. It postulates that RNA or a similar biopolymer, being both an information-carrying molecule, as well as an enzymatic one, must have played a central role in initiating self-replication [2–4]. But formidable difficulties remain for developing this narrative into a complete and rigorous theory of the origin of life [5–8]. Both theoretical and experimental investigations show that well-mixed populations of RNA or similar biopolymers often suffer from calamitous pitfalls, including the error catastrophe for replicases [9] and parasitism for cooperative enzymes [10–13]. Moreover, the complexity of long RNA sequences that could serve as efficient catalysts creates a challenge for explaining their spontaneous prebiotic synthesis [14]. Indeed, despite decades of efforts in prebiotic chemistry (and some exciting progress, e.g. [15,16]), building efficient, stable and prebiotically plausible replicases (sometimes called the holy grail of the RNA world) has remained a challenge [17,18].

In modern cells, lipid membranes compartmentalize information-carrying and enzymatic molecules akin to those sought after by RNA world researchers. Hence, at some point in the development of life, either before, during or after the emergence of self-replicating genetic elements, such compartmentalization must have occurred. There is evidence in support of the prebiotic availability of lipid membranes. It has been shown that amphiphilic molecules, like simple fatty acids that are building blocks for the lipid membrane, can be produced in a prebiotically plausible manner [19]. Alternatively, lipids could have been imported to earth by chondrite meteorites [20–22]. Hence, such molecules were likely abundantly present on the prebiotic Earth [23–27]. These molecules are able to spontaneously assemble into lipid vesicles in aqueous conditions [27,28], forming compartments, which in this context are known as *protocells*.

Protocells alleviate some of the pitfalls that can impede the transition from prelife to life. The contents of protocells are held near each other and share the same fate. This results in increased interactions within the protocell and decreased interactions with the outside environment. It also means that the protocell can

house a segmented genome, i.e. the information within the protocell need not be stored in one contiguous polymer (e.g. [23,29]). It can also dampen the effects of side reactions for any auto-catalytic cycles that may be required to start and maintain a metabolism [30]. Protocells can also divide into new protocells that inherit parts of their contents [31,32]. These properties of protocells enable them to help in selection for cooperative polymers, in particular replicases [10,11,33–38]. In addition to enclosing information and dividing, protocells are able to merge, thereby sharing their contents [30,39–41]. In biology, sharing information content between two individuals is considered a defining property of sex.

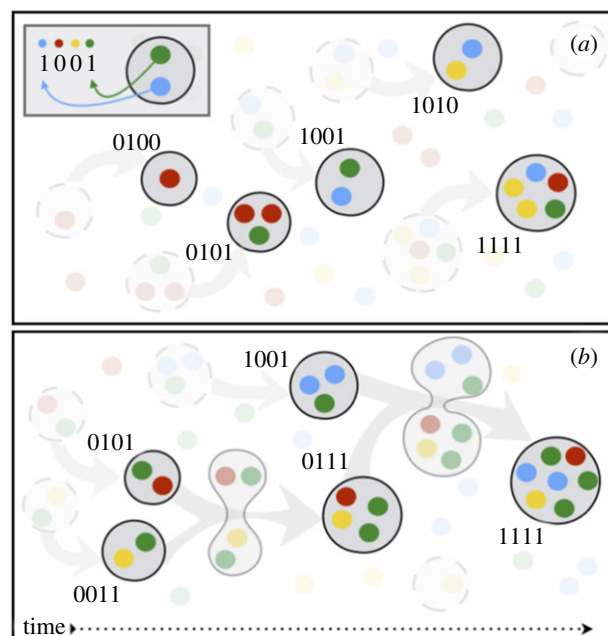
The implications of this information-sharing ability among protocells, which is a form of ‘primordial sex’, have not received much attention. For reasons outlined in the rest of this study, we suggest that the ability for these compartments to merge categorically changes the time required to produce an evolvable protocell. Hence, we propose that early presence of membranes, possibly even before the advent of replication, could have vastly improved the chances of producing complicated cells by luck. In such cases, it would not be unreasonable to assume that the starting set of molecules from which an evolvable cell emerges could be large. This assumption is rarely made in origin of life models, because it is considered a probabilistic miracle.

## 2. Model

To test this hypothesis, we investigate a simple first-passage process [42,43]. We assume that in order to be evolvable, a protocell needs to contain a certain number,  $N$ , of component types (i.e. distinct molecules of various complexity) [29,44–47]. In early life, these could be molecules as simple as ions, activated monomers, molecules that stabilize the membrane or more complicated polymers, like oligopeptides, and even elementary ribozymes and simple unlinked genes [16,23,29,30,34,35,48–53]. More precisely, the target set should result in an auto-catalytic network that results in a evolvable cell with non-negligible probability. Such a scheme has been proposed since Oparin, and has been defended more recently [53]. We term the smallest set of necessary and sufficient components from which an evolvable protocell can be made a *minimal evolvable protocell*.

We can accordingly represent the functional (or genetic) content of each protocell as a binary string of length  $N$ . For simplicity, we ignore the redundancy (or dose) of each component in the protocell, and are only concerned with each component’s presence. If a protocell contains a particular component  $i$ , then the string will have a value of 1 at the  $i$ th position and 0 otherwise. Whenever a protocell randomly assembles, we assume that it contains each of the  $N$  component types independently (components do not compete for positions) with probability  $p$ . That is, protocell assembly uniformly samples each type (with sufficient abundance) from the environment with probability  $p$ . Whenever two protocells merge, the value of the resulting string at every position  $i$  is simply determined by a bitwise OR operation on the  $i$ th bits of the two parent protocells (i.e. if either of the original cells contain a component, the resulting cell will also contain it). This is shown schematically in figure 1.

The dynamical process is as follows. On the first step, the *accumulator*—the object of our attention—consists of a



**Figure 1.** Merging occurs between randomly assembled protocells. (a) Each colour (and a ‘1’ bit at each corresponding position on a protocell’s representative binary string) indicates the presence of one of the four components needed for the protocell to be evolvable (here,  $N = 4$ ). Randomly assembled lipid membranes form around the components. (b) Whenever two protocells merge, they share their contents. Sharing of contents is computed as a bitwise OR operation between each of the two parent strings of length  $N$ . (Online version in colour.)

randomly assembled protocell. If less than  $N$  components are enclosed, then one of two things can happen: with probability  $\delta$ , the accumulator loses its contents, and on the second step, the accumulator consists of a new randomly assembled protocell, with the accumulation process starting over. The accumulator can lose its contents if, for example, its membrane’s integrity is lost, it is infected by a parasite, or it divides, and the parameter  $\delta$  accounts for all such possibilities. Or with probability  $1 - \delta$ , on the second step, the accumulator merges with a randomly assembled protocell from the environment, possibly gaining additional components. In this case, if the accumulator still has less than  $N$  components after merging, then one of two things can happen: with probability  $\delta$ , the accumulator loses its contents, and on the third step, the accumulator consists of a new randomly assembled protocell, with the accumulation process starting over. Or with probability  $1 - \delta$ , on the third step, the accumulator merges with another randomly assembled protocell from the environment, possibly gaining additional components. This process continues until the accumulator gains all  $N$  components necessary for evolvability. The total number of steps (or time units),  $Z$ , needed to gain all  $N$  components is equal to the total number of random assembly and merging events in the accumulation process.

The time,  $Z$ , needed to form a minimal evolvable protocell is thus a random variable that depends on the particular accumulator being tracked. If we track many such accumulators, then what is the mean first-passage time,  $E[Z]$ , for an accumulator to achieve all  $N$  components necessary for evolvability?

Begin by considering the simple case  $\delta = 1$  (no merging occurs). If the accumulator consists of a randomly assembled

protocell that has all  $N$  components, then the minimal evolvable protocell has been achieved. But if there are less than  $N$  components, then the accumulator is reset without merging. Thus, the expected number of such random assembly events required to accumulate all  $N$  components necessary for evolvability,  $E_{\delta=1}[Z]$ , grows exponentially with  $N$ , i.e.

$$E_{\delta=1}[Z] = \left(\frac{1}{p}\right)^N.$$

Under this model of independent components, for large values of  $N$ , the spontaneous generation of a minimal evolvable protocell would be a probabilistic miracle. This is the worst-case scenario. We now focus our attention on understanding how  $E[Z]$  grows with  $N$  when  $0 < \delta < 1$ .

In what follows, it is convenient to use the parameter  $q \equiv 1 - p$ . Denote by  $S(q, \delta, N)$  the probability that, starting from a randomly assembled protocell, the accumulator achieves all  $N$  components before being reset. We determine  $S(q, \delta, N)$  as follows. First, assume that there is no death of the accumulator. Then  $1 - q^z$  is the probability that, after  $z$  steps, the accumulator has achieved a particular component. Therefore,  $1 - (1 - q^z)^N$  is the probability that the accumulator has not achieved all  $N$  components after  $z$  steps. It follows that  $(1 - q^z)^N - (1 - q^{z-1})^N$  is the probability that the accumulator achieves all  $N$  components in exactly  $z$  steps. Then, considering death of the accumulator, since the probability that the accumulator survives for  $z$  steps without being reset is simply  $(1 - \delta)^{z-1}$ , we have

$$S(q, \delta, N) = \sum_{z=1}^{\infty} (1 - \delta)^{z-1} [(1 - q^z)^N - (1 - q^{z-1})^N].$$

This can be simplified as

$$S(q, \delta, N) = \frac{\delta}{1 - \delta} \sum_{z=1}^{\infty} (1 - \delta)^z (1 - q^z)^N. \quad (2.1)$$

Denote by  $T(z; q, \delta, N)$  the probability mass function for the number of steps,  $z$ , needed for the accumulator to gain all  $N$  components (i.e. reach its target) when starting from a randomly assembled protocell, given that all  $N$  components are accumulated before being reset. We have

$$T(z; q, \delta, N) = \frac{(1 - \delta)^{z-1} [(1 - q^z)^N - (1 - q^{z-1})^N]}{S(q, \delta, N)}. \quad (2.2)$$

Roughly speaking,  $T$  is the distribution of how many steps it takes to succeed in hitting the target set starting from a reset without being reset again. Denote by  $R(z; q, \delta, N)$  the probability mass function for the number of steps,  $z$ , taken before the accumulator is reset when starting from a randomly assembled protocell, given that the accumulator is reset before gaining all  $N$  components. We have

$$R(z; q, \delta, N) = \frac{\delta(1 - \delta)^{z-1} [1 - (1 - q^z)^N]}{1 - S(q, \delta, N)}. \quad (2.3)$$

Roughly speaking,  $R$  is the distribution of the number of steps the accumulator takes until it resets starting from a reset without hitting the target. In what follows, we omit explicitly writing the functional dependencies on  $q$ ,  $\delta$  and  $N$  for notational convenience.

For all  $0 < \delta < 1$ , the mean first-passage time,  $E[Z]$ , needed to form a minimal evolvable protocell is calculated

directly from

$$E[Z] = \frac{\sum_{z=1}^{\infty} z[ST(z) + (1 - S)R(z)]}{S}. \quad (2.4)$$

Substituting equations (2.1)–(2.3) into equation (2.4) and simplifying, we obtain

$$E[Z] = \frac{1}{S\delta} - \frac{1 - \delta}{\delta}. \quad (2.5)$$

To extract the large- $N$  behaviour of  $E[Z]$  from equation (2.5), we simplify the summation in equation (2.1) for large  $N$  using the following procedure. For a smooth function  $f(x)$ , we use the notation  $f^{(i)}(x) = d^i f(x)/dx^i$ . We can express an integration of  $f^{(i)}(x)$  with respect to  $x$  from 0 to  $\infty$  as

$$\int_0^{\infty} dx f^{(i)}(x) = \frac{1}{N} \sum_{z=0}^{\infty} \int_0^1 dy f^{(i)}\left(\frac{z+y}{N}\right).$$

Next, we write a Taylor expansion of  $f^{(i)}((z+y)/N)$  in powers of  $y/N$  and perform the integration over  $y$ . We have

$$\int_0^{\infty} dx f^{(i)}(x) = \sum_{m=0}^{\infty} \frac{1}{(m+1)! N^m} \left[ \frac{1}{N} \sum_{z=0}^{\infty} f^{(i+m)}\left(\frac{z}{N}\right) \right]. \quad (2.6)$$

Substituting equation (2.6) into equation (2.1) to express the summation as an integration, substituting the integral form of the Beta function,  $B(x, y) = \int_0^1 dt t^{x-1} (1-t)^{y-1}$ , and henceforth using  $\sim$  to denote asymptotic equivalence as  $N \rightarrow \infty$ , we obtain

$$S \sim \frac{-\delta}{(1 - \delta) \log(q)} B\left(\frac{\log(1 - \delta)}{\log(q)}, N + 1\right). \quad (2.7)$$

Substituting equation (2.7) into equation (2.5), expressing the beta function using gamma components,  $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ , using Stirling's formula for the gamma function,  $\Gamma(x) \sim x^x e^{-x} \sqrt{2\pi/x}$ , and simplifying for large  $N$ , we find that  $E[Z]$  grows asymptotically as

$$E[Z] \sim \alpha N^k, \quad (2.8)$$

where

$$\alpha = \frac{-(1 - \delta) \log(1 - p)}{\delta^2 \Gamma(k)}$$

and

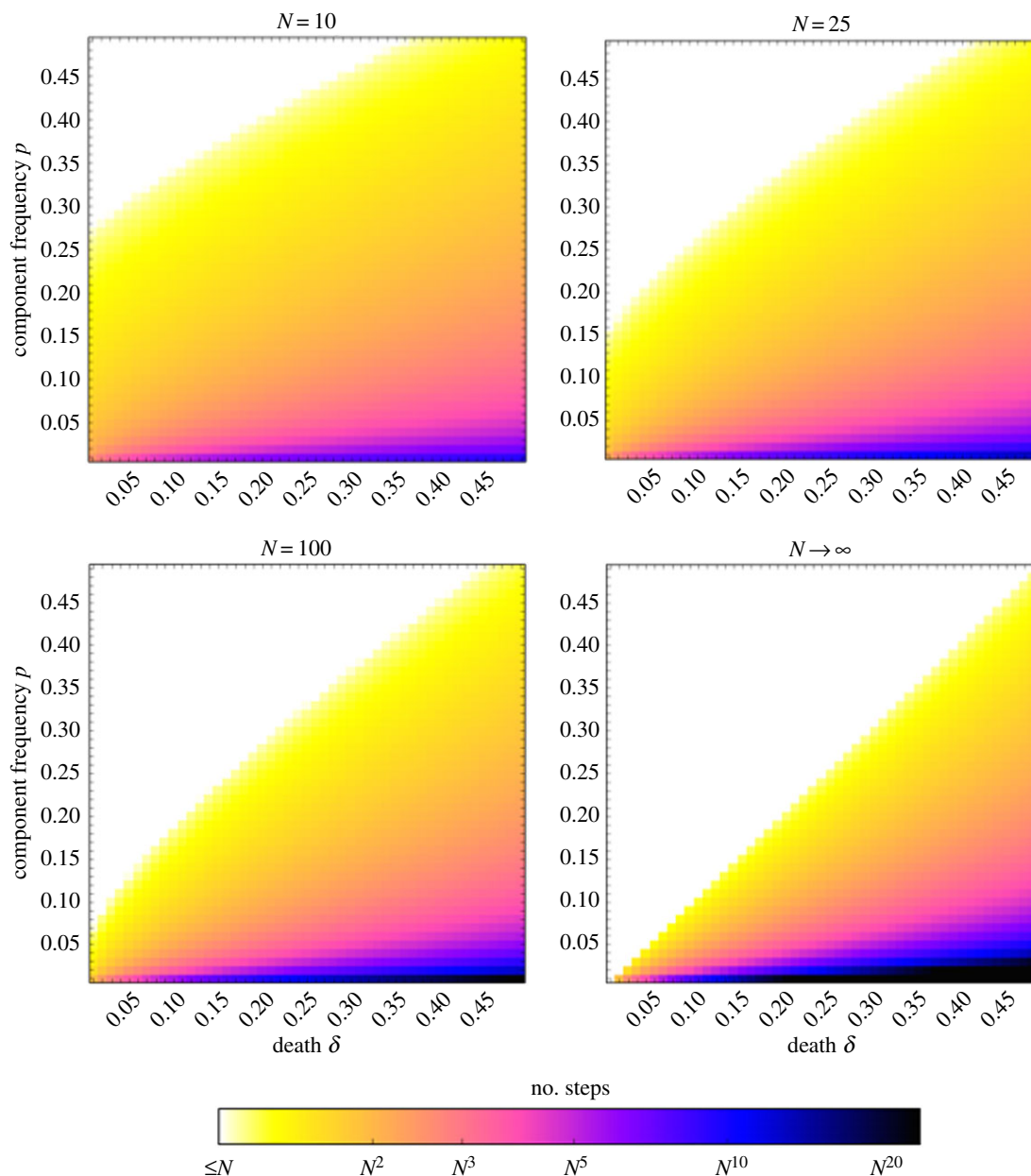
$$k = \frac{\log(1 - \delta)}{\log(1 - p)}.$$

The time complexity of concurrence of components is thus fundamentally altered: for any slight amount of merging, i.e. for any value  $0 < \delta < 1$ ,  $E[Z]$  grows algebraically with  $N$ . Intriguingly, for many values of  $p$  and  $\delta$ ,  $E[Z]$  grows only as a small power of  $N$ , and for many other values of  $p$  and  $\delta$ ,  $E[Z]$  grows only sublinearly with  $N$  (figure 2).

For the particular case in which  $\delta \ll 1$ ,  $p \ll 1$  and  $\delta$  is not too large relative to  $p$ , equation (2.8) admits a simple approximation:

$$E[Z] \approx \frac{1}{\delta} N^{\delta/p}. \quad (2.9)$$

The exact form of  $E[Z]$  for all values of  $N$  given by equation (2.5),  $E[Z]$  measured using a Monte Carlo simulation of the accumulation of components, the exact asymptotic result for  $E[Z]$  given by equation (2.8), and the approximation for  $E[Z]$  given by equation (2.9) are plotted in figure 3 for several values of  $p$ ,  $\delta$  and  $N$ .



**Figure 2.** Minimal evolvable protocells are achieved in polynomial time for the vast majority of parameter space. For  $N = 10$ ,  $N = 25$  and  $N = 100$ , we perform Monte Carlo simulations of the accumulation of components, and we plot  $N^{\log_{10}(E[Z])}$  as a function of  $p$  and  $\delta$ . For  $N \rightarrow \infty$ , we plot  $N^k$  as a function of  $p$  and  $\delta$ . (Online version in colour.)

For the particular case  $\delta = 0$ , the accumulator begins as a randomly assembled protocell, and it is never reset. For this case, the mean first-passage time,  $E_{\delta=0}[Z]$ , grows logarithmically with  $N$ , i.e.,<sup>1</sup>

$$E_{\delta=0}[Z] = \sum_{i=1}^N \binom{N}{i} \frac{(-1)^{i+1}}{1 - (1-p)^i} \sim \frac{\log(N)}{-\log(1-p)}$$

Also of interest for the biologically realistic case  $0 < \delta < 1$  is the probability mass function,  $P(Z = z)$ , for the number of steps needed to achieve a minimal evolvable protocell.  $P(Z = z)$  is given by

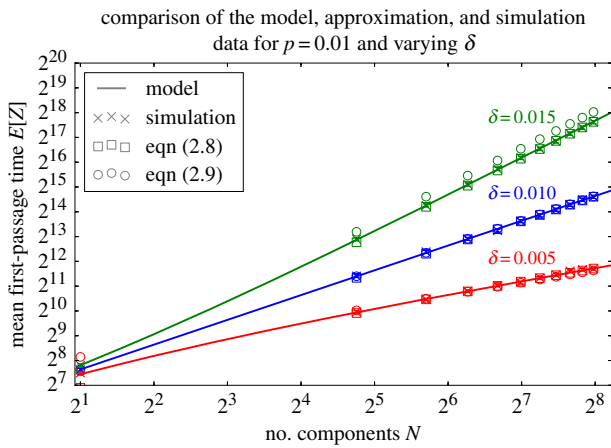
$$P(Z = z) = S \sum_{i=1}^z (1-S)^{i-1} \sum_{\sum_{j=1}^i z_j = z} T(z_1) \prod_{j \neq 1} R(z_j). \quad (2.10)$$

In essence, this equation splits the total number of steps  $z$  into the runs in which the accumulator resets without success ( $i - 1$  times), and the final run in which the accumulator hits

the target set. For simplicity, the number of steps taken in the successful run is labelled as  $z_1$ , and the rest are denoted by  $z_j$ .

If  $N$  is small, then there is typically a small number of resets before the accumulator gains all components, which corresponds to each  $z_j$  being comparable in magnitude to  $z$  in the summations in equation (2.10). But if  $N$  is large, then there is typically a large number of resets before the accumulator gains all components, which corresponds to having  $z_j \ll z$  for all  $j$  in the summations in equation (2.10). In this case, the total number of steps,  $Z$ , is the sum of many independent and identically distributed random variables.

To provide a sense of how good of an estimator  $E[Z]$  is for the variable  $Z$ , we look at its concentration  $\tilde{Z} = Z/E[Z]$ . Denote  $\mu$  as the average number of steps before an accumulator resets given that the accumulator resets before gaining all  $N$  components. Denote  $\sigma^2$  as the variance in the number of steps before an accumulator resets given that the accumulator resets before gaining all  $N$  components. We have  $\mu = \sum_{z=1}^{\infty} zR(z) \sim 1/\delta$  and  $\sigma^2 = (\sum_{z=1}^{\infty} z^2 R(z)) - \mu^2 \sim (1-\delta)/\delta^2$ . Since both  $\mu$  and  $\sigma^2$  are



**Figure 3.** For  $p = 0.01$  and  $\delta = 0.005$ ,  $\delta = 0.01$  and  $\delta = 0.015$ , we plot the exact form of  $E[Z]$  for any value of  $N$  given by equation (2.5) (solid lines),  $E[Z]$  measured using a Monte Carlo simulation of the accumulation of components (cross), the asymptotically exact form of  $E[Z]$  given by equation (2.8) (square) and the approximation for  $E[Z]$  given by equation (2.9) (circle). (Online version in colour.)

finite, the central limit theorem enables a simplification of equation (2.10) for large values of  $N$ : we obtain the probability density function for  $\tilde{Z}$ :

$$P(\tilde{Z} = \tilde{z}) \sim E[Z] \sum_{i=1}^{\infty} \frac{S(1-S)^i}{\sqrt{2\pi i \sigma^2}} \exp\left(-\frac{(E[Z]\tilde{z} - i\mu)^2}{2i\sigma^2}\right). \quad (2.11)$$

The moments of  $\tilde{Z}$  are directly computed from equation (2.11)

$$E[\tilde{Z}^m] \sim \int_0^{\infty} d\tilde{z} \tilde{z}^m P(\tilde{Z} = \tilde{z}) \sim m!$$

Since the corresponding moment-generating function is that of the exponential distribution, it follows that  $Z$  is exponentially distributed about  $E[Z]$ :

$$P(Z = z) \sim \frac{1}{\alpha N^k} \exp\left(-\frac{z}{\alpha N^k}\right).$$

For large  $N$ , the natural production of minimal evolvable protocells via random assembly and repeated fusion is therefore simply a Poisson process.

### 3. Discussion

It is noteworthy that  $E[Z]$ , equation (2.5), provides an upper bound on the time to construct a minimal evolvable protocell for many natural variations of this process. There are many ways in which the first-passage time can be shortened. For instance, the expected time to reach the target set of  $N$  components is reduced if cells divide (and retain some components) instead of losing all components through death. Redundancy in components, where a protocell might have one or more backup copies of each component, can have a similar effect. Additionally, random assembly does not necessarily take exponential time if the components are not independently sampled. Yet, regardless of the worst-case time complexity, we show that fusion helps in reducing the hitting time. Our simple model is specified by only three parameters. Our model is, therefore, robust for exploring the time complexity of a myriad of compartmentalization scenarios by simply tuning the values of  $p$ ,  $\delta$  and  $N$ .

Doing so will help in understanding several biological questions and relates our work to other studies that are interested in the timescale of evolutionary events. For instance,

Wilf & Ewens [54] arrive at exactly the same formula for  $\delta = 0$  when looking for the time it takes for evolution (on a smooth landscape with a single peak, hence  $\delta = 0$ ) to reach a target set of genes. This analysis is also favourable to viewing sex as the default biological state. Computational and experimental analyses of sex suggest that it may make evolutionary search over a landscape more efficient [55–58]. Our analysis adds that this advantage could be present, and aid, in starting cellular replication itself. While biologists have considered the possibility of early sex before [40,59], it was soon observed that parasitism could be a serious problem [30]. However, our exact asymptotic analysis instead suggests that sex is a good strategy, even in the presence of parasites.

Oceanic currents in early earth could have brought together primitive protocells with disparate components, which subsequently merged and eventually spawned an evolvable protocell. In this scenario, protocell formation and merging act as a necessary bridge between physically and chemically heterogeneous prebiotic environments for biological construction. Indeed, there is exciting, ongoing experimental work on creating ‘self-sustaining’ protocells, which can divide and subsequently restore their viable composition via fusion for a few generations [60]. Similarly, recent experiments have demonstrated predation-like behaviour and phagocytosis in communities of semi-permeable colloidal objects which encapsulate biological molecules [61,62]. As these mechanisms mirror merging, they provide a potential platform where protocells with novel functionality can be searched for through parallelization and random merging.

Our mathematical model is also appropriate for investigating the biological activation of modern viruses. In particular, our model captures a process known as *multiplicity reactivation*. In this process, multiple non-functional, mutant viruses of the same strain combine, thereby covering each other’s loss-of-function mutations and producing a functioning virus. Our analysis readily provides the expected number of such viral particles required that would reactivate a virus (although in many cases  $N$  is too small for the speed-up to be of interest). In a similar scenario, in multi-compartment viruses, multiple distinct components need to co-infect the same host in order to produce a new virion. In numerous plant viruses, such as the genus *Tymovirus*, the infection occurs when two or more functionally distinct virions infect the same host [63,64]. The occurrence of this type of combinatorial reproduction in many RNA viruses, which are thought to be ancient, is consistent with the thesis that primordial sex played an integral role in early life [65,66].

Research into minimal synthetic cells has shown that cells with a few hundred genes are able to self-sustain in complex media [67–69]. This suggests that even for low values of  $p$ , in this case the probability of required genes versus random protein-coding genes, novel self-sustaining cells (and possibly viruses) could be produced, either in the laboratory or in early life, by a feasible number of fusions. The feasibility of finding novel viable combinations through a merging process in the laboratory should also be of help in order to understand the density of viable solutions within the fitness landscape, a problem of interest in algorithmic perspectives of evolution and origin of life [55,70,71].

We may never know with certainty what path has resulted in the emergence of life on earth. There are likely many possible paths to evolvability, none of which have been fully delineated to this date. So far, virtually all

models of protocells assumed a small initial set size, precisely because co-occurrence of many components together is unlikely. We show that even if the number of required components  $N$  is large, there are tenable paths to construct such an assembly. The merging mechanism is not as critical if  $N$  is small, but in the presence of merging compartments we are no longer restricted to this scenario. Here, we have devised and analysed a model that captures a general set of possibilities for an evolvable protocell to emerge. It is noteworthy that our model remains agnostic about whether template-directed replication or metabolism emerged first and it can apply in both scenarios as well as different levels of complexity (from chemicals to enzymes and genes) in the underlying components.

To the best of our knowledge, our study is the first to provide a rigorous and quantitative blueprint for comparing the plausibility of a subset of paths to life: those that involve compartmentalization.

**Data accessibility.** Monte Carlo simulations were used to validate the model. The Python code for these simulations is available upon request.

## References

- Nowak MA. 2006 *Evolutionary dynamics*. Cambridge, MA: Harvard University Press.
- Crick FH. 1968 The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379. (doi:10.1016/0022-2836(68)90392-6)
- Orgel LE. 1968 Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393. (doi:10.1016/0022-2836(68)90393-8)
- Woese C. 1967 *The genetic code*. New York, NY: Harper and Row.
- Fontana W, Buss LW. 1994 'The arrival of the fittest': toward a theory of biological organization. *Bull. Math. Biol.* **56**, 1–64. (doi:10.1016/S0092-8240(05)80205-8)
- Nowak MA, Ohtsuki H. 2008 Prevolutionary dynamics and the origin of evolution. *Proc. Natl Acad. Sci. USA* **105**, 14 924–14 927. (doi:10.1073/pnas.0806714105)
- Smith E, Morowitz HJ. 2016 *The origin and nature of life on earth: the emergence of the fourth geosphere*. Cambridge, UK: Cambridge University Press.
- Chen IA, Nowak MA. 2012 From prelife to life: how chemical kinetics become evolutionary dynamics. *Acc. Chem. Res.* **45**, 2088–2096. (doi:10.1021/ar2002683)
- Eigen M. 1971 Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523. (doi:10.1007/BF00623322)
- Bansho Y, Furubayashi T, Ichihashi N, Yomo T. 2016 Host–parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc. Natl Acad. Sci. USA* **113**, 4045–4050. (doi:10.1073/pnas.1524404113)
- Markvoort AJ, Sinai S, Nowak MA. 2014 Computer simulations of cellular group selection reveal mechanism for sustaining cooperation. *J. Theor. Biol.* **357**, 123–133. (doi:10.1016/j.jtbi.2014.04.029)
- Fontanari JF, Santos M, Szathmáry E. 2006 Coexistence and error propagation in pre-biotic vesicle models: a group selection approach. *J. Theor. Biol.* **239**, 247–256. (doi:10.1016/j.jtbi.2005.08.039)
- Hogeweg P, Takeuchi N. 2003 Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Orig. Life. Evol. Biosph.* **33**, 375–403. (doi:10.1023/A:1025754907141)
- Bernhardt HS. 2012 The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol. Direct.* **7**, 23. (doi:10.1186/1745-6150-7-23)
- Attwater J, Wochner A, Holliger P. 2013 In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **5**, 1011–1018. (doi:10.1038/nchem.1781)
- Lincoln TA, Joyce GF. 2009 Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232. (doi:10.1126/science.1167856)
- Pross A, Pascal R. 2013 The origin of life: what we know, what we can know and what we will never know. *Open Biol.* **3**, 120190. (doi:10.1098/rsob.120190)
- Higgs PG, Lehman N. 2015 The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17. (doi:10.1038/nrg3841)
- McCullom TM, Ritter G, Simoneit BR. 1999 Lipid synthesis under hydrothermal conditions by Fischer–Tropsch-type reactions. *Orig. Life. Evol. Biosph.* **29**, 153–166. (doi:10.1023/A:1006592502746)
- Yuen G, Blair N, Des Marais DJ, Chang S. 1984 Carbon isotope composition of low molecular weight hydrocarbons and monocarboxylic acids from Murchison meteorite. *Nature* **307**, 252–254. (doi:10.1038/307252a0)
- Lawless JG, Yuen GU. 1979 Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. *Nature* **282**, 396–398. (doi:10.1038/282396a0)
- Deamer DW. 1985 Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317**, 792–794. (doi:10.1038/317792a0)
- Segré D, Ben-Eli D, Deamer DW, Lancet D. 2001 The lipid world. *Orig. Life. Evol. Biosph.* **31**, 119–145. (doi:10.1023/A:1006746807104)
- Lane N, Martin WF. 2012 The origin of membrane bioenergetics. *Cell* **151**, 1406–1416. (doi:10.1016/j.cell.2012.11.050)
- Luisi PL, Walde P, Oberholzer T. 1999 Lipid vesicles as possible intermediates in the origin of life. *Curr. Opin. Colloid Interface Sci.* **4**, 33–39. (doi:10.1016/S1359-0294(99)00012-6)
- Paleos C. 2015 A decisive step toward the origin of life. *Trends. Biochem. Sci.* **40**, 487–488. (doi:10.1016/j.tibs.2015.06.001)
- Deamer DW. 1986 Role of amphiphilic compounds in the evolution of membrane structure on the early earth. *Orig. Life. Evol. Biosph.* **17**, 3–25. (doi:10.1007/BF01809809)
- Yamamoto S, Maruyama Y, Hyodo S-A. 2002 Dissipative particle dynamics study of spontaneous vesicle formation of amphiphilic molecules. *J. Chem. Phys.* **116**, 5842–5849. (doi:10.1063/1.1456031)
- Segré D, Ben-Eli D, Lancet D. 2000 Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl Acad. Sci. USA* **97**, 4112–4117. (doi:10.1073/pnas.97.8.4112)
- Szathmáry E, Santos M, Fernando C. 2005 Evolutionary potential and requirements for minimal protocells. In *Prebiotic chemistry, Topics in*

## Endnote

<sup>1</sup>A reader with interest in algorithms may recognize this result as the expected number of levels in a skip list with  $N$  elements, where  $1 - p$  is the probability of each element's appearance in the next level of the skip list [72,73].

- current chemistry*, vol. 259 (ed. P Walde), pp. 167–211. Berlin, Germany: Springer.
31. Markvoort A, Smeijers A, Pieterse K, vanSanten R, Hilbers P. 2007 Lipid-based mechanisms for vesicle fission. *J. Phys. Chem. B* **111**, 5719–5725. (doi:10.1021/jp068277u)
  32. Fanelli D, McKane AJ. 2008 Thermodynamics of vesicle growth and instability. *Phys. Rev. E* **78**, 051406. (doi:10.1103/PhysRevE.78.051406)
  33. Bianconi G, Zhao K, Chen IA, Nowak MA. 2013 Selection for replicases in protocells. *PLoS Comput. Biol.* **9**, e1003051. (doi:10.1371/journal.pcbi.1003051)
  34. Zhu TF, Szostak JW. 2009 Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.* **131**, 5705–5713. (doi:10.1021/ja900919c)
  35. Budin I, Debnath A, Szostak JW. 2012 Concentration-driven growth of model protocell membranes. *J. Am. Chem. Soc.* **134**, 20 812–20 819. (doi:10.1021/ja310382d)
  36. Adamala K, Szostak JW. 2013 Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100. (doi:10.1126/science.1241888)
  37. Hanczyc MM, Fujikawa SM, Szostak JW. 2003 Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302**, 618–622. (doi:10.1126/science.1089904)
  38. Kinsler G, Sinai S, Lee NK, Nowak MA. 2017 Prebiotic selection for motifs in a model of template-free elongation of polymers within compartments. *PLoS ONE* **12**, e0180208. (doi:10.1371/journal.pone.0180208)
  39. Krapivsky PL, Redner S, Ben-Naim E. 2010 *A kinetic view of statistical physics*. Cambridge, UK: Cambridge University Press.
  40. Bernstein H, Byerly HC, Hopf FA, Michod RE. 1984 Origin of sex. *J. Theor. Biol.* **110**, 323–351.
  41. Santos M, Zintzaras E, Szathmáry E. 2003 Origin of sex revisited. *Orig. Life. Evol. Biosph.* **33**, 405–432. (doi:10.1023/A:1025759024888)
  42. Redner S. 2001 *A guide to first-passage processes*. Cambridge, UK: Cambridge University Press.
  43. Chou T, D'Orsogna MR. 2014 *First passage problems in biology*, ch. 13, pp. 306–345. Singapore: World Scientific. (doi:10.1142/9789814590297\_0013).
  44. Szathmáry E, Demeter L. 1987 Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**, 463–486. (doi:10.1016/S0022-5193(87)80191-1)
  45. Gánti T. 1975 Organization of chemical reactions into dividing and metabolizing units: the chemotons. *BioSystems* **7**, 15–21. (doi:10.1016/0303-2647(75)90038-6)
  46. Kauffman SA. 1986 Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1–24. (doi:10.1016/S0022-5193(86)80047-9)
  47. Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. 2012 Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72–77. (doi:10.1038/nature11549)
  48. Mansy SS, Schrum JP, Krishnamurthy M, Tobé S, Treco DA, Szostak JW. 2008 Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125. (doi:10.1038/nature07018)
  49. Fishkis M. 2007 Steps towards the formation of a protocell: the possible role of short peptides. *Orig. Life Evol. Biosph.* **37**, 537–553. (doi:10.1007/s11084-007-9111-4)
  50. Goldford JE, Hartman H, Smith TF, Segrè D. 2017 Remnants of an ancient metabolism without phosphate. *Cell* **168**, 1126–1134. (doi:10.1021/ar2002683)
  51. Black RA, Blosser MC, Stottrup BL, Tavakley R, Deamer DW, Keller SL. 2013 Nucleobases bind to and stabilize aggregates of a prebiotic amphiphile, providing a viable mechanism for the emergence of protocells. *Proc. Natl Acad. Sci. USA* **110**, 13 272–13 276. (doi:10.1073/pnas.1300963110)
  52. Bar-Yam Y. 1997 *Dynamics of complex systems*, vol. 213. Reading, MA: Addison-Wesley.
  53. Vasas V, Fernando C, Santos M, Kauffman S, Szathmáry E. 2012 Evolution before genes. *Biol. Direct* **7**, 1. (doi:10.1186/1745-6150-7-1)
  54. Wilf HS, Ewens WJ. 2010 There's plenty of time for evolution. *Proc. Natl Acad. Sci. USA* **107**, 22 454–22 456. (doi:10.1073/pnas.1016207107)
  55. Chatterjee K, Pavlogiannis A, Adlam B, Nowak MA. 2014 The time scale of evolutionary innovation. *PLoS Comput. Biol.* **10**, e1003818. (doi:10.1371/journal.pcbi.1003818)
  56. Livnat A, Papadimitriou C, Dushoff J, Feldman MW. 2008 A mixability theory for the role of sex in evolution. *Proc. Natl Acad. Sci. USA* **105**, 19 803–19 808. (doi:10.1073/pnas.0803596105)
  57. Livnat A, Papadimitriou C. 2016 Sex as an algorithm: the theory of evolution under the lens of computation. *Commun. ACM* **59**, 84–93. (doi:10.1145/2934662)
  58. McDonald MJ, Rice DP, Desai MM. 2016 Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236. (doi:10.1038/nature17143)
  59. Haldane JBS. 1929 The origin of life. *Rationalist Annu.* **148**, 3–10.
  60. Kurihara K, Okura Y, Matsuo M, Toyota T, Suzuki K, Sugawara T. 2015 A recursive vesicle-based model protocell with a primitive model cell cycle. *Nat. Commun.* **6**, 8352. (doi:10.1038/ncomms9352)
  61. Qiao Y, Li M, Booth R, Mann S. 2017 Predatory behaviour in synthetic protocell communities. *Nat. Chem.* **9**, 110–119. (doi:10.1038/nchem.2617)
  62. Rodríguez-Arco L, Li M, Mann S. 2017 Phagocytosis-inspired behaviour in synthetic protocell communities of compartmentalized colloidal objects. *Nat. Mater.* **16**, 857–863. (doi:10.1038/nmat4916)
  63. Smith JM. 1989 *Evolutionary genetics*. Oxford, UK: Oxford University Press.
  64. Rao A. 2006 Genome packaging by spherical plant RNA viruses. *Annu. Rev. Phytopathol.* **44**, 61–87. (doi:10.1146/annurev.phyto.44.070505.143334)
  65. Nasir A, Caetano-Anollés G. 2015 A phylogenomic data-driven exploration of viral origins and evolution. *Sci. Adv.* **1**, e1500527. (doi:10.1126/sciadv.1500527)
  66. Koonin EV, Senkevich TG, Dolja VV. 2006 The ancient virus world and evolution of cells. *Biol. Direct* **1**, 29. (doi:10.1186/1745-6150-1-29)
  67. Mushegian AR, Koonin EV. 1996 A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA* **93**, 10 268–10 273. (doi:10.1073/pnas.93.19.10268)
  68. Gil R, Silva FJ, Pereto J, Moya A. 2004 Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–537. (doi:10.1128/MMBR.68.3.518-537.2004)
  69. Hutchison III CA. 2016 Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253. (doi:10.1126/science.aad6253)
  70. Walker SI, Davies PC. 2013 The algorithmic origins of life. *J. R. Soc. Interface* **10**, 20120869. (doi:10.1098/rsif.2012.0869)
  71. Kaznatcheev A. 2013 Complexity of evolutionary equilibria in static fitness landscapes. (<http://arxiv.org/abs/1308.5094>)
  72. Pugh W. 1990 Skip lists: a probabilistic alternative to balanced trees. *Commun. ACM* **33**, 668–676. (doi:10.1145/78973.78977)
  73. Kirschenhofer P, Martínez C, Prodinger H. 1995 Analysis of an optimized search algorithm for skip lists. *Theoret. Comput. Sci.* **144**, 199–220. (doi:10.1016/0304-3975(94)00296-U)