CrossMark

Chemical Evolution and the Evolutionary Definition of Life

Paul G. Higgs[1]

Abstract Darwinian evolution requires a mechanism for generation of diversity in a population, and selective differences between individuals that influence reproduction. In biology, diversity is generated by mutations and selective differences arise because of the encoded functions of the sequences (*e.g.*, ribozymes or proteins). Here, I draw attention to a process that I will call *chemical evolution*, in which the diversity is generated by random chemical synthesis instead of (or in addition to) mutation, and selection acts on physicochemical properties, such as hydrolysis, photolysis, solubility, or surface binding. Chemical evolution applies to short oligonucleotides that can be generated by random polymerization, as well as by template-directed replication, and which may be too short to encode a specific function. Chemical evolution is an important stage on the pathway to life, between the stage of "just chemistry" and the stage of full biological evolution. A mathematical model is presented here that illustrates the differences between these three stages. Chemical evolution leads to much larger differences in molecular concentrations than can be achieved by selection without replication. However, chemical evolution is not open-ended, unlike biological evolution. The ability to undergo Darwinian evolution is often considered to be a defining feature of life. Here, I argue that chemical evolution, although Darwinian, does not quite constitute life, and that a good place to put the conceptual boundary between non-life and life is between chemical and biological evolution.

Introduction

Evolution is often considered to be a defining feature of life (Baross 2007). One of the most frequently used definitions of life states that it is "a self-sustaining chemical system capable of undergoing Darwinian evolution" (Joyce 1994). There has been a lot of argument both for and against the evolutionary definition of life (e.g., Ruiz-Mirazo et al. 2002; Koshland 2002; Cleland and Chyba 2007; Benner 2010), but most of those arguing against it would like to add further requirements in addition to evolution, rather than to remove the requirement for evolution. After analyzing the words used in over 100 published definitions of life, Trifonov (2011) concludes that reproduction and evolution are the minimal set of key concepts, and suggests a consensus definition: "life is self-reproduction with variations."

There is a certain amount of fatigue in the origins of life field regarding definitions. For example, Szostak (2012a) argues that attempts to define life do not help to understand life's origins. He argues that there is a continuous pathway of increasing complexity from chemistry to biology and that deciding where to put the boundary line along this pathway is just an arbitrary decision. In this paper, I am motivated by trying to understand the steps on the pathway from chemistry to biology, and while I agree that definitions are not an end in themselves, I think that having clear definitions can actually help us to understand the processes involved in the origin of life.

The aim of this paper is to consider a stage that I will call chemical evolution that is intermediate between non-

✉ Paul G. Higgs
  higgsp@mcmaster.ca

[1] Origins Institute and Department of Physics and Astronomy, McMaster University, Hamilton, ON, Canada

Table 1 Summary of factors that distinguish selection without replication, chemical evolution, and biological evolution

	Selection without replication	Chemical evolution	Biological evolution
Selection	Physicochemical properties	Physicochemical properties	Encoded function
Replication	No	Yes	Yes
Darwinian	No	Yes	Yes
Chemical synthesis	Yes	Yes	No
Mutation	No	Yes or No	Yes
Timescale	Fast	Fast	Slow
Repeatable	Yes	Yes	No
Open-ended	No	No	Yes
Is it life?	No	Depends on the definition	Yes

living chemistry and fully fledged biological evolution. I will consider a simple mathematical model that establishes why chemical evolution really is Darwinian evolution and not just chemistry, but also establishes how chemical evolution differs from the usual kind of biological evolution that applies to genes and proteins in modern organisms. Darwinian evolution requires a mechanism of replication that passes on the properties of the parent, a mechanism of selection that allows fitter individuals to survive and/or reproduce faster than less fit individuals, and a mechanism for generation of diversity in the population. In biological evolution, diversity is generated by mutation. Mutation usually refers to a point mutation in a sequence, but other processes that generate sequence variants (like insertions and deletions) could also be classed as mutations. The key point for mutations is that new sequences are variants on the old ones rather than being generated independently of the old ones. The key point that distinguishes chemical evolution, as I define it here, is that the molecular diversity on which evolution acts is generated by random chemical synthesis in addition to mutation, or instead of mutation.

Once a genetic system, such as DNA, is operating, occasional errors are bound to occur during replication. Hence, new sequences are continually generated that create variation on which selection can act. Cells today synthesize DNA by copying an existing strand. Cells do not synthesize long random nucleic acid sequences without a template because the chance of a random strand having a useful function would be very small. DNA polymerases are necessary to copy the long DNA sequences used by organisms today. However, at the time of the origin of life, non-enzymatic template-directed synthesis may have been important (Deck et al. 2011; Szostak 2012b; Leu et al. 2013). I envisage that this could operate on oligomers that were relatively short, and which did not need to have any specific function other than being a template and being stable for sufficiently long for template-directed replication to occur. Template-directed replication would have been occurring alongside random polymerization without a

template. Thus there would be a continued source of sequence diversity generated by random polymerization, in addition to diversity generated by error-prone replication of existing oligomers. If replication was slow and inefficient originally, then a majority of diversity could have been due to random chemical synthesis rather than mutation.

In this paper, I will present a simple mathematical model to study the way that selection might operate on molecular diversity that is generated by chemical synthesis. I will show that there are three conceptually different stages that may be considered as stages on the path to life. These are summarized in Table 1. The first stage—selection without replication—involves selection acting on molecular diversity without any process of replication or reproduction. Since there is no replication, this cannot be considered as evolution. Nevertheless, there is a change in frequencies of different molecules as a result of the selection process. The second stage is chemical evolution. This is true Darwinian evolution, since it incorporates replication and heredity, but the molecules involved are simple enough to be generated by random chemistry, and the diversity generated by random chemistry is significant in comparison to mutational diversity. The third case is biological evolution. This involves evolution of molecules that are too long to be synthesized chemically from scratch, and can only be synthesized by copying existing sequences. Diversity is generated by mutation, not random synthesis. I will return to discuss Table 1 in detail after the presentation of the model.

There are several kinds of processes that could lead to selection at the chemical level. Hydrolysis opposes the polymerization of nucleic acids, proteins, and many other polymers. Cycling of wet and dry conditions may be relevant as a means to drive formation of longer polymers in prebiotic conditions (Mamajanov et al. 2014; Da Silva et al. 2015; Forsythe et al. 2015; Higgs 2016). Thus stability against hydrolysis is likely to be an important property that would cause selection among different oligomers. Other properties could be the stability to photolysis by UV light (Mulkidjanian et al. 2003), the strength of interaction

with a mineral surface (Deck et al. 2011), or the ability to be encapsulated in a lipid vesicle (Damer and Deamer 2015). For all these reasons, selective differences will arise between oligomers, and if template-directed replication occurs, chemical evolution will generate large differences in the concentration of different oligomers. This can occur even when the oligomers are too short to be able to encode a specific function, such as being a ribozyme catalyst in the RNA World, or being a protein-coding gene in a modern organism. Hence, another important difference between chemical and biological evolution, as summarized in Table 1, is that chemical evolution acts on physicochemical properties, while biological evolution acts on encoded function.

Pross (2012) has emphasized the continuity between chemistry and biology, and introduced a concept of dynamic kinetic stability that is an analogy with fitness in biology. He emphasizes the importance of replication as the foundation for life, and argues that we need a better understanding of the way replication and selection function in chemical systems in order to understand the origin of life. Although I generally agree with his approach, Pross (2012) deliberately avoids giving quantitative examples of dynamic kinetic stability. The model in this paper is intended to be a simple, but precise example of how selection, replication and evolution can operate at the chemical level.

A simple 'hill-climbing' analogy may be useful at this point. Selection without replication gets you one step up the slope; chemical evolution gets you close to the top of the nearest hill; and biological evolution allows you to explore the whole mountain range. The chemical evolution stage is important during the origin of life, and is probably necessary to get true biological evolution going. In the discussion, I will consider the implications of chemical evolution for the evolutionary definition of life. I will suggest that, even though chemical evolution is Darwinian evolution, it does not satisfy my own preconceptions of what should count as life. Therefore, if I go against the advice of Szostak (2012a) and decide to put a boundary between non-life and life on the continuous pathway from chemistry to biology, I would like to put that boundary between chemical evolution and biological evolution.

A Simple Model Illustrating Chemical Evolution

I will consider a simple model of formation and selection of oligomers of length L made from two different monomers. There are $N = 2^L$ possible oligomers. We suppose that these oligomers can be formed by spontaneous chemical synthesis at rate s and that they can replicate at rate r. I am using the same notation here as in several previous papers related to RNA replication (Wu and Higgs 2012; Shay et al. 2015; Kim and Higgs 2016), where the distinction between spontaneous synthesis without a template (s), non-enzymatic template-directed replication (r), and ribozyme-catalyzed replication (k) is important. This paper considers s and r without k, as I am envisaging chemical evolution occurring at a stage before the origin of ribozymes or other biological catalysts.

For simplicity, we suppose that s and r are equal for all oligomers. However, the oligomers differ in their stability. We can label the oligomers with an integer i, in the range $0 \leq i \leq N-1$. Let b_i be the breakdown rate of oligomer i, which can be different for each oligomer. Let X_i be the concentration of oligomer i, and let the total concentration be $X_{tot} = \sum_i X_i$. The concentrations satisfy the following differential equations:

$$\frac{dX_i}{dt} = (s + rX_i)(1 - X_{tot}) - b_iX_i. \tag{1}$$

The factor $(1-X_{tot})$ represents the fact the formation of oligomers is limited by some resource such as monomer supply or space; hence, the total concentration of oligomers cannot rise above a carrying capacity, which we have set to one. If either s or r is large enough, the total concentration will rise very close to the carrying capacity.

Firstly, consider the case where there is no replication ($r = 0$). In this case, the equilibrium concentrations are

$$X_i = \frac{s}{b_i}(1 - X_{tot}). \tag{2}$$

Summing these, we obtain a condition for X_{tot}:

$$X_{tot} = s(1 - X_{tot}) \sum_i \frac{1}{b_i} = \frac{sN}{\hat{b}}(1 - X_{tot}), \tag{3}$$

where $\hat{b}$ is the harmonic mean of the breakdown rates, defined as

$$\hat{b} = \left(\frac{1}{N}\sum_i \frac{1}{b_i}\right)^{-1}. \tag{4}$$

Rearranging (3), we obtain

$$X_{tot} = \frac{sN}{sN + \hat{b}} \tag{5}$$

The mean concentration of individual oligomers is $\bar{X} = X_{tot}/N$. Finally, from (2) and (5), we obtain a convenient form for the equilibrium concentration of each oligomer:

$$X_i = \frac{\hat{b}}{b_i}\bar{X}. \tag{6}$$

This equation simply says that if selection acts on the stability of oligomers in the absence of replication, then

the concentrations of the oligomers are inversely proportional to the breakdown rates. An example of this is shown in Fig. 1. We considered hexamers ($L = 6$, $N = 64$). The breakdown rate of each hexamer was set to a random number in the range 1–2, and s was set to one. The dashed line shows the mean frequency, $\bar{X}$, calculated for this particular choice of b_i, and the solid line shows the curve with concentration inversely proportional to b_i, which is to be expected from Eq. (6). As a check, we also used the differential Eq. (1) to find the stationary solution for each oligomer. These are shown as open circles in Fig. 1. The open circles fall exactly on the solid curve, as expected.

The main point so far is that selection without replication leads to only a modest change in the oligomer concentrations. In this example, all the oligomers would have equal frequency in the absence of selection, and the selection on stability leads to concentrations that are inversely proportional to the breakdown rates. However, all the oligomers are still present in the mixture. Since the breakdown rates fall within a factor of two of each other, the concentrations also fall within a factor of two. Selection without replication simply changes one mixture into another mixture with slightly different concentrations, and does not create a high concentration of the 'fittest' sequences. In contrast, when replication occurs as well as selection, then we have chemical evolution. This leads to very much larger differences in concentrations of the sequences, as will now be shown.

If the replication rate r in Eq. (1) is not zero, then the stationary solution for X_i can be written as

$$X_i = \frac{s(1 - X_{\text{tot}})}{b_i - r(1 - X_{\text{tot}})}, \tag{7}$$

where X_{tot} satisfies

$$X_{\text{tot}} = \sum_i \frac{s(1 - X_{\text{tot}})}{b_i - r(1 - X_{\text{tot}})}. \tag{8}$$

The solution of (8) cannot be written in a simple closed form; therefore, it is easier to find the stationary solutions for X_i by numerically integrating the differential Eq. (1) forward in time till a steady state is reached. From (1), it can be seen that replication becomes important relative to spontaneous formation when $rX_i \sim s$ or greater. As the total concentration is close to one (the carrying capacity), the typical concentration of one oligomer is close to $1/N$. Thus, we expect that replication should have a noticeable effect on concentrations when $r \sim sN$ or greater.

We define $s_{\text{tot}} = sN$, which is the total spontaneous synthesis rate of all the oligomers together. In Fig. 1, $s = 1$ and $s_{\text{tot}} = 64$. The figure shows the equilibrium concentrations for $r = s_{\text{tot}}$, and for $r = 10s_{\text{tot}}$, with the same values of the b_i as before. When $r = s_{\text{tot}}$, replication and spontaneous synthesis are roughly equal. It can be seen that the range of concentrations is now roughly twice what it was in the absence of replication, with high concentration sequences being further increased in concentration and low concentration sequences being further decreased. When $r = 10s_{\text{tot}}$, replication is significantly faster than spontaneous synthesis. The range of concentrations is now very much larger than in the absence of replication. Most of the sequences have a lower concentration than they would without replication, but the few most stable sequences (those with the lowest breakdown rates) have very much higher concentrations. In the limit where $r \gg s_{tot}$, concentration of the most stable sequence would be close to 1, and the others would all have negligible concentrations.

This example demonstrates the power of evolution and natural selection. The fittest sequences are those with the lowest breakdown rates. Selection is caused by differences in breakdown rates. However, the effect on sequence concentrations only becomes strong when r is high. When r is non-zero, there is heredity in this problem, because high-fitness sequences replicate to create additional high-fitness sequences. If replication occurs in addition to selection on breakdown rates, this is a true evolutionary problem, and we see the survival of the fittest sequences in high concentrations. This is an example of chemical evolution, because spontaneous chemical synthesis is relevant in this problem, and because it contains replication and selection, which are the essential features of Darwinian evolution. In order to be an evolutionary system, we require r to be non-zero. If r is zero, there is no heredity, selection is weak, and evolution cannot occur.
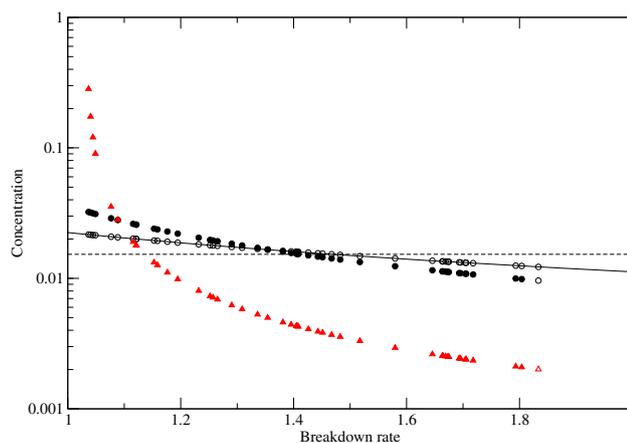
Fig. 1 Concentration of oligomers as a function of their breakdown rate. *Dashed line*—mean concentration $\bar{X}$; *Solid line*—exact solution for the model with selection only; *Open circles*—simulation results for the model with selection only; *Black circles*—simulation results for the chemical evolution model with $r = s_{\text{tot}}$; *Red triangles*—simulation results for the chemical evolution model with $r = 10s_{\text{tot}}$

The above example illustrates why chemical evolution is different from the case of selection without replication, which may be considered as "just chemistry" without evolution. In the following section, we consider the way that chemical evolution relates to more usual examples of biological evolution.

The Relationship Between Mutation and Spontaneous Chemical Synthesis

Mutations are essential in most evolutionary models because mutations generate the diversity on which natural selection acts. In the chemical evolution model above, mutation was not necessary to generate diversity, because spontaneous synthesis was generating a diverse mixture of random oligomers. For simplicity, we therefore left out mutation in the previous section. However, replication processes at the time of the origin of life must have been error prone, hence it is important to consider the effects of mutation in addition to chemical synthesis.

In order to do this, we need to associate each oligomer i with a specific sequence in the binary sequence space, *e.g.*, sequence $i = 24$ is the binary sequence 011000 and sequence $j = 57$ is the binary sequence 111001. The Hamming distance $d(i,j)$ between two sequences is the number of point mutations required to convert one to the other, *e.g.*, $d(24,57) = 2$. We suppose that mutations occur at the time of replication, with a probability u per base of making an error, and a probability $1-u$ of correctly replicating. The probability of producing sequence j by replication of sequence i is therefore

$$Q_{ij} = u^{d(i,j)}(1-u)^{L-d(i,j)}. \tag{9}$$

Adding mutation into Eq. (1), we obtain

$$\frac{dX_i}{dt} = \left(s + r\sum_j Q_{ji}X_j\right)(1 - X_{tot}) - b_iX_i. \tag{10}$$

This model is similar to the molecular quasispecies model (Eigen et al. 1988, 1989; Bull et al. 2005) in that the binary sequence space is used, with mutations depending on the Hamming distance between sequences. In the chemical evolution model, we have introduced selection via differences in breakdown rates, whereas in the quasi-pecies model, selection is usually introduced via differences in replication rates; however, differences in both these factors could be included in either of these models. In the chemical evolution model, the total concentration is limited by the carrying capacity term, which represents resource limitation, whereas in the quasispecies model, the concentration is limited by diluting the system; however, both mechanisms insure that there is a finite total

concentration, and that each sequence is being selected relative to the average of the population. The most important difference between the models is that the term representing spontaneous chemical synthesis of oligomers, s, is included in the chemical evolution model, but not in the quasispecies model. The aim of this section is to show that there is a similarity between the effects of chemical synthesis and mutation.

Figure 2 shows a comparison of the cases with chemical synthesis alone, both chemical synthesis and mutation, and mutation alone. The case with chemical synthesis alone (red triangles) is the same as Fig. 1. We chose $s_{tot} = 64$, corresponding to $s = 1$ for each oligomer, and $r = 10s_{tot} = 640$, as before. The case with both factors present (open circles) has a mutation probability $u = 0.01$ per base, with all the other parameters the same. Adding mutation as an extra source of diversity reduces the effects of selection: the concentrations of the sequences move close to the average, and there is a smaller range of concentrations than in the case with only chemical synthesis. The third case in Fig. 2 has mutation only, with $u = 0.01$, $s_{tot} = 0$, and $r = 640$ (black circles). Removing chemical synthesis increases the effect of selection dramatically: the range of concentrations is broader by two orders of magnitude.

When there is chemical synthesis only, the concentrations decrease as a smooth function of breakdown rate. This is no longer the case when mutation is added, because the concentrations depend on the fitness landscape, i.e., the relative positions of high and low fitness sequences in the sequence space. For the random choice of breakdown rates that we chose here, it happens that the sequences with the second and third lowest breakdown rate are separated by a
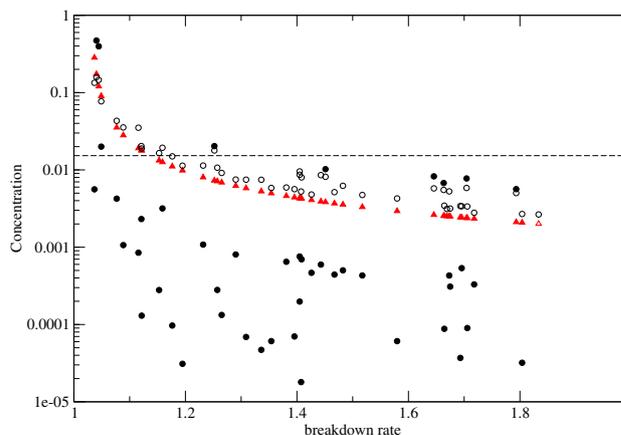
Fig. 2 Concentration of oligomers as a function of their breakdown rate. *Dashed line*—mean concentration $\bar{X}$ in absence of replication; *Red triangles*—simulation results for the chemical evolution model with $s_{tot} = 64$ and $u = 0$. *Open circles*—simulation results for the model with $s_{tot} = 64$ and $u = 0.01$; *Black circles*—simulation results for the model with $s_{tot} = 0$ and $u = 0.01$. The replication rate is $r = 640$ in all three cases

single mutation, and mutation between these two sequences reinforces their concentrations. The sequence with the lowest breakdown rate is isolated from these two. In Fig. 2 it can be seen that the concentrations of the second and third lowest sequences are actually higher than the concentration of the lowest sequence when mutation is present. This is similar to the so-called 'survival of the flattest' effect (Wilke et al. 2001; Sardanyés et al. 2008), where a single high-fitness sequence can be outcompeted by a group of lower fitness sequences that are close to one another in sequence space.

In the quasispecies theory, the case of the master sequence landscape has been studied in detail (Eigen et al. 1988; 1989). This is a simple choice of fitness landscape where there is one master sequence with a high fitness, and all the other sequences have an equal fitness that is lower than the master sequence. This case gives rise to the error threshold phenomenon: for low mutation rates, there is a finite fraction of the master sequence in the population, whereas for mutation rates greater than a threshold value, the concentration of the master sequence becomes negligible. We will now show that the chemical evolution model has an error threshold that depends on the rate of chemical synthesis as well as the mutation rate.

Let there be a single master sequence with low breakdown rate b_0, and let all the other sequences have a higher breakdown rate b_1. Let X_0 be the concentration of the master sequence, and let X_1 be the combined concentration of all the other sequences. The fidelity of replication of the master sequence is $(1 - u)^L$. Hence, the rate of mutation from the master sequence to all the other sequences is $M = 1 - (1 - u)^L$. The rate of back mutation from the other sequences to the master sequences can be assumed to be negligible if L is large enough. From (10) we can write down two equations for X_0 and X_1.

$$\frac{dX_0}{dt} = (s + rX_0(1 - M))(1 - X_{tot}) - b_0 X_0 \tag{11}$$

$$\frac{dX_1}{dt} = (s_{tot} - s + rX_0 M + rX_1))(1 - X_{tot}) - b_1 X_1 \tag{12}$$

Exact solutions can be obtained for these equations by numerical integration forward to the steady state. However, it is also possible to obtain an approximate solution analytically. We already assumed that L was large enough to neglect back mutations; therefore we can also neglect s relative to s_{tot}, since $s_{tot} = 2^L s$. At the stationary state, from Eqs. (11) and (12) we have

$$1 - X_{tot} \approx \frac{b_0}{r(1 - M)} = \frac{b_1 X_1}{s_{tot} + rX_0 M + rX_1}. \tag{13}$$

If r is large enough, the total concentration is close to one, so $X_1 \approx 1 - X_0$. Rearranging (13) we obtain an approximate solution for X_0:

$$X_0 \approx \frac{b_1 r(1 - M) - b_0(s_{tot} + r)}{(b_1 - b_0)r(1 - M)}. \tag{14}$$

The error threshold occurs when the numerator of (14) becomes zero. In terms of the mutation rate, the master sequence concentration goes to zero when $(1 - M) = b_0(s_{tot} + r)/b_1 r$, i.e., when

$$u \approx \frac{1}{L} \ln\left(\frac{b_1 r}{b_0(s_{tot} + r)}\right). \tag{15}$$

This shows that the maximum error rate per base is roughly inversely proportional to the sequence length, as in the usual quasispecies theory. Additionally, however, Eq. (14) shows that there is an error threshold as a function of the chemical synthesis rate. For a given mutation rate, the master sequence concentration goes to zero when

$$s_{tot} = (b_1(1 - M) - b_0)r/b_0. \tag{16}$$

Figures 3 and 4 show the exact solution for X_0 from (11) and (12) in comparison to the approximate solution from (14). In Fig. 3, the mutation rate is varied with $s_{tot} = 0$, and with s_{tot} fixed at 50. This is the usual error threshold as a function of mutation rate. In Fig. 4, the synthesis rate is varied with $u = 0$, and with u fixed at 0.01. This makes it clear that the same error threshold phenomenon occurs as a function of the synthesis rate and as a function of mutation rate.

The model above demonstrates that spontaneous chemical synthesis in chemical evolution models functions in a similar way to mutation in biological evolution models. Both factors generate diversity of sequences on which natural selection can act. The concentrations of sequences that occur at equilibrium depend on the relative strength of
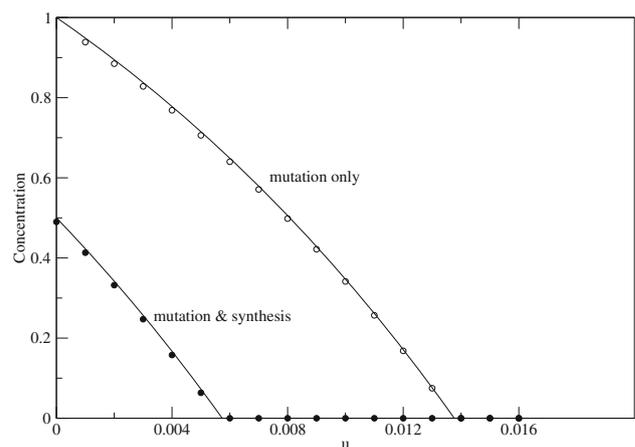
Fig. 3 Concentrations of the master sequence as a function of the per-base error rate, u. Mutation only—$s_{tot} = 0$; Mutation & synthesis—$s_{tot} = 50$. Points are numerical solutions of Eqs. 11 and 12. Solid lines are predictions of Eq. 14. In all cases, $L = 50$, $r = 100$, $b_0 = 1$, and $b_1 = 2$
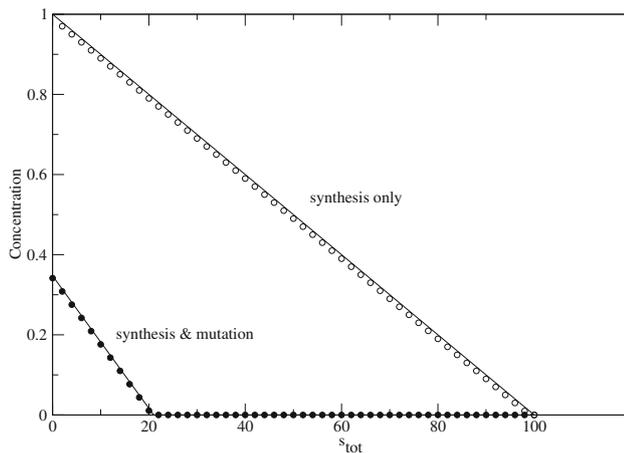
Fig. 4 Concentrations of the master sequence as a function of the total chemical synthesis rate, s_{tot}. Synthesis only—$u = 0$; Synthesis & mutation—$u = 0.01$. Points are numerical solutions of Eqs. 11 and 12. Solid lines are predictions of Eq. 14. In all cases, $L = 50$, $r = 100$, $b_0 = 1$, and $b_1 = 2$

selection against these two diversifying processes. Both these processes give rise to an error threshold in the master sequence landscape.

Discussion: Refining the Evolutionary Definition of Life

Having considered the results of the model above, the distinctions between the three cases in Table 1 should be clear. Selection is present in all cases, but selection acts directly on the physicochemical properties of the oligomers in the case of chemical evolution, and on the function encoded in the sequence in the case of biological evolution. In the chemical case, all short sequences have physicochemical properties, whereas in the biological case, only a small fraction of long sequences have an encoded function. For example, only a small fraction of random DNA sequences would encode a functional protein in a modern organism, and only a small fraction of random RNA sequences would encode a functional ribozyme in the RNA World. For a protein-coding gene, it is clear that selection depends on the ability of the protein to perform its role, rather than possible small differences in the DNA that encodes the protein. For a ribozyme in the RNA World, there might be selection on physicochemical properties as well as on function. For example, the structure of a ribozyme might affect its stability against hydrolysis. Nevertheless, the catalytic ability of the ribozyme would be the essential selective feature that distinguished it from other RNAs whose structure was comparable in stability to the ribozyme but which had no catalytic function.

Replication is present in both chemical and biological evolution in Table 1. For this reason, both chemical evolution and biological evolution are Darwinian, whereas selection without replication is not Darwinian and is not really evolution at all. Chemical synthesis must be present for selection without replication and for chemical evolution. Once the molecules involved become large enough, random chemical synthesis cannot occur and replication is the only way to make a second copy. The existence of molecules that can only be produced by a replication process is a defining feature of biological evolution, in the sense I am using it here, and I argue that this is an essential feature of life itself. Mutation is clearly absent for selection without replication, and clearly present in biological evolution. Mutation is probably present in chemical evolution as well, because replication processes are bound to have occasional errors, but mutation is not an essential feature in the chemical evolution case, because chemical synthesis also generates diversity on which selection can act.

The three cases in Table 1 also differ in time scale. Selection without replication and chemical evolution occur on a 'fast' chemical time scale comparable with the lifetime of a molecule. The equilibrium distribution of concentrations will be reached fairly rapidly. It is not necessary to wait for long times for mutations to generate unusual sequences, because all the sequences are being generated continually by random chemical synthesis. We are used to thinking of biological evolution as a slow process because the mutation rate is very small in modern organisms. But even if mutation rates were much larger in the early stages of life, it is clear that biological evolution is slow compared to chemical evolution. Biological evolution is occurring in a very large sequence space, and a long time is necessary to search this sequence space by the occurrence of random mutations. On the other hand, chemical evolution is occurring in a much smaller sequence space, and no waiting is necessary because all possible variants are in existence from the start. The length of the sequences is important: it is possible for all short oligomers to be present at the same time in a mixture of a reasonable volume, but for sequences of length comparable to known ribozymes and protein-coding genes, this is no longer possible. For the same reason, chemical evolution is repeatable but biological evolution is not. In a small sequence space, if we start with the same chemical mixture, the same evolutionary process will generate the same outcome. But if we have a large sequence space, the finite size of the population is relevant, and fixation of specific mutations may or may not occur—hence, the outcome will be different every time. As was stated in the introduction, chemical evolution will get close to the top of the first hill in a predictable way, but biological evolution will take a different path into the mountains every time.

The final category in Table 1 is open-endedness. This is another distinguishing feature between biological evolution and chemical evolution. The diversity of molecules generated by random chemical sequences may be large, but it is always finite, and barring some major change in the environment, the range of molecules going into a chemical evolution system will not change. As chemical evolution occurs within a predefined sequence space, it is not open-ended. In contrast, biological evolution occurs in a sequence space that is large enough that it is never exhausted. For as long as the Earth is still in existence, it will always be possible for Earth-based life to find new genes and new protein sequences that have never existed before.

Let us return to the question of defining life. If we operate with the definition that life is something 'capable of Darwinian evolution,' then systems showing chemical evolution or biological evolution would both be living. This leaves me feeling uncomfortable, because what would seem like life to me would be something showing biological evolution and not chemical evolution. Thus I would like to put the dividing line between life and non-life at the point between the chemical and biological evolution. If we refine the definition such that life is something 'capable of biological evolution,' rather than just Darwinian evolution, there may seem to be a hint of circularity at first, given that biology is, by definition, the study of life. However, in fact, I have already defined biological evolution and chemical evolution in a non-circular way. Biological evolution is open-ended, non-repeatable, requires mutation as a source of diversity, and works with complex entities that can only be synthesized by replicating an existing entity. Chemical evolution operates in a predefined space of possibilities, is repeatable, requires random chemical synthesis as a source of diversity, and works with entities that are simple enough to be synthesized from scratch as well as by replication. It appears to me that when previous authors used evolution as a defining feature of life, they were thinking of biological evolution all along, and not chemical evolution.

I acknowledge the point of Szostak (2012a) that researchers always want to put the boundary for life close to their own research area. As I am interested in the RNA world and the origin of replication, then I have always been happy that the evolutionary definition of life would consider an RNA world system to be alive. A polymerase ribozyme that was able to replicate itself and other RNA strands of arbitrary sequence would be able to support open-ended evolution. I would be satisfied to call this life, even if it lacked some features that some other definitions of life would require (e.g., cell membrane, metabolism, ability to respond to the environment, autopoiesis, etc.). Even though it may be a somewhat arbitrary decision whether to require additional features in the definition of

life, and whether to insist on biological evolution as a criterion for life, instead of just Darwinian evolution, I emphasize that the distinction between biological and chemical evolution is clear and non-arbitrary. I also hope that the concept of chemical evolution, as I have defined it in this paper, is a useful one.

The stage of chemical evolution is an important step on the path to life. The first replicating molecules must have arisen in a complex mixture of organic molecules, including molecules that are now common building blocks in biology, like ribonucleotides and amino acids, and also many other molecules that are chemically similar but did not end up being part of the biological repertoire of life as we know it. This is often called the problem of the 'prebiotic clutter' (Joyce 2002), and it is an important part of the problem of how an RNA World might have originated. It is sometimes argued that prebiotic synthesis of RNA was too difficult, and that an RNA world must have been preceded by a pre-RNA system, based on a simpler kind of polymer or other replicating system (Hud et al. 2013). It has also been argued that the properties of RNA have evolved to make it a good replicator, or a good template for non-enzymatic template-directed synthesis (Krishnamurthy 2014, 2015). One way to envisage this is that life might begin with a replicating pre-RNA system, the pre-RNA would evolve a means of synthesis of RNA, and the RNA would evolve a means of synthesis of DNA and proteins. In such a scenario, the pre-RNA would have to evolve specific sequences that replicated pre-RNA strands, and other specific sequences that catalyzed the synthesis of RNA. This would be a process of biological evolution, because it would involve searching for catalytic sequences in the pre-RNA sequence space. However, if we suppose that abiotic chemistry could generate RNA as a component of the prebiotic clutter, then the problem is understanding how RNA is selected from the mixture, not understanding the synthesis of RNA *per se*. We can envisage this occurring by chemical evolution, rather than biological evolution.

We have recently considered a model in which oligomers of nucleotides of various types can form by spontaneous random polymerization, and in which oligomers can act as templates favoring non-enzymatic ligation of shorter oligomers of the same kind (Tupper et al. 2017). We have shown that this can explain the emergence of three uniform properties observed in RNA: the use of nucleotides of a single chiral enantiomer, a single kind of sugar in the backbone, and a single kind of 3′-5′ bond. These properties arise because uniform oligomers (i.e., oligomers in which the monomers are all the same chirality, all the same kind of nucleotide, or all the same bond structure in the backbone) are better templates than mixed oligomers. This scenario operates at the level of oligomers, without requiring evolution of strands long enough to be specific

catalysts, i.e., it is chemical evolution, not biological evolution. One problem with scenarios involving pre-RNA is that it is difficult to see how a function can be transferred from one kind of polymer to another. Although a pre-RNA strand might hybridize with an RNA strand and pass on the sequence, the RNA strand is unlikely to have the function possessed by the pre-RNA, because function would depend on the details of the three-dimensional structure, and not just the base sequence. The transfer of information from RNA to DNA, which occurs at the end of many RNA World scenarios, does not suffer from this problem, because the DNA only acquires the genetic role, and the function remains in the RNA. If RNA emerges by chemical evolution at the oligomer level, as we suggest in this article and in more detail elsewhere (Tupper et al. 2017), a replicating RNA system can emerge directly from the prebiotic mixture, and it is not necessary to transfer functions between different types of polymer.

In previous papers modeling the RNA World (Wu and Higgs 2009, 2011, 2012; Shay et al. 2015) we have shown that many reaction systems involving autocatalytic replication have two different stable states, and we have referred to these as living and dead states. The distinction is that in the living state, there is a high concentration of ribozymes, and replication of RNA strands is catalyzed by ribozymes, whereas in the dead state, there is a very low concentration of ribozymes, and catalytic replication is negligible with respect to the slow chemical synthesis of random RNAs. The details of these earlier papers differ, but all deal specifically with ribozyme-catalyzed replication, which we call the k reaction in Shay et al. (2015); Wu and Higgs (2012). In contrast, in the chemical evolution model discussed here, there are no ribozymes and replication is non-enzymatic template-directed synthesis, which we call the r reaction in the previous papers. If replication is dependent on the existence of a ribozyme with a specific sequence, then this is already biological evolution by the terms of the present discussion, and what we called the living state previously is alive by the terms of the present paper too. The interesting question is whether a system with the r reaction but no k reaction can also be alive. Firstly, we note that the existence of the two separate stable states for some parameter values in the previous papers depends on the presence of the k reaction. Without the k reaction there is only one stable state for each set of parameters. Nevertheless, it is possible to consider this state as being a living one if the r rate is fast enough. When r is fast, sequences are maintained by heredity up to a finite error threshold (as with the error threshold case discussed above), whereas when r is slow, each sequence will produce less than one descendent on average, and so there is no heredity of sequences even if replication is perfectly accurate. By the terms of the present paper, a model involving the r reaction but no k reaction is a living state exhibiting biological evolution only if r is sufficiently fast and accurate to maintain functional sequences by heredity and if the diversity generated by mutation is more significant than the diversity generated by random synthesis (s reaction).

Several other computational models illustrate the stages of selection without replication, and chemical evolution. The model of Guttenberg et al. (2017) considers selection of subsets of molecules from a prebiotic mixture without including replication. This model includes interactions between the molecules involved, and hence has a more complex fitness landscape than the case I considered here; nevertheless, there is no replication, and no Darwinian evolution. On the other hand, the GARD model (Segré et al. 2000; Markovitch and Lancet 2014) describes evolution of populations of lipid assemblies. Diversity is generated by incorporation of molecules from a supply of chemically synthesized lipids, and although there is no sequence information in the GARD model, there is heredity of compositional information; hence this corresponds to what I am calling chemical evolution. Walker et al. (2012) have also studied a model of chemical evolution at the level of oligomer sequences. They use the term 'universal sequence replication' to describe the fact that all sequences are able to replicate by template-directed replication at equal rates in their model. The model of Walker et al. (2012) fits the definition of chemical evolution I have used here, because diversity is generated by spontaneous polymerization of new sequences, rather than by mutation of existing ones.

Several papers on the required criteria for evolution have some close parallels with the arguments I have given above. Szathmary and Maynard Smith (1997) discuss 'units of evolution,' which have the properties of multiplication (M), heredity (H), and variability (V). Griesemer (2000) adds the property of fitness differences (F), which is required for natural selection to operate in the usual Darwinian sense. A system with M, H, and V but no F would exhibit neutral evolution and random genetic drift without selection. Interestingly, Griesemer (2000) discusses the case of systems with V and F but no H or M, which have an opportunity for selection but do not have the capacity to respond to selection. This corresponds to the case I have called Selection without Replication in this paper. Szathmary and Maynard Smith (1997) comment on the distinction between replicators with limited and unlimited heredity. I have used the word open-ended here instead of unlimited, and I agree that this is a key property that distinguishes biological from chemical evolution. Both Griesemer (2000) and Szathmary and Maynard Smith (1997) draw attention to the fact that not all units of evolution are replicators. For example, a whole organism (even

a single-celled organism) is not a replicator, in the sense that each part of the offspring organism is not synthesized by replicating the corresponding part of the parent, but rather as a result of a developmental process. A strict replicator would be something like a nucleic acid, where there is direct transfer of information from one molecule (or could one conceive of some other structure?) to another. The distinction of reproducers and replicators is important in some contexts, although it does not seem critical in the context of this paper. Anything as complex as a whole cell or a multicellular organism is clearly alive by any definition. Here, I am discussing processes at the molecular level before we get to whole cells. Replication is the most appropriate word to discuss the copying and transfer of information between molecules.

Chessari and Luisi (2012) have considered the question of how prebiotic chemistry might have synthesized macromolecular sequences in many identical copies. They give an example in which peptides with high solubility are selected from a mixture of random peptides, which mostly have low solubility. This fits the case I call selection without replication. In reply, I have argued (Higgs 2012) that the reason many identical copies arise is because of replication. If non-enzymatic template-directed replication of nucleic acid sequences is occurring, then it will give rise to many copies of sequences and their complements that are very similar to one another. The fact that proteins cannot act as templates in the way that nucleic acids do is a primary reason for preferring nucleic acids-first scenarios for the origin of life over proteins-first scenarios. If there is no replication, then selection can narrow down the sequence space of random peptides (or nucleic acid oligomers) to a certain extent, as Chessari and Luisi (2012) show, but the diversity of the sequences that remain is still likely to be high. The model presented in this paper shows clearly why chemical evolution leads to a much stronger degree of selection than just selection without replication. Thus, while I agree that some degree of selection can occur even if there is no replication, I maintain that without replication it is possible to go only a very limited way along the path from chemistry to biology. The process of chemical evolution, as identified in this paper, is an important stage that gets us further along this path. But only when the stage of biological evolution is reached does it seem appropriate to say that we have truly arrived at life.

References

Baross JA (2007) Evolution: a defining feature of life. In: Baross JA, Sullivan WT (eds) Planets and life: the emerging science of astrobiology. Cambridge University Press, Cambridge, pp 213–221

Benner SA (2010) Defining life. Astrobiology 10:1021–1030

Bull JJ, Meyers LA, Lachmann M (2005) Quasispecies made simple. PLoS Comput Biol 1:e61

Chessari S, Luisi PL (2012) On evidence: the lack of evidence for prebiotic macromolecular synthesis. Orig Life Evol Biosph 42:412–415

Cleland CE, Chyba CF (2007) Does 'life' have a definition? In: Baross JA, Sullivan WT (eds) Planets and life: the emerging science of astrobiology. Cambridge University Press, Cambridge, pp 119–131

Da Silva L, Maurel MC, Deamer D (2015) Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. J Mol Evol 80:86–97

Damer B, Deamer D (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. Life 5:872–887

Deck C, Jauker M, Richert C (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. Nature Chem 3:603–608

Eigen M, McCaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92:6881–6891

Eigen M, McCaskill J, Schuster P (1989) The molecular quasi-species. Adv Chem Phys 75:149–263

Forsythe JG, Yu SS, Mamajanov I, Grover MA, Krishnamurthy R, Fernandez FM, Hud NV (2015) Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic earth. Angew Chem Ind Ed 54:9871–9875

Griesemer J (2000) The units of evolutionary transition. Selection 1:67–80

Guttenberg N, Virgo N, Packard N (2017) The 'selection first' pathway to life. Presented at Astrobiology Science Conference

Higgs PG (2012) Creating many identical copies of a macromolecular sequence. Orig Life Evol Biosph 42:416–419

Higgs PG (2016) The effect of limited diffusion and wet-dry cycling on reversible polymerization reactions: implications for prebiotic synthesis of nucleic acids. Life 6:24

Hud NV, Cafferty BJ, Krishnamurthy R, Williams LD (2013) The Origin of RNA and "My Grandfather's Axe". Chem and Biol 20:466–474

Joyce GF (1994) Foreword. In Deamer D, Fleischaker G (eds) Origins of Life: The Central Concepts, Jones and Bartlett, Boston, pp xi-xii

Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221

Kim YE, Higgs PG (2016) Co-operation between polymerases and nucleotide synthetases in the RNA world. PLoS Comput Biol 12:e1005161

Koshland DE (2002) The seven pillars of life. Science 295:2215–2216

Krishnamurthy R (2014) RNA as an emergent entity: an understanding gained through studying its non-functional alternatives. Synlett 25:1511–1517

Krishnamurthy R (2015) On the emergence of RNA. Israel J Chem 55:837–850

Leu K, Kervio E, Obermayer B, Turk-MacLeod R, Yuan C, Luevano J-M, Chen E, Gerland U, Richert C, Chen IA (2013) Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. J Am Chem Soc 135:354–366

Mamajanov I, MacDonald PJ, Ying J, Duncanson DM, Dowdy GR, Walker CA, Engelhart AE, Fernandez FM, Grover MA, Hud NV, Schork FJ (2014) Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules 47:1334–1343

Markovitch O, Lancet D (2014) Multispecies population dynamics of prebitic compositional assemblies. J Theor Biol 357:26–34

Mulkidjanian AY, Cherepanov DA, Galperin MY (2003) Survival of the fittest before the beginning of life: selection of the first oligonucleotide-like polymers by UV light. BMC Evol Biol 3:12

Pross A (2012) What is Life? How Chemistry Becomes Biology. Oxford University Press, Oxford

Ruiz-Mirazo K, Pereto J, Moreno A (2002) A universal definition of life: autonomy and open-ended evolution. Orig Life Evol Biosph 34:323–346

Sardanyés J, Elena SF, Solé RV (2008) Simple quasispecies models for the survival-of-the-flattest effect: the role of space. J Theor Biol 250:560–568

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

Shay JA, Huynh C, Higgs PG (2015) The origin and spread of a cooperative replicase in a prebiotic chemical system. J Theor Biol 364:249–259

Szathmary E, Maynard Smith J (1997) From replicators to reproducers: the first major transitions leading to life. J Theor Biol 187: 555–571

Szostak JW (2012a) Attempts to define life do not help to understand the origin of life. J Biomol Struct Dyn 29:599–600

Szostak JW (2012b) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3:2

Trifonov EN (2011) Vocabulary of definitions of life suggests a definition. J Biomol Struct Dyn 29:259–264

Tupper AS, Shi K, Higgs PG (2017) The role of templating in the emergence of RNA from the prebiotic chemical mixture. Life - submitted

Walker SI, Grover MA, Hud NV (2012) Universal sequence replication, reversible polymerization and early functional biopolymers: a model for the initiation of prebiotic sequence evolution. PLoS ONE 7(4):e34166

Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331–333

Wu M, Higgs PG (2009) Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world. J Mol Evol 69:541–554

Wu M, Higgs PG (2011) Comparison of the roles of nucleotide synthesis, polymerization, and recombination in the origin of autocatalytic sets of RNAs. Astrobiology 11:895–906

Wu M, Higgs PG (2012) The origin of life is a spatially-localized stochastic transition. Biol Direct 7:42