Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

Can synthetic biology shed light on the origin of life?

Christophe Malaterre

Université Paris 1 - Panthéon Sorbonne, IHPST, Université Libre de Bruxelles
christophe.malaterre@gmail.com

Abstract. It is a most commonly accepted hypothesis that life originated from inanimate matter, somehow being a synthetic product of organic aggregates, and as such, a result of some sort of prebiotic synthetic biology. In the past decades, the newly formed scientific discipline of synthetic biology has set ambitious goals by pursuing the complete design and production of genetic circuits, entire genomes or even whole organisms. In this paper, I argue that synthetic biology might also shed some novel and interesting perspectives on the question of the origin of life, and that, in addition, it might challenge our most commonly accepted definitions of life, thereby changing the ways we might think about life and its origin.

Keywords: synthetic biology, origin of life, definition of life, prebiotic chemistry

The origin of life remains one of the most puzzling unanswered questions of science. Several decades after Schroedinger's *What is life?* (1944), Haldane's *The origin of life* (1929) and Oparin's book of the same title (1924), newer insights have been gained into this ever-more challenging question, but nothing close to a definite answer. Today, synthetic biology presents itself as a novel scientific discipline, somehow at the borderline between biology and engineering, and whose objective is to engineer biological systems in radically novel ways: «synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature» (Serrano 2007: 1). By engineering living systems, by deconstructing and reconstructing novel forms of life, synthetic biology comes close to the frontier of living and non-living matter. For some, «synthetic biology […] attempts to recreate in unnatural chemical systems the emergent properties of living systems» (Benner & Sismour 2005: 533). Could synthetic biology thereby shed some new light on the old question of the origin of life? In this contribution, I will argue that this is indeed the case, even if these new insights might still be far from providing a full answer. To start with, I will propose a delineation of the discipline of synthetic biology, broadly construed, by looking at the types of problems it aims to solve. I will then consider the question of the origin of life so as to make explicit the types of problems that need solving from this perspective. This will make it possible to understand how, and to which extent, synthetic biology might contribute to solving this most fascinating puzzle. I will also argue that synthetic biology, because it is committed to

Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

creating novel forms of life, is likely to make another indirect contribution to this question, namely that of challenging our definition of life.

1. Delineating Synthetic Biology

Contemporary synthetic biology certainly has little to do with the synthetic biology of the late 19th century as advocated by Moriz Traube, Stéphane Leduc and the like (Keller 2002). Methods and tools have changed, no doubt. Yet the name remains. And probably also one common objective: the synthesis of complete living organisms from non-living matter. By chemically trying to reproduce the biological phenomenon of mitosis, Leduc, for instance, was pursuing the objective of synthesizing cellular artificial organisms. Such a synthesis remains an objective of contemporary synthetic biology, even if it has somehow been side-tracked by shorter-terms ones. The synthetic biology of today somehow encompasses, redefines and broadens biotechnology (Koide et al. 2009): some of its ultimate goals are to design and build complete bio-molecular and genetic systems that react to specific signals, process them, and produce desired outputs including informational signals, chemical compounds, molecular structures, energy, nutrients, thereby potentially improving food production, enhancing human health and preserving the environment. For some, «the term [synthetic biology] is used to describe the wholesale engineering of genetic circuits, entire genomes, and even organisms» (Lucentini 2006: 30). I will use this definition as a starting block to delineate synthetic biology, as it is done today, along three major dimensions that capture the objects manipulated by this discipline as well as the types of problems it pursues. I therefore propose to identify three types of synthetic biology.

Synthetic Biology as Engineering of Genetic Circuits (Type I synthetic biology)
By interacting with one another though the mediation of other molecular entities such as RNAs or proteins, genes can form complex webs of interactions that may be referred to as «genetic networks» or «genetic circuits». Like electronic circuits, genetic circuits may have positive or negative feedback loops, as well as linear or non-linear interactions. Unlike electronic circuits, genetic circuits are not wired as molecular interactions take place in aqueous solution. In any case, one of the peculiarities of such circuits is the way they behave with time, that is to say their dynamics, which is often time quite difficult to decipher from the networks' mere static description. There is obviously a long tradition in biology to study such networks as they appear in nature, and their behavior over time (e.g. Jacob and Monod 1961). Yet, the *de novo* design and production of genetic networks is a much more recent endeavor and a major focus of synthetic biology, the objective of which is no longer the discovery and explanation of naturally occurring genetic circuits, but the complete engineering of networks that behave according to plan. This is what I propose to label 'Type I synthetic biology'.

Examples include the design and production of synthetic oscillatory networks (e.g. Elowitz & Leibler 2000; Sprinzak & Elowitz 2005; Stricker et al. 2008, Tigges et al. 2009), of toggle and bi-stable switches (e.g. Gardner, Cantor & Collins 2000; Kim, White & Winfree 2006), or even

the exploration of stabilizing features such as auto-regulatory loops (e.g. Becskei & Serrano 2000). Sometimes also an existing genetic circuit is being 'rewired', that is to say made to respond to another molecular signal (e.g. Dueber et al. 2005). Generally, the desired function is realized by inserting specific genes into an existing cell, for instance a bacteria: the inserted genes are then expected to make this cell behave in a specific way that corresponds to the dynamics of the newly created genetic circuit. Take the example of an oscillatory network that make bacteria periodically synthesize fluorescent proteins, on top of all their regular activity: the artificially engineering network that is added to the genetic circuitry of the bacteria makes the bacteria perform the function of a visible clock (Elowitz & Leibler 2000).

The driving motivation behind such synthetic biology research is to be able to design a genetic circuit from scratch that will perform a specific task within a given living organism, and without interfering with the regular functioning of the organism, typically its self-sustaining metabolic activity and reproduction capacity. Hence the goal of engineering «biobricks» (e.g. Endy 2005) that can perform sets of given functions, and might be used as modules to give rise to even more complex behaviors, somehow similarly to the way electronic components, such as transistors, can be combined into an integrated circuit. There is therefore a strong analogy between this type of synthetic biology and electronics that is reinforced by the desire to built catalogs of standardized parts such as the Registry of Standardized Parts. In addition, there is a need for well-understood organisms within which such biobricks can be inserted and be made to work in the desired way. This is the role of «chassis organisms» that are flexible and versatile enough to express a wide variety of foreign genes, while somehow retaining their core functional integrity (e.g. Metzgar et al 2004).

Synthetic Biology as Engineering of Entire Genomes (Type II synthetic biology)

Synthetic biology also covers research work that does not concern specific genetic circuits but complete genomes. The objective in this case is no longer the design and implementation of a given function within an organism thanks to sets of biobricks, but the *de novo* synthesis of whole genomes that can then be made to work, typically by inserting them into a cell whose nucleus has been emptied of its original genetic material. I will refer to this type of work as 'Type II synthetic biology'. In such cases, a complete synthetic genome is produced *in vitro* from readily available industrial nucleic acids that are assembled into a sequence very similar to the sequence of wild-type organisms.

Examples range from the *de novo* chemical synthesis of the 7500 base-pair RNA genome of the smallpox virus (Cello, Paul & Wimmer 2002) to that of the 580000 base-pair DNA genome of *Mycoplasma genitalium* (Gibson et al. 2008). And these projects are also to be related to complete genome exchanges between species of organisms (e.g. Lartigue et al. 2007) and to genome simplification and redesign (e.g. Chan et al. 2005). In the case of *Mycoplasma genitalium* for instance, the synthesis involved five major steps: the chemical synthesis of a hundred oligonucleotides of some 6000 base-pairs each, the patching these oligonucleotides four by four into 24000 base-pair DNA strands by *in vitro* polymerase chain reaction (PCR) amplification and cloning into *E. Coli*, further patching into 72000 and 144000 base-pair strands

by the same techniques, and the final patching of these longer strands by transformation associated recombination (TAR) cloning in the yeast *S. cerevisiae*. The main challenges of such projects appear to consist in the assembly, manipulation and cloning of such large DNA sequences. Yet, what is lurking behind is the capability to chemically synthesize living organisms from scratch. Even if one might be reluctant to qualifying viruses as living organisms, and even if the synthesis of the genome of *Mycoplasma genitalium* is not the synthesis of the complete organism, it remains that the possibility of completely synthesis living systems appears closer than it was before. Incidentally, such research projects also open up the possibility of investigating the viability of organisms whose genomes have been significantly altered and potentially reduced, thereby of examining the conditions required for carrying out the essential vital functions, as would be the case with a minimally reduced *Mycoplasma genitalium* genome of some 350 genes (Hutchison et al. 1999).

Synthetic Biology as Engineering of Organisms (Type III synthetic biology)

In addition to devising genetic circuits and synthesizing complete genomes, synthetic biology also includes research that aims at engineering complete novel living systems from scratch. I will refer to this type of synthetic biology as 'Type III'. In this case, the objective is not really to copy nature and built up organisms that would be duplicates or modifications of naturally occurring ones, but rather to investigate the self-organizational properties of matter at the transition from inanimate matter to life. The goal that this synthetic biology pursues is to be able to *in vitro* assemble chemical systems that would be capable of metabolic activity and self-maintenance, of reproduction, and potentially of variation. In a way, such chemical systems would be partly alive, depending on the functions they would able to carry out.

Examples include research on self-assembly amphiphile molecules, lipids or fatty acids that can spontaneously self-assemble into micelles or bilayer vesicles, the properties of which might include growth, budding, division, fusion, catalysis of the formation of other vesicles (e.g. Bachman, Luisi & Lang 1992; Monnard & Deamer 2002), or even catalysis of the synthesis of RNA-like polymers, potentially showing how an early prebiotic coupling might have appeared between lipids and a form of primitive genetic polymer (Rajamani et al. 2008). Other examples of research in this area include the tentative assembly of protocells by having catalytic RNAs encapsulated into vesicles (e.g. Szostak, Bartel & Luisi 2001), and potentially starting from a cell-free protein synthesis (Noireaux et al. 2005), as well as the assembly of lipid aggregates that might be capable of integrating at the same time proto-genes and a proto-metabolism (e.g. Rasmussen et al. 2003).

In such instances, Type III synthetic biology appears to manipulate simpler, smaller molecular objects than in the previous cases (Type I and Type II). It also manipulate these objects with the objective of bridging nonliving and living matter in radically novel ways: it is neither with the objective of inserting biobricks or networks of biobricks into existing life forms, nor with the goal of chemically synthesizing complete genomes and organisms copied from current ones. Rather, Type III synthetic biology aims at investigating the chemical synthesis of the simplest possible protocells from readily available organic compounds. To do so, it relies on

two major activities: the design and engineering of such potential protocells or chemical systems on the one hand, and on the other the tentative implementation of these designs and processes into 'wet chemistry'.

2. Questions on the Origin of Life

Research on the origins of life aims at explaining the transition from inanimate matter to life that is supposed to have taken place in the early ages of our planet some four billion years ago. One would think that the rock record would provide unique evidence of what has happened at that time. Yet its interpretation does not give unambiguous answers. Indeed, Archean molecular fossils remain puzzling and their putative biological origin hard to establish. Nevertheless, despite unsettled controversies about the origin of such fossils and differing interpretations, it appears reasonable to believe that living systems already existed on the Earth some 3.5 billion years ago (Schopf 2006; Brasier et al. 2006). Whether fossil records can say more remains uncharted territory. In any case, this leaves a few hundred million years for life to appear on Earth. Now the question: how?

Contemporary research on the origins of life encompasses a fairly broad spectrum of approaches, from prebiotic chemistry (chemistry that is supposed to be compatible with the environmental conditions of the primitive Earth) to molecular biology and artificial cell engineering, including also modeling from theoretical biology as well as contributions from geology, micro-paleoontology or planetology that define historical and environmental constraints. Setting aside divine intervention and panspermia, the main challenge remains to bridge nonliving and living matter in conditions compatible with those that are estimated to be those of the early Earth (e.g. Kasting 1993, 2005). In other words, how can we explain that the simple molecules that were available in the cosmos and on the primitive Earth ended up generating quite complex functional sets of chemical compounds capable of life? I argue that this quest for explanation rests on three major components: (1) the identification of 'prebiotic molecular entities', (2) the specification of 'prebiotic evolutionary processes' that explain how the different molecular entities have been produced, and (3) the specification of 'functioning mechanisms' that explain how primitive living organisms carry out the different functions that would make them alive.

Prebiotic molecular entities

The successive sets of molecular entities that are thought to have existed from those available on the primitive Earth up to the appearance of fully functional protocells are precisely those that I propose to refer to as 'prebiotic molecular entities'. These objects are nothing but molecules and sets of molecules that happen to play a crucial role in the explanation of the appearance of life on Earth. In particular, four major classes of such life-relevant molecules can be defined, by order of chemical size, complexity and appearance.

Obviously, one has to start with the molecules that are supposed to have been readily available on the very early Earth or in space at that time. These *cosmic molecules* typically encompass simple compounds such as methane, ammonia or water.

From these, one should be able to explain the appearance of *prebiotic bricks*: these are the first organic molecules synthesized in prebiotic conditions. They include amino-acids, simple peptides, sugars, bases, nucleic acids, nucleotides, lipids and so on.

A second step should then explain the appearance of *functional molecules*, typically polymers and assemblies of *prebiotic bricks* into larger compounds that possess puzzling properties such as that of cross-catalysis. Such molecules would include oligo-peptides, short RNA strands, as well as potentially other genetic and catalytic polymers that might have preceded RNA.

A third step would then lead to the very first functional molecular organizations, some would say *protocells*. There supra-molecular assemblies are sets of interacting functional molecules: they might, for instance, consist in auto-catalytic networks potentially coupled with lipid vesicles; as such, they might be taken as the first signs of life.

In order to bridge nonliving and living matter, any explanation of the origin of life needs to appeal to each of these four classes of molecular entities. One of the requirements is also that such molecular entities should be compatible with the chemical and environmental conditions of the primitive Earth, for instance in terms of atmospheric composition, pH, temperature, water solvency and so forth.

Prebiotic evolutionary processes

The second component required for a satisfactory explanation of the origin of life consists in being able to put forward evolutionary processes that can account for the transitions from one type of molecular entities to the next, up to the very first protocells, and with the constraint that these processes ought to be totally abiotic and compatible with the environmental conditions of the early Earth. At least three major classes of evolutionary processes need to be put forward in order to gradually step from cosmic molecules to protocells.

The first one consists in sets of *prebiotic chemical processes* that would account for the synthesis of prebiotic bricks from readily available cosmic molecules. Prebiotic chemical processes of this kind typically consist in complex sets of chemical reactions, organized in networks of linked reactants and products, also determined by specific chemical conditions. Examples include Miller's synthesis of amino-acids by applying an electric discharge to a mixture of CH_4, NH_3, H_2O, and H_2 (Miller 1953; Bada & Lazcano 2003) or the synthesis of nucleic bases such as adenine from a mixture of cyanidric acid HCN and ammonia NH_3 (Oró & Kimball 1961; Ferris & Orgel 1966; Orgel 2004) or in eutectic solutions of HCN (Schwartz, Joosteenn & Voet 1982), among many others.

A second one consists in a *principle of prebiotic evolution* that might justify the appearance of more complex organic molecules with specific functional properties. In other words, such principle would explain how prebiotic functional molecules appeared from sets of prebiotic bricks. As the *in vitro* synthesis of artificial catalytic RNAs, also called 'ribozymes'

(Bartel & Szostak 1993; Johnston et al. 2001) indicates, such a principle of prebiotic evolution might consist in rounds of random chemical synthesis of assemblies of prebiotic bricks, selection of some of them for specific catalytic activities or for increased stability, followed by their multiplication either by cross-catalysis (Lifson 1997) or by differential molecular survival (de Duve 1987), potentially coupled with chemical variation processes.

A third explanatory component consists in *prebiotic self-organization processes.* Such processes aim at explaining the appearance of the first signs of organization, be they structures (e.g. vesicles) or functional systems (like self-maintained auto-catalytic networks), under prebiotic conditions. Examples of structural self-organization processes include the physical explanation of the spontaneous formation of vesicles in solutions containing amphiphile molecules under certain conditions of concentration, pH and temperature (Hargreaves & Deamer 1978; Bachman, Luisi & Lang 1992; Monnard & Deamer 2002); explanations have also been put forward to account for properties of such membranes such as the appearance of selective molecular exchanges (Sacerdote & Szostak 2005), of growth, budding, fusion or fission properties (Hanczyc & Szostak 2004), or of surface catalysis properties (Rajamani et al. 2008). Examples of functional self-organization include cross-catalytic nucleic acids (Yjivikua, Ballester & Rebek 1990), cross-catalytic RNAs (Sievers & von Kiedrowski 1994; Kim & Joyce 2004), and cross-catalytic sets of oligo-peptides (Lee et al. 1996; Yao et al. 1998; Ashkenasy et al. 2004). Even if they are not yet successfully realized *in vitro*, it is believed that such self-organization processes, potentially combined with one another, coupled and integrated, might lead to an explanation of the appearance of fully functional protocells, i.e. chemical systems somehow capable of self-maintenance and reproduction with variation.

Functioning mechanisms

A third component necessary to explain the origin of life is an account of how a protocell carries out the different properties that make it 'alive': only an account of the functioning mechanisms of such protocells will make it possible to fully understand what it takes to be alive. In other words, in addition to explaining how one bridges nonliving and living matter on longer time scales thanks to prebiotic evolutionary processes, it also appears necessary to understand how a living protocell functions on its own shorter time-scale.

Such functioning mechanisms might include complete models of protocells once these protocells have been successfully produced *in vitro*. These functioning mechanisms might then resemble protocell models like those of the 'chemoton' or of the 'autopoietic system' (Ganti 1971; Maturana & Varela 1973) that describe how protocells might be able to function. The chemoton, for instance, is a minimal cell model that is composed of three stoichiometrically coupled autocatalytic subsystems: a metabolic system that produces the molecular compounds required for the self-maintenance of the cell; a template replication system that duplicates the information required for metabolism and regulation; and a continuously renewed and growing membrane that encloses the two other subsystems. Such functioning mechanisms should make it possible to understand how a protocell carries out all the different properties that somehow make it alive.

Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

3. Synthetic Biology and the Questions on the Origin of Life

Can synthetic biology shed light on the origin of life? To answer this question, let us assess to which extent each one of the three types of synthetic biology might provide keys to each of the three challenges related to the question of the origin of life.

Prebiotic molecular entities seen from synthetic biology

In most of synthetic biology, there is little or no interest in the origin of the molecules and molecular assemblies that are used as experimental starting blocks in this discipline. What matters above all is the functioning of the biological systems that are assembled or synthesized; the origins of the parts, that is to say the molecular entities that compose such synthetic systems, are of little importance. As a matter of fact, these molecular entities can have multiple origins: they might be extracted from living organisms, or artificially synthesized thanks to man-made chemical processes. In synthetic biology, their origin is of little concern compared to the functional success or failure of the biological systems that they are components of: as far as molecular entities are concerned, 'anything goes' provided it works. As a consequence, the prebiotic relevance of these same molecular entities is also of little or no interest. In other words, if some molecules can do the job they were intended to, the fact that there might be synthesized by prebiotic chemical processes or not simply is not relevant. This is definitely so for some of the most significant projects carried out in synthetic biology. For instance, in their design and production of a genetic oscillatory network, Elowitz and Leibler cloned known nucleotide sequences via polymerase chain reaction (PCR) so as to build specific plasmids that were subsequently introduced into a given strain of *E. coli* (Elowitz & Leibler 2000). Obviously, the steps of this experimental process as well as the molecular entities upon which it rests have no prebiotic relevance: the process of PCR is totally artificial; the DNA strands of interest have specific sequences that come from extant living organisms; the final assemblies consist in living bacteria that have been modified by introduction of plasmids. None of these elements might be qualified as prebiotically relevant, and in this respect, Type I synthetic biology sheds no light on the origin of life.

In a similar fashion, when Venter, Hutchison and Smith synthesized the entire genome of *Mycoplasma genitalium*, they made use of modern technologies to assemble 'cassettes' of several thousand base-pairs each, of *in vitro* PCR amplification and cloning into *E. coli* for patching the cassettes together, and of transformation associated recombination (TAR) cloning in *S. cerevisiae* for final assembly of the genome. Obviously, in this example of Type II synthetic biology, none of these experimental steps might have spontaneously happened on the primitive Earth. On the other hand, one could argue that the sets of genes that would be identified as necessary and sufficient for a minimal living cell base on the same biochemistry as that of current living organisms (e.g. Hutchison et al. 1999), would bring valuable insight with regards to the question of the origin of life: of course, being already an extremely sophisticated mechanism of some 250-300 genes , such a minimal genome would not be the backbone of some

of the earliest forms of life; nevertheless, it might provide a most relevant evolutionary milestone in between the origin of life *per se* and the simplest and most primitive life forms we currently know of.

Type III synthetic biology is probably the area of synthetic biology that is most concerned with the prebiotic relevance of the molecular entities it manipulates. For instance, when Szostak, Bartel and Luisi propose to synthesize protocells by encapsulating catalytic RNAs into vesicles, one of their concerns is the potential relevance of their assemblies as hypothetical steps in the transition from non-living to living matter (Szostak, Bartel & Luisi 2001). Indeed, even if they carefully warn that «solutions found in the laboratory need not be chemically similar or even directly relevant to the actual molecular assemblies that led to the origin of life on Earth» (2001: 387), they place their work within the framework of the 'RNA world' hypothesis and conclude by saying that their «experimental possibilities could provide fascinating insights into what is now a complete black box of early evolution» (2001: 390). Their protocells include two types of molecular entities: on the one hand an RNA replicase, and on the other lipid molecules that can self-assemble spontaneously into vesicles. Both types might bear prebiotic relevance, the first ones within the hypothesis of the 'RNA world' according to which self-replicating RNA strands might have constituted the very first forms of life (Gilbert 1986), the second ones within the hypothesis of the 'lipid world' that emphasizes the appearance of amphiphile molecules on the primitive Earth, and their self-assembly into vesicles that can grow, bud, divide and fuse (Segré et al. 2001). In such cases therefore, synthetic biology might be able to shed some light on the origin of life, not so much by explaining the origin of the molecular entities that constitute the building blocks of its experimental research, but by giving an account of how such molecular entities and their self-assembly might have resulted in primitive life forms, while ensuring that these building blocks were compatible at least with some scenarios of prebiotic chemistry.

Evolutionary processes seen from synthetic biology

The identification of the prebiotic evolutionary processes that might have contributed to the gradual transition from non-living to living matter is probably also not the prime focus of synthetic biology as a whole, even though, as we will see, some work in this area might partly contribute to this aim, especially from Type III synthetic biology.

Indeed, it is probably fair to say that Type I and Type II synthetic biology shed no much light on such prebiotic evolutionary processes, and that the discovery of these processes is not one their goals. For instance, the research teams that work on engineering genetic circuits (Type I synthetic biology) do not provide much insight on any hypothetical prebiotic evolutionary process. When Elowitz and Leibler (2000), Stricker and collegues (2008), or Tigges and colleagues (2009) engineer synthetic genetic oscillatory networks, their research does not point to any such evolutionary process. Rather, their focus consists in designing and successfully implementing specific functions, in these cases genetic oscillators, independently of any constraint or research question that might have to do with the origin of life. The only 'evolutionary processes' that are of interest in such research projects are the dynamic evolutionary trajectories of the genetic systems they have engineered. For instance, the focus of

the project described by Stricker and colleagues (2008) consists in the design of a genetic oscillatory network and the study of its dynamics over time: of particular interest are the large-amplitude fluorescence oscillations that persist throughout observation runs, as well as the oscillatory period of the network that can be tuned by altering inducer levels, temperature and media source. This dynamic behavior is also at stake when comparisons are made with computational models: such comparisons might point to key design principle for constructing robust oscillators, for instance a time delay in the negative feedback loop, or the effect of positive feedback as means of increasing the robustness of the oscillations or implementing greater tunability (Stricker et al. 2008). Such dynamic behaviors however are characterized over rather short time-periods and do not entail drastic changes to the systems they describe. As such, they do not correspond to the longer-term evolutionary processes that aim at explaining the progressive transition from non-living to living matter and which, by doing so, would entail strong changes to the systems they might apply to (let us recall that such prebiotic evolutionary processes aim at explaining how organic molecules might have emerged from simpler widely available chemical compounds, how functional polymers might have appeared from sets of random monomers, or more generally how supra-molecular assemblies with life-like properties might have self-organized from mixtures of prebiotic molecules).

In a similar fashion, Type II synthetic biology (engineering of entire genomes) is also not much concerned with prebiotic evolutionary processes. When synthesizing genomes, Cello, Paul and Wimmer (2002) like the research team of Venter, Hutchison and Smith (Gibson et al. 2008) focus on the production of an end-result: a target genome. What matters therefore is whether the sequencing of the synthetic genome matches or not the target genome. To this end, any experimental process can be used to assemble nucleotides into oligonucleotides and subsequently into DNA strands and complete genomes: polymerase chain reaction (PCR) amplification, plasmid introduction and cloning into living bacteria, transformation associated recombination (TAR) cloning and so forth. Obviously, such processes bear no relevance to the prebiotic processes that might have occurred on the primitive Earth and led to the first forms of life.

The perspective is slightly different, I will argue, in the case of Type III synthetic biology (engineering of organisms). Indeed, when scientists investigate how lipids or fatty acids might spontaneously self-assemble into micelles or bilayer vesicles that are then capable of growth, division or fusion (e.g. Bachman, Luisi & Lang 1992; Monnard & Deamer 2002), they often have in mind the potential prebiotic relevance of such processes. For instance, for Bachman, Luisi and Lang, «because of the simple mechanisms underlying their spontaneous formation, aqueous micelles are plausible candidates for the first self-replicating bounded structures» (1992: 59). Similarly, Monnard and Deamer qualify the process of spontaneous self-assembly of bilayer vesicles as «steps towards the first cellular life» (2002: 1996). In such cases, the experimental processes that lead to the formation of micelles or vesicles are thought to be good candidates for prebiotic processes that would explain the appearance on the primitive Earth of similar structures. Also, when Deamer and his team show how RNA-like polymers can be synthesized non-enzymatically from mononucleotides in lipid environments and how such RNA-like polymers might end-up encapsulated within lipid vesicles, they claim that «this process provides

a laboratory model of an early stage of evolution toward an RNA World» (Rajamani et al. 2008). And when Szostak, Bartel and Luisi propose to synthesize protocells by encapsulating catalytic RNAs into vesicles, they claim that «by supplying a population of [proto]cells with random RNA sequences, one might observe the process of evolving complexity in real time» (Szostak, Bartel & Luisi 2001: 390). Examples of Type III synthetic biology can therefore be found that stress the relevance of their work in terms of processes that might shed light on some of the prebiotic evolutionary mechanisms.

Functioning mechanisms seen from synthetic biology

There seems to be two competing approaches in synthetic biology with regards to understanding the functioning mechanisms of the biological systems that are under investigation. On the one hand, some appear to favor a 'black box' approach: in this case, what matters is not so much the understanding of the mechanisms that produce a given result as this very result itself. As Madrigal puts it: «it's not what you learned, but what you made" (2008: 1). An example might be that of Venter and his team who harnessed a certain number of techniques to achieve their goal of artificially synthesizing the complete genome of *Mycoplasma genitalium* (Gibson et al. 2008). One of this techniques is to use the yeast *S. cerevisiae* for the final assembly of four huge DNA strands into the final genome of *Mycoplasma genitalium*: while this process works, little is known about the mechanisms that make it work and typically also why this type of assembly is not possible within the bacterium *E. coli* that was used in earlier steps of the same experiment to stitch together smaller segments of DNA. Similar examples can be found all across synthetic biology. For instance, one could argue that when a genetic oscillator is identified and works within a given organism, one does not fully understand why it does so, and why in contrast the same genetic oscillator would not work in another closely related organism (Serrano 2007). And in a similar fashion, when Deamer and his team show that RNA-like polymers can be synthesized non-enzymatically from mononucleotides in lipid environments and that such polymers can become encapsulated within lipid vesicles (Rajamani et al. 2008), they show that such things do indeed work the way they do, yet they do not claim to fully master the underlying explanations of these phenomena. Therefore, wherever such a 'black-box' attitude pertains, synthetic biology is unlikely to yield a detailed explanation of the functioning mechanisms of potential primitive living systems, and this is likely to be the case in any of the types of synthetic biology I have identified.

A competing view can however be found: the view that takes after Feynman and according to which "what I cannot create, I do not understand" (see Hawking 2001). In this case, understanding how things work the way they do is also a key component of the research agenda. Take for instance the projects that aim at assembling protocells from catalytic RNAs and lipid vesicles (e.g. Szostak, Bartel & Luisi 2001). Of course, the maim objective is to produce such protocells and make sure that they work, that is to say grow, divide, evolve and so forth. Nevertheless, a secondary objective is also to understand how these protocells work. Three elements may contribute to this understanding. First, a detailed knowledge of the parts that constitute the system that is to be built, in this case specific strands of RNA, and specific lipid

molecules. Second, knowledge of the assembly process and conditions that lead to this very system, for instance in terms of molecular concentration, pH, temperature but also in terms of chemical and thermodynamic processes that govern the self-assembly of the lipid molecules into micelles or vesicles of different shapes and the engulfment of RNA strands and so forth. And third, knowledge of how the different molecular entities dynamically interact and evolve over time once the protocells have been assembled, leading for instance to their growth, budding, fission or fusion. If such projects are indeed pursued and successfully realized, synthetic biology will indeed be able to provide very valuable insights on the functioning mechanisms of specific types of protocells, and thereby on the potential functioning mechanisms of hypothetical primitive life forms, provided, of course, that the molecular components of these protocells, and the experimental conditions that lead to their formation, be compatible with the chemistry of the prebiotic Earth.

4. Indirect Perspectives on Life

Overall it so appears that synthetic biology might be able to shed *some* light on the origin of life, and that contributions are somehow more likely to come not so much from Type I synthetic biology (engineering of genetic circuits) but from Type II (engineering of complete genomes) and most of all from Type III synthetic biology (engineering of organisms). In this respect, insights might be provided on some of the later stages of the emergence of the first living organisms, yet not so much on the earlier stages, that is to say the prebiotic synthesis of the necessary molecular compounds and the range of evolutionary processes that could explain the complete transition from non-living to living matter. In addition, and because synthetic biology is all about novel life forms, I argue that it can bring another major indirect contribution to the quest of the origin of life, namely that of expanding life and redefining its boundaries.

Novel living forms

If synthetic biology is about the 'engineering of biology' and the 'synthesis of novel functions' (Serrano 2007), by so doing, it is also, *de facto*, about the creation of novel life forms, that is to say of life forms that do not exist in nature and that are not the result of Darwinian evolution. When novel genetic circuits (Type I synthetic biology) are implemented in existing organisms, be they *E. coli, S. cerevisiae* or *Acinetobacter* to name but a few, when complete genomes are synthesized (Type II synthetic biology) and genomes of certain organisms replaced by those of others, or when novel organisms are the overt objective of research projects (Type III synthetic biology), novel living organisms are *de facto* created. Of course, the extent to which these organisms are 'novel' depends on the degree of their modification when compared with the most closely related natural organisms. In this respect, an *E. coli* that has been supplemented with a fluorescent 'repressilator' (Elowitz & Leibler 2000) will appear less novel than the much desired protocell (Szostak, Bartel & Luisi 2001). In any case, the point simply is that synthetic biology keeps creating novel life forms that might result from addition or modification of genetic

circuits, from genome simplification or more drastically from complete novel synthesis. As a consequence, synthetic biology is expanding the realm of known life.

This expansion can be pictured as going in three directions. The first is 'sideways': when current living organisms are modified by insertion of engineered genetic circuits (e.g. Elowitz & Leibler 2000), or by a rewiring of existing pathways (e.g. Dueber et al. 2004), the novel organisms are organisms that are somehow related to the natural organisms they once were, and their degree of complexity, that is to say the number of functions they might carry out, is also similar. The second is 'downward' in the sense of simpler complexity and ancestry: this is for instance the case when genome reductions are carried out to investigate the functional limits of minimal genome sets (e.g. Hutchison et al. 1999); such genome reductions expand the realm of life by offering insights into simpler organisms than those that are known to exist today, thereby into organisms that could potentially be related to ancestral DNA-based cellular life forms. The third direction that synthetic biology might expand known life is what might be called 'fast-downward', somehow moving the barrier of life beyond minimal DNA-based cells into even simpler protocells (e.g. Szostak, Bartel & Luisi 2001), and in this case, potentially offering glimpses into the very first forms of life that appeared on Earth.

Redefining life

By expanding the realm of life in these three directions, synthetic biology is also bound to contribute to a broader philosophical question, that of defining or 'redefining' life (many definitions of life have already been proposed, see for instance Sagan 1970; Luisi 1998; Palyi et al. 2002: 15-56; Popa 2004: 197-205). Indeed, by engineering novel life forms, synthetic biology offers insights into novel ways of delineating what is alive from what is not. In particular, the engineering of minimal organisms, be they simplified versions of current ones or radically novel systems, might challenge some of the traditional dichotomous definitions of life, typically those according to which any given system is either alive or not alive.

As a matter of fact, there is already disagreement about whether certain biochemical systems should qualify as alive or not. For instance, some argue that viruses should not count as living systems in so far as they lack metabolic activity (e.g. Luisi 1998; Ruiz-Mirazo, Pereto and Moreno 2004) whereas others argue they should, in particular when they form 'viral factories' (e.g. Raoult & Forterre 2008). Also, some argue that self-replicating strands of RNA as hypothesized in the 'RNA world' scenario (Gilbert 1986) may count as being alive since they are capable of replication and variation (e.g. Luisi 1998), whereas others argue the contrary for lack of metabolic activity and membrane enclosure (e.g. Shapiro 1986; Segré et al. 2001), and the debate goes on. In a similar fashion, one could question whether minimal protocells as those proposed by Szostak, Bartel and Luisi (2001) or Libchaber and his team (Noireaux et al. 2005), might more properly qualify as alive than, for instance, those pursued by Rasmussen and colleagues (Rasmussen et al. 2003).

Beyond a dispute about where the 'true' demarcation between living and non-living matter really is, the creation of novel border-line biochemical systems might also indicate that such a clear-cut demarcation simply does not exist. Rather than being an 'all-or-none' issue, the

transition from non-living to living matter might very well be more properly qualified as a 'more-or-less' question. In other words, the predicate 'to be alive' would be better captured in 'fuzzy logic' as a continuum along a zero-to-one scale rather than in first order 'classical logic' (Bruylants, Bartik and Reisse in press). Furthermore, synthetic biology could end up producing novel organisms that might be, not only 'more-or-less alive', but also 'more-or-less alive in different ways', that is to say along several dimensions or modes that might correspond to the functioning diversity of such living systems. For instance, a system could be more-or-less successful at reproducing itself, or more-or-less successful at metabolizing components or energy tokens from given sets of available compounds, or even more-or-less successful at individuating itself by means of more-or-less robust and sophisticated membranes.

As a result, by expanding the domain of known living organisms, synthetic biology might end-up producing different 'types' of living systems, and by the same token, might point to different scales and modes for biochemical systems to be alive. Life might then not be a matter of 'yes-or-no', but a matter of 'modes of being alive' and of 'degrees' along these modes. Even if synthetic biology is not there yet, this discipline may recast the way we think about life and help redefine this fundamental concept.

5. Conclusion

In this contribution, I proposed to review the extent to which synthetic biology might shed light on the origin of life. Synthetic biology has been mapped out as consisting of three main types of activities: engineering of genetic circuits (Type I), engineering of complete genomes (Type II), engineering of organisms (Type III). On the other hand, I have argued that the question of the origin of life can be split along three different dimensions: (1) the prebiotic relevance of the molecular entities that are manipulated, (2) the identification of prebiotic evolutionary processes and (3) the specification of functioning mechanisms. Overall it appears that synthetic biology is not much concerned with the origin of prebiotic organic compounds; at best some projects within Type III synthetic biology can be considered as taking into consideration the prebiotic relevance of the molecular components of their biochemical systems. It is also mainly Type III synthetic biology that might offer some glimpses about potential prebiotic evolutionary processes in so far as such processes might be derived from the type of *in vitro* self-assembly and molecular evolution that pertain to the projects Type III synthetic biology is interested in. Finally, some hypotheses about the functioning mechanisms of primitive life forms might be derived from work on minimal genomes within Type II synthetic biology, as well as on protocells like targeted by some Type III projects. Overall therefore, and even if this is not its prime focus, synthetic biology might indeed be able to shed some new light on the origin of life. Its major contribution however might be that of challenging our current dichotomous delineation of life from non-life: indeed, by synthesizing novel forms of life, synthetic biology is bound to produce an increasing number of 'border-line' biochemical systems that might show, if anything, that life is not a matter of 'yes-or-no' but a matter of 'modes of being alive' and of 'degrees' along these modes.

Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

Even if the direct contribution of synthetic biology to the question of the origin of life might appear somehow limited, radically changing the way we view life is no little achievement.

Acknowledgments. I am very grateful to the participants to the *Historical and Philosophical Foundations of Synthetic Biology* conference organized in Paris in April 2008, for very stimulating discussions, and in particular to Michel Morange and Jacques Reisse. Support from the *CNRS interdisciplinary program "Origines des planètes et de la vie"* and from the *Fondation Louis D.* of the *Institut de France* is gratefully acknowledged.

References

Ashkenasy G, Jagasia R, Yadav M, Ghadiri MR (2004) Design of a directed molecular network. Proceedings of the National Academy of Sciences USA 101(30): 10872-10877.

Bachmann PA, Luisi PL, Lang J (1992) Autocatalyic self-replicating micelles as models for prebiotic structures. Nature 357: 57-59.

Bada J, Lazcano A (2003) Prebiotic soup – revisiting the Miller experiment. Science 300: 745-746.

Bartel D, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261: 1411-1418.

Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. Nature 405: 590-593.

Benner SA, Sismour AM (2005) Synthetic biology. Nature Reviews Genetics 6: 533-543.

Brasier M, McLoughlin N, Green O, Wacey D (2006) A fresh look at the fossil evidence for early Archaean cellular life. Philosophical Transactions of the Royal Society B 361(1470): 887-902.

Bruylants G, Bartik K, Reisse J (in press) Is it useful to have a clear-cut definition of life? On the use of fuzzy logic in prebiotic chemistry. Origins of Life and Evolution of Biospheres.

Cello J, Paul A, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. Science 297: 1016–1018.

Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. Molecular Systems Biology 1: 0018.

de Duve C (1987) Selection by differential molecular survival: A possible mechanism of early chemical evolution. Proceedings of the National Academy of Sciences USA 84: 8253-8256.

Dueber JE, Yeh BJ, Bhattacharyya RP, Lim WA (2004) Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. Current Opinion in Structural Biology 14: 690-699.

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335-338.

Endy D (2005) Foundations for engineering biology. Nature 438: 449-453.

Ferris JP, Orgel LE (1966) An unusual photochemical rearrangement in the synthesis of adenine from hydrogen cyanide. Journal of the American Chemical Society 88: 1074.

Ganti T ([1971] 2003) The Principles of Life. Oxford: Oxford University Press.

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339-342.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA III, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319(5867): 1215–1220.

Gilbert W (1986) The RNA world. Nature 319: 618.

Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

Haldane JBS (1929) The origin of life. The Rationalist Annual 148: 3-10; republished in Bernal J.D. (1967) The Origin of Life. London: Weidenfeld and Nicolson: 242-249.

Hanczyc M, Szostak JW (2004) Replicating vesicles as models of primitive cell growth and division. Current Opinion in Chemical Biology 8(6): 660-664.

Hargreaves WR, Deamer DW (1978) Liposomes from ionic, single-Chain amphiphiles. Biochemistry 17: 3759-3768.

Hawking S (2001) The Universe in a Nutshell. London: Bantam Press.

Hutchison CA III, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal *Mycoplasma g*enome. Science 286: 2165-2169.

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3: 318-356.

Johnston W, Unrau P, Lawrence M, Glasner M, Bartel D (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292: 1319-1325.

Kasting JF (1993) Earth's early atmosphere. Science 259(5097): 920-926.

Kasting JF (2005) Methane and climate during the Precambrian era. Precambrian Research 137(3-4): 119-129.

Keller EF (2002) Making Sense of Life : Explaining Biological Development with Models, Metaphors, and Machines. Cambridge MA: Harvard University Press.

Kim DE, Joyce GF (2004) Cross-catalytic replication of an RNA ligase ribozyme. Chemistry and Biology 11(11): 1505-1512.

Kim J, White KS, Winfree E (2006) Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. Molecular Systems Biology: 68.

Koide T, Pang WL, Baliga NS (2009) The role of predictive modeling in rationally re-engineering biological systems. Nature Reviews Microbiology 7: 297-305.

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA III, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. Science 317: 632-638.

Lee D, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382: 525-528.

Lifson S (1997) On the crucial stages in the origin of animate matter. Journal of Molecular Evolution 44: 1-8.

Lucentini J (2006) Is this life?. The Scientist 20(1): 30-34.

Luisi PL (1998) About various definitions of Life. Origins of Life and Evolution of the Biosphere 28: 613-622.

Madrigal A (2008) Synthetic biology: it's not what you learned, but what you made. http://www.wired.com/wiredscience/2008/01/biology-moving/

Maturana H, Varela F (1973) Autopoiesis: The Organization of the Living. Dordrecht: Reidel.

Metzgar D, Bacher JM, Pezo V, Reader J, Döring V, Schimmel P, Marlière P, de Crécy-Lagard V (2004) *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. Nucleic Acids Research 32(19): 5780-5790.

Miller S (1953) A production of amino acids under possible primitive Earth conditions. Science 117: 528-529.

Monnard PA, Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. The Anatomical Record 268: 196–207.

Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A (2005), Toward an artificial cell based on gene expression in vesicles. Physical Biology 2: P1-P8.

Oparin AI (1924) The Origin of Life, trad. A. Synge, republished in Bernal JD (1967) The Origin of Life. London: Weidenfeld and Nicolson: 199-234.

Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Critical Reviews in Biochemistry and Molecular Biology 39(2): 99-123.

Oro J, Kimball AP (1961) Synthesis of purines under possible primitive earth conditions: Adenine from hydrogen cyanide. Archives of Biochemistry and Biophysics 94: 217-227.

Preprint. Final version published in *Biological Theory* (Fall 2009), 4(4): 357-367
http://www.mitpressjournals.org/doi/abs/10.1162/BIOT_a_00002

Palyi G, Zucchi C, Caglioti L, eds (2002) Fundamentals of Life. Paris: Elsevier.
Pohorille A, Deamer D (2002) Artificial cells: prospects for biotechnology. Trends in Biotechnology 20(3): 123-128.
Popa R (2004) Between Chance and Necessity: Searching for the Definition and Origin of Life. Heidelberg: Springer-Verlag.
Rajamani S, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Origins of Life and Evolution of Biospheres 38(1): 57-74.
Raoult D, Forterre P (2008) Redefining viruses: lessons from Mimivirus. Nature Reviews Microbiology 6: 315-319.
Rasmussen S, Chen L, Nilsson M, Abe S (2003) Bridging nonliving and living matter new browser window will open. Artificial Life 9: 269-316.
Ruiz-Mirazo K, Pereto J, Moreno A (2004) A universal definition of life: autonomy and open-ended evolution. Origins of Life and Evolution of Biospheres 34(3): 323-346.
Sacerdote MG, Szostak JW (2005) Semipermeable lipid bilayers exhibit diastereo-selectivity favoring ribose. Proceedings of the National Academy of Sciences USA 102: 6004–6008.
Sagan C ([1970] 1986) Life. Encyclopedia Britannica, 15th ed, Vol. 22. Chicago: Encyclopedia Britannica Inc.: 985-1002.
Schopf JW (2000) Solution to Darwin's dilemma: Discovery of the missing Precambrian record of life. Proceedings of the National Academy of Sciences USA 97(13): 6947–6953.
Schopf JW (2006) Fossil evidence of Archaean life. Philosophical Transactions of the Royal Society B 361(1470): 869-885.
Schrödinger E (1944) What Is Life? Mind and Matter. Cambridge: Cambridge University Press.
Schwartz AW, Joosteenn H, Voet AB (1982) Prebiotic adenine synthesis via HCN oligomerization in ice. Biosystems 15: 191-193.
Segré D, Ben-Eli D, Deamer DW, Lancet D (2001) The lipid world. Origins of Life and Evolution of the Biosphere 31: 119-145.
Serrano L (2007) Synthetic biology: promises and challenges. Molecular Systems Biology 3: 158-162.
Sievers D, Von Kiedrowski G (1994) Self-replication of complementary nucleotide oligomers. Nature 369: 221-224.
Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438: 443-448.
Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456: 516-519.
Szostak JW, Bartel D, Luisi PL (2001) Synthesizing Life. Nature 409: 387-390.
Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. Nature 457: 309-312.
Yao S, Ghosh I, Zutshi R, Chmielewski J (1998) Selective amplification by auto- and cross-catalysis in a replicating peptide system. Nature 396: 447-450.
Yjivikua T, Ballester P., Rebek J Jr (1990) A self-replicating system. Journal of the American Chemical Society 112: 1249-1250.