A synthetic approach to abiogenesis

James Attwater & Philipp Holliger

Synthetic biology seeks to probe fundamental aspects of biological form and function by construction (resynthesis) rather than deconstruction (analysis). Here we discuss how such an approach could be applied to assemble synthetic quasibiological systems able to replicate and evolve, illuminating universal properties of life and the search for its origins.

Four billion years of evolution on Earth has yielded a biosphere packed with exquisitely optimized molecular components that enable life to evolve and thrive. Traditionally, molecular biologists seek to analyze their function and interconnectedness in order to better understand the nature of living systems. Yet ultimately these components and systems are all representatives of a single interrelated biology deriving from the last universal common ancestor (LUCA), a breakthrough organism (or set of genes[1]) already reminiscent of modern prokaryotes. In the absence of the discovery of other biologies (on Earth or beyond[2]) not related to LUCA, it is challenging to establish universal principles and laws of biology. To paraphrase Carl Sagan: our biology, although amazingly diverse, is 'provincial'—in contrast, for example, to the laws of physics, whose generality can be observed throughout the cosmos. However, by going beyond the simple analysis and deconstruction of extant life and building 'new biologies' through modification, reconstruction and *de novo* construction, synthetic biology promises a fresh perspective and ultimately a better understanding of the unifying principles of living systems.

Our ability to modify existing biology has grown rapidly since the advent of genetic engineering. Biological tools allow the insertion, deletion and modification of genes at will, enabling the rewiring of pathways to generate new phenotypes, produce new metabolites[3] or carry out computation[4]. Advances in DNA solid-phase synthesis and assembly have culminated in the first synthesis of a genome of a unicellular organism[5]. Such technologies can explore the arrangement and context of genes in genomes on an unprecedented scale, opening up biotechnological opportunities through our growing mastery over the central dogma (**Fig. 1a**).

This naturally leads to questions as to how far we can stray beyond life's present molecular paradigms. New chemical functionalities such as unnatural amino acids and carbohydrates can be added to both unicellular and multicellular organisms[6,7], and in one case, one of the bases of the genome has been entirely replaced by an unnatural analog[8]. Ribosome engineering[9] or global recoding[10] allows reassignment of the genetic code, and expansion of the genetic alphabet itself might be possible[11]. The challenge of all these approaches in supporting stable augmentation is to fit the new functionalities into preexisting biological networks through either replacement or the establishment of orthogonal pathways or chemistries[12]. Such efforts promise to reveal the limits (if any) to life's tolerance of expansion into new chemical and informational space.

These approaches, however, remain defined by and embedded in preexisting biology and thus are likely to reflect its constraints. The underlying principles of our biology may be more clearly exposed by a fully synthetic approach: constructing simple cells comprising a limited number of essential components derived from biology. A number of strategies have recapitulated transcription, translation and metabolic activity within model membranes[13–16] but cannot yet regenerate the proteins or translation factors needed to support replication. To achieve this, a minimal cell would need to encapsulate a sufficient set of cellular replication, transcription and translation machinery components together with their encoding genome (**Fig. 1b**). This could define the core set of natural components required for propagation and expression of genetic information, establishing heredity and providing a stable chassis for bioengineering[17]. However, although clearly much simpler than any extant cellular organism, such a semisynthetic cell is still estimated to require between 100 and 150 genes expressed from a >100-kilobase genome. Is such complexity a prerequisite of life, or is it merely a consequence of using components firmly rooted in extant biology? In other words, is life more complicated than it needs to be?

An appreciation of the minimal requirements of life as a self-perpetuating phenomenon may emerge from even more radical strategies. Constructing systems exhibiting key traits of life, using synthetic components not found in biology, will challenge our preconceptions of what constitutes life. Abandoning the superbly refined molecular machinery of extant biology may seem an inauspicious step to take but may prove both instructive and liberating. Much of the central complexity of biology arises from the need to support and integrate three separate biopolymer systems: protein, RNA and DNA. If we can step back from this paradigm of cooperating biopolymers, simple, more streamlined forms of life might be feasible.

Philipp Holliger and James Attwater are at the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.
e-mail: ph1@mrc-lmb.cam.ac.uk

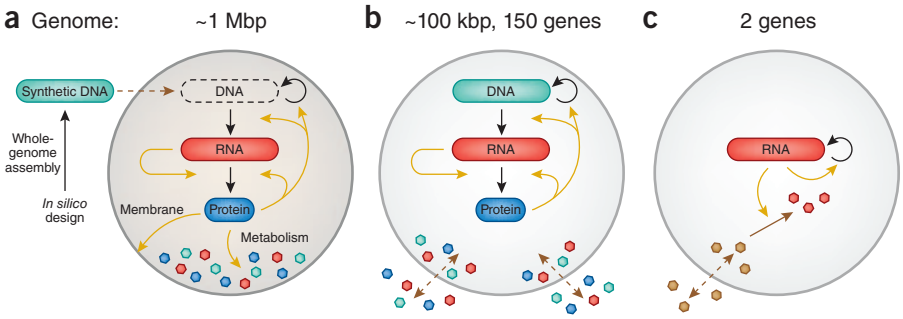

Figure 1 | Synthetic biological systems of increasing simplicity. (**a**) DNA as software: a bacterium reprogrammed by transformation with a synthetic genome[5]. Changes are transmitted through the central dogma, implementing a new phenotype by influencing informational systems, metabolism and the cell membrane (black arrows represent information transfer; orange arrows show catalysis). (**b**) A proposed minimal heterotrophic cell, lacking metabolism and comprising components dedicated solely to maintaining DNA replication, transcription and translation[17]. (**c**) A putative maximally simple RNA organism: a synthetic protocell founded on a single biopolymer (RNA) inhabiting a dynamic membranous vesicle[20]. An RNA replicase copies both itself and a metabolic ribozyme (synthase) that provides building blocks (by activating or trapping permeable precursors). The synthase is dispensable if activated building blocks are membrane permeable; alternatively, the replicase is dispensable if activated building blocks are capable of non-enzymatic RNA synthesis.

One might begin by asking which critical requirements these must satisfy. A unique property of life is its capacity for self-replication and open-ended improvement, through evolution enabled by a heritable molecular memory. In the search for plausible molecular strategies to implement this capability in a synthetic system, one conceptually attractive approach is to take inspiration from research into abiogenesis, the emergence of life on the early Earth, which provides useful constraints on both chemical and biophysical parameters[18,19]. The generation of a simple evolving system guided by these parameters in turn would have clear implications for the plausibility of proposed transitions from inanimate to animate matter at the origin of life; thus, advances in either field might drive progress in the other.

Extant biology is thought to have been preceded by a primordial biology lacking DNA and encoded proteins, where RNAs served as both genetic material and metabolic enzymes. Inspired by this 'RNA world' conjecture, a simple synthetic 'protocell' has been proposed comprising a self-replicating RNA system (**Fig. 1c**) that also promotes growth and division of a dynamic membrane envelope[20], with faster replicating species driving protocell division and random assortment heredity not dissimilar to L-form bacterial growth[21]. This concept provides a framework to address a fundamental parameter of life: the minimal level of complexity needed for a system to be able to evolve. To access and explore such a scenario, synthetic strategies must generate and integrate novel components: for instance, a replicating RNA genome.

Can self-replication be implemented using RNA as the sole informational component? Dramatic exponential growth has already been observed for a pair of engineered cross-catalytic RNA ligases wherein each ribozyme catalyzes the other's assembly from two pieces, allowing emergence of new ligase phenotypes through recombination[22]. Cooperative networks of self-assembly have also been described developing from pools of diversified variants of the *Azoarcus* self-splicing intron[23]. Such life-like 'growth' of RNAs raises hopes for RNA replication, but the evolutionary potential of these systems remains circumscribed by the sequence constraints of the prefabricated RNA components required for assembly.

The capacity of RNA systems to evolve freely could be enhanced by harnessing near-complete information transfer from a template via base-pairing, enabling RNA self-replication from monomer building blocks. Perhaps the most remarkable replicative activity of this kind is an RNA polymerase ribozyme evolved *in vitro* from a random RNA sequence pool[24]. Initially able to copy up to 14 nucleotides, it has undergone numerous cycles of engineering and evolution to yield variants able to synthesize other enzymatically active RNAs[25] or RNAs longer than the polymerase ribozyme itself (>200 nucleotides on a favorable template[26]), demonstrating the synthetic potential of such ribozymes.

One might ask if relatively complicated 'replicases' are needed to implement simple forms of self-replication and evolution. Work on replication of RNA using a purely non-enzymatic, chemical strategy was initiated in the 1970s[27] and has seen significant recent progress through the use of non-native linkage and base chemistries[28], stepwise removal of hydrolysis products[29] and discovery of cofactors enabling some template-directed RNA synthesis inside a model protocell[30]. Non-enzymatic replication systems, though, remain limited in rate and fidelity and, consequently, the genome size that they would support.

Furthermore, both non-enzymatic and enzymatic approaches must still overcome a set of formidable challenges before a robust replicating and evolving RNA system can be achieved[31]. Among the most pressing issues are the inhibitory effects of template secondary structures and the related need for strand separation. These require the definition of either enzymatic activities or physicochemical regimes able to unfold template RNA duplex structures while supporting RNA folding and base-pairing for synthesis. A counterintuitive yet elegant concept that has emerged from recent work is that nucleic acid replication (as well as assembly and division of protocellular membranes[32]) may be aided by a degree of 'helpful heterogeneity'[33] (**Fig. 2**). Chemical heterogeneity in the form of sporadic incorporation of transient 2′ modifications[34] or alternative RNA backbone linkages (2′-5′ vs. 3′-5′)[35] would destabilize product duplexes. Compositional heterogeneity in the substrate pool—moving beyond the paradigm of controlled monomer extensions to dynamic multisubstrate systems, as initially explored in the *Azoarcus* system[23]—could yield helpful emergent properties through oligomer interactions[33]. Indeed, such 'systems chemistry' approaches have driven fundamental advances in the prebiotic synthesis of RNA building blocks[36] and may be the key to enabling RNA self-replication.

If efficient RNA replication were to be realized in a compliant membrane, a simple protocellular system could be constructed from synthetic components that could exhibit many life-like properties such as growth, division, heredity and evolution[20]. Great progress has been made in the past decade in the study of such potential membranous protocells with regard to protocell

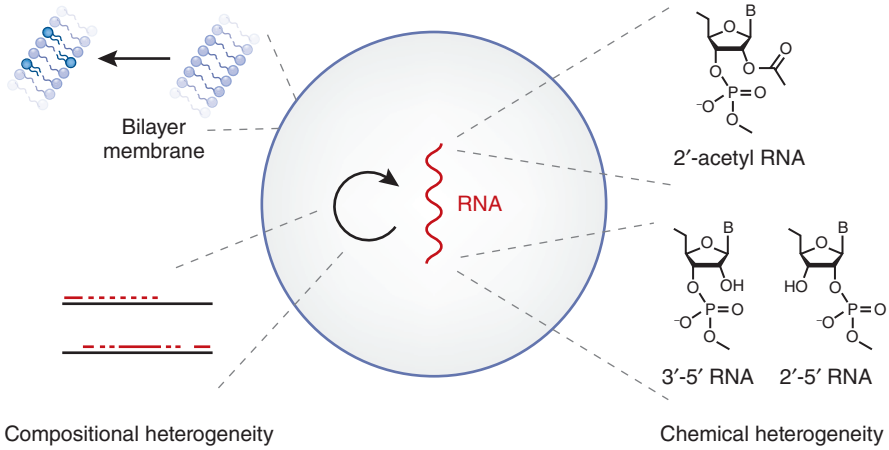

Figure 2 | Helpful heterogeneity in an RNA-based protocell. Clockwise from top left: phospholipid incorporation can facilitate fatty acid–based membrane growth and competition[33]. 2′-acetylation of ribonucleotide building blocks can aid RNA synthesis by enhancing regiospecificity and yields product duplexes that are easier to melt[34]. This is also promoted by sporadic incorporation of 2′-5′ linkages, and molecules with more such linkages could serve as better templates[35]. Initiation of RNA synthesis in both directions at multiple points on a template, instead of from a single primer, might yield faster, more robust replication[33].

growth and division[32], permeability, stability[37] and compatibility with non-enzymatic nucleic acid replication[30]. An RNA replicase in a protocell could also transcribe other ribozymes[25], facilitating a primitive metabolism—such as synthesis of RNA building blocks[38]—that in turn promotes RNA replication (**Fig. 1c**). Such a protocell would provide a powerful framework for evaluation and evolutionary optimization of system-level properties of all interacting and cooperating components. This would yield a test bed for fundamental questions of early molecular evolution, such as the role of lateral gene transfer[1] and the relative benefits of harnessing stable information storage offered by DNA versus the enhanced catalytic power provided by proteins.

Membranes, replication and metabolism are universal features of modern biology and may be sufficient to construct a simple synthetic cell as outlined above. But are these three canonical elements all necessary for life? Simpler systems are certainly conceivable. For example, a replicase in a protocell could dispense with metabolism if suitable permeable building blocks were provided externally to support a heterotrophic lifestyle. Without a metabolism to contain, do we even need a membrane? Colocalization of replicating components is a prerequisite for Darwinian evolution, by ensuring improved phenotypes selectively benefit their encoding genotype[20]. Compartmentalization within a membrane can keep related molecules together and discourages spread of inactive and parasitic molecules[39]. However, a range of alternative physical compartmentalization strategies (e.g., minerals[40], ice[41] and coacervates[42]) might fulfill this minimal 'container' function even in the absence of membranes. Indeed, although membranous protocells are attractive conceptually owing to their analogy with extant biology, on occasion some organisms can cope without membranes[43].

In principle, neither metabolism nor genetics nor indeed cellular structure need be constrained by natural chemistry: inorganic and supramolecular chemistries might generate cell-like structures and growth strategies radically different from those seen in biology[42,44,45]. Likewise, systems might be built using the opposite stereochemistry to that of today's life-forms (mirror life[2]) or using unnatural nucleic acids[11,28,46] for genetic information storage and unnatural amino acids for proteins or alternative coded polymers[18]. Such approaches diverge from historical aspects of the origins of life yet might reveal to what extent (if any) natural protein and nucleic acid chemistries possess exceptional properties, represent compromises between complexity and fidelity[47], or instead reflect prebiotic chemistry or evolutionary history.

Although this is a young field, some principles likely to be relevant for all forms of life have already begun to reveal themselves. These include the primacy of the genome and genotype[5], the plasticity of chemical makeup compatible with both essential nucleic acid and protein function[9,46], and emergent properties of simple protocellular membranes[32], replication systems[22,23] and replicases[48]. Indeed, *in vitro* evolution methodologies developed to isolate synthetic components, together with the advent of high-throughput sequencing technologies, promise a better understanding of evolution itself, offering unprecedented insight into adaptive fitness landscapes, routes to the emergence of functional phenotypes, connectedness of adaptive peaks, and the roles of recombination and drift[49,50].

The RNA world conjecture represents an attractive framework for the assembly of a synthetic cell owing to RNA's tractability to evolutionary refinement and its likely relevance to early biology. However, success in the assembly of a simple synthetic system with life-like properties such as heterotrophic growth, self-replication, heredity and evolution will be instrumental in revealing fundamental concepts of biology and abiogenesis, regardless of its precise molecular architecture.

ACKNOWLEDGMENTS

This work was supported by a Homerton College, Cambridge, Junior Research Fellowship (J.A.) and by the Medical Research Council (programme number U105178804).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Woese, C. *Proc. Natl. Acad. Sci. USA* **95**, 6854–6859 (1998).
2. Davies, P.C. *Philos. Trans. A Math. Phys. Eng. Sci.* **369**, 624–632 (2011).
3. Paddon, C.J. *et al. Nature* **496**, 528–532 (2013).
4. Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. *Nat. Nanotechnol.* **5**, 666–670 (2010).
5. Gibson, D.G. *et al. Science* **329**, 52–56 (2010).
6. Greiss, S. & Chin, J.W. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011).
7. Prescher, J.A., Dube, D.H. & Bertozzi, C.R. *Nature* **430**, 873–877 (2004).
8. Marlière, P. *et al. Angew. Chem. Int. Ed. Engl.* **50**, 7109–7114 (2011).
9. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J.W. *Nature* **464**, 441–444 (2010).
10. Lajoie, M.J. *et al. Science* **342**, 357–360 (2013).
11. Hirao, I., Kimoto, M. & Yamashige, R. *Acc. Chem. Res.* **45**, 2055–2065 (2012).
12. Rackham, O. & Chin, J.W. *Nat. Chem. Biol.* **1**, 159–166 (2005).
13. Noireaux, V., Maeda, Y.T. & Libchaber, A. *Proc. Natl. Acad. Sci. USA* **108**, 3473–3480 (2011).
14. Kurihara, K. *et al. Nat. Chem.* **3**, 775–781 (2011).
15. Nourian, Z., Roelofsen, W. & Danelon, C. *Angew. Chem. Int. Ed. Engl.* **51**, 3114–3118 (2012).
16. Stano, P., D'Aguanno, E., Bolz, J., Fahr, A. & Luisi, P.L. *Angew. Chem. Int. Ed. Engl.* **52**, 13397–13400 (2013).
17. Forster, A.C. & Church, G.M. *Mol. Syst. Biol.* **2**, 45 (2006).
18. Benner, S.A., Ricardo, A. & Carrigan, M.A. *Curr.*

Opin. Chem. Biol. **8**, 672–689 (2004).
19. Budin, I. & Szostak, J.W. *Annu. Rev. Biophys.* **39**, 245–263 (2010).
20. Szostak, J.W., Bartel, D.P. & Luisi, P.L. *Nature* **409**, 387–390 (2001).
21. Mercier, R., Kawai, Y. & Errington, J. *Cell* **152**, 997–1007 (2013).
22. Lincoln, T.A. & Joyce, G.F. *Science* **323**, 1229–1232 (2009).
The first demonstration of autonomous and exponential RNA self-copying.
23. Vaidya, N. *et al. Nature* **491**, 72–77 (2012).
24. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. *Science* **292**, 1319–1325 (2001).
The first demonstration of a general RNA-catalyzed templated replication activity.
25. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. *Science* **332**, 209–212 (2011).
26. Attwater, J., Wochner, A. & Holliger, P. *Nat. Chem.* **5**, 1011–1018 (2013).
An RNA polymerase ribozyme that can synthesize RNAs longer than itself.
27. Ninio, J. & Orgel, L.E. *J. Mol. Evol.* **12**, 91–99 (1978).
28. Zhang, S., Blain, J.C., Zielinska, D., Gryaznov, S.M. & Szostak, J.W. *Proc. Natl. Acad. Sci. USA* **110**, 17732–17737 (2013).
29. Deck, C., Jauker, M. & Richert, C. *Nat. Chem.* **3**, 603–608 (2011).
30. Adamala, K. & Szostak, J.W. *Science* **342**, 1098–1100 (2013).
Templated RNA synthesis within a protocell facilitated by citrate.
31. Szostak, J.W. *J. Syst. Chem.* **3**, 2 (2012).
32. Zhu, T.F. & Szostak, J.W. *J. Am. Chem. Soc.* **131**, 5705–5713 (2009).
33. Szostak, J.W. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2894–2901 (2011).
34. Bowler, F.R. *et al. Nat. Chem.* **5**, 383–389 (2013).
35. Engelhart, A.E., Powner, M.W. & Szostak, J.W. *Nat. Chem.* **5**, 390–394 (2013).
36. Powner, M.W., Gerland, B. & Sutherland, J.D. *Nature* **459**, 239–242 (2009).
A prebiotic synthesis of the pyrimidine ribonucleotides, providing key building blocks for the RNA world.
37. Mansy, S.S. & Szostak, J.W. *Proc. Natl. Acad. Sci. USA* **105**, 13351–13355 (2008).
38. Moretti, J.E. & Müller, U.F. *Nucleic Acids Res.* doi:10.1093/nar/gkt1405 (21 January 2014).
39. Ichihashi, N. *et al. Nat. Commun.* **4**, 2494 (2013).
40. Baaske, P. *et al. Proc. Natl. Acad. Sci. USA* **104**, 9346–9351 (2007).
41. Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A. & Holliger, P. *Nat. Commun.* **1**, 76 (2010).
42. Koga, S., Williams, D.S., Perriman, A.W. & Mann, S. *Nat. Chem.* **3**, 720–724 (2011).
43. Kim, G.H., Klotchkova, T.A. & Kang, Y.M. *J. Cell Sci.* **114**, 2009–2014 (2001).
44. Rasmussen, S. *et al. Science* **303**, 963–965 (2004).
45. Cooper, G.J. *et al. Angew. Chem. Int. Ed. Engl.* **50**, 10373–10376 (2011).
46. Pinheiro, V.B. *et al. Science* **336**, 341–344 (2012).
47. Szathmáry, E. *Nat. Rev. Genet.* **4**, 995–1001 (2003).
48. Attwater, J. *et al. Chem. Sci.* **4**, 2804–2814 (2013).
49. Ameta, S., Winz, M.L., Previti, C. & Jäschke, A. *Nucleic Acids Res.* **42**, 1303–1310 (2014).
50. Jiménez, J.I., Xulvi-Brunet, R., Campbell, G.W., Turk-MacLeod, R. & Chen, I.A. *Proc. Natl. Acad. Sci. USA* **110**, 14984–14989 (2013).