

On the origin of systems

Systems biology, synthetic biology and the origin of life

David Deamer

During the past decade, systems biology and synthetic biology have emerged as two new subdisciplines of biology. Although seemingly disparate, there are distinct overlaps that are worth exploring, particularly in terms of the origin of life. Systems biology is guided by the growing understanding that most cellular processes occur in the form of networks controlled by sensors, signals and effectors—properties that find analogies in electronic control systems. A few important life processes could even be characterized as digital, such as the genetic code, ribosome function and the on–off firing of action potentials in neurons. Systems biologists now try to embed these processes in computational models that could provide predictive insights into biological processes at the cellular level.

By contrast, other important cellular functions occur through spontaneous self-assembly processes that are not under precise regulatory control, such as the insertion of lipids and certain proteins into membranes. Most other regulatory signals are essentially diffusion processes that exert their effects by summation over time and space. This includes, for example, the diffusion of ions through channels to produce the resting and action potentials of individual neurons, the diffusion of neurotransmitters across the synaptic cleft and the diffusion of growth factors that govern the differentiation of cells into specific tissues within the developing embryo.

These regulatory functions are not in any sense digital, although they work well—indeed, life depends on them. The origin of life therefore clearly illustrates how a living system can emerge from a chaotic environment in the absence of anything that could be called digital control. Instead, the origin of life is better understood in terms of synthetic biology, which is currently defined

in engineering terms with a primary goal of developing a ‘toolkit’ that will make it possible to manipulate the genetic blueprint of living cells. The main point I wish to make here is that the definitions of systems biology and synthetic biology can be enlarged to encompass research on two fundamental questions faced by biologists today: how did life begin and can a laboratory version of cellular life be fabricated?

This is not an easy undertaking: even the simplest living cells, bacteria, contain thousands of interacting molecules organized into intricate networks called systems. New words such as genomes, proteomes, transcriptomes and metabolomes are being coined to describe these networks at various levels, and show where the understanding of living systems is headed—the suffix ‘ome’ has the sense of ‘whole’ or ‘total’. Pioneering researchers are now mapping out the interactions between the major protein components of living cells, and the resulting maps are called, of course, interactomes.

The word ‘system’ is derived from the Greek word *systema*, which describes a set of entities that interact in an orderly and organized fashion. Today, the word is widely applied to everything from political systems to solar systems. However, here I use the specific biological definition: in living cells, systems are complex sets of molecular components that interact in order to carry out a specific function and are regulated by a range of control mechanisms.

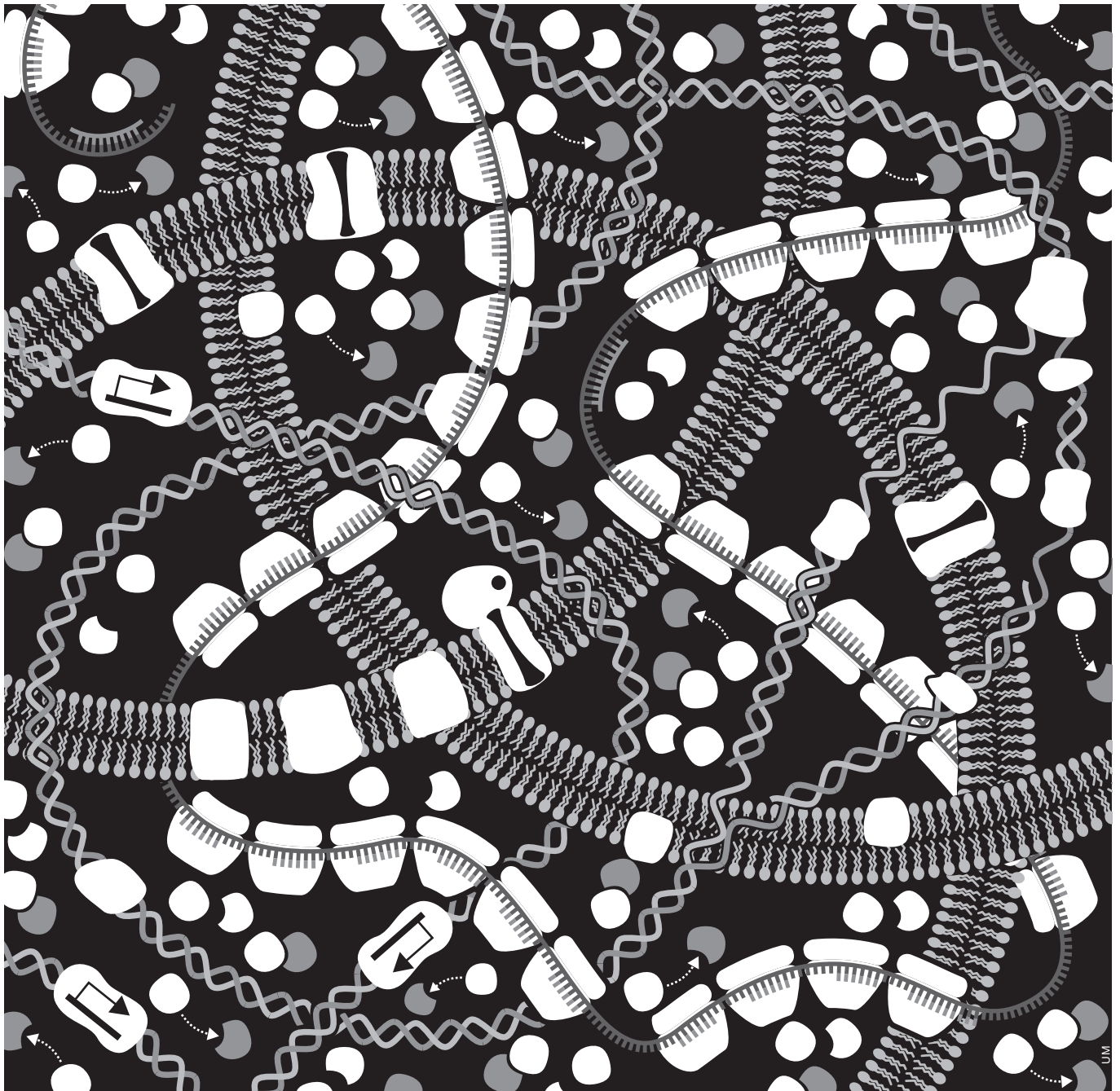
Four general systems are fundamental to all living organisms: a system of enzymes that catalyse and guide metabolic reactions, a system of enzymes and membranes that produce energy for the cell, a system of enzymes and ribosomes that synthesize proteins using the genetic information stored in nucleic acids, and a system of enzymes that

The origin of life therefore clearly illustrates how a living system can emerge from a chaotic environment in the absence of anything that could be called digital control

replicate the nucleic acids so that genetic information can be passed on to the next generation. There are many other cellular systems, of course, such as those responsible for transporting nutrients across membranes, cell division, sensory response and motility, but the four outlined above are probably the most fundamental to the definition of life. Some would add cell division and evolution to this list; however, many kinds of cells are alive—for instance, adult neurons—but never divide or evolve.

Primitive versions of these four systems must have comprised the first unit of life, presumably in the form of cellular compartments that maintain the essential interactions among encapsulated components. How such systems could spontaneously arise is a question that few have yet addressed in origins of life research, and so it is wide open for future investigations.

The concepts of systems and complexity go hand in hand. Systems are real and can be defined, but it is more difficult to measure complexity and reduce it to a numerical value. This limitation is particularly obvious in biological systems. For instance, there is no equation to calculate whether the brain of a bird is more complex than the brain of a bee, even though it might seem intuitively obvious. By contrast, it might be possible to use a comparison of calculators and computers as a guide for numerically estimating degrees of biological complexity. To describe how



much more complex a personal computer is than a calculator, one counts the number of components in each device, then compares the number of possible interactions between components—the structure of the logic circuits—and how fast functional interactions can occur. Perhaps something similar can be done with a living system, because one can count the number of different kinds of molecule in a living cell, the number of metabolic pathways and the number of regulatory interactions between

components of the pathways. It follows that one might be able to understand life in terms of the relative complexity of the systems that allow life to exist. One can then ask a fundamental question: what is the simplest system of molecules that can be called alive? In other words, what is a minimal cell with the fewest systems and the lowest level of complexity? The answer will provide insight into the systems of molecules in the prebiotic environment that eventually gave rise to life.

The emergence of life on the early Earth involved the self-assembly of certain organic compounds into increasingly complex microscopic structures. The term biocomplexity is useful for describing qualitatively how interacting molecular components gave rise to the first forms of life, which were further shaped by evolutionary processes that ultimately generated the contemporary biosphere. For instance, the sterile surface of the early Earth probably became gradually more complex with the addition of organic compounds by

either synthesis or accretion. In turn, the mixture of organic solutes became more complex over time as organic molecules underwent chemical evolution and began to assemble into molecular aggregates. Examples include the chemical synthesis of random polymers from suitable monomers such as amino acids, and the assembly of membranous vesicles from amphiphilic molecules. The nascent biosphere was even more complex at the time of the origin of life ~4 billion years ago, when one or more of these self-assembled structures happened to have properties that made it possible to use energy to accumulate simpler molecules from the environment and to assemble these into reproductions of the original structure. After life began, biocomplexity further increased as macromolecular structures became organized into systems within the cellular unit of life in order to catalyse metabolic pathways and to transmit information from one kind of molecule to another.

I have compared the relative complexity of calculators and computers in terms of the number of components and the number of their functional interactions. Now, this analogy can be used to get a sense of the relative structural and functional complexity of the systems of molecules within a living cell. The number and kind of interacting structural and functional units can be reasonably well estimated for a bacterial cell such as *Escherichia coli* (Table 1), which is found to have >25 million components.

This is an astonishing example of the power and complexity of biological systems. A personal computer has only ~1,000 components in a volume of perhaps 5l; however, life has found a way to pack millions of components into a volume of a few cubic micrometres. Not only that, but the functions of these components are tightly regulated, and a cell can also grow and reproduce itself, something that a robotic device with a computer brain can do only in science fiction.

This estimate of the number of components can be used to consider the number of interactions between components. New techniques have made it possible to analyse the encounters that take place between proteins. This in turn has given rise to the concept of interactomes, which is an extension of the concepts of genomes and proteomes. The interactions are defined in terms of the total number of protein species in a living cell, each of which is functionally linked by one or more interactions with other proteins in the cell.

Table 1 | Number of molecules in a single *Escherichia coli* cell

Molecular components	Number of molecules
Kinds of proteins	1,850 (mostly enzymes)
Total number of proteins	2.36 million
RNA in ribosomes	18,700
Transfer RNA	205,000
Messenger RNA	Variable depending on growth cycle
DNA	One circular double helix
Lipid	22 million (mostly in cell membrane)
Lipopolysaccharide	1.2 million
Peptidoglycan	One (forms cell wall)
Glycogen	4,360 (energy storage of cell)

An illustration of a typical interactome shows thousands of lines between small coloured balls, each connected by one or more lines to other neighbouring balls. The balls represent the individual proteins of a system, and the colours indicate the specific system of which it is a member. If a ball has only a single line, it means that only a single interaction has been established. Other balls might be at the centre of multiple lines, indicating that they are interacting with a dozen or more other proteins in the cell. Whenever a unit within a system has more than one interaction with other units, the complexity of the system increases exponentially. Interactomes are at the cutting edge of systems biology at the cellular level, and have now been reported for the yeast *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*.

One of the essential properties of life is that most functional interactions are regulated. As a result, living cells are maintained in a homeostatic steady state within certain set limits. The regulation is mostly done by feedback loops, defined by a sensor, a signal and an effector, with the sensor having an output that loops back to control the effector. A common example of feedback control is the thermostat that controls the temperature in a home. The sensor is a thermometer in the thermostat that measures the temperature of the air. The thermostat generates an electrical signal that closes a switch when the temperature falls below a certain set limit and opens the switch when the temperature is above that limit. When the switch is closed, the furnace turns on to heat the house, and the switch opens when the desired temperature

is reached. This means that the actual temperature oscillates a few degrees above and below the desired value, which is characteristic of feedback loops. Another familiar, but more complex, feedback loop is the cruise control that maintains the forward motion of an automobile at a certain speed.

Systems are real and can be defined; however, it is more difficult to measure complexity and reduce it to a numerical value

Both thermostats and cruise controls are simple on–off mechanisms. The furnace actively heats a house but does not cool it, and the cruise control actively accelerates a car but does not slow it. At the cellular level, most of the feedback loops involve braking, in which one or more enzymes in a system are inhibited by a product. In more complex living systems, physiological processes typically have both an accelerator and a brake to provide for more precise control. For instance, the level of glucose circulating in the blood is regulated by two hormones called insulin and glucagon. Insulin lowers the blood sugar level by increasing the rate at which glucose is transported into cells, whereas glucagon raises the blood sugar level by increasing the rate at which glucose is released from the liver.

I have made the point that the origin of life could also be considered within the context of synthetic biology as the origin of molecular systems having certain properties. On the prebiotic Earth, countless natural experiments—cellular compartments containing random mixes of polymers—took part in a

single massive experiment of combinatorial chemistry. Life began when a few of these membranous compartments happened to contain a specific mix of macromolecules that allowed them to grow by energy-driven polymerization and then to replicate the macromolecules. On the basis of what is known about biological systems today, can a hypothetical first living system be developed?

One can begin by enumerating the parts of such a system, then seeing how they might work together. The organic compounds that are assumed to be present as nutrients for the emergent system include a mixture of amino acids, a mixture of compounds resembling nucleotides and a mixture of amphiphiles able to self-assemble into membranous compartments. By using energy available in the environment, a metabolic subsystem within the compartments chemically changes nutrients into usable monomers, then activates the latter so that they are able to polymerize by a catalysed reaction. Two kinds of polymer are required: one species catalyses the polymerization of the activated monomers into a second species of macromolecule having a genetic function, and the second species directs the replication of the first species. The polymers are maintained in a membranous compartment that grows by the spontaneous addition of amphiphilic molecules. The boundary of the compartment allows small monomers into the internal volume, but retains any polymers that might be synthesized from the monomers.

Now, one can define the control points. There must be feedback control between the polymerization reaction and the replication reaction, otherwise too much of one or the other macromolecule will be synthesized. Another feedback system regulates growth of the membrane and growth of the polymers. A third feedback system regulates the synthesis of activated monomers. If these three processes are not regulated and synchronized with each other, the organism will inevitably grow too fast or lag behind, and fall apart.

No one has yet attempted to develop an experimental system that incorporates all of the above components and controls, so one can only speculate about how control systems might have developed in early forms of life. One obvious point in the network offers a place to start. Small nutrient molecules must get across the membrane boundary, and so the rate at which this happens will clearly control the overall process of growth.

Sidebar A | Further reading

Synthetic biology

For an overview of contemporary research in synthetic biology, see the website of the Drew Endy laboratory at Stanford University, California, USA (http://openwetware.org/wiki/Endy_Lab)

Origin and early evolution of life

See Sullivan & Baross (2007)

Source of prebiotic organic compounds

See Chyba & Sagan (1992)

Self-assembly processes: prebiotic chemistry

See Walde (2005)

Self-assembly processes: protocells

See Rasmussen *et al* (2008)

Self-assembly processes: interactomes

See Gandhi *et al* (2006)

I propose that the first control system in the origin of life involved an interaction of internal macromolecules with the membrane boundary. The interaction represents the signal of the feedback loop, and the effector is the mechanism that governs the permeability of the bilayer to small molecules. As internal macromolecules were synthesized during growth, the internal concentration of small monomeric molecules would be used up and growth would slow. However, if the macromolecules disturbed the bilayer in such a way that permeability was increased, this would allow more small molecules to enter and support further growth, representing a positive-feedback loop. The opposing negative feedback would occur if the disturbed bilayer could add amphiphilic molecules more rapidly, thereby reducing the rate of inward transport by stabilizing the membrane. This primitive regulatory mechanism is hypothetical, of course; however, it could be a starting point for research on how control systems were established in the first forms of life.

The origin of life is best understood as occurring within a hierarchy of increasingly complex systems of molecules governed by chemical and physical laws. For life to begin, the core catalysts and information carriers were necessarily part of a system that included a container, a transporter and the ability to capture chemical energy from its surroundings. Beyond the simple fact that systems are organized into linear and branched networks, they also are controlled by regulatory processes that involve feedback loops. Another property of biological systems is that specific protein components of the system undergo constant interactions.

These can now be established by biochemical and genetic methods, resulting in a map of the interactions that is referred to as an interactome. A challenge for origins of life research is to understand the minimal interactome that will allow life to begin as a functional system of compartments and large molecules capable of catalysis and replication, together with feedback loops that regulate their functions. Finally, if life is really understood in terms of systems biology, it should also be possible to fabricate artificial versions of life in the laboratory. This is where systems biology meets synthetic biology, and the result will surely change the way in which life on Earth is viewed, not to mention the real possibility that life has arisen on other planets by similar processes.

See Sidebar A for further reading.

REFERENCES

- Chyba C, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**: 125–132
- Gandhi TKB *et al* (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* **38**: 285–293
- Rasmussen S, Bedau MA, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF (eds) (2008) *Bridging Nonliving and Living Matter*. Boston, MA, USA: MIT Press
- Sullivan WT, Baross J (eds) (2007) *Planets and Life: The Emerging Science of Astrobiology*. Cambridge, UK: Cambridge University Press
- Walde P (ed) (2005) *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models. Topics in Current Chemistry 259*. Berlin: Springer



David Deamer is at the Department of Biomolecular Engineering, University of California, Santa Cruz, California, USA.
E-mail: deamer@soe.ucsc.edu

doi:10.1038/embor.2009.117