Driving forces in the origins of life.

Permalink
https://escholarship.org/uc/item/6mc952bj

Journal
Open biology, 11(2)

ISSN
2046-2441

Authors
Dill, KA
Agozzino, L

Publication Date
2021-02-03

DOI
10.1098/rsob.200324

Peer reviewed
Driving forces in the origins of life

K. A. Dill and L. Agozzino

Subject Area:
biochemistry/biophysics/genetics/
structural biology

Keywords:
origin of life, survival of the fittest, autocatalysis, fitness

Author for correspondence:
K. A. Dill
e-mail: dill@laufcenter.org

What were the physico-chemical forces that drove the origins of life? We discuss four major prebiotic ‘discoveries’: persistent sampling of chemical reaction space; sequence-encodable foldable catalysts; assembly of functional pathways; and encapsulation and heritability. We describe how a ‘proteins-first’ world gives plausible mechanisms. We note the importance of hydrophobic and polar compositions of matter in these advances.

1. What forces drove the origins of biology?

How did life begin? What drove the transition, more than 3 billion years ago, from physical chemistry to biology (Pchem 2Bio)? We seek the origins of biology’s forces of sustainability and persistent innovation. To be clear, this is not the same as seeking mechanisms of self-replication. Here is a metaphor. Consider an imaginary self-replicating mouse trap. This device is outfitted so that it can reach into a bin of metal and wood parts and assemble a copy of itself. But what happens when the bin runs out of parts? Self-replication, by itself, is not a sustaining force. Nor does it explain how it’s self-replication abilities arose from physico-chemical stochastic processes in the first place. Here, we are interested in the causative actions that could have driven physical chemistry (Pchem) to discover biology (Bio), with its unique abilities to propagate in ways that are resourceful, adaptive and persistent.

First, an overview of related research. The origins field has a long history, dating back, at least, to Darwin’s idea in 1871 of a ‘warm little pond’ [1,2] and then of a ‘primordial soup’ [3,4]. Many are studies of prebiotic chemistry, including prominent early ones by Urey [5] and Miller [6] in the early 1950s, and Orgel in 1968, [7], which have sought molecules and conditions that were plausible on the early earth and their possible reactions. Others have focused on what biological precursor molecules might have come from space, for example, in the Murchison and other meteorites [8]. There have been speculations on chicken-and-egg ‘what-came-first’ problems. Metabolism first [9]? Proteins and functionality? Nucleic acids and information? An RNA world first [10,11]? A world of encapsulated replicating RNAs [11]? A lipid world [12]? What interactions might have led to the genetic code [13–15]? For general reviews, see [16–19]. And since there are no definitive experiments yet, much work is speculation using theory and modelling, such as of primitive replication, in Eigen’s quasi-species models [20,21], the GARD model [22] and others [23–25]. The present work is aimed in a different direction: to seek plausible origins of biology’s drive towards persistence and long-term innovation. Here are our starting points.

2. Our premises about prebiotic chemistry

Life originated on Earth. We assume that origins happened on Earth. While amino acids and simple organics are found on meteorites from space, we are interested in more life-like complexity, which is unlikely to have come from panspermia (i.e. originating in space before coming to Earth [26,27]).
Both living and non-living matter have dynamical processes in biological evolution are different than in physical chemistry.

3. Distinguishing between life and non-life

To scrutinize the transition, we first ask what distinguishes living from non-living systems. Living systems metabolize (i.e. take in resources), grow and duplicate. But some non-living systems also do these things. Candle flames can take in fuel, oxidize it, grow bigger fires and light new fires. Oil droplets can grow and duplicate. Related processes occur in self-replicating computer codes or in human institutions that compete for resources. For our purposes, a living system:

- is ‘wet’ (i.e. made of molecules);
- has units of agency, such as cells;
- metabolizes, taking in matter and energy;
- grows and replicates independently; and
- has lineages and heritable variation.

This definition excludes fires and oil drops (no heritability), viruses (no independent growth) and self-replicating computer codes or human institutions (not wet). It includes plasmodia, which are multi-nucleated but bounded. Others have defined life to include Darwinian selection [29] or computer codes such as Artificial Life [30–32].

3.1. The dynamics is different: persistence versus relaxation to equilibria

The biological dynamics we consider is evolutionary change. Both living and non-living matter have dynamical behaviours that entail stochastic searching of degrees of freedom (DOF), sampled by the actions of random forces and driven toward macrostates that can be predicted by a variational principle. But the details are very different; see table 1. For one thing, biology’s evolutionary tendencies are not a drive toward equilibrium. For more than 3 billion years, life has been in a stable nonequilibrium. Survival of the fittest (SOF) is a principle of long-term sustained dynamics, not equilibrium. For another thing, different dynamical processes dominate biological evolution versus chemistry. In Pchem, atoms and molecules search positions, velocities and conformations, sampled by random thermal forces. In biology, cells search different growth rates, sampled by random changes in monomer sequences in proteins and nucleic acids. And, the nature of disorder is different; their corresponding entropies do not even have the same units. How did Pchem come upon, and enable, biology’s processes and forces?

4. Survival of the fittest is a persistence principle

4.1. Evolution is sustained by positive feedback

What is the nature of SOF as a dynamical variational principle? Much of textbook physical chemistry describes systems subject to negative feedback: they are stable, subject to restoring forces, having states of equilibria to which they return after perturbation. By contrast, the centerpiece of biology’s evolutionary dynamics—SOF—is a principle of persistence, i.e. a sustained capacity for a particular type of positive feedback or what, in simpler chemical systems, would be called autocatalysis [33,34]. One example of autocatalysis is \( A + B \rightarrow 2B \). Another is a forest fire, where burning is cooperative among fuel elements that are at high density. Here, we refer to this positive feedback as bootstrapping, taken from the expression: lifting yourself up by your bootstraps. Biological evolution is sustained by SOF. What physical chemistry begat that principle?

4.2. The SOF principle, described in general terms

Suppose you have some operational device that has persistent input and output; for example, a cell, a machine or a company. You can tweak the inner workings of the device to alter its productivity. Fitness is a measure of how effectively (by some metric) the input resources are converted to output. A company can tweak its process to make more product from less resource. In this context, survival measures the amount of...
input resource the device takes in. If a company makes product more efficiently, then the company gains a bigger market. This gives it access to even more resources, allowing it to outcompete other such companies for resources. In SOF, there is a feedback loop: advantageous actions are rewarded by new capacity to take more actions. The better the performance, the greater the access to even more resources, creating a virtuous cycle of improvement and dominance over the resource pool.

4.3. Biology implements SOF in a specific, clever and convoluted way

The pawn that the hand of evolution moves is not the cell, but cell lineages. The metric of survival is the population of a cell lineage relative to others. The ‘knob’ that evolution turns to change that population is the growth rates of cells. Evolution ‘turns that knob’ by random mutations of proteins (and also recombination, lateral gene transfer, plasmids and gene duplication today). A cell’s growth rate is largely determined by its rate of protein production. Hence, here is how the SOF positive feedback loop is implemented in biology: a change such as a mutation increases a cell’s growth rate, causing the cell to duplicate faster, increasing the population of that cell’s lineage of ancestors relative to other lineages. This gives that cell’s lineage greater access to resources in the next generation. This positive feedback principle leads to some of biology’s most marvellous features, described below.

4.4. SOF acts by advantages, not by averages

Positive feedback processes can be controlled by small fluctuations. Compare to a river. A river’s flow properties are dominated by the largest and deepest channels, not the small tributaries, because the typical observables are averages, which are dominated by the biggest flows. By contrast, a key feature of positive feedback is that it can become dominated by the very smallest metaphorical tributaries, provided that those flows are somehow advantageous to the process [35,36]. It allows for ratcheting of advantage. It raises up winners: the few and the good can bootstrap up to dominate over the many and the average. If a single individual cell happens to be well fit for its environment, it grows rapidly. Its lineage can come to dominate the population. This positive feedback manifests as adaptability, innovation, improved match for environments and apparent goal directedness. We note that once an improbability process such as SOF is discovered, there are no limits to the marvellous intricacies it can lead to.

In the Pchem₂Bio transition, how did stochastic physical dynamics ‘discover’ stochastic biological dynamics? How did polymer chain sequences emerge as the searchable degrees of freedom? What random processes searched and sampled them? And what autocatalytic chemical or physical process could have bootstrapped its way to becoming cellular SOF? Below are four important ‘discoveries’ that Pchem made to reach biology, three of which are positive-feedback bootstrap processes.

5. Major discoveries in Pchem₂Bio

Below we list key discoveries made by physico-chemical processes on the road to biology. (1) Coupling drivers to chemistry. Non-equilibria (NEQ) sampled and drove chemical reactions and molecular processes. (2) Proteins as mobile programmable catalysts. Monomer sequences in proteins became searchable degrees of freedom, giving programming catalysts and molecular machines. (3) Assembling biochemical pathways. Functionally similar reactions associated into spatially localized pathways. (4) Creating individuals and lineages. Encapsulation into cells allowed for a distinction of SELF and competition. A genetic code, memory and heritability allowed for survival of the fittest. Our proposition here is that they needn’t have happened all at once. A first step of (1)–(3) would require only proteins. Even today, the existence of horizontal gene transfer implies that linear heritability is not an obligatory early step.

5.1. Dynamical processes can sample and drive molecular processes

Was there some special aspect of dynamics in general that created or enabled life [38]? We consider two roles of dynamics, per se, in origins: (i) as a mixer and random driver of chemical reactions, and (ii) through specific mechanisms that can drive particular relevant innovations.

5.1.1. Forces of disorder can explore chemical reaction space

In general, NEQ per se, is not a driver towards order. The sun, winds, waves and volcanoes drive randomness, mixing and disorder. Even so, disordering can give predictable outcomes.
For example, thermal forces that randomize the velocities of gas atoms lead to the ideal gas law, a precise relationship. But the randomization that matters on the road to biology is over a very different space than that of gas velocities; it is over the space of chemical reactions. Early earth dynamics could drive different molecules together randomly, sometimes reacting with each other, sometimes catalysed by surfaces, and continually producing product wherever there are continual inputs of appropriate energy and matter [39]. And although organic-molecule reaction space is very large [40], the space of today’s biochemical reactions is relatively small and simple [41] (figure 1), hence ancestral versions of them must have been similar [42,43]. There is no reason to believe there was a specific goal-driven force to select out those reactions that would become biochemistry. But geophysical mixing dynamics could at least have searched and sampled some simple reactions, which, through particular dynamical mechanisms described below, could have led to biology.

5.1.2. Far from equilibrium drivers toward persistence and innovation; not just restoring forces

Prigogine and colleagues popularized the view that biology-like spatio-temporal patterning—in chemical oscillators like the Belousov–Zhabotinsky reaction, for example—can arise from NEQ processes [44,45]. Non-equilibrium forces are special; they differ in at least two ways from equilibrium forces. First, non-equilibrium forces are zero at equilibrium. For example, while bar magnets have a static pull, electromagnets have no pull when the electric field is turned off. In Fick’s law, particles stop flowing when there is no concentration gradient. Also, hurricanes operate only when the underlying thermal conditions drive them. Non-equilibrium structures and organization are sustained by non-equilibrium inputs of matter and energy. Second, NEQ differs by push versus pull, i.e. by supply versus demand. Near-equilibrium processes are pulled toward equilibria, a tendency towards a state of minimum free energy. They are governed by the second law of thermodynamics. By contrast, FFE is pushed by input energy and matter that are out of equilibrium. Imagine a flood that carves a new river bed; it does not aim to go any particular place, it just pushes water, which flows through a path of least resistance. Evolution does not steadily march towards predetermined goals [46], like second law equilibrium restoring processes do. The NEQ realm is broad and innovative, through particular mechanisms, many of which are not yet fully understood, and two of which are described below.

5.2. The foldcat bootstrap: protein foldamers as programmable catalysts

5.2.1. The importance of proteins as programmable catalysts

Biology would be impossible without its machines and catalysts, protein enzymes. On the one hand, Orgel and others argued that there is severe difficulty in achieving biochemistry-like reactions with only prebiotically available catalysts [18,47]. On the other hand, important recent experiments have achieved significant reactions using prebiotically available catalysts [48–52]. Even so, chemistry in the prebiotic era was hostile to chemical innovation. The catalysts for those reactions were mineral surfaces or metal ions, many of which were spatially immobile (not accessible to substrates), capable only of catalysing limited reactions, each only under limited and different conditions, and only where substrates were sufficiently concentrated.

Biology is more innovative than prebiotic chemistry. Biology’s catalysts—mostly proteins—are mobile and can go where the substrates are; can be altered to work in different environments, including just in water, or in membranes; can operate at whatever ambient temperature is needed for the organism; and are readily tunable to any degree that is needed to fit within whole reaction pathways and cycles. Protein catalysts could be called programmable, in the sense that their extraordinarily wide range of capabilities can be controlled by just a simple single kind of process, namely mutating amino acid sequences.
This importance of this breakthrough—of discovering programmable catalysts—can be illuminated with a metaphor. Compare a fictitious prebiotic organic chemist ‘demon’ (i.e. working with random processes) to a corresponding biology demon. The Ochem demon cannot create a complex multistep process without many different specific catalysts, each chosen for different conditions, some with intermediate products produced in particular ways. This is sufficiently challenging that academic organic chemists can publish research papers about them! By contrast, the Bio demon just spins some dials on a big dashboard, picking a reaction type, picking the solvent and temperature conditions, picking the desired acceleration and linking multiple reactions together by stringing together pathways of multiple enzymes. Of course, much trial and error is needed for both demons. The early discovery, by physical chemical processes, of catalysts that are explorable and optimizable through random changes of sequences of monomers in a polymer chain is arguably one of the most important steps made during the origins of life because of its capacity for rapid trial-and-error invention of complex chemical processes and diverse functionalities, all brought together under single conditions. Our term ‘programmability’ here does not refer to heritability or a genetic encoding; rather, it is simply intended to express that changing an amino acid sequence can change a molecule’s functional capability.

Here, we describe a mechanism for the origins of proteins as programmable catalysts, controllable through their amino acid sequences. We call it the foldcat bootstrap mechanism. It is an autocatalytic process by which short peptides become elongated, sequence selective and develop primitive versions of the today’s protein enzymes and machines. It addresses the following question: what physical process might drive particular subpopulations of chain sequences to self-amplify at the expense of other subpopulations? In this mechanism, random peptides fold and help catalyse the elongation of others in a primitive ribosome-like way. In this way, short-chain peptides grow longer and more plentiful, growing protein mass.

There are many plausible prebiotic processes that can polymerize individual amino acids into peptides, or nucleic acids into short DNA or RNA molecules. But these polymerizations all suffer from the so-called Flory problem, namely that the resultant chains are mostly very short (≈2-8-mers); longer chains are exponentially less probable (figure 2a).

Known prebiotic polymerizations also do not address how the randomness in polymerized sequences leads to ordered and informational sequences, and how such processes became autocatalytic, leading to stable steady states of production of long-chain informational-sequence polymers.

The foldamer catalyst hypothesis [53] offers an explanation. In this hypothesis, chains are polymerized using two types of monomers: hydrophobic (H) and polar (P), as modern-day proteins are. When H and P monomers are linked into long chains, like today’s proteins, different HP sequences spontaneously fold in water to different ‘native’ structures [55] (figure 2b). The structures are driven by the oil–water principle that hydrophobic monomers seek to minimize contact with water.

According to this hypothesis, some short-chain HP sequences will compactify in aqueous solutions into structures that have some exposure of their hydrophobic residues on their surface. Call those hydrophobic surfaces landing pads, and those chains catalysts. If a second short peptide chain lands its own H monomers on the sticky hydrophobic surface of the first one, a catalyst, then the second chain will undergo an enhanced rate of covalent elongation because of the sticky localization of the chain and an H monomer to be added (figure 3a).

The HP foldcat mechanism gives the three properties sought above. First, exact enumeration in the HP lattice model shows that this mechanism leads to amplified populations of longer chains (figure 3b). It also leads to reduced subspace of HP sequences, initiating a process of converting random sequences to informational polymers. And, it generates an autocatalytic set that continues propagating other sequences in that set; see figure 4. The following paragraphs give arguments for the plausibility of this mechanism.

5.2.2. Evidence for folding in HP polymers

Today’s protein folding code is dominated by the binary HP patterning in the sequence [53,55]. This is proven in experiments where proteins that have been massively mutated, in ways that preserve only a given HP pattern, still fold to their appropriate native structures [56–59]. Moreover, HP foldability does not even require that the polymer backbone be a peptide. Peptoid chains (polymers of N-substituted glycines) can also fold into HP-sequence-dominated structures [60]. Further evidence for the early role of hydrophobicity is that ancestral proteomes are more hydrophobic [61,62].
5.2.3. Catalysis and binding are ubiquitous in peptides and proteins

Functional peptides are ubiquitous in today’s biology (the *Handbook of Biologically Active Peptides* [63] is more than 2000 pages long!). Short proteins function as hormones, signalling molecules, growth factors, venoms, antibiotics and more. Enzymatic activities are known in chains even as short as dipeptides [64–66], and including ATP binding activity [67]. 7-mer amyloid peptides can catalyse reactions and auto-catalyse their own formation [68,69]. So, amyloid structures might have been prebiotic catalysts [70]. Moreover, proteins are highly promiscuous binders. For example, half the yeast proteome has protein–protein binding affinities stronger than 1 kcal mol\(^{-1}\) [71]. And regarding whether simple peptides could help elongate others, we note that non-ribosomal peptide extension and chemical modification is done on peptide scaffolds [72]. Furthermore, once a protein has a binding site, that site often readily mutates to become an active site [73].

5.2.4. Perspectives on the foldcat mechanism

Here, we note some caveats and suggest some experimental tests. First, we are not aware of any evidence yet for simple peptides folding and catalysing chain elongation in other peptides. But we are also not aware of any tests of it. The value, as we see it, in the present theoretical speculation, is in giving a mechanism that is sufficiently detailed that it can be tested through experiments.

Second, what are the limitations of the model? While figures 2–4 illustrate the foldcat mechanism with graphic simplifications—to two dimensions, to a code that is only binary (H and P), and to conformations that are confined to a lattice—extensive studies with larger code alphabets and in 3D [56,57] have shown that this simple model recapitulates important behaviours of real proteins. The 2D HP model has its equivalent of secondary and tertiary structures; the thermodynamic behaviours of short chains in 2D resembles longer chains in 3D because of the dominance of surface-to-volume ratios and hydrophobic interactions; and as noted above, the sequence-to-structure code degeneracy in real proteins is known from experiments to be close to binary [55]. For understanding the nature of both conformational and sequence spaces, microscopic atomic details often matter much less than an ability to do coarse-grained enumerations, which is readily done in simple models. At present, it is not possible to draw unbiased inferences about the nature of sequence space with more atomistically detailed models than HP lattices. And, while the mechanism illustrated here adds only H monomers, driven only by hydrophobicity, this is just an illustration because any broader distribution of amino acids that would have been used in primitive proteins would have likely harnessed additional interactions as well.

Third, while the example above of the Foldcat mechanism illustrates ‘inventing’ primitive ribosomes, it also follows that there would be broad random coverage of sequence-structure space, so other (weak) protein machines would be generated too. We infer that proteins and functional diversity could have been a first step in Pchem2Bio, followed by encapsulation, heritability and memory.

5.3. The catpath mechanism assembles functional pathways

Imagine the prebiotic stew above, of small molecules and catalysts. How could that stew have been divided up and encapsulated into individual cells? Physico-chemical actions would only aggregate them together randomly into vesicles...
5.3.1. In the catpath mechanism, function dictates structure

The catpath process contrasts with two standard situations: (1) two independent particles will simply diffuse away from each other, or (2) two particles with mutual affinity will come together and bind each other. The catpath attraction is not based on a binding affinity, A–B; rather, it is an example of function driving structure*: processes that have a common mediator come together. Unlike simple A–B binding or droplets. That would not lead to biology. Each cell needs assemblies of reactions that form functional pathways, cycles and hypercycles (i.e. interlinked cycles)? What would cause different enzymes with related functions to come together in space, like bucket brigades, in which the output of one reaction is close enough to become the input of another reaction? Here, we describe such a process.

The catpath mechanism is a non-equilibrium reaction-diffusion mechanism that brings reactions together in space based on their related functionalities [74]. In this process, a catalyst A, fixed at a given location, draws a catalytic B in its spatial neighbourhood; the effective attraction between the catalysts (cats) is mediated by a common substrate or product, on which they both act.

Figure 5 (top) shows the catpath mechanism. The square-box objects in the figure are catalysts, such as enzymes. The letter inside each catalyst box is an identifier of the reaction it catalyses. The catalysts are mobile and free to diffuse, towards or away from other such catalysts. The circular objects are the substrates and products, typically small molecules. Inside the circles are numbers that identify or label them. The arrow in each icon shows the direction of catalysis, from substrate to product.

In the catpath mechanism, a mobile catalyst molecule B, which converts 1s to 2s, diffuses toward the position of a catalyst molecule A, which converts 1s to 2s; see figure 5 (bottom). This attraction is a reaction-diffusion process [75]. Because the A cats are continuously supplied with 1s, so they continuously produce 2s. These product 2s will diffuse away from the parent A at some rate, but will concentrate around A for certain relative speeds. The B cats have a binding affinity for their substrates, 2s in this case. So Bs will diffuse toward the 2s, thus toward the A cats. In this way, A and B cats are attracted to each other, mediated by a small molecule substrate/product in common.

5.3.2. The catpath mechanism could assemble transducers and machines

Critically important in biology is energy transduction coupled to chemical reactions. Often one domain of a protein performs an energetically uphill reaction, driven by an energetically downhill reaction in another domain, typically by converting ATP to ADP or by flows of protons or ions down their concentration gradients. Without such coupling, it would be impossible to metabolize food to synthesize biomolecules, to run molecular motors, chaperones, ribosomes or other machines, to perform signalling, or to synthesize biomolecules such as proteins and nucleic acids. Today’s processes are well understood through the physical chemistry of binding events coupled to conformational changes in proteins; see figure 7. These processes, such as in ATPases and GTPases, entail multiple protein domains that are bound together into a complex: one domain performs the uphill action and the other domain converts to ATP to ADP to ‘pay the energetic price’ for the uphill step. A crucial ‘discovery’ during origins of life must have been the combining of two protein domains in such transduction processes [79]. The innovation this allowed, on the road from chemistry to biology, was the ability to power energy cycles and biochemical circuits. Protein domains may have been driven to assemble by the catpath mechanism, but there are no studies yet as far as we know.

5.3.3. The catpath mechanism can drive SOF-like bootstrapping

Where does the SOF principle come from? Might its prebiotic precursor have been some simple autocatalytic chemical cycle, such as shown in figure 8? Here is what we are seeking to explain. If a chemical process is changed in a way that causes it to run faster (in biological language, a mutation increases the fitness), how does that lead the process to recruit more resources for itself (more survival)? For the autocatalytic cycle in figure 8, the catpath Mechanism can link survival to fitness. Catalyst A converts substrate 1s and substrate 2s to product 3s. Catalyst B converts substrate 3s and substrate 4s to product 1s. The two catalysts are linked as a cycle: the head of each reaction is the foot of the other. The substrates and products, 1s and 3s, are common to the two reactions. Mutating catalyst A to a better one, A’ increases the cycle speed. Because of the catpath force, the greater cycle speed drives greater attraction to B of A’ relative to A. The machine A’B is more stable and persistent than the machine AB, hence is the more reliable consumer of new resources.
5.4. The heritability bootstrap: replicating the ‘self’

Achieving SOF requires informational linkage between how fast a cell replicates, on the one hand, and the size of the population of its lineage, on the other hand. This requires, first, that living systems come in discrete units, i.e. individual agents such as a cell (call it ‘the self’). This compartmentalization is enforced by lipid bilayers and related boundaries. The cell must contain information about how it achieves its growth speed. And it also requires a mechanism for transmitting information down generations, from parents to daughters. Below, we just make brief points about the physical chemistry of encapsulation and heritability.

5.4.1. Encapsulation distinguishes individuals and lineages, enabling competition

In the origins of life, compartmentalization could have arisen from oil droplets or vesicles in a lipid world [12,80,81]. They readily grow and divide. Droplets or vesicles or containers can grow in proportion to the amount of material inside them, providing the first step in a growth-based SOF mechanism. Natural surface-to-volume forces will cause such compartments to split into two when they get big enough, giving a physico-chemical basis for the divide and replicate aspects of SOF. The interiors of such primitive cells would be concentrated proteins, as in today’s cells. Their growth could come from the foldcat mechanism, for example. It would be interesting to see more detailed modelling.

5.4.2. Genomes implement memory for precise heritability

SOF requires accurate information transmission: of cell growth rates to lineage populations. This is achieved today by covalent memory in RNA and DNA genomes. A plausible explanation for the physico-chemical origin of the genetic code is the stereochemical hypothesis [13,82–86]. In this view, the genetic code arose from weak stereochemical binding affinities between nucleic acids and peptides, ultimately leading to codons and anticodons in today’s more complex
machinery. Here are the lines of evidence supporting that mechanism. mRNA coding sequences undergo co-aligned binding to protein sequences [87]. In pyrimidine solvents, amino acids bind to pyrimidine and purine bases in proportion to their hydrophobicities [88]; see figure 9. Nucleic acid base-stacking is driven by hydrophobic interactions and hydrogen bonding [90]; nucleic acids at high concentrations assemble into non-covalent base stacks even without a backbone [91]; free histidine binds an RNA aptamer when selected for affinity [92] and adenine binds to peptide backbones [93]. Evidence of physical affinities also appears in the identity recognition elements by which AA-tRNA synthetases recognize cognate tRNAs [94].

6. First a soup of protein machines, then encapsulation and lineages

We have postulated two stages in Pchem2Bio: forming a nearbiotic soup requiring only peptide foldamers and metabolites, followed by cellular encapsulation and informational molecules. Here, we give additional context.

6.1. Unlike an RNA world, a ‘proteins-first’ world has a plausible sustainability mechanism

The final step to biology, whatever it may have been, would have required all ingredients: proteins, informers, metabolites and encapsulation. But dividing Pchem2Bio into two steps, either real or conceptual, allows us to model possible mechanisms in more granular detail. The principal argument here for proteins first is simply that we can identify a possible mechanism. Here, we have argued for the importance and primacy of establishing the driving forces. By contrast, if instead, nucleic acids as a vehicle for memory and information, were to have come first, what would have driven it? We know of no principle or force that would have caused it to happen. Moreover, if memory were first, what machinery would construct it? It is not clear how or why it would construct itself in a self-sustaining way.

6.2. Why not an RNA world first?

The idea that origins could have started with RNA came after the discovery by Cech [95] and Altman [96] of ribozymes, the discovery by Cech and Altman of ribozymes, which can catalyse the same reaction, such as an RNase, which breaks down RNA molecules, the protein version is 100 000-fold better than the hammerhead ribozyme [17]. And RNA-based catalysts are limited, mostly phosphoryl transferases, such as RNA polymerases, ligases and RNA nucleases. The catalytic power of proteins, with 20 amino acids of very different chemical moieties, is much broader than that of RNA molecules, with only the four bases, with recognition driven largely by hydrogen bonding.

Third, the most common reaction products from many prebiotic syntheses of small molecules are amino acids, possibly because with only around 15 atoms each, they are easier to synthesize than nucleic acid bases, having around 35 atoms each. And, the yields of the different amino acids in those experiments resembles the compositions in today’s proteins [21]. Fourth, Carter & Wills show that aa-tRNA synthetases came before ribozymes, not the other way around [107]. Fifth, and more importantly, the implication of the Guseva mechanism [53] is that the foldability of polymer chains is the crucial ingredient that enables the autocatalytic explosion of functionality in PchemBio. Foldability is mainly a property of proteins, not RNA molecules.

6.3. Proteins are better for function; DNA is better for information

There is a plausible explanation for biology’s current division of labour in which proteins are functional and DNA is informational. For functionality, you need sequence-structure relations: changing the sequence, changes the structure, changes the function. The physics that enables this is folding. Proteins fold better—and for essentially all sequences—than RNA does. For information, and for memory-like actions, you specifically want the opposite. You want a type of molecule that can store all information the same, with no preferences, with the absolute minimum possible sequence structure relationships. DNA is an almost perfect informational molecule: it is very stiff, has no fold and its double-strandedness protects either strand from binding to external agents (apart of course, from transcription and such.)

6.4. A full story of PchemBio would entail informers and proteins emerging together

After a nearbiotic soup, the emergence of a genetic code requires both proteins and informational molecules to develop together [108]. Here is evidence for their concurrent development. For one thing, nucleic acids and amino acids can both arise in common from the same prebiotic processes [52]. For another thing, RNA and peptides have binding complementarity, like hands in gloves [86]. So, if a peptide-first world already drives preferences for some peptides over others, it’s easy to imagine them coupling with companion informational molecules. Interestingly, frameshifting at the mRNA/DNA level leads to protein sequences with largely unchanged hydrophobicity profiles [109], indicating how even coarse-grained hydrophobic composition alone, in the absence of specific sequences, could have carried information. In addition, Carter has shown that ‘urzymes’, which are shrunken cores of amino-acid-tRNA synthetase (aaRS) proteins, and which may have been evolutionary precursors, are unstructured small proteins having hydrophobic cores in the absence of specific sequences, could have carried information. In addition, Carter & Wills have argued that aminoacylated-tRNA molecules must have evolved in parallel with the proteins that they are responsible for helping to make, not preceding them [107].
6.5. First, a single happy pond; later, bickering individual lineages

Modelling has suggested that the origins of life started from a single community in a cauldron, something like a localized pond, before becoming individual competing cellular lineages, perhaps through an autocatalytic phase-transition-like event [112]. Community in a cauldron as a first step has the advantage that it can be communally supportive since there are no predators yet. The pond doesn’t need to compete, just to survive. Crick speculated [13] that the community-first mechanism explains today’s single genetic code, i.e. that ‘all life evolved from a single organism (more strictly, from a single closely interbreeding population)’. Although there are now counter-examples and non-universality in codes, for example in mitochondria and some nuclear genomes, the differences are small [113].

7. Conclusion

By what stochastic physical chemistry did dead matter ‘invent’ live matter? We cannot look to equilibrium principles because life has remained far from equilibrium (FFE) for 3 billion years. Unlike equilibria, which are pulled by goal-like end states, FFE dynamics are driven by the pushing flows of available matter and energy. Fitness is a tendency towards matching to environments, a driver for effective utilization of resources.

What mechanisms might have led to the autocatalysis and SOF? We describe three bootstraps. In the foldcat bootstrap, proteins became controllable catalysts, programmable through their sequences. In the catpath bootstrap, different enzymes come together in space to form pathways. In the encapsulation/heritability bootstrap, biochemistry becomes encapsulated and compartmentalized into cells, and outfitted with genetic memory to link past to future. Proteins and biochemistry, through the first two bootstraps, could have been stably self-sustaining, prior to encapsulation and heritability. Of course, this is presently just a speculation. But, there is no evident alternative mechanism by which nucleic acids could achieve persistent sustainability prior to proteins. A thread through these mechanisms is the antipathy between hydrophobic and polar interactions, in protein chains, in folding, in encapsulation, and in protein-nucleic acid interactions.

Endnotes

1 That is, a tendency for the leveraging of advantage. We use the term broadly, allowing for genetic drift, where the advantage can be small or in the form of an increased genetic diversity.
2 For our purposes, ‘autocatalysis’ is too specific to chemistry and ‘survival’ is too specific to biological populations.
3 This contrasts with the notion of irreducible complexity [37], that living systems are so complex they could not have arisen by random physical processes. Our point here is that an improbability process, even one that operates by trial and error, need not be limited in the intricacies it can achieve.
4 Here is a metaphor from industrial innovation, another process of stochastic discovery: the 60-year trajectory that led to the iPhone in 2007. It started with computers and solid-state electronics in the 1950s. But the drive in the 1950s was not the pull-driven goal to create an iPhone. Rather, the drive was push-driven by commercial opportunities of the day for computers and integrated circuits. The iPhone was just a late-stage consequence.
5 H and P correspond roughly to what Carter calls class I and class II, buried and surface residues [54].
6 Of course, modern proteins use all 20 types of amino acids for both folding and function. This model just indicates that an early code could have been crude, simple and weak.
7 It is the converse of the well-known principle in chemistry and biology that structure determines function. Here, function, in proteins, refers to an enzyme’s chemical action, and structure refers to the spatial localization of functional activities.

References

1. Darwin C. 1964 On the origin of species: a facsimile of the first edition. Cambridge, MA: Harvard University Press.
2. Follmann H, Browson C. 2009 Darwin’s warm little pond revisited: from molecules to the origin of life. Naturwissenschaften 96, 1265–1292. (doi:10.1007/s00114-009-0602-1)
3. Haldane J. 1929 The origin of life. Ration. Annu. 148, 3–10.
4. Oparin AI. 1938 The origin of life. New York, NY: MacMillan.
5. Urey HC. 1952 On the early chemical history of the earth and the origin of life. Proc. Natl Acad. Sci. USA 38, 351. (doi:10.1073/pnas.38.4.351)
6. Miller SL. 1953 A production of amino acids under possible primitive earth conditions. Science 117, 528–529. (doi:10.1126/science.117.3046.528)
7. Ferris JP, Sanchez RA, Orgel LE. 1968 Studies in prebiotic synthesis. III. Synthesis of pyrimidines from cyanacetyle and cyanate. J. Mol. Biol. 33, 693–704. (doi:10.1016/0022-2836(68)90314-8)
8. Shimoyama A, Ponnamperruma C, Yanai K. 1979 Amino acids in the Yamato carbonaceous chondrite from Antarctica. Nature 282, 394–396. (doi:10.1038/282394a0)
9. Wachtershauser G. 1988 Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452. (doi:10.1128/MBR.52.4.452-484.1988)
10. Gilbert W. 1986 Origin of life: the RNA world. Nature 319, 618–618. (doi:10.1038/319618a0)
11. Joyce GF, Szostak JW. 2018 Protocells and RNA self-replication. Cold Spring Harbor Perspect. Biol. 10, a034801. (doi:10.1101/cshperspect.a034801)
12. Segré D, Ben-Eli D, Deamer DW, Lancet D. 2001 The lipid world. Orig. Life Evol. Biosph. 31, 119–145. (doi:10.1023/A:1006746607104)
13. Crick FH. 1968 The origin of the genetic code. J. Mol. Biol. 38, 367–379. (doi:10.1016/0022-2836(68)90392-6)
14. Orgel LE. 1992 Molecular replication. Nature 358, 203–209. (doi:10.1038/358203a0)
15. Orgel LE. 1998 The origin of life—a review of facts and speculations. Trends Biochem. Sci. 23, 491–495. (doi:10.1016/S0968-0004(98)01300-0)
16. Deamer D. 2012 First life: discovering the connections between stars, cells, and how life began. Berkeley, CA: University of California Press.
17. Plaxco KW, Gross M. 2011 Astrobiology: a brief introduction. Baltimore, MD: Johns Hopkins University Press.
80. Monnard PA, Deamer DW. 2011 Membrane self-assembly and control of protein stability in protein-protein interaction networks. PLoS Comput. Biol. 7, e1001323.

81. Mohajerani F, Zhao X, Somasundar A, Velegol D, Feng M, Gilson MK. 2020 Enhanced diffusion and efficient free energy transduction by reciprocally-coupled gating. FEBS Lett. 592, 2901–2916. (doi:10.1002/febs.26311)(16)

82. Carter CW, Kraut J. 1974 A proposed model for interaction of polypeptides with RNA. Proc. Natl Acad. Sci. USA 71, 283–287. (doi:10.1073/pnas.71.2.283)

83. Hlavnjak M, Polysyarsky AA, Zagrovic B. 2012 Sequence signatures of direct complementarity between miRNAs and cognate proteins on multiple levels. Nucleic Acids Res. 40, 8874–8882. (doi:10.1093/nar/gks679)

84. Polysyarsky AA, Hlavnjak M, Zagrovic B. 2013 Analogue encoding of physicochemical properties of proteins in their cognate messenger RNAs. Nat. Commun. 4, 1–11. (doi:10.1038/ncomms3784)

85. Polysyarsky AA, Zagrovic B. 2013 Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. Nucleic Acids Res. 41, 8434–8443. (doi:10.1093/nar/gkt618)

86. Feng B et al. 2019 Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. Proc. Natl Acad. Sci. USA 116, 17669–1774. (doi:10.1073/pnas.1909221116)

87. Smith GP, Fracica TP, Todisco M, Zanchetta G, Zhu C, Hayden E, Bellini T, Clark NA. 2018 Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson–Crick selectivity. Proc. Natl Acad. Sci. USA 115, E7658–E7664. (doi:10.1073/pnas.1721369115)

88. Majerfeld I, Puthenvedu D, Yaras M. 2005 RNA affinity for molecular L-histidine: genetic code origins. J. Mol. Evol. 61, 226–235. (doi:10.1007/s00239-004-0360-9)

89. Narusny A, Kessel A, Solan R, Alva V, Kolodny R, Ben-Tal N. 2020 On the evolution of protein–adenine binding. Proc. Natl Acad. Sci. USA 117, 4701–4709. (doi:10.1073/pnas.1911349117)

90. Carter Jr CW, Wolfenden R. 2016 tRNA acceptor-stem and anticodon bases embed separate features of amino acid chemistry. RNA Biol. 13, 145–151. (doi:10.1080/15476286.2015.1124888)

91. Kruger K, Grabowski PI, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982 Self-splicing RNA: autoexonuclease and autocatalysis of the ribosomal RNA intervening sequence of Tetrahymena. Cell 31, 147–157. (doi:10.1016/0092-8674(82)90414-7)

92. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983 The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35, 849–857. (doi:10.1016/0092-8674(83)90117-4)

93. De Pauw Plana LR, Turner RJ, Steer BA, Schimmel P. 1998 Genetic code origins: tRNAs older than their synthetases? Proc. Natl Acad. Sci. USA 95, 11295–11300. (doi:10.1073/pnas.95.19.11295)

94. Koonin EV, Novozhilov AS. 2009 Origin and evolution of the genetic code: the universal enigma. IUBMB Life 61, 99–111. (doi:10.1002/iub.146)

95. Koonin EV, Novozhilov AS. 2017 Origin and evolution of the universal genetic code. Annu. Rev. Genet. 51, 45–62. (doi:10.1146/annurev-genet-120116-047113)

96. Zagrovic B, Bartonek L, Polysyarsky AA. 2018 RNA-protein interactions in an unstructured context. FEBS Lett. 592, 2901–2916. (doi:10.1002/febs.26311)(16)

97. Hart C, Brown P, Wills PR. 2018 Interdependence, diversity, and control of protein stability in protein-protein interactions. Mol. Biol. Evol. 35, 229–243. (doi:10.1093/molbev/msy046)

98. Smith E, Morowitz HJ. 2016 The origin and nature of life on earth: the emergence of the fourth geosphere. Cambridge, UK: Cambridge University Press.

99. Petrov AS et al. 2015 History of the ribosome and the origin of translation. Proc. Natl Acad. Sci. USA 112, 13936–13941. (doi:10.1073/pnas.1509761112)

100. Chen IA, Roberts RW, Szostak JW. 2004 The emergence of competition between model protocells. Science 305, 1474–1476. (doi:10.1126/science.1100577)

101. Vaidya N, Manapat ML, Chen IA, Xuvi-Brunet R, Hayden EJ, Lehman N. 2012 Spontaneous network formation among cooperative RNA replicators. Nature 491, 72–77. (doi:10.1038/nature11549)

102. Janzen E, Blanco C, Feng H, Kendel J, Chen IA. 2020 Promiscuous ribozymes and their proposed role in prebiotic evolution. Chem. Rev. 120, 4879–4897. (doi:10.1021/acs.chemrev.9b00620)

103. Bowman JC, Hud NV, Williams LD. 2015 The ribosome challenge to the RNA world. J. Mol. Evol. 80, 143–161. (doi:10.1007/s00239-015-9669-9)

104. Wills PR, Carter Jr CW. 2018 Insupposable problems of the genetic code initially emerging in an RNA world. Biosystems 164, 155–166. (doi:10.1016/j.biosystems.2017.09.006)

105. Koonin EV. 2011 The logic of chance: the nature and origin of biological evolution. Upper Saddle River, NJ: FT Press.

106. Carter CW, Wills PR. 2018 Did gene expression co-evolve with gene replication? In Origin and evolution of biodiversity (ed. P. Pontarotti), pp. 293–313. Berlin, Germany: Springer.

107. Carter CW, Wills PR. 2018 Interdependence, reflexivity, fidelity, impedance matching, and the evolution of genetic coding. Mol. Biol. Evol. 35, 269–286. (doi:10.1093/molbev/msx265)

108. Cech TR. 2009 Crawling out of the RNA world. Cell 136, 599–602. (doi:10.1016/j.cell.2009.02.002)

109. Barta et al. 2020 Framework providing preserves key physicochemical properties of proteins. Proc. Natl Acad. Sci. USA 117, 5907–5912. (doi:10.1073/pnas.1911203117)

110. Sapienza PJ, Li W, Williams T, Lee AL, Carter Jr CW. 2016 An ancestral tryptophanyl-tRNA synthetase precursor achieves high catalytic rate enhancement without ordered ground-state tertiary structures. ACS Chem. Biol. 11, 1661–1668. (doi:10.1021/acschembio.10a01101)

111. Carter CW, Wolfenden R. 2015 tRNA acceptor stem and anticodon bases form independent codes related to protein folding. Proc. Natl Acad. Sci. USA 112, 7489–7494. (doi:10.1073/pnas.1507596112)

112. Vetsigian K, Weese C, Goldenfeld N. 2006 Collective evolution and the genetic code. Proc. Natl Acad. Sci. USA 103, 1069–1071. (doi:10.1073/pnas.0603780103)

113. Osawa S, Jukes TH, Watanabe K, Muto A. 2019 Recent evidence for evolution of the genetic code. Microbiol. Mol. Biol. Rev. 56, 229–264. (doi:10.1128/MMBR.56.1.229-264.1992)