

Energy Flow and the Organization of Life

Understanding the emergence and robustness of life requires accounting for both chemical specificity and statistical generality. We argue that the reverse of a common observation—that life requires a source of free energy to persist—provides an appropriate principle to understand the emergence, organization, and persistence of life on earth. Life, and in particular core biochemistry, has many properties of a relaxation channel that was driven into existence by free energy stresses from the earth's geochemistry. Like lightning or convective storms, the carbon, nitrogen, and phosphorus fluxes through core anabolic pathways make sense as the order parameters in a phase transition from an abiotic to a living state of the geosphere. Interpreting core pathways as order parameters would both explain their stability over billions of years, and perhaps predict the uniqueness of specific optimal chemical pathways. © 2007 Wiley Periodicals, Inc. Complexity 13:51–59, 2007

1. INEVITABLE LIFE?

Life is universally understood to require a source of free energy and mechanisms with which to harness it. Remarkably, the converse may also be true: the continuous generation of sources of free energy by abiotic processes may have forced life into existence as a means to alleviate the buildup of free energy stresses. This assertion—for which there is precedent in nonequilibrium statistical mechanics and growing empirical evidence from chemistry—would imply that life had to emerge on the earth, that at least the early steps would occur in the same way on any similar planet, and that we should be able to predict many of these steps from first principles of chemistry and physics together with an accurate understanding of geochemical conditions on the early earth. A deterministic emergence of life would reflect an essential continuity between physics, chemistry, and biology. It would show that a part of the order we recognize as living is thermodynamic order inherent in the geosphere, and that some aspects of Darwinian selection are expressions of the likely simpler statistical mechanics of physical and chemical self-organization.

The principles that suggest life was inevitable have not yet been applied quantitatively to biochemistry, but they are commonplace and well understood for simpler systems encountered in meteorology, materials science, and other fields. A lightning bolt is a plasma channel created when air suffers dielectric breakdown in response to a charge separation between the upper atmosphere and the ground. The ionized plasma is kept far from equilibrium with the surrounding air by the driven

HAROLD MOROWITZ AND ERIC SMITH

Harold Morowitz is with the Krasnow Institute for Advanced Study, George Mason University, Fairfax, Virginia 22030. Eric Smith is at the Santa Fe Institute, Santa Fe, New Mexico 87501 (e-mail: desmith@santafe.edu)

motion of its own charged particles. Convective storms such as hurricanes constitute nonequilibrium states of transport, evaporation, and condensation sustained by temperature differences at different places on the globe, particularly in the oceans.

The common feature of ionized air in a lightning bolt and wall storms in a hurricane is that both create *channels* to transport currents of matter and energy¹ between two reservoirs at different potential. For lightning the potential is voltage and the current is charge, and for convective weather the potential is temperature and the current is heat. Without lightning or hurricanes, charge or heat could still move by diffusion, but the resistance to their motion through near-equilibrium states is much greater and the transport much slower than through the channel state. We understand well how voltage or temperature differences can drive these nonequilibrium channels to form and stabilize them under perturbations, and we have ways to predict the main features of the channel states from the properties of the systems in which they arise.

Whereas weather is a diffuse phenomenon primarily involving mass transport and physical state change, life creates transport channels in the chemical domain, employing the more concentrated energy flows associated with molecular re-arrangements. Two major abiotic processes lead to free energy sources sufficiently concentrated to maintain chemical order. Most organisms have evolved to depend primarily on energy from either one process or the other, creating a dichotomy in the

basic strategies for energy metabolism. Some organisms draw energy exclusively from the transfer of electrons from donor to acceptor molecules² created by geothermal chemistry. Other organisms convert visible solar light to infrared thermal radiation, in the process generating the oxidants and reductants necessary to biochemistry. Geochemical redox energy originates in nuclear fission reactions, while solar flux results from nuclear fusion. In the absence of life, energy from either source remains to some extent unused because the energy carriers are not easily accessed.

For example, the reaction of molecular hydrogen (a reductant) with carbon dioxide (an oxidant), though energy-yielding, does not happen spontaneously at an appreciable rate. If these molecules are not consumed by organisms, they tend to accumulate wherever they are generated by geochemical processes. Similarly, either inelastic absorption of light is prohibited by quantum selection rules (primarily applicable to small molecules in the atmosphere) or it triggers a photo-dissociation that eliminates the electronic transition responsible for absorption after a single event. George Wald has suggested [1] that achieving repeated absorption of high-energy visible photons poses a difficult chemical problem, solved only twice in the course of evolution with the emergence of the rhodopsins and the chlorophylls. The major converter of light to heat on earth is multiple weak inelastic scattering in the oceans, which is the primary driver of the global weather system but which cannot concentrate energy sufficiently to activate chemical transitions on a large scale.

It has not been quantitatively demonstrated that these free energy

stresses can force complex chemical order into existence or stabilize it under perturbations, nor are we sure which elements of living order could be predicted from first principles if this were so. However, guided by intuition from systems like lightning and weather, we can recognize many features in core intermediary metabolism that suggest metabolism formed in this way, and that it may be a unique or at least optimal channel for accessing these particular sources of energy. We can start with a fresh look at the structure of metabolism, and at its organization in ecosystems and along phylogenetic lines today.

2. STRUCTURE AND UNIVERSALITY OF BIOCHEMISTRY

The first thing to appreciate is that life on earth as a whole, and every ecosystem to the extent that it can be considered in isolation, synthesizes and utilizes all necessary organic molecules within a closed network of reactions. In organisms this property of biochemical completeness is called *autotrophy*, meaning “self-feeding,” so an autotrophic organism is one that requires only inorganic material inputs and a source of energy to live, grow, and reproduce. All self-sufficient ecosystems are by definition autotrophic, while only some organisms are. Organisms not capable of autotrophy are designated *heterotrophic*, or “different feeding,” because they must draw organic inputs from their environments.

Second, we know from analysis of entire genomes³ that the complete metabolic chart of autotrophs has a *universal core*, based on a set of fewer than 500 small—less than 400 Da molecular weight—organic molecules. These include sugars, amino acids, nucleotides, fatty acids, and a few more complex cofactors for handling organic functional groups or metal ions, such as coenzyme A, NAD, pyridoxal phosphate, biotin, and heme.

¹Here we distinguish internal energy, represented in thermodynamics by U , from the free energy $U + pV - TS$, which contains contributions from pressure p , volume V , temperature T and entropy S . The increase of entropy attending the exchange of energy between two reservoirs can cause a free energy decrease even when energy in the reservoirs is conserved.

²Electron donors are called reductants, and electron acceptors are called oxidants. A donor-acceptor pair that yields free energy from the transfer of electrons from donor to acceptor is known as a redox couple.

³Source <http://www.genome.ad.jp/kegg/>

Within core metabolism, we recognize two major categories of function. *Anabolism* comprises the set of reactions that build organic compounds, while *catabolism* is the breaking down of organic compounds for energy or materials. Anabolism is essentially a *reductive* process, meaning that it consumes energy-rich electrons to create molecular bonds. It is possible for an organism to exist with anabolic reactions alone, if suitable electron donors are provided by its environment, and many major clades of anaerobic organisms that are thought to have very ancient lineage are autotrophs living on geochemical inputs (called *chemoautotrophs*) whose metabolism is almost entirely anabolic [2].

Catabolism is associated with organization into ecosystems, in which heterotrophs consume other organisms or products thereof, and break down their organic molecules for energy or biologically available nitrogen or carbon. Catabolic reactions appear to function as support for anabolism, providing the same inputs as those used by reductive autotrophs but not changing the essential anabolic reaction network. The existence of a universal anabolic core is the biochemical basis of trophic ecology. For at least the last two billion years, most inputs to catabolism have originated with photosynthesis, and most catabolic reactions are either neutral (such as fermentation) or oxidizing (such as respiration). Organisms that are photoautotrophic, obtaining energy from sunlight, have also come to store organic molecules and to access them catabolically.

Only the anabolic core of intermediary metabolism is truly universal, existing in the same form in all ecosystems. Catabolic strategies vary among ecosystems and can be almost absent in communities of chemo-autotrophs. The greater variability of catabolic networks, together with the isolation of chemoautotrophic anabolism in the apparently most ancient clades, suggests to us that the first living systems were reductive chemo-autotrophs and may have been

autotrophic at the level of individual cells.

At the core of the anabolic network are the 11 carboxylic acids of the citric acid cycle. Though most familiar as the Krebs cycle in oxidizing heterotrophs, which use it to break down organic molecules to CO₂, the same reactions run in the opposite direction⁴ in many autotrophs, reducing environmental CO₂ to build organic acids. The biosynthetic pathways for all compounds in the anabolic core originate in the citric acid cycle, and when run in the reductive direction this cycle can duplicate its own members from abiotic CO₂ and electrons, a property designated *network autocatalysis*. Thus the reductive citric-acid cycle appears, at the level of the biosynthetic network, to be a self-contained engine of synthesis for all biochemical precursors [3, 4].

The rest of the core anabolic network unfolds from the citric acid cycle in a kind of cascade. The addition of ammonia and reductant yields the simpler biological amino acids directly from citric-acid cycle intermediates, many in five or fewer simple reaction steps. The 3-carbon sugars and their 6- and 5-carbon derivatives including ribose are built from the citric-acid cycle intermediate pyruvate, also in relatively few steps. The saturated fatty acids that form cell membranes are produced from another intermediate—acetate—in a recursive reaction that further reduces the carbon-oxygen bonds of acetate to hydrocarbon. More elaborate pathways leading to the complex amino acids, nucleic acids, and cofactors, follow from these elementary steps in a

dense and surprisingly simple web of reactions.

The biosynthetic reactions from environmental CO₂ and a reductant such as molecular hydrogen are exergonic as a whole [4], though many of the individual steps require input of energy to overcome free energy “hills” if they are to occur at appreciable rates. The energy provider for most of these reactions, as well as much of the re-arrangement chemistry of sugars, is the hydrolysis of the high-energy ester bonds of phosphate polymers. Today these are provided by ATP, but the ubiquity of phosphate in core metabolism, together with the improbability that ATP could have been synthesized until several other chemicals were available as precursors, leads us to surmise that energetic phosphates must have been ubiquitous in the geochemical settings where life began.⁵

The cascade of metabolism thus functions overall as a relaxation channel

⁵The status of phosphates on the early earth is still unclear. Several authors [3–6] have proposed inorganic polyphosphate as a precursor and analogue to ATP in the earliest phases of emergent life. Polyphosphates are found in volcanic condensates and other dehydrating settings today [6, 7], but hydrothermal vent fluids appear depleted in orthophosphate relative to the seawater from which they form [8], and vent polyphosphates are not reported. Christian de Duve has argued [9] that the ubiquitous use of thioesters as intermediates in the formation and hydrolysis of phosphate ester bonds is another ancient feature of biochemistry, which could have led to the production of pyrophosphate and other phosphate polymers from reduced sulfur minerals in a pre-biotic age. Building on this argument and on physiological universals of the acetogenic bacteria and methanogenic archaea, William Martin and Michael Russell [10] have proposed acetyl phosphate as the earliest continuously produced source of activated phosphates.

⁴This reductive direction for the reactions is unfortunately known as the “reverse citric acid cycle.” As it was probably the first form of the cycle to exist, and may have been the first ordered chemical pathway in the emergence of life, it would be more natural to call the reductive cycle the forward direction, and the later, oxidizing Krebs cycle the “reverse” direction.

for two major sources of geochemical energy: electron transfer from reductants and the hydrolysis of phosphates. The network structure suggests a channel that could have formed spontaneously and unfolded in stages, with some of the simplest (2- to 4-carbon) compounds in the citric-acid cycle arising by thermal fluctuations and then self-amplifying by virtue of the autocatalysis of the network. The carbon atoms in citric-acid cycle intermediates are not completely reduced, and the formation of fatty acids from acetate releases another increment of redox energy by producing hydrocarbon groups and water. The formation of sugars, amino acids, and other compounds creates pathways to hydrolyze phosphate esters or to incorporate ammonia. Apart from some uncertainty about phosphates, all of these species are known to be produced in the hydrothermal vent environments of the modern earth. Vent energy sources support communities of organisms today [11] in the same way we propose they supported the earliest proto-metabolic chemical systems, except that modern organisms do not require energetic phosphates from the environment because they can recycle phosphate internally using redox energy and the machinery of oxidative phosphorylation.

The anabolic network is dense, and leaves no energetic waste products. Said another way, core metabolism converts all carbon from a fully oxidized abiotic input (CO_2), either into its most reduced form (methane), or into constituents of biomass such as the fatty acids, sugars, amino acids, nucleotides, and cofactors, which then provide the platform for further metabolism. This feature would also be expected of a network that was *driven* into existence by pressures to transfer electrons from high-energy to low-energy bonds, and to return phosphate from a high-energy polymer state to low-energy orthophosphate.

If modern metabolism recapitulates the process of biogenesis—and we will

discuss why it should—the energetic cascade through the anabolic network suggests a cascade in the emergence of life, one that began in metabolism. The earliest and simplest stage in the cascade, the reductive citric acid cycle, is most accessible by random fluctuations and is also the stage coupled directly to the abiotic carbon source, CO_2 . This cycle looks simple enough to have emerged spontaneously in an unsupervised network of small-molecule reactions, but only captures about one third of the redox energy available in transforming CO_2 to its most-reduced and lowest-energy intermediate, acetate. Fatty acid synthesis captures about another third of the available energy, but can only occur from a starting pool of acetate, and is not network-autocatalytic for carbon intake. However, it does provide an outlet for citric-acid-cycle carbon, maintaining the disequilibrium that drives the cycle.

The remaining third of available redox energy not released by fatty acid synthesis is obtained by converting hydrocarbons to methane, or more directly by converting acetate⁶ to methane as some methanogens do today. Direct acetogenesis and methanogenesis do not have the autocatalytic network topology of reductive citric acid cycling, which we will argue was responsible for the concentration of carbon flux through the small collection of citric acid cycle intermediates during the emergence of life. Furthermore, these competing pathways through acetate are associated with a distinctive set of cofactors [2]—some

of them complex—and methanogenesis has the strongest isotope selectivity measured for a primary organosynthetic pathway [12]. These considerations have led us to view acetogenesis and methanogenesis, as they occur in modern organisms, as later redox innovations in the reducing world.⁷ The emergence of biochemical intermediates would then have been a sort of bootstrap process in which simple networks near the input were elaborated to more diverse networks downstream, as further reductant, ammonia, or phosphate reacted with the starting compounds.

The chemical ensembles in which life exists are structured by kinetics as well as energetics, and the provision and structure of catalysis will be central to understanding how a chemical network could have been organized by energy flows. Our understanding of kinetics is being changed by a growing realization that catalysis is not limited to complex polymers of amino acids or RNA with sequences adapted to highly specific reactions. Even the biological monomers and small polymers have been found to possess catalytic activity [14, 15], and we have proposed [16] that catalysis was continuously provided during the emergence of life by a

⁶The precursor to methane in methanogenic archaea, and also the output of citric-acid cycling, is actually the thioester acetyl-CoA. We are suppressing here the important but complex discussion of sulfur chemistry in core anabolism, in order to emphasize the network structure along the path for reduction of carbon in the primary organic backbones.

⁷The relations among these three ancient processes (direct acetogenesis and methanogenesis, and reductive citric acid cycling) in the emergence of life may, however, have been complex and subtle. The earliest organic synthesis networks must have been stochastic and unsupervised, though we think that factors like network autocatalysis made them more orderly than de Duve's notion of the gemisch [9, 13] suggests. Martin and Russell, in particular, argue persuasively [10] that synthesis of methyl thiols or some similar activated species by geological processes was essential to the first energy metabolism, and that the reductive citric-acid cycle reactions cannot be understood outside the context of such energetic support.

hierarchy of molecules, starting in the monomer world and passing through stages of longer and more organized polymerization.

Presenting metabolism from a perspective of energy flow—that is, starting with reductive autotrophy and adding complexity as it was likely added during the emergence of life—yields a remarkably hierarchical, compact, and *simple* picture, but one limited to descriptive chemistry. To understand whether the emergence of life was a kind of “lightning bolt” through the graph of geochemical reactions, and whether the spontaneously generated path through that graph can be predicted from first principles, requires a quantitative statistical framework for handling flows through chemical networks.

3. FLUX AND ORDER

Such a framework was suggested in its broadest outline by E. T. Jaynes, who more than anyone else has argued that statistical physics is simply the application of principles of statistical inference to systems whose boundary conditions cannot completely specify their internal configurations [17]. The entropy of physics and chemistry is the entropy of information theory, formalizing Laplace’s principle of “insufficient reason.” Entropy maximization fixes the infinitely many free moments of a distribution only partly specified by boundary conditions without restricting what the conditions can be. The properties of the resulting distribution then reflect required consequences of the boundary constraints, whatever else may or may not be knowable about the system. These methods for extracting the consequences of partial constraint clearly are not limited to cases of thermodynamic equilibrium. Entropy can depend on currents as well as on configurations. When it does, the principle of free energy minimization for open systems, which is derived from entropy maximization, can be extended to driven systems. The free energy is then generalized to depend on the richer

boundary conditions responsible for driving.

In principle Jaynes’s observation opens a whole new domain for statistical mechanics, but in practice few instances that usefully reach beyond equilibrium have been derived. Chemistry is still studied with the classical state variables and entropy of Gibbs,⁸ and the application of information theory to life at a system level has largely been limited to bioinformatics of gene or protein sequences and a few problems of binding affinity. The use of equilibrium entropy makes life seem paradoxical [18]: the so-called equilibrium entropy of biomass is less than that of the same matter in an abiotic state, and yet the core molecules and pathways of life have apparently been stable despite continual perturbation for almost four billion years. The theory of stability in statistical mechanics therefore suggests that an earth with life must be more probable than an earth distributed according to thermal equilibrium and that the “equilibrium entropy” within the biosphere is not the proper measure of likelihood on a driven earth.

The paradox is not entirely naive, though Boltzmann saw already that it was based on an incomplete description. Many have recognized that the increase in equilibrium entropy of the nonliving environment can allow the equilibrium “entropy” of biomass to decrease (refrigerators do this all the time), but this recognition misses an essential connection and leaves us still unable to calculate or make predictions. Life mediates *flows* in its environment, which in some cases can correctly be associated with *rates* of entropy change [17, 19]. Yet life is carried by *states* of matter, and a proper statistical theory should at least allow us to predict the amount of order in a state that can be

stabilized by a given rate of environmental entropy increase. The answer to this question will tell us whether life could have formed as a spontaneous chemical channel.

The failure of equilibrium entropy to provide a valid maximization principle for driven systems has led to attempts [20–22] to formulate principles based on “entropy production”, which necessarily invoke dynamical as well as statistical assumptions. Though these may inherit some validity from well-formulated entropy maximization in special cases, they ultimately fail as general principles because they are based on the application of the equilibrium expression for entropy to driven systems where that expression no longer measures the actual uncertainty in the systems’ states.

Examples have been worked out that avoid these problems simply by enriching the description of a system’s state space to include measures of order in its processes as well as in its configurations. In classical mechanics, this leads to an entropy that depends on the net drift currents created by non-equilibrium boundaries in space, as well as the concentrations of chemical species [23]. The same generalization can be made in quantum mechanics—the natural language for describing systems of chemical bonds and reactions—where the entropy becomes a function of reaction currents as well as configurations [24]. The resulting entropy-maximization principles can predict the spontaneous *formation* of currents, whereas the equilibrium entropy is maximized on currentless states. These are the appropriate entropy principles underlying Ohm’s law for small electric currents and Fourier’s law for heat flow. The very existence of a current constitutes a restriction on the system and hence a decrease in its entropy, but these near-equilibrium models are too simple to capture the possibility of driving the *configurations* into significantly more ordered states to create better channels for energy flow—a feature of life we set out to explain.

⁸or their extensions to equilibrium quantum-mechanical ensembles for intramolecular properties or reaction transition states.

Applied to nonlinear systems, these same ideas establish an important mathematical connection. The presence of positive feedback in a current-carrying system can create a threshold for the sudden formation of macroscopic order, and the crossing of this threshold is a *phase transition* equivalent in all statistical respects to equilibrium phase transitions [25, 26]. The ordered state creates a channel between the environment's input and output reservoirs with much better conductance than the equilibrium state. Order in turn is maintained by energy extracted from the current between the reservoirs. The entropy principle used in these examples is not equivalent to maximization of the inter-reservoir current subject to some constraint, because the system state is not constrained with respect to current capacity. Rather the ordered state results from a balance of many contributions to the true entropy, some of which may be understood approximately as the entropy of the internal system state, some in terms of entropy transport between reservoirs, and some in terms of entropy production from dissipation [24].

The foregoing examples will fall short of an integrated framework for understanding life, because the chemical substrate of life has a complex structure and informational richness exceeding that of other physical systems. However, these models do provide in patchwork a set of appropriate tools for mathematically relating the origin and robustness of life to simpler classes of self-organization. They show, when the nonequilibrium entropy is properly formulated, how the initiation of currents can be consistent with entropy maximization and can even be required by it in open systems. They also show how positive feedback, leading to configurational phase transitions that enhance current flow, can arise when an internal cycle of matter or energy in the system is coupled to the transport between its reservoirs. The way such coupling enters these examples of physical self-organization [25, 26] corresponds

closely to the reaction structure of reductive citric acid cycling [27]. On the basis of this correspondence we infer that network autocatalysis was fundamental to concentrating carbon fluxes through specific pathways, against the randomizing effects of dissipation. These statistically favored fluxes become the *order parameters* in the phase transitions to life, an identification that reconciles their robustness with their origin in fully stochastic and unsupervised reaction networks.

In the attempt to understand organization in driven systems, particularly regarding their coupling to the environment, the potential forms of order *not observed* are as informative as those observed. The forms of order we do not observe in biochemistry are the many organic molecules with a complexity and free energy comparable to that of core metabolites, yet which are not produced by metabolism. The metabolic core comprises a surprisingly small and well-ordered collection of molecular species, suggesting limitations in the kinds of organic molecules relevant to its emergence, and in the pathways capable of extracting energy from these sources. For instance, the citric acid cycle is distinguished as a simple pathway for organosynthesis from CO₂ and reductant, but it is of no apparent significance in networks driven by high concentrations of the reactive molecule formaldehyde—which undergoes aldol condensations to form complex mixtures of diverse sugars [28]. Similarly, the biological amino acids are among the most natural elaborations of citric-acid cycle intermediates in an environment of ammonia and reductant [29]. These amino acids are not distinguished in synthetic reactions from gas-phase free radicals [30], and they are not distinguished in the deep-space photochemistry of interstellar and cometary ices or meteors [31].⁹

⁹It appears to be a common assumption in the experimental literature on "prebiotic chemistry" that the production of

4. THE COLLAPSE TO LIFE

The creation of energy channels by means of phase transition provides a way to understand how the core biochemistry of life can have been stable throughout the age of the earth: a state of the geosphere that includes life becomes *more likely* than a purely abiotic state. The nonliving earth would have been *metastable* under conditions of continuous geochemical free energy production. Its "collapse" to greater stability was the emergence of life.

We must be careful, however, in describing the more stable state to which earth collapsed. It is probably wrong to regard biomass as a thermodynamic "state" of terrestrial matter. A more accurate picture from well-studied systems might be phase separation in binary alloys [32], in which the whole alloy is in a thermodynamic state consisting of separated "phases" of different chemical composition that interact at interfaces. The phase separation of the cellular milieu from its abiotic surrounding is clearly more complex than phase separation in alloys, and the interactions between the phases—both at interfaces and mediated by them—has qualitative differences because life is among other things a transport phenomenon. Still, alloys provide a model for the emergence of combined spatial and chemical structure—clearly essential to understanding metabolism.

Cells use membranes to concentrate organic reagents and exclude toxins. They use them to create pH and voltage differences converting redox to phosphate energy and they use them to segregate polar and nonpolar environments

metabolites by abiotic processes is the main problem to be overcome, and that later stages of evolution somehow separated the observed universal metabolism from a rich primordial soup. Our explanation would isolate more of the observed metabolic pathways at precellular stages, where chemical selection more closely resembles thermodynamic relaxation than it does cellular evolution.

in which different biochemical reactions are favored. It is likely that cells came into existence as the most productive platform for metabolism, after which the characteristics of individuality and heritable variation so central to Darwinian selection emerged as a byproduct of cellular form.

The fact that the biosphere must be considered as but one in a collection of geospheres was appreciated by Rankama and Sahama [[33], p. 320], who puzzled over how biomass—so small a fraction of terrestrial matter—can have exercised such a large influence on the earth's processes, particularly in the oxygenation of the atmosphere and oceans. We have already noted that the inelastic absorption of light by photosynthetic organisms provides a channel—*independent* of other physical processes—for the transduction of light energy between the visible and thermal infrared bands. An elementary calculation [34] shows that absorption in place of elastic scattering increases heat-transporting capacity by a factor of $\sim 10^9$ for each atom acting as an absorber. In other words, every atom effectively rendered capable of photo-absorption by organosynthesis becomes a billion times more efficient as a transporter of energy from sunlight to black-body radiation than that same atom in an abiotic state. The enhancement of this energy capacity is mitigated by the large volume of biomass apparently required to create each photosynthetic absorption site.

An emergence of life by collapse does not imply that the collapse was a single event, or even that it has run to completion.¹⁰ The concept of phase transition

unifies the emergence of life with the major transitions in evolution that were clearly chemical and apparently irreversible, and it also emphasizes that life is not necessarily a single statistical phenomenon.

There were presumably distinct chemical “batteries” in the early geosphere (redox, phosphate, perhaps pH), and different modules within metabolism appear to be stabilized independently to some degree by chemical potentials from the different sources. The unification of these essentially geochemical processes within cells could have required a later innovation, such as oxidative phosphorylation for the recycling of phosphates, which are difficult to transport across nonpolar cell membranes. Nitrogen fixation appears as a still later innovation in response to changes in ocean chemistry [35], apparently occurring after the last universal common ancestor (LUCA) had emerged. The emergence of the LUCA was likely not even energetically the end of the collapse of the abiotic earth, though it does represent the synthesis of remarkably many forms of order. An anabolic, reductive, nonphotosynthetic LUCA could have incorporated all of the major structures and functions needed for cells to evolve and flourish, as reductive autotrophs do today.

The innovation of photosynthesis could then have proceeded atop this sophisticated platform of genes, enzymes, and compartments, with photosynthesis from hydrogen sulfide preceding that from water [36]. The amount of free energy available from sunlight relative to that available from geotectonic chemistry is large enough that energetically the collapse to photosynthetic life may have created a more significant channel than the emergence of cells. This is why photosynthesizing life appears to account for most biomass, either directly, or indirectly through byproducts on which other ecosystems depend. Other apparent novelties following on photosynthesis were the reductive pentose phosphate

cycle for carbon fixation, endosymbiosis, and ultimately multicellularity—the last two of which seem to have been driven particularly by the rise of molecular oxygen in the atmosphere and oceans.

This collapse to life is an ongoing process. Its most energetically significant and accessible steps probably followed one another rapidly in the initial stages, with further refinements coming later and more sporadically—either because they followed from less likely events or because they offered less energy-flow advantage. However they occurred, we would expect the truly chemical transitions to be distinguished by *irreversibility*, as the emergence of photosynthesis, endosymbiosis, and multicellularity apparently are.

This point of view emphasizes that life is a confederacy—of chemical constituents, pathways, and functions—and that the modular form of its emergence continues to be reflected in its modern organization. The essentially thermodynamic succession of a biogeosphere through increasingly stable states also explains why metabolism seems to show a clear and deterministic progression, while species and their ecologies appear to have been subject to continual waves of extinction.

5. A NEW UNDERSTANDING OF LIFE WITHIN THE GEOSPHERE

However tightly integrated they are in modern cells, the functions of life are heterogeneous and retain somewhat independent dynamics. It may be that we should view core metabolism as an outgrowth of geochemistry rather than a totally novel innovation of biology. As trees in a dry landscape tend to grow along the riverbeds, the structures of life have formed themselves around the pathways of primordial metabolism. It is precisely there that significant innovation was *not* required. Rather than view core metabolism as a subset extracted from the primordial soup by later evolution we would characterize

¹⁰In very large systems capable of highly intricate order, such as the biosphere, it is difficult to assign a practical meaning to even the concept of “completion.” Presumably most of the variables in organism and ecosystem structure are nearly neutral with respect to the (scalar) criterion of energy transport, and the evolutionary dynamics of these variables must be understood in other terms.

extant life as the downstream elaboration of possibilities created by a metabolic core much like the present one.

The emergence of cells in this picture was a process of enfolding and controlling geochemical flows. The premise that cells produced all organic molecules from abiotic precursors right from the beginning has sometimes been called an “autotrophic” theory of the origin of life, in order to distinguish it from the idea of “heterotrophic” dependence on a primordial soup [37]. The terminology is unfortunate as it conflates different concepts. We picture the emergence of life as a sort of chemical bootstrapping process from abiotic precursors in steady state, as opposed to a process initiated by a primordial soup and then required to develop anabolic pathways to provide its own inputs as the soup was exhausted. The notions of autotrophy and heterotrophy emerged later, when cellularization and speciation arose to facilitate metabolism and evolution. The dynamics at this stage determined how cells and species partitioned the metabolic functions to form ecological structure.

The metabolism of ecosystems is empirically more universal than their partitioning according to the roles and interactions of species. This fact makes sense if metabolism is the bridge between geochemistry and the novel elements in cell physiology and trophic ecology. The major chemical or metabolic transitions in evolution have a different character from the record of extinctions, invasions, and ecological successions. The latter are governed by the dynamics of repartitioning rather than of primary production. Finding a stable ecological division of labor requires solving a complex network of constraints. The species making up an ecology must jointly implement a universal metabolic chart, while predation, saprophytic, parasitism, and symbiosis, create a flux balance of organic molecules among the

individuals. The magnitudes of metabolic fluxes are in turn constrained by the body sizes of individuals and by environmental mass transport [38]. In principle, many distinct ecologies can satisfy these constraints with comparable overall energy throughput, though they may be transformable into one another only by complex avalanches of species replacement and trophic restructuring.

Other evidence for the difference between species dynamics and the dynamics of metabolism comes from the preservation of large parts of the metabolic network during major chemical transitions in evolution. It is easy to imagine that with the rise of oxygen reductive anabolic pathways might have been so poisoned by this waste product as to bring life to an end [39]. Instead they were essentially preserved, with core networks like the citric-acid cycle reversing direction to break down compounds for energy—which in the reducing world they had synthesized—while still providing a complete set of biosynthetic precursors. Innovations such as the light/dark cycle now used by plants emerged, suppressing anabolic activity during the sunlight hours when molecular oxygen is produced, and re-activating it at night when oxygen partial pressure is lower.

The prior existence of thermodynamically ordered pathways for organosynthesis opens the possibility for a natural sequence of steps from protometabolism to the first macromolecular phase of life. Geochemically ordered primordial pathways would favor the preservation of any randomly formed polymers that increased pathway fluxes, thus producing more of the raw materials from which the polymers were made. The capacity for molecular replication could have been completely sequence-independent in its very earliest stages. If the first polymers were rare in a world where a nascent metabolic order already existed, their sequences would originally have been selected on the basis of their

interactions with metabolites.¹¹ Only at a later stage, when many such sequences were present, would we expect them to have taken on relations to each other that aided in their replication. Under this order of selective forces the commitment to metabolism would have been made before replication became a coordinated macromolecular process, and we would expect the resulting metabolism to preserve the prior geochemical pathways. The fixing of proto-metabolic pathways by coded macromolecules created initial conditions from which these pathways would be inherited by all subsequent generations of cells.

Darwinian competitive exclusion is rooted in the chemical competitive exclusion of metabolism, whether through differential rates of growth or differential resource capture. The congruence of cell physiology with the statistical paths of least resistance will inevitably be a determinant of fitness. Through it the forces that drove the emergence of life will be expected to have stabilized its metabolic core throughout earth’s history. If the emergence and maintenance of metabolism have indeed both been brought about by the same thermodynamic forces, then universality due to descent from common ancestors becomes indistinguishable from universality due to evolutionary convergence.

The organization of energy flow through metabolic pathways allows us to recognize many forms of continuity absent in conventional thinking. We have good reason to believe that the

¹¹We thus dissociate the source of selective advantage from the process of molecular replication, to the point where the two could have been completely independent. Note that this premise is directly opposite the usual premise adopted for an RNA world [40]: that differential ability of macromolecules to self-replicate was the first source of selective advantage.

first emergent metabolism was similar in many respects to modern universal core anabolism. Metabolism itself becomes a bridge from driven geochemistry to the foundations of cell physiology and trophic ecology. If our story is correct, the thermodynamic forces responsible for the emergence of anabolism within prebiotic chemistry have ensured its stability throughout the ensuing history of life. Energy flow embeds

life within the geosphere not just mechanistically but conceptually as an inevitable form of driven geochemical order.

Acknowledgments

We are thoroughly indebted to Cormac McCarthy for advice on the manuscript and to Shelley Copley for shaping our views of biochemistry and catalysis. We

are also grateful to Michael Russell and William Martin for comments and ongoing correspondence and to Doug Erwin and Vijayasarathy Srinivasan for comments in the course of preparation. We are supported in this work by grant FIBR: The Emergence of Life: From Geochemistry to the Genetic Code Award 0526747, National Science Foundation. DES acknowledges Insight Venture Partners for individual support.

REFERENCES

- Wald, G. Fitness in the universe: Choices and necessities. *Orig Life Evol Biosphere* 1974, 5, 7–27.
- Lengeler, J.W.; Drews, G.; Schlegel, H.G. *Biology of the Prokaryotes*; Blackwell Science: New York, 1999.
- Russell, M.J.; Hall, A.J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J Geol Soc London* 1997, 154, 377–402.
- Smith, E.; Morowitz, H.J. Universality in intermediary metabolism. *Proc Nat Acad Sci USA* 2004, 101, 13168–13173. (SFI preprint # 04-07-024).
- Baltscheffsky, H.; von Stedingk, L.-V.; Heldt, H.-W.; Klingenberg, M. Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science* 1966, 153, 1120–1122.
- Michael, R.W.B.; Kornberg, A. Inorganic polyphosphate in the origin and survival of species. *Proc Nat Acad Sci USA* 2004, 101, 16085–16087.
- Yamagata, Y.; Watanabe, H.; Saitoh, M.; Namba, T. Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* 1991, 352, 516–519.
- Kelley, D.S.; Baross, A.J.; Delaney, J.R. Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annu Rev Earth Planet Sci* 2002, 30, 385–491.
- Christian de Duve Singularities; Cambridge U. Press: New York, 2005.
- Martin, W.; Russell, M.J. On the origin biochemistry at an alkaline hydrothermal vent. *Phil Trans Roy Soc B* doi:10.1098/rstb.2006 2006, 361, 1–39.
- Hügler, M.; Wirsén, C.O.; Fuchs, G.; Taylor, C.D.; Sievert, S.M. Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the ε subdivision of proteobacteria. *J Bacteriology* 2005, 187, 3020–3027.
- Pruess, A.; Schauder, R.; Fuchs, G. Carbon isotope fractionation by autotrophic bacteria with three different CO₂ fixation pathways. *Z Naturforsch* 1989, 44, 397–402.
- de Duve, C. *Blueprint for a Cell*; Neil Patterson: Burlington, NC, 1991.
- Huang, F.; Bugg, C.W.; Yarus, M. RNA-Catalyzed CoA, NAD, and FAD Synthesis from Phosphopantetheine, NMN, and FMN. *Biochemistry* 2000, 39, 15548–15555.
- Pizzarello, S.; Weber, A.L. Prebiotic amino acids as asymmetric catalysts. *Science* 2004, 303, 1151.
- Copley, S.D.; Smith, E.; Morowitz, H.J. The origin of the RNA world. In preparation 2006.
- Rosenkrantz, R.D.; Jaynes, E.T. *Papers on Probability, Statistics and Statistical Physics*?Reidel, D.: Dordrecht, Holland, 1983.
- Schrödinger, E. *What is Life?: The Physical Aspect of the Living Cell*; Cambridge U. Press: New York, 1992.
- Tykodi, R.J. *Thermodynamics of Steady States*; Macmillan: New York, 1967.
- Kondepudi, D.; Prigogine, I. *Modern Thermodynamics: From Heat Engines to Dissipative Structures*; Wiley: New York, 1998.
- Onsager, L. Reciprocal Relations in Irreversible Processes, Vol. 37, 1931.
- Onsager, L. Reciprocal Relations in Irreversible Processes. II *Phys Rev* 1931, 38, 2265–2279.
- Corbet, A.B.; Morowitz, H.J. Information-theoretic derivation of a nonequilibrium velocity distribution. *Phys Rev A* 1972, 6, 2298–2301.
- Smith, E. Thermodynamic dual structure of linearly dissipative driven systems. *Phys Rev E* 2005, 72, 36130. (SFI preprint # 05-08-033).
- Smith, E. Carnot's theorem as noether's theorem for thermoacoustic engines. *Phys Rev E* 1998, 58, 2818–2832.
- Smith, E. Statistical mechanics of self-driven carnot cycles. *Phys Rev E* 1999, 60, 3633–3645.
- Smith, E.; Morowitz, H.J.; Copley, S.D. Core metabolism as a self-organized system.
- Weber, A.L. Sugars as the optimal biosynthetic carbon substrate of aqueous life throughout the universe. *Orig Life Evol Biosphere* 2000, 30, 33–43.
- Copley, S.D.; Smith, E.; Morowitz, H.J. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc Nat Acad Sci USA* 2005, 102, 4442–4447.
- Miller, S.L. Production of amino acids under possible primitive earth conditions. *Science* 1953, 117, 528–529.
- Pizzarello, S. Chemical evolution and meteorites: An update. *Orig Life Evol Biosphere* 2004, 34, 25–34.
- Okamoto, H. *Desk Handbook: Phase Diagrams for Binary Alloys*; ASM International: Materials Park, OH, 2000.
- Rankama, K.; Sahama, T.G. *Geochemistry*; U. Chicago Press: Chicago, IL, 1950.
- Hoelzer, G.A.; Pepper, J.W.; Smith, E. On the logical relationship between natural selection and self-organization. *J Evol Biol* 2006, 19, 1785–1794.
- Raymond, J.; Seifert, J.L.; Blankenship, R.E. The natural history of nitrogen fixation. *Mol Biol Evol* 2004, 21, 541–554.
- Blankenship, R.E. *Molecular Mechanisms of Photosynthesis*; Blackwell Science: Malden, MA, 2002.
- Fry, I. *The Emergence of Life on Earth: A Historical and Scientific Overview*; Rutgers U. Press: New Brunswick, NJ, 2000.
- West, G.B.; Brown, J.H. The origin of allometric scaling laws in biology from genomes to ecosystems: Towards a quantitative unifying theory of biological structure and organization. *J Exp Biol* 2005, 208, 1575–1592.
- Margulis, L.; Sagan, D. *Microcosmos: Four Billion Years of Evolution from Our Microbial Ancestors*; U. California Press: Berkeley, CA, 1997.
- Eds. Gesteland, R.F.; Cech, T.R.; Atkins, J.F. Eds. *The RNA World*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006.