A Theory of Biochemical Organization, Metabolic Pathways, and Evolution

Harold J. Morowitz

Robinson Professor of Biology and Natural Philosophy

and Director of the Krasnow Institute for Advanced Study

George Mason University

Fairfax, Virginia 22030

Abstract

Biogenesis and evolution are viewed from the perspective of the universality of the metabolic chart with respect to primary metabolism and the phylogenetic specificity of secondary metabolism. This analysis is developed within the context of the evolution of the universal ancestor through hierarchical networks of chemical reactions.

The theme of this paper is the examination of biological and prebiological evolution in terms of the universal present day graph of metabolic pathways. Since the core features of metabolism date back to the universal ancestor some 3.9 billion years ago, this perspective looks for continuity between biogenesis and evolution and seeks for a deep logic of evolution within the reductionist roots of quantum mechanics, the periodic table, and the boundary conditions of the planet Earth. This is not to deny the existence of frozen accidents and random events at any stage but is to look at the framework in which they happen. This infrastructure provides certain profound deterministic aspects to the process of biogenesis.

The constructs to be utilized are energy, mass, and structure (information, organization, complexity, etc.) We lack a universally agreed on term or concept for the third feature, but it can be understood in a general way. Matter is not in fact the structureless mass of hydrodynamics nor the point mass of classical mechanics. Matter in the temperature range of biological objects consists of atoms, atomic nuclei and electrons. Atoms, of course, have an electronic structure imposed by allowable quantum mechanical states and the Pauli exclusion principle. As a result each type of atom has unique properties which determine its interactions with like atoms and those of other types. All of form in the size range of nanometers or less is dependent on the details of atomicity. Structural information is thus dependent on the differences among the atoms of the periodic table. In biochemistry, information is not a mathematical abstraction, it is associated with molecular structures. Information in this context cannot be isolated from the details of chemistry.

There has been and continues to be a confusion between Shannon information which is a measure on a normalized probability distribution and anecdotal uses of the word information.

What the previous exercise makes clear is that we can in principle develop a metric for structural complexity in terms of the constructs of covalent bond chemistry which is after all at the root of biochemistry. In principle there is no reason why the method cannot be extended to other forms of bonding, in which case the fixity of the structures needs to be modified by essentially energetic and kinetic considerations. This will complicate the situation to a considerable degree.

Bioevolution, in its most global aspects, is the transformation of the earth's chemistry from its prebiotic state to cellular life to the development of the rich biota and biogenic products represented by the many millions of species, present and past, and the fossil chemicals.

Chemical reactions which could be called prebiotic took place on earth about four billion years ago. The processes grew more complex and gave rise to life, and the universal ancestor. That life has persisted and evolved with increasing complexity since that time. These statements stem from empirical generalizations in geochemistry, paleontology, biochemistry and systematics and phylogeny. Every organism is a lineal descendent of the universal ancestor.

The purpose of this monograph is to provide a theoretical understanding of the development of biochemical networks and subsequent biological complexity and evolutionary unfolding. This is done through postulating a scheme of nested shells of reaction networks of increasing complexity and sophistication. At the most reductionist and universal levels of biological chemistry these shells can be delineated with some precision and the nature of the emergence by which new shells arise can be outlined in some detail. It has been noted (Smith a hot aqueous system with a steamy atmosphere. Occasional meteoritic impacts volatilized large quantities or in some cases all of the water. We assume the same impact profile for each ensemble member. The atomic composition of the surface more or less had the same value as present except for an unknown value for hydrogen which constantly escapes from an aqueous planet of the size, density, and temperature of the earth.

In this discussion, only the prenoetic domain of biology is under consideration. It seems clear that the emergence of the mind in animal evolution represents the appearance of radically new features, which will not be dealt with in this essay that focuses on the pre hominid period. In any case the "gedanken" construct of the ensemble allows us to distinguish the average features of the collection from the historical variation of the individual members.

Many of the aspects of life on earth follow in a necessary fashion from biota arising and existing in an aqueous environment. The properties of water (Edsall and Wyman, 1958), particularly in the liquid and solid states, mainly follows from: the large dipole moment due to charge asymmetry along the oxygen hydrogen bond, the non-covalent hydrogen bonds, and the spatial distribution of electrons about the hydrogen and oxygen. All ensemble members of planets will have watery surfaces.

The theory we are developing is reductionist in that it depends on atomic properties outlined in the periodic table. Indeed, a major goal of a general biochemical theory of life is a mapping from the periodic chart of the elements unto the chart of metabolic pathways. The theory requires the introduction of constructs suggested by both physics and biology. These constructs are:

1. Far from equilibrium entity.

3. Post replicative memory

If a pre-replication entity has certain properties that distinguish it from other such entities and its post replication entities both have these properties to some degree, we refer to the process as having post replicative memory.

4. Competition for resources

If two entities use common nutrients or energy sources, they are said to be in competition for resources. This usually applies to replication, but may also apply to maintenance. Identifiable entities form a necessary basis for Darwinian competition.

5. Structural complexity

Complexity is the length of the shortest algorithm capable of generating a description of an entity at some point in time that is suitable for some predictive statement. The measure of complexity is thus dependent on the domain of the predictive statement that is the use for which the description will be applied, and the search for the shortest algorithm. One possible formalism has been discussed briefly above.

With the preceding definitions in mind, we note that a feature of earth life as we know it is that it consists of stable replicating entities, separated from the homogeneous phase system and preserving individual integrity. This is particularly important because competition between all replicating entities, in a Darwinian sense, is a primary feature of life on earth. It presumably began as soon as identifiable stable self-replicating entities existed in the system.

Entities must exist and be distinguishable in order for Darwinian competition to occur.

micrometers (μ) or volumes the order of 10^{-3} μ^3 to 8 X 10^{-3} μ^3. Thus there are the order of 10^9 atoms in a minimal cell. At the atomic level, no two cells can be identical; they are simply too complex. Using ideas that he traces back to Neils Bohr, Elsasser argued that such objects cannot be subject to exact quantum mechanical analysis. We cannot start with a number of objects all in the same state and then follow the statistics.

It is central to atomic physics that no two electrons or two protons are distinguishable. At the core of evolution is the idea that all organisms are distinguishable. Biological behavior is an emergent property at a certain level of complexity and phase separation. Each organism is a historical phenomenon carrying evidence of its four billion year old past. In any case complexity assures us that all the objects of biology (cells and organisms) are different.

Replication, the arising of offspring similar to the organisms from which they have come assures us of a class of objects similar enough to use the notion of a class or species. If all offspring were identical we could have no evolution. If all offspring were radically different we could have no evolutionary continuity, no speciation. At some intermediate range, the evolutionary development of a planet is possible.

Competition organizes the selection among not identical offspring and provides the pruning algorithms for evolution. Thus complexity, replication, and competition provide the basic driving forces for diversity, but leaves out the unity among diversity so characteristic of biology.

From the point of view of physical chemistry, being an entity requires phase separations in the sense implied in the Gibbs phase rule. In aqueous environments the principle feature governing phase separation is electrical asymmetry usually noted by such attractor leading to the breakdown of entities. To maintain a non-equilibrium entity requires as a necessary condition the flow throught of energy from an effectively high energy (high frequency) source to an effectively low energy (low frequency) sink. (Morowitz, 1968)

There are two principle sources of energy flow in earth-like planetary systems. The first is the flux of solar energy with a color temperature of the order of 5750K. The second is the release of magmatic reductants from beneath the crust. The leak of hydrogen from the planet leaves sufficient oxidants so that the planetary oxidants and reductants form couples whose reactions lead to the release of energy. These two sources seem to account for all biotic energy flux. The sink for both of these processes is the 3K cold of outer space. The planet is constantly radiating infrared to this cosmic sink.

To think of a model of the equilibrium state to which a living system will decay, take some reference living system such as an organism, a terrarium, and aquarium or any of the microcosms and mesocosms that have been studied and surround that system with a wall which will permit the flow of heat and is imperable to the flow of matter and light. Place the isolated system in an isothermal reservoir at its initial temperature and allow it to age. It will slowly decay to equilibrium and maximum entropy following the second law of thermodynamics which applies to such isolated systems. Dayhof et al (1964) have developed an algoithm to determine the chemical composition of a specified system of carbon, hydrogen, nitrogen and oxygen atoms when it reaches the equilibrium state. This is done by minimizing the Gibbs free energy subject to the constraints.

The equilibrium systems are dominated by small molecules such as $C0_2$, H_20, N_2, NH_3, etc. This decay toward equilibrium, which is a true global attractor is also taking place in

Amphiphilic molecules in water form a number of highly specific microphases, including various coacervates. Of special interest in biogenesis is a phase consisting of a planar biomolecular layer with the apolar portions in contact with each other and the polar portions in contact with the surrounding aqueous medium. Such structures can form into another microphase, vesicles, with an aqueous interior, a biomolecular leaflet membrance and an aqueous exterior.

Consider next contemporary biological structures. There are three basic methods of structure building: macromolecules, membranes, and combinations of the above. The first two are composed of molecular subunits. In biological macromolecules these are the monomers which may form either precise or statistical polymers. Nucleic acids and proteins are precise polymers and carbohydrates may exist in straight chain or branched polymers which do not have completely defined sequences although there is a statistical regularity in nearest neighbor frequencies.

In membranes the subunits analogous to monomers are the amphiphiles. The membrane structure forms as a local free energy minimum with only the most general of rules governing the composition in terms of the amphiphilic subunits. This is because the subunits are not covalently bonded to each other. In short there are two types of fundamental structures, each made of subunits in the molecular weight range 80-600 daltons. In the first class of structures the sequences and other details are precisely defined as in the overall structure. In the second the governing principle is less informational and more thermodynamic so that a much greater degree of variation is permitted in the components and their relation to each other. Biological entities are predominantly bounded by the imprecise or sloppy variety within the range of chemical interest where concentrations of 10^{-45} moles/liter are truly homeopathic.

Other types of entities are of course possible including crystals, glasses, and coacervates. These do not lend themselves to the roles of reaction vessels and to the requirement of self replication. Although coacervates have been suggested (Oparin 1957, Fox), the fact that vesicles map over directly into the contemporary structure of biological cell plasma membranes is another argument for favoring them as the initial entities. The principle of continuity is an important methodological construct (Morowitz 1992, Fry 1995). Other surfaces such as clay and pyrite have also been suggested but they present problems in continuity.

Some self replicating far from equilibrium entities can in principle act in the most general way by acquiring material from the environment and processing it to give rise to two or more similar entities each with the ability to continue the process.

Associated with self replicating far from equilibrium systems is an energy requirement. There is a constant second law driven thermal degradation toward equilibrium that must be countered by the expenditure of work. Thus there is the requirement of a proper energy source and the ability to convert that energy into a form that is useful in maintaining the structure and synthesizing similar structures.

In order for a vesicle entity to become a self replicating entity it must exhibit two features: the ability to synthesize membrane amphiphiles from substances in the environment and the ability to process available energy to develop an anabolism to counter the catabolic decay inherent in the thermal degradation which has been noted above is closely associated transfers involving the hydrolysis of phosphate ester bonds and oxidation-reduction reactions. The two classes are themselves coupled at membranes chemo-osmotically (the Mitchell hypothesis) so there is great unity to all biotic energy processing. Restricting the chemical forms of energy flow in a reacting system restricts the possible chemical networks in that system. Thus the universal features of energy processing limit the available biochemicals in an emerging metabolism. These are very significant constraints.

Given a group of complex self replicating far from equilibrium entities, differences will arise between the entities due to chance occurrences. If these differences persist through the process of replication, the phenomenon of post replication memory generates recognizable species of entities. In contemporary biology, memory is compartmentalized in the DNA genome and species differences are by and large genetic differences. The phenomenon of memory need not always have been reified in such a precise and elaborate way. For example, some biologists have postulated an earlier state of evolution when RNA was both the message and the genetic structure. This requires considerably less molecular apparatus than the present structure. Memory can also exist without specific macromolecules but may be instantiated in chemical networks, catalytic loops and reflexive autocatalysis. Any of these simple ways also leads to speciation, that is to populations of distinguishable replicators with different chemical processes, efficiencies, and other physicochemical differences. Memory does not require macromolecules.

Replicators by their very nature must use environmental material and energy sources. The inherent finiteness of resources of necessity puts distinguishable species of replicators in competition with each other. Thus all the Darwinian features of survival and fitness emerge.

dioxide as the starting material. In trying to reconstruct prebiotic chemistry we thus focus on autotrophic pathways. This is in line with the postulate that biogenesis proceeds from chemical simplicity to chemical complexity. We can hierarchically order compounds by counting how many chemical reaction steps they are from carbon dioxide. There will be an index number assignable to each carbon in a compound. Note that in highly reducing environments carbon dioxide may be replaced by formic acid via the redox reaction.

$$CO_2 + H_2 \rightleftharpoons H - \overset{\overset{\displaystyle O}{\|}}{C} - OH$$

3. At some early stage phase separation was required. This may be archived in one of two general ways: (a) absorption on a surface or (b) trapping in a coacervate. For the former suggestions have been made of clay water interface, air water interface and pyrite water interface. For the latter oil droplets, proteinoids, and amphiphile vesicles have been suggested. Of all of these possibilities, only vesicles show a high degree of continuity with the universal cell structure. Vesicles require at the outset amphiphiles in the environment. This does put some demands on the chemical composition of the environment. The requirement is nonspecific. Vesicles can be made from complex mixtures of amphiphiles. Only small amounts of such surface agents are required. In any case, very early in the biogenesis of the kinds of cells we know, a major task must have been the synthesis of amphiphiles. This will be true regardless

excited electronic state to a chemically useful form before it causes random chemical reactions which tend to be destructive. An example of this occurs in the photodynamic effect where visible energy absorbed by vital stains leads to cell death. The fact that there are only two classes of photosynthetic pigments in all of biology, chlorophylls and retinals, suggests that there may be something very special about an appropriate energy transduction.

5. At the prenoetic level before thought appeared, biological processes can be analyzed and reduced to constructs of quantum mechanics, thermodynamics and kinetic theory as embodied in the periodic table and chemical bonding theory. Biogenesis should be reduceable to the level of experimental laboratory chemistry.

6. The development of biological complexity of which biogenesis is a part can be analyzed through a hierarchy of nested reaction networks involving increasing complexity. Nests are of course hierarchical and represent the evolutionary course of biogenesis. At the core the behavior tends to be governed by deterministic physical chemistry and as one move out frozen accidents play an ever increasing role in the historical unfolding of biology. As has been noted (Smith and Morowitz, 1978) biology stands between physics and history.

Metabolic Pathways

Since the subsequent analysis is based predominantly on an analysis of the metabolic pathways chart, we will outline some of its general features. It is a vast empirical generalization based on a century and a half of work by an army of biochemists who have labored to characterize all of the chemical reactions taking place in living cells. It begins with

This is distinguished from ordinary graph theory in which all connections are of the class a.

The chart as usually drawn in biochemistry texts reflects both its historical origins and the stress on human metabolism and medicine which has been a driving force for the science of biochemistry. For our purposes we shall make the distinction between primary and secondary metabolism. The former is that part of the chart which is shared by all autotrophs. Heterotrophs will be missing those portions of primary metabolism whose products are obtained from the environment by absorption or ingestion. Primary metabolism consists of those reactions that transduce energy, synthesize necessary monomers, cofactor and other small molecules and assemble the molecules of cellular functions: proteins, nucleic acids, carbohydrates, and lipids. The number of relatively low molecular weight molecules involved in the core metabolism is limited in the order of a few hundred compounds. Primary metabolism cuts across all taxa and is clearly close to or at the evolutionary origin.

Secondary metabolism encompasses a vast number of taxon specific reactions that have evolved along the tree of life and are the biochemical part of evolutionary radiation. An example of a secondary metabolite is cholesterol which occurs only in the *Animalia*. Another example of secondary metabolism is involved in the excretion of nitrogen by animals. In the process of eating foods for energy, animals ingest far more nitrogen in protein than they require. This excess must be excreted. Marine animals tend to do this as ammonia. Most mammals excrete urea and birds and reptiles tend to uric acid for nitrogen elimination. These secondary pathways are clearly not required by plants and bacteria. We will confine our subsequent discussion to primary metabolism.

carbon dioxide into metabolic intermediates we look at present day carbon fixation to locate the core more exactly. Among the autotrophic bacteria there are three CO_2 fixation pathways.

1. The major entry of CO_2 in the present day biosphere is by a reaction which may be simplified.

   Ribulose 1,5 bisphosphate + $CO_2 \rightarrow$ 2 molecules of 3 Phosphoglycerate

The energy to generate Ribulose 1,5 bisphosphate comes in the form of ATPs and NADHs which in photosynthetic organisms are generated in the light reactions. The enzyme that catalyzes the reaction shown above is ribulose 1,5 bisphosphate carboxylate which is also known as Rubisco.

Ribulose 1,5 bisphosphate is synthesized in the rather complex Calvin cycle involving fructose 6 phosphate, glyceradehyde 3 phosphate, xyulose 5-phosphate, erythrose 4 phosphate, dihydroxyacetone phosphate, sedoheptulose 7-phosphate and ribose 5-phosphate and the enzymes transketolase, aldolase, phosphatase, phosphopentose epimerase and phosphoribulose kinase. Thus CO_2 incorporation through this pathway requires a complex set of enzymes as well as a source of pyrophosphate and reducing power as NADH or similar compounds. These can come from photosynthesis or reactions of environmental oxidants and reductants for some lithoautotrophs. Rubisco is a dominant protein in most ecosystems.

2. A second family of CO_2 fixation methods is referred to as the reductive acetyl CoA pathway. The reaction is characterized by one highly reductive pathway going to CH_3- while another CO_2 goes to the carboxy acid level and the two combine to acetyl CoA. This system uses complex cofactors of the tetrahydrofolate system. Further CO_2 incorporation takes place in going from acetyl CoA to pyruvate to oxaloacetate either

This cycle has the following properties:

(a) It is autocatalytic since one molecule of citrate leads to the synthesis of two molecules of citrate

(b) The entire process involves only eleven kinds of molecules.

(c) The reaction direction is determined by a strong reducing environment and CO_2 as an oxidant leading to citrate and water.

(d) The cycle lies at the core of biosynthesis and contains the precursors of all the major synthetic pathways of autotrophs.

(e) Only two molecules in the network are chiral: malate and isocitrate, and neither of these is a precursor along the main line synthetic pathways.

<u>The Conjecture</u> - We postulate that the reductive citric acid cycle is part of the earliest biological chemistry. We further postulate that the earliest biochemistry functioned without enzymes. The reductive citric acid cycle plus the pathway to amphiphile synthesis fulfills the logic of a self replicating system (Morowitz et al 1988) in that it presents a continuous synthesis which also synthesizes membrane so that the entity feature can be maintained. The reductive TCA mode is the simplest engine of synthesis in terms of accessory molecular hardware and one which starts with the thermodynamic ground state molecules CO_2 and reductant.

The key to moving from simplicity to complexity is thus generating larger molecular structures and requires the synthesis of C-C bonds between carbons in small precursor self replicating systems in the sense that we have discussed above. If they are in an environment with appropriate energy sources and sufficient low molecular weight precursors for the synthesis on amphiphiles the membranes will grow and the vesicles will eventually divide.

A second postulate of the theory is that the evolution of self reproducing systems proceeds by adding layers of complexity to the existing reaction network while maintaining the essential core structure. The best understood instantiation of this idea can be seen in the reaction networks of intermediary metabolism. We will analyze these briefly in terms of shells of reactions in autrotrophs and later return to a more detailed view. Because new functions add at the periphery we will further postulate that the intermediary metabolism of autotrophs recapitulates biogenesis.

The Metabolic Chart - Shells A, B, and C

The central shell of the metabolic chart consists of the citric acid cycle, the pathway from acetate to long chain fatty acids and the pathway from acetate to glucose which in the other direction is called glycolysis. Although numerous cofactors and enzymes are used at the substrate level the shell consists of compounds of CH and O. The role of phosphorus will be discussed later. Sulfur and thioesters appear as important intermediates even at the earliest time (deDuve 1995).

The inner core of metabolism is the citric acid cycle in which, under aerobic conditions, acetate is oxidized to two carbon dioxides with the concomitant reduction of three NAD and one FAD from the energy released. The reduced compounds are central in the an experimental question and is currently being investigated.

The entry of nitrogen into metabolic processes has certain features of interest. At present it enters the metabolism of the biosphere in the most reduced form, ammonia. It enters by a very small number of pathways with one dominant, which involves glutamate and glutamine. This statement is of ecological as well as biochemical significance. Although nitrogen may enter the cells as atmospheric N_2 or various oxides it is always reduced to NH_3 before entering the synthetic pathways.

In the overall scheme of metabolism, shells are connected by gateways. In designating shells we only deal with the substrate level molecules, not with cofactors and enzymes.

The first shell then consists of glycolysis, the citric acid cycle and the pathways of fatty acid synthesis. The gateway to the second or B shell includes of those reactions which convert keto acids to amino acids. The second shell encompasses most pathways to amino acid synthesis. The third shell (C) involves incorporating sulfur into the amino acids cysteine and methionine. The gateways to the fourth or D shell involves ring closure and synthesis of nitrogen and dinitrogen heterocycles. The fourth shell consists of purines and pyrimidines and their derivatives and a large number of cofactors. Most water soluble vitamins are shell D molecules.

Other shells can be assigned in terms of sterol synthesis and diverse reaction families. The first shell is concerned with energetics, amphiphile synthesis and precursors to biosynthesis. The second shell is concerned with amino acids and catalysis. It is the point where chirality becomes central in biosynthesis. The third shell provides the basis for more sophisticated protein structure. The fourth shell produces molecules to fine tune reactions in-