The Jigsaw Model: An Explanation for the Evolution of Complex Biochemical Systems and the Origin of Life

(January 18, 2000)

John F. McGowan, Ph.D.
Desktop Video Expert Center
NASA Ames Research Center
Mail Stop 233-18
Moffett Field, CA 94035-1000
E-Mail: jmcgowan@mail.arc.nasa.gov
Telephone: (650) 604-0143

ABSTRACT

Living organisms contain complex biochemical systems of co-adapted molecules, usually proteins, that are difficult to account for through random variation and natural selection. Living organisms also contain complex biomechanical systems of co-adapted parts that are difficult to account for through random variation and natural selection. Similarly, even the simplest living organisms appear to be complex biochemical systems of co-adapted molecules. It is extremely difficult to explain how such a first living organism could have been assembled by random combination of molecules in the hypothetical prebiotic soup. A mechanism is proposed to account for the appearance of complex biochemical systems of co-adapted molecules, essentially machines, both in living organisms and in the prebiotic soup. This mechanism is not consistent with current understanding of the genetic code and, if true, would indicate that current understanding of the genetic code is incomplete. The mechanism may account for the appearance and function of supposedly non-coding or "junk" DNA in the genome. Some speculations on the practical value for medicine of finding and elucidating this mechanism, if it exists, are presented.

1. INTRODUCTION

The argument from design infers the existence of a designer from the apparently designed character of living things. A watch implies a watchmaker. Darwinian evolution, now widely accepted, accounts for the apparently designed character of life through inheritance of random variations, mutations, and natural selection[1].

Darwinian evolution is expected to be a slow, gradual process in which gross changes are produced by the accumulation of small changes over hundreds of thousands or millions of years. These changes are produced by random mutations in the DNA within the genome. Since the mutations are random most mutations are expected to be negative mutations that do not enhance the ability of the life form to survive and reproduce. These are eliminated by natural selection. The few mutations that enhance the ability of the life form to survive and reproduce are selected by natural selection. Over time these mutations accumulate, generating new species and complex structures such as wings and the camera eye.

Although Darwinian evolution is widely accepted, the theory does have problems and a few scientists have questioned it since the time of Darwin, including, for example, Louis Agassiz, St. George Mivart[2], Richard Goldschmidt, Otto Schindewolf, Pierre-Paul Grasse[3], Michael Denton[4], and Michael Behe[5]. There are two main obstacles to Darwinian evolution – the fossil record and complex organs, what Michael Behe has claimed are “irreducible complexity” in living things.

Darwinian evolution predicts the existence of many intermediary forms leading to current life forms. The fossil record should show a slow, gradual, continuous transformation through a series of intermediary forms from earlier life forms into creatures such as man. While some examples of this appear to exist, the actual fossil record contains mass disappearances of species, mass proliferation of new species in geologically short time periods such as the Cambrian explosion of 600 million years ago, and long periods of stasis. A number of “living fossils” have been discovered that appear identical to ancestors millions of years ago. This has usually been explained by arguing that the fossil record is incomplete. Stephen Jay Gould and Niles Eldredge proposed the theory of "punctuated equilibrium” in which evolution occurs rapidly in small isolated populations to account for these discrepancies[6]. A radical interpretation would be that some non-Darwinian process is rapidly transforming species or creating new species in a geologically short time.

The second problem is organs of extreme complexity or irreducible complexity in the language of Michael Behe. An irreducibly complex system is a system of several parts that must work together to produce a useful result in which omission or even a small change to any part results in failure of the system. Irreducibly complex systems are difficult, although not impossible, to produce through Darwinian evolution, that is random variation and natural selection. Since the system only works with all parts, it cannot be built up one part at a time by Darwinian evolution. Since the parts are tightly

coupled, one must change all parts simultaneously to improve the system. If only one part is slightly changed, the system will fail. Since the time of Darwin, organs such as the camera eye have been argued to be irreducibly complex. Although the language irreducibly complex was not used, the thinking was essentially the same. Michael Behe has argued that several biochemical systems such as the blood clotting system are irreducibly complex.

It is believed that the genetic material of living things contains instructions for constructing each part in an irreducibly complex system, a code. In particular, there is believed to be a one-to-one relationship between genes and distinct proteins. One gene codes for one and only one protein. In this textbook version of biochemistry, distinct proteins are coded independently in the genome. This is significant. Michael Behe's arguments hinge on this view. Examples of irreducible complexity in living systems are cascades or networks of distinct proteins that work together to produce a useful result such as blood clotting. With random variations, most variations or mutations will produce a defective or inferior code for a part of a living system such as a protein. For example, radiation induced mutations in fruit flies have almost never produced a positive mutation. The probability of a positive mutation must be extremely small, less than 1 in 100,000 mutations. The immune system is believed to try over 100,000 different antibodies to find one that is effective against an intruder. Further it is doubtful whether the immune system is performing a purely random search through the space of possible antibodies. To create or modify an irreducibly complex system, where the parts are independently coded in the genetic material , at least two separate mutations must occur simultaneously, changing two or more parts together, so that they work together to produce a useful result. This is not impossible, but highly unlikely.

Darwinian evolution may account for these complex systems if they are not irreducibly complex. No rigorous means of proving that a system is irreducibly complex exists as yet. A second argument is that the irreducibly complex systems or their major components may have developed for some other purpose, for which they were not irreducibly complex. Then, some chance mutations, such as a chance combination of two systems evolved for other purposes, generated the final irreducibly complex system. Some irreducibly complex systems could be chance combinations of several mutations, flukes.

Another important possibility is that the parts forming irreducibly complex systems in living creatures are not independently coded. In this case, a single variation, a single mutation, could simultaneously change two or more parts while preserving their ability to work together.

The problem of the fossil record and the problem of irreducibly complex systems in life forms are complementary. At first glance, irreducibly complex systems need to be created in a single discontinuous jump, not by means of a slow accumulation of features. Thus, if some unknown non-Darwinian process is generating the irreducibly complex systems that may be present in living things, then one would expect to see discontinuous, abrupt changes in the fossil record. Accordingly, the possibility that some unknown non-Darwinian process acts as a designer within the development of life should be considered.

2. SYSTEM-LEVEL GENETIC CODING AND THE ORIGIN AND EVOLUTION OF COMPLEX BIOCHEMICAL SYSTEMS

Thus far, the discussion has assumed that the different parts of irreducibly complex systems are independently coded. For example, distinct proteins are coded by independent genes. One can call these theories of the genetic code part-level genetic codes. In part-level genetic codes, classical Darwinian natural selection cannot account for irreducibly complex systems in a naturalistic manner without resort to extremely improbable coincidences. Multiple independent parts must change simultaneously in what appears to be a highly intelligent manner to create an irreducibly complex system or transform one irreducibly complex system into another irreducibly complex system.

If one discards the notion that different parts in an irreducibly complex system are independently coded, then much less intelligence is required to create and modify such systems. The typical characteristic of irreducibly complex systems is that the parts fit together at a boundary, such as a key in a lock. A lock and key is an irreducibly complex system. Both the lock and key are required to produce a useful result. Secondly, any small change in either the lock or the key results in failure of the system; the key cannot open the lock. This is achieved by a close matching or fitting between the edge of the key and the edges of the components within the lock. The sharp transition at a physical edge gives rise to a very tight tolerance on the system.

Similarly two or more proteins are physically adapted to one another. They fold into complex three dimensional shapes that fit together like a lock and key. In theory, a mutation of the gene coding for a protein results in a different shaped protein that no longer can work with its partner. A mutation can produce a completely different protein shape. The change is not localized to the region of the modification in the protein chain. If the two proteins are coded independently, then two simultaneous mutations are required to modify the system successfully.

A system-level genetic code is a genetic code, a system of instructions for generating a living thing, that codes systems of co-adapted parts. Parts are not coded independently. A mutation in a system-level genetic code changes several parts simultaneously. The system-level genetic code prefers or exclusively represents systems of co-adapted parts.

One can code for mechanical parts such that they preferentially or exclusively form co-adapted systems that are irreducibly complex. A simple toy, the jigsaw puzzle, is a good example of this. A jigsaw puzzle can be very simple, two pieces, or very complex, any number of pieces. Consider two pieces of a jigsaw puzzle that fit together perfectly. The whole jigsaw puzzle is an example of an irreducibly complex system. If we modify or omit any single piece of the puzzle separately, the puzzle is broken.

If each piece of the jigsaw puzzle is coded independently, for example as an ordered list of vertices of the piece, then a single mutation – add or subtract a vertex or change the location of a single vertex in the code for one piece – can only change one piece in the puzzle. In a jigsaw puzzle coded this way any single mutation will break the puzzle. The mutated piece will not fit in the puzzle. To successfully modify the jigsaw puzzle so that

it is still a jigsaw puzzle, one must modify the codes for at least two pieces simultaneously. In this coding scheme, all single mutations and the vast majority of multiple simultaneous mutations result in a broken puzzle.

The reason that a single mutation invariably breaks the puzzle is that each piece is coded independently of the other pieces in the jigsaw puzzle. This is not actually how jigsaw puzzles are made. As the name indicates, jigsaw puzzles are made by starting with a single larger piece and cutting it into pieces with a figurative or an actual jigsaw. This procedure always produces a set of pieces that fit together. One can code for the jigsaw puzzle as: start with a single larger piece with three or more vertices. The initial piece is coded as an ordered list of the vertices. Then give instructions for cutting, typically repeatedly bisecting, the puzzle, with a jigsaw. For example cut down 4 centimeters, turn 90 degrees, cut 3 centimeters, turn 45 degrees, cut 6 centimeters, and so forth. This can construct a very complex system, the jigsaw puzzle, that is irreducibly complex. If the jigsaw puzzle is coded this way, any single mutation will change the entire puzzle, not just one piece. However, all of the pieces of the jigsaw puzzle will fit together regardless of what mutation or mutations occur. No matter how the instructions for cutting the puzzle are changed, pieces that fit together are always produced. Random mutations in this coding scheme never produce a system that is not irreducibly complex. A working puzzle is always produced.

The actual mechanism by which morphological features such as bones that fit together in living creatures are produced is not understood. It is possible that mechanical parts of living things are coded in the genetic material using a system-level genetic code conceptually similar to the jigsaw example. The different parts could be coded in an interdependent way. Indeed this seems rather likely. Many studies – for example, gene knockout studies in mice where a single gene is disabled - indicate that a single mutation usually affects more than one, indeed many, organs within an organism simultaneously. This is to be expected if a system-level genetic code biased toward coding complex systems is used.

Systems of proteins that work together to produce a useful result are more difficult to explain. It is generally believed that each protein is coded independently of its partners in a complex biochemical system, for example the blood clotting system. The textbook rule is one gene, one protein. It is for this reason that Michael Behe's Darwin's Black Box uses examples of biochemical systems to challenge Darwinian evolution, rather than physical systems such as the eye or wing. The book implicitly assumes that system-level genetic codes are conclusively ruled out in biochemical systems. A clear exception to this is pleiotropy where longer proteins are assembled from shorter proteins. Several different longer proteins in several different parts within an organism are all assembled from a shorter protein apparently coded by a single gene. Change this gene and several different parts are simultaneously changed.

How could a system of proteins that work together be produced? Start with a very large, very long, possibly entirely random protein. In water, the protein will fold together to form a tangled ball of protein. This folding is not a random process. The sections of the protein that are co-adapted, have a chemical affinity for one another, will attach to one

another during the folding process. Even though we start out with a purely random, white noise, protein, it self-organizes into a tangled structure where parts that fit together are adjacent to one another.

Next, envision a jigsaw molecule that bisects the tangled ball protein. This could be a physical cutting of the tangled ball. The jigsaw molecule acts like a physical saw or a pair of scissors. Alternatively, the jigsaw molecule could use a physical or chemical mechanism to preferentially attack open loops within the tangled ball where the amino acids on the open loop do not exhibit a chemical affinity. The jigsaw molecule would not attack regions, amino acid sequences, that bind together. This would break the tangled ball into separate pieces with a chemical affinity for each other. Once the tangled ball is bisected or fragmented, it becomes two or more proteins depending on how the tangled ball protein is folded and where the jigsaw molecule cuts. These proteins are not produced randomly. These are proteins that fit together. This two-step process can produce a system of co-adapted proteins that tend to react with one another.

If the process is random, the system of proteins generated in this way will usually not be very effective. It would be an example of an irreducibly complex system that doesn't work or works poorly. However, a single variation, random or not, would generate another, different, system of co-adapted proteins. Some of these systems would work, do something useful collectively. This jigsaw mechanism could search the space of irreducibly – the reader may substitute extremely complex – complex systems by a single mutation search. There would be no need for multiple simultaneous mutations to produce a modified system that works. A single mutation would suffice – either in the original tangled ball protein or in the jigsaw molecule, that is the rules for cutting the tangled ball. This single mutation would be biased by the hypothetical mechanism toward producing a system of partially or fully co-adapted proteins, rather than a random collection of proteins. Furthermore, this single mutation at a genetic level could correspond to a systemic macromutation or saltation, a gigantic change in the living creature. At a genetic level, it is a small change that could be passed on to future generations. A "hopeful monster" produced in this way would still be able to interbreed with the rest of its species.

A genetic code in which different parts of complex systems are not independently coded is not sufficient to resolve the evolutionary conundrum. If this code can represent systems of parts that do not fit together with equal probability as systems of parts that do fit together, then the probability of a mutation yielding a working irreducibly complex system is still tiny. If each part is coded independently, then the probability of a mutation, actually two or more simultaneous mutations, is astronomically small. Thus, any genetic code that codes the parts of a complex system independently is unlikely to produce irreducibly complex systems through random variation. However, many possible genetic codes that code a complex system in an interdependent manner where any change in the coded representation changes more than one part of the system also have this problem. Only genetic codes that preferentially or exclusively represent systems of parts that fit together resolve the problem of generating irreducibly complex systems through Darwinian evolution. In this case, the random variations search a restricted subset of the possibilities, the space of all systems with co-adapted parts instead

of the space of all systems, for a viable system. In this paper, these genetic codes are called system-level genetic codes.

The tangled ball and jigsaw molecule scheme is one example of a mechanism by which proteins in extremely complex systems could be coded in an interdependent way. Other mechanism may exist. If the proteins are not coded independently as currently believed, then the evolution of irreducibly complex, or extremely complex, systems of proteins may be accounted for. Given that such irreducibly complex biochemical systems appear to be observed, it seems prudent to look for evidence of such complex coding methods in genetics.

2.1 HOW HUMANS DESIGN AND CODE COMPLEX SYSTEMS AS A MODEL FOR A SYSTEM-LEVEL GENETIC CODE

Humans code irreducibly complex systems all the time. The ways in which human beings design and encode complex systems may explain how system-level genetic codes function in living organisms. The end product of most human designs of complex systems is a part-level description of the complex system, such as a blueprint or computer program. In this part-level description, the parts that must work together appear to be coded independently. Yet the parts are perfectly co-adapted to work together. For example, a signal encoder and decoder such as a video compression system typically consists of an encoder program and a separate decoder program. Both must be created in parallel to work together. Clearly human beings are able to keep the two distinct parts in sync.

Human design is a mystery, only partially understood. Typically top-down design is used to create a complex system. The classic example is to start by representing the entire system as a single block, e.g. "MARS TO EARTH COMMUNICATIONS LINK", in a diagram. One cannot manufacture a working system from this single block design. It cannot directly code for manufactured parts. Then the human designers segment the top level block into smaller parts, e. g. "MARS TRANSCEIVER" and "EARTH TRANSCEIVER". Then the subdivision is repeated, e.g. "MARS TRANSCEIVER" is divided into "MARS TRANSMITTER" and "MARS RECEIVER". "EARTH TRANSCEIVER" is divided into "EARTH TRANSMITTER" and "EARTH RECEIVER". So far, the design still cannot code directly for manufactured working parts. In a very rigorous system design, an interface between the parts may be specified at each stage. The subdivision process is iterated until a low level design where each block has a simple one to one relationship with a simple part, a fundamental building block that cannot be subdivided, is reached. Then, the low level design is translated into a part-level description of the system, such as a blueprint, a synthesizable chip design in VLSI chip design, or a computer program. The exact algorithm, if it is an algorithm, by which humans subdivided the high level blocks is not known. Therein lies one of the mysteries of human intelligence.

The actual code for a complex system designed by humans consists of all levels of the design. Only the final part-level description is ultimately used or absolutely necessary to build the actual working system. Human designers usually retain the documentation for

all levels of the design. In a hypothetical system-level genetic code, the part-level description would correspond to the obviously coding sections of the DNA known to molecular biologists. Yet DNA contains large sections that do not seem to code directly for anything. These are currently interpreted as junk DNA that doesn't do anything.

The high level designs cannot code directly for the living organism. They are related to the coding design by a cascade of subdivision operations that break the organism down into manageable parts in an interrelated way. If this is an appropriate model of the hypothetical system-level genetic code, then there must exist a mechanism to subdivide the high level blocks into coding DNA sequences. This mechanism has either not been observed or not recognized for what it is. A mutation in the subdivision of a high level block can affect many different coding genes simultaneously in a coordinated way. A mutation in the part-level description, the coding genes, would almost always break the system. Ironically, only a mutation in the "non-coding genes" could produce a new working complex system.

In practice, despite the existence of formal top-down design methodologies such as Critical Path Management (CPM), Program Evaluation and Review Technique (PERT), structured software design, and object oriented software design, actual human design and representation of a system is more complex and less well understood. Actual designers go through a complex top-down, bottom-up, intuitive leap, and so forth process to produce actual working designs. The simple model above is at best a crude approximation to the actual mechanism.

In this hypothesis, the genome contains a system-level description for the organism, DNA that does not code directly for the organism, and a part-level description that directly codes for the organism. Molecular biology has partially elucidated the part-level description but largely missed the system-level description.

3. THE ORIGIN OF LIFE

The jigsaw mechanism and system-level genetic codes may explain the origin of life. One of the principal problems with the origin of life is that even the simplest living thing, based on current knowledge, would have to be a complex system of parts, rather than a single self-replicating part, for example a single molecule. All existing life forms, even the simplest, are complex systems of parts, a single cell organism, and not self-replicating molecules.

Most origin of life research has focused on some variant of a prebiotic soup in which complex organic molecules are randomly generated by some process, such as the electric discharges in the famous Urey-Miller experiments. This essentially random chemical process eventually generates a single self-replicating molecule. This molecule has variously been speculated to be DNA which seems very unlikely, a self-replicating RNA molecule, a protein, or various hypothetical organic molecules no longer present. The prebiotic soup provides the nutrients, the building blocks, for this self-replicating molecule to make copies of itself. Eventually, over hundreds of millions of years, this

self-replicating molecule evolved into a simple one-celled organism by Darwinian evolution.

Even if one could find some plausible pre-biotic soup of chemicals that generated DNA, RNA, or proteins in some random or pseudo-random manner, no example of a self-replicating DNA molecule, RNA molecule, or protein exists. One is left with the apparent need for several different molecules to come together, in what can only be described as a miracle, to form the first self-replicating organism.

The focus on finding a means to randomly generate a self-replicating molecule reflects the apparent difficulty of generating a system of cooperating co-adapted molecules. Intuitively, this does not seem like something that could be produced randomly. Rather it seems to require intelligent design. The self-replicating molecule focus also reflects a strong bias in chemistry toward the design and synthesis of single chemicals, such as drugs. The dream in drug design is typically a single "magic bullet" drug that selectively targets a defect, such as cancer cells, and leaves everything else in the living system alone. Antibiotics are the prototype of this model. The prevalence of side-effects in real world drugs illustrates the complex co-related system nature of living systems.

The tangled ball organic polymers such as proteins and jigsaw molecule mechanism might provide an explanation for the origin of life, particularly if a relatively simple molecule or physical process such as an acid could act as a crude jigsaw in the pre-biotic soup. Then, the combination of the randomly generated long chains of organic polymers such as proteins and the jigsaw molecule randomly generates many systems of partially or fully co-adapted organic molecules. This mechanism is more plausible if the first single-cell organism was comprised of proteins alone. The DNA and RNA mechanism of information storage and retrieval was evolved subsequently.

Since the jigsaw molecule can break down the organic molecules, the source of long organic polymers should be physically separated from the jigsaw molecule source, for example at two ends of a volcanic lake or ocean inlet. Then the long chain polymers, the tangled balls, diffused through the lake toward the source of the jigsaw molecules. The tangled balls encounter the jigsaw molecules and the systems of shorter co-adapted polymers were produced. Most of these systems would not be alive and would soon fall apart and dissipate in the pre-biotic soup. However, one system coalesced into the first single-cell organism or into a self-perpetuating or self-reinforcing system that eventually mutated into a truly self-replicating system. The organism might begin to replicate, spreading through the lake or ocean. Initially, the organisms diffusing into the region where the jigsaw molecule is produced would die while the others diffusing away would survive. The jigsaw molecule might be incorporated into this first organism, meaning that the system-level genetic code might date to the origin of life, permitting Darwinian evolution of irreducibly complex systems from the very start of life. Alternatively the jigsaw mechanism might have been discarded at some point, leaving the seeming current genetic code where each protein appears to be coded by one and only one gene.

4. CONCLUSIONS

The existence of complex biochemical systems in many living organisms and the seeming appearance of jumps in the fossil record suggests that a system-level genetic code exists not only in the genes controlling the development of embryos and morphology but also in the genes controlling the formation of complex biochemical systems such as the blood clotting cascade. The apparent origin of life on Earth may suggest that a chemical or physical mechanism such as the proposed jigsaw mechanism existed on the early Earth. This mechanism produced systems of complex co-adapted molecules, most probably through bisection or fragmentation of larger, more complex molecules.

If a system-level genetic code exists, the medical value of identifying and understanding the system-level description almost cannot be overstated. One of the principal problems in drug design and medical treatment design is the complex interrelationship between the many chemical and physical parts of the organism. The system-level description of the human body would explain these relationships, allowing understanding, prediction, and control or even elimination of dangerous side effects. Cancer cells may be an example of a systemic macromutation caused by a mutation in the system-level genetic code. Essential aspects of the aging process, the ultimate cause of most leading diseases in the industrial world, may be coded in the system-level genetic code.

5. ABOUT THE AUTHOR

John F. McGowan is an engineer and researcher in the field of digital video at NASA Ames Research Center. He has worked on video and still image quality metrics, perceptual optimization of JPEG still images, and MPEG digital video decoders. He has a Ph.D. in Physics from the University of Illinois at Urbana-Champaign and a B.S. in Physics from the California Institute of Technology.

6. REFERENCES

[1] Darwin, Charles, Origin of Species, 6th ed., New York University Press, New York, 1872

[2] Mivart, St. George, On the Genesis of Species, Macmillan and Co., London, 1871

[3] Grasse, Pierre-Paul, Evolution of Living Organisms, Academic Press, New York, NY, 1977

[4] Denton, Michael, Evolution: A Theory in Crisis?, Adler and Adler, 1986

[5] Behe, Michael J., Darwin's Black Box, Simon and Schuster, New York, 1996

[6] Eldredge, Niles and Gould, Stephen Jay, "Punctuated Equilibria: An Alternative to Phyletic Gradualism" in Models in Paleobiology, ed. T.J.M. Schopf, Freeman, Cooper, and Co., San Francisco, 1973