Intelligent Design:
Required by Biological Life?

February 19, 2008

K.D. Kalinsky

I. Introduction:

This article is meant to stimulate thought and discussion. As that discussion unfolds, I expect that this article will be revised over time in the same way that a paper submitted to a journal is often revised during the process of review. The purpose of this article is to attempt to bring some clarity to the discussion of intelligent design and the origin and diversity of biological life. Essentially, we have two options. Either biological life required intelligent design or it did not. As with most problems in science, it is difficult to prove one option or another with absolute certainty. Instead, options can be evaluated against each other in an attempt to estimate which option is more likely. Even then, the fact that one option may be more likely than another does not 'prove' that it is actually the case. Instead, I will propose a way in which both options can be evaluated against each other. The results indicate that it seems highly likely that intelligent design was required for biological life.

II. Defining some terms and concepts

There is considerable confusion over what intelligent design is. Indeed, the concept is often used in different ways. It is sometimes used to describe a cause and other times used to describe an effect. For example, someone can ask if a laptop computer *requires* intelligent design or they can ask if it is an *example* of intelligent design. In the first use, 'intelligent design' is being used in the causal sense; it is a necessary cause for a laptop computer. In the second use, it is being used as a result or effect; intelligent design is the *result* of a prior cause, presumably a mind. For the purpose of this article, I will treat intelligent design as an effect. In other words, the question, 'does this laptop computer *require* intelligent design?, can be replaced by the question, 'is this laptop computer an *example* of intelligent design?' Of course, if one is inclined to be more exhaustive in their definitions then, as is often the case in lexicons, two or more definitions or senses of the term can be offered.

If we take intelligent design to be an effect, then we can define it as an effect that *requires* a mind. If we take intelligent design to be causal, then we can define it as the *process* of producing an effect that requires a mind. Since the common denominator in both uses is *an effect that requires a mind*, I will define intelligent design as follows:

Intelligent Design: an effect that requires a mind.

It follows from this that a necessary requirement for intelligent design is a mind. Of course, natural processes could also be necessary as well but, in this case, insufficient to produce the effect. Thus, at the very least, intelligent design requires a mind but may also require natural processes as well. In other words, natural processes may be necessary for intelligent design, but they are not sufficient; a mind is also necessary. The other option is the hypothesis that intelligent design is not required for a given effect. This second option must assume that natural processes are not only necessary to produce the effect, but they are also sufficient. A mind is not necessary. Thus, to be perfectly clear, this second option entails that *mindless* natural processes are *necessary and sufficient* to produce the given effect.

To illustrate the two options, let us imagine that the SETI Institute obtains a signal from outside the solar system that contains the first 50 prime numbers. If they were to conclude that it was more likely that a mind would be necessary to produce the signal than that mindless natural processes were sufficient to produce the phenomenon, then the signal would be a *possible* example of intelligent design. It would only be a *possible* example, due to the nature of scientific investigation; we could not be certain. No matter how improbable, it is still logically possible that the signal could have been generated by mindless natural processes. The best we could do is to weigh the probability that a mind could produce such a signal against the probability that mindless natural processes could do it and draw a conclusion as to which option was more likely. We know that a mind can generate the first 50 prime numbers, so the probability that a mind could produce that information is 1. If the probability that natural processes could generate the first 50 prime numbers is less than 1, then one can compare the two probabilities to decide how much more likely intelligent design is than mindless natural processes. If it turns out that intelligent design is ten times more likely, or a thousand more times more likely, then it becomes increasingly irrational to invoke mindless natural processes, and increasingly rational to invoke intelligent design.

> **Causal Hypothesis:** For any effect, either mindless natural processes are sufficient to cause the effect, or a mind is required.

The problem arises in estimating which of the two options is more likely. We need something that we can use to distinguish between examples of intelligent design and mindless natural processes. One possibility is the following hypothesis:

> **Intelligence Hypothesis:** an attribute that distinguishes a mind from mindless natural processes is the ability of a mind to produce effects requiring significant levels of functional information.

The above Intelligence Hypothesis allows that mindless natural processes can accidentally produce functional information within, say, the background noise of a physical system, but the information will not achieve a significant level. It also allows for the fact that a mind can mimic mindless natural processes by producing effects that do not require a significant level of functional information. We are left with the following questions:

1. How is functional information measured?
2. What constitutes a significant level of functional information?

Before we look at these questions, we will take a brief look at the role of intelligent design in science.

III. The role of intelligent design in science

Intelligent design plays an important role in at least three areas of science, including intelligent design detection, reverse engineering, and applied design (e.g., human intelligence applied to experimental design). In this section, I shall limit the discussion to the role of intelligent design detection in science. Intelligent design detection can be defined as follows:

> **Intelligent design detection:** the discipline of examining an effect and determining if it is an example of intelligent design.

An area of science where intelligent design detection is front and center is in the SETI Institute's ongoing search for extra-terrestrial intelligence. Radio and optical signals from deep space are monitored and analyzed to determine if the signal may have come from an extra terrestrial intelligence or not. Intelligent design detection is also important for archeology, where a distinction must be made between artifacts and effects due to natural processes. For example, ground penetrating radar can be used to search for ancient building sites and artifacts. The results must be continually analyzed to determine if what is being seen by radar is the result of mindless natural processes, or the product of intelligent design. Intelligent design detection is also central to forensic science, which concerns itself with whether the crime was carried out by an intelligent agent, in this case, human, or was due to natural causes. With advances in genetics and cell biology, and the discovery of molecular machines, molecular computers and functional sequence complexity encoded in the genomes of life, intelligent design detection has now become necessary in biology. Furthermore, the J.Craig Venter Institute's creation of a synthetic *M. genitalium* genome presents us with a genome that is known to have been built by intelligent design, and that contains five 'watermarks'.[1] Strictly speaking, it is the 'watermarks' that are known to be a result of intelligent design. Some have asserted that intelligent design has no place in science, but of course intelligent design detection is firmly entrenched and essential to SETI, archeology, and forensic science. Those who insisted that ID has no place in biology will have to admit that now that synthetic genomes are being constructed, with 'watermarks', intelligent design detection is now an issue in biology as well. The job of science is to come up with a general, scientific approach to intelligent design detection. One possible approach that has the potential of being general enough to be applied to SETI, archeology, forensic science, and biology is suggested by the Intelligence Hypothesis: examine the functional information required to produce the effect and then to decide if it is more likely than not that intelligence was required to produce that degree of functional information.

---

[1] http://blog.wired.com/wiredscience/2008/01/venter-institut.html

IV. Functional Information

The Intelligence Hypothesis suggests that intelligence can produce effects that require a significant amount of functional information. To proceed, we need a method to measure functional information and, second, we need to decide what constitutes a *significant* level of functional information.

Measuring functional information

A method to measure functional information has recently been published by Hazen *et al.* whereby functional information is defined as

$$I(E_x) = - \log_2[M(E_x)/N] \qquad (1)$$

where E_x is the degree of function x, $M(E_x)$ is the number of different configurations that achieves or exceeds the specified degree of function x, $\geq E_x$, and N is the total number of possible configurations.[2] To illustrate, suppose we inherit grandfather's safe that has a combination lock that requires three numbers, each within the range of 0 to 99. Since each number has 100 possibilities, and there are three numbers, $N = 100^3 = 1{,}000{,}000$ possible combinations. Let us suppose that the mechanism has a little slop to it such that one need only get within 1 digit of each of the three numbers. In other words, for each of the three numbers in the combination, there are actually three functional options. Therefore, the total number of functional combinations that will open the safe is $M(E_x) = 3^3 = 27$ functional combinations. The amount of functional information required to open grandfather's safe is therefore

$$I(E_x) = - \log_2[27/1{,}000{,}000] = 15 \text{ bits of functional information.}$$

As Hazen *et al.* point out, 'functional information quantifies the probability that, for a particular system, a configuration with a specified degree of function will emerge', where the probability is denoted by $M(E_x)/N$. Strictly speaking, the probability that Hazen *et al.* speak of is the probability P_f of achieving the function in a single sampling, or

$$P_f = M(E_x)/N. \qquad (2)$$

As more trials are attempted, the probability of achieving the function improves.

Estimating I_{sig}

This raises the second question; what constitutes a *significant* level of functional information (I_{sig})? The Intelligence Hypothesis suggests that the attribute that distinguishes intelligence from mindless natural processes, is the ability to produce significant levels of functional information. Mindless natural processes can accidentally produce effects requiring a low level of functional information. For example, if we were

---

[2] Hazen, R.M., Griffen, P.L., Carothers, J.M. & Szostak, J.W. (2007) 'Functional information and the emergence of biocomplexity', *PNAS* **104**, 8574-8581.

to move grandfather's safe down to the riverbank, and attach a water driven turbine to the dial, and install the turbine in a turbulent portion of the current, where the turbine could be turned either direction by the current, it is possible that, after a long enough time, the variable current may actually open grandfather's safe. Of course, the number of trials may vastly exceed 1,000,000 if the same combinations are mindlessly tried more than once.

Recall, as Hazen *et al.* point out, that probability is at the core of the equation to measure functional information. We must establish a relationship between the number of trials mindless natural processes have for the particular problem, and P_f.

A search by mindless natural processes is essentially a random walk, where the search proceeds in no set direction and, for any point in the search, it can be returned to any number of times. This is not to be confused with an evolutionary search that is directed by a fitness function or a fitness landscape, which will be discussed later. We must first establish I_{sig} for a mindless natural search. In such a search, the probability that a given sampling will *not* be successful is $1 - P_f$. For a search involving R trials, the probability that it will not be successful is $(1 - P_f)^R$. Therefore, the probability that the search will be successful is simply $1 - (1 - P_f)^R$. Let us assume that a search will be successful if the search performs enough trials to raise the probability of success to 0.5, or

$$0.5 = 1 - (1 - P_f)^R.$$

Simplifying, we get

$$P_f = 1-(1-0.5)^{1/R} \qquad (2)$$

Eqn. (2) gives us an estimate for the most improbable functional event that a blind search could reasonably expect to find, given R trials. That being the case, the highest level of functional information that natural processes could reasonably be expected to produce for a given function would be the case where only one functional configuration would reasonably be found in R trials, or

$$I_{nat} = - \log_2[1-(1-0.5)^{1/R}]. \qquad (3)$$

The requirement for I_{sig} is that it must be greater than I_{nat}. For example, if the turbine method of trying to open grandfather's safe was capable of 500,000 trials before the system wore out, then the turbine-safe system could reasonably be expected to produce as much as 13 bits of functional information (I_{sig} = 13 bits). Since a functional combination requires 15 bits of functional information, one could not reasonably expect the turbine system to open the safe without any intelligent design so far as finding the right combination is concerned. Therefore, if such a system were built and the safe successfully opened, we could on reasonable grounds accuse the engineer of having biased the system to find the right combination, for the physical system was unlikely to have done it without any intelligently designed bias built in. Due to the nature of probability, however, it is possible that the river current could open grandfather's safe on the very first try, or it might never open the safe. It is also possible that the engineer did not build in an intelligently designed bias to find the right combination, we were just fortunate. We could never be absolutely sure, therefore, whether there was a built in intelligently designed bias or not. Since at the core of functional information is probability, we can never arrive at a definitive conclusion, only a likely, probable, or plausible conclusion. This leads to the following considerations.

Probability considerations

1. I_{nat} is not a cutoff for the amount of functional information natural processes can produce. Rather, the probability that natural processes can produce x amount of functional information decreases exponentially as the amount of functional information increases beyond I_{nat}. For example, if I_{nat} = 32 bits of functional information, using Eqn. (1), this corresponds to a probability of approximately 10^{-10} whereas 64 bits of functional information corresponds to a probability of approximately 10^{-19}. In other words, 64 bits of functional information is only twice as much information as 32 bits, but one billion times more difficult to find in a search.

2. Our observations indicate that there does not seem to be any known limit to the amount of functional information that intelligence can produce. It seems to be capable of producing anywhere from 0 bits and up.

3. In view of the previous two points, we can only speak of the likelihood that an effect required intelligent design, where the greater the difference between the functional information required for the effect and I_{nat}, the more likely it is that intelligent design was required. This would hold true for SETI, archeology, forensic science, and biological life.

Method to gauge the likelihood of intelligent design

Given that there is no known upper limit for the amount of functional information a mind can produce, for any effect requiring or producing functional information, intelligent design is the more likely explanation if

$$I(E_x) > I_{nat}. \qquad (4)$$

The greater the difference between $I(E_x)$ and I_{nat}, the more likely it is that intelligent design was required. It will be assumed, for simplicity, that the probability that mindless natural processes can achieve I_{nat} is 1 and decreases probabilistically for $I(E_x) > I_{nat}$. The probability that intelligent design can achieve $I(E_x)$ will be assumed to be 1 for any finite amount of functional information. This is a reasonable assumption, given our observations of what intelligence can do and the apparent absence of any upper limit. This method can, in principle, be applied within the fields of forensic science, archeology, SETI, and biology, as well as in areas outside of science, such as lottery gaming investigations, plagiarism investigations, and the justice system, to name a few.

V. Application to Biological Life

Now that we have a method to identify examples of intelligent design, we are now equipped to apply it to biological life to see what the likelihood is that it was designed. We shall first discuss the relationship between natural selection and functional information. We will then derive a generous estimate of I_{nat} for an evolutionary search. We shall then estimate $I(E_x)$ for several test cases and, applying the method suggested in the previous section, see if it is likely that biological life is an example of intelligent design.

Natural selection, fitness functions, and functional information

It is usually assumed that the origin and diversification of life is not a blind search. Actual mutations, insertions, deletions, and genetic drift may be chance events, but natural selection essentially guides the search and, hence, the search is not blind. On the one hand, it is assumed that natural selection explains how life could appear and diversify without requiring any intelligence, but on the other hand, terms that that are usually applied to intelligence, such as 'design' and 'selecting' are commonly applied to natural selection. It is very common to read articles where the author marvels at what natural selection is capable of. Of course, this raises the question, does natural selection, itself, require intelligent design? The fatal mistake made by many who appeal to natural selection is the assumption that natural selection, itself, does not require intelligent design. It is bad science that does not test its assumptions, so we must apply intelligent design detection to natural selection itself.

Although natural selection is credited with somehow discovering the right combination of nucleotides to code for, say, proteins like SecY or RecA, there is a great deal of vagueness about how it actually is supposed to do this, and not just for two proteins, but for thousands. Not only must it somehow locate the proper sequences that are determined *a priori* by physics that will produce a stable 3-D structure, but it must also be able to assemble the information that will produce more impressive things like molecular machines, molecular computers and, ultimately, the cell and the organism itself. Fortunately, the field of genetic algorithms or evolutionary algorithms can be used to introduce rigor to the concept of natural selection.

Every evolutionary search process, whether we are discussing natural selection, or a computational evolutionary algorithm, requires a fitness function. Without a fitness function, the search degenerates into a blind search, at best. The *fitness function* can be defined as follows:

> **Fitness function** (evaluation function): represents "the requirements to adapt to. It forms the basis for selection, and thereby it facilitates improvements. More

> accurately, it defines what improvement means. From the problem-solving perspective, it represents the task to solve in the evolutionary context."[3]

Since the fitness function, whether it is found in nature, or in a genetic algorithm, must contain the requirements to adapt to, or that defines the desired outcome, it must contain at least as much functional information as the desired outcome. If the functional information contained in the fitness function is *less* than the functional information required for the desired outcome, then the deficit must be made up for in a blind search, which falls prey to the probability problems that emerge.

Natural selection requires a fitness function. If a given protein is a product of natural selection operating within a fitness landscape, then sufficient functional information required to find that protein in an evolutionary search must be encoded within the fitness function. If a few hundred, or several thousand proteins are required, then a great deal more functional information must be encoded within a much more complex fitness function. If molecular machines are also desired, then additional functional information must be included within the fitness function.

No one actually knows where this amazing fitness function is in nature such that we can measure the amount of functional information that it contains. However, if we assume that natural selection is responsible for, say, the origin of gene coding proteins, then we can estimate the amount of functional information the fitness function of nature contains by measuring how much functional information a given protein requires. We are then in a position to see if intelligent design is required for natural selection to produce the given protein by comparing the degree of functional information that must be encoded within the fitness function of nature and comparing it with I_{nat}. Intelligent design detection methods must be applied to natural selection to see if intelligent design is required to encode the appropriate amount of functional information into the fitness function of natural selection. There is no escape; the functional information within a fitness function must be measured and evaluated, and a test performed to see if it requires intelligent design.

For example, if an evolutionary algorithm is attempting to produce even more complex software commands, then the fitness landscape includes the operating system within which those commands will survive or fail and the functional information required to produce that operating system must be measured. If the fitness function is outside the actual algorithm, say within a database, then the functional information contained in the database must also be included. To summarize; if natural selection or a fitness function are credited with producing a given amount of functional information, then if that functional information exceeds I_{nat}, by the method proposed in this article, ID is required to properly configure the fitness function.

Estimating I_{nat}:

---

[3] Eiben, A.E. & Smith, J.E. (2003) *Introduction to Evolutionary Computing*, Springer-Verlag Berlin Heidelberg, 19.

It is estimated that there may be somewhere between 500 and 900 different protein folds,[4,5] that form roughly 4,000 to 7,000 different protein families.[6] The stable folds are determined by physics, not biology. This requires that any evolutionary process must perform a search of sequence space to locate those areas where physics produces a stable, 3-D structure. Origin of life theorists are not decided as to what processes could lead to the minimal genome. Regardless of whether one prefers a genetic approach or a metabolic approach, we do know that at some point, proteins must be produced, or at least the information coding for stable, folded proteins must be achieved. We can, therefore, take all origin of life scenarios and put them into a 'black box' which performs an evolutionary search and outputs the stable folded proteins that are permitted by physics. It is not necessary to know what the processes within this black box do, all we need to know is the output. The output can be evaluated two ways, one way is to assume that the black box is performing a blind search which, of course, requires no intelligent design, and the other way is to assume that some sort of fitness function is operating within the black box which may or may not require intelligent design, depending upon how much functional information is required for the output. To estimate I_{nat} for a pre-biotic, origin of life search, we must estimate the number of trials available for a blind search. We will then be in a position to estimate I_{nat} and compare it with the functional information required to produced a minimal genome to see if a fitness function would be necessary that would require intelligent design. Since we do not know what processes could perform the search, let us be extremely generous.

Taylor *et al.* have estimated that the mass of the earth would equal about 10^{47} proteins, of 100 amino acids each.[7] If we suppose that the entire set of 10^{47} proteins reorganized once per year over a 500 million year interval (about the estimated time period for pre-biotic evolution), then that search permits about 10^{55} options to be tried. Using Eqn. (3), $I_{nat} \approx$ 185 bits of functional information. Of course, this scenario is much more generous than any scenario under consideration, but at least we will not be underestimating I_{nat}. If $I(E_x)$ requires more than I_{nat}, then we can assume that either a fitness function requiring intelligent design must be included in the black box, or intelligent design is operating in some other fashion to properly encode the functional information.

We are now ready to examine four test cases.

Case One: the Venter Institute's synthetic genome for *M. genitalium*:

---

[4] Taylor, W. (2002) 'A periodic table for protein structures', *Nature*, **416**, 657-660.

[5] Zhang, C. & DeLisi, C. (1998) 'Estimating the number of protein folds', *JMB*, **284**, 1301-1305.

[6] Wolf, Y., Grishin, N. & Koonin, E. (2000) 'Estimating the number of protein folds and families from complete genome data', *JMB* **299**, 897-905.

[7] Taylor, S., Walter, K., Kast, P. & Hilvert, D. (2001) 'Searching sequence space for protein catalysts' *PNAS* **98**, 10596-10601.

The five 'watermarks' in the synthetic Venter genome are formed by choosing base pairs that, when translated into amino acids and using the amino acid single letter symbols, spell out the following five words:

VENTERINSTITVTE
CRAIGVENTER
HAMSMITH
CINDIANDCLYDE
GLASSANDCLYDE.

Hazen *et al.* point out that the number of functional options can vary according to the degree of efficiency required by the system. This is true for both human languages and biopolymer sequences. In this case, however, we will assume that the Venter Institute wants their watermarks correctly spelled according to the above sequences. Given that there are 20 options for each site in each word, using Eqn. (1), $I(E_x) = 259$ bits of functional information. Since I_{nat} has been estimated at 185 bits of functional information, $I(E_x) > I_{nat}$. These results indicate that it is about 10^{22} times more probable that the watermarks required ID than that they could be produced by mindless natural processes. Therefore, by the method proposed here, we can conclude that the 'watermarks' are likely produced by ID, in this case, the Venter Institute.

Case Two: a folded, functional protein domain:

Axe has estimated that the frequency of occurrence of stable, folded functional protein domains, a structurally independent component of a protein, is somewhere between 10^{-64} to 10^{-77}.[8] These values correspond to $M(E_x)/N$ in Eqn. (1). The functional information required, therefore, to code for a stable, folded protein domain is 213 to 256 bits. Since we have estimated I_{nat} at a generous 185 bits, which is much too low to achieve the amount of functional information required to produce a folded, functional protein domain, $I(E_x) > I_{nat}$ and it is at least 10^{19} times more probable that ID can produce a folded functional domain than mindless natural processes. The method of ID detection proposed in this article, therefore, reveals that ID is highly likely to be required to produce folded, functional protein domains. If the sequences coding for a stable fold are the product of a pre-biotic black box that contains a fitness function, then the fitness function will require intelligent design.

Case Three: an average 300 amino acid protein:

The functional information required to produce an average, 300-amino acid protein, can be estimated by analyzing the set of aligned sequences for SecY and RecA. These two proteins are particularly interesting because they are also universal proteins, found throughout organic life. It is inferred, therefore, that they would be required in a minimal genome. Analyzing a set of 1,553 aligned sequences for RecA and 469 aligned sequences

---

[8] Axe, D. (2004), ‘Estimating the prevalence of protein sequences adopting functional enzyme folds’, *JMB*, 1295-1315.

for SecY reveals that 832 bits of functional information are required for RecA and 688 bits for SecY.[9] It is reasonable, therefore, to estimate the functional information required for the average 300 amino acid protein to be around 700 bits of information. $I(E_x) > I_{nat}$ and ID is 10^{155} times more probable than mindless natural processes to produce the average protein. Again, if natural selection is invoked to explain the origin of proteins, a fitness function will be necessary that requires intelligent design.

Case Four: the simplest life form:

It is estimated that the simplest life form would require at least 382 protein-coding genes.[10] Using our estimate in Case Four of 700 bits of functional information required for the average protein, we obtain an estimate of about 267,000 bits for the simplest life form. Again, this is well above I_{nat} and it is about $10^{80,000}$ times more likely that ID could produce the minimal genome than mindless natural processes. Again, if one wishes to explain the origin of the simplest life form by natural selection, a fitness function will be required that is capable of generating 267,000 bits of functional information, well into the area that requires intelligent design.

VI. Conclusion:

In this article, I have proposed a method for the detection of intelligent design. Applying that method to the Venter 'watermarks' in their synthetic genome, the method successfully identifies the watermarks as highly likely to require ID. The same method applied to protein domains, average proteins, and the minimal genome, also indicates that ID is required for even the simplest life form. If life is the product of natural selection and an extremely complex fitness landscape, then we can conclude that it is extremely likely that intelligent design was required to configure the fitness function. Whether there actually is such an impressive fitness function encoded in nature is outside the scope of this discussion. I have not proved that intelligent design was required for biological life. Instead, I have shown that given intelligent design is easily capable of generating functional information on the level of what is required for biological life, and given that the functional information required for biological life is far beyond what we could reasonably expect nature to generate, intelligent design is the most probable explanation, by many orders of magnitude, for biological life. Intelligent design would also be the most probable explanation for any fitness function operative in natural selection that could successfully locate the folding proteins required for life.

---

[9] Durston, K., Chiu, D., Abel, D. & Trevors, J. (2007) 'Measuring the functional sequence complexity of proteins', *Theoretical Biology and Medical Modeling*, **4**:47.

[10] Glass, J., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M., Maruf, M., Hutchison III, C., Smith, H., & Venter, J. (2006) ‘Essential genes of a minimal bacterium’ *PNAS*, **103**, 425-430.