Spring June 3rd, 2016

Biological Information and Its Origins: The Curious Case of RNA

Grace E. Andrews
Seattle Pacific University

BIOLOGICAL INFORMATION AND ITS ORIGINS: THE CURIOUS CASE OF RNA

by

GRACE ANDREWS

DR. W. GRABOW, FACULTY ADVISOR

DR. S. VOKOS, SECOND READER

A project submitted in partial fulfillment

of the requirements of the University Scholars Program

Seattle Pacific University

2016

Approved___Dr. Jeff Keuss _____

Date_____

Abstract

Information is an important pre-requisite for the onset of life, which means that any study of the origins of life must also address the origin of information. Biological information today is generally conceptualized in terms of the Central Dogma of Biology, with DNA as a digitized code within the cell. However, here I propose that biological information - particularly in prebiotic conditions, where information would have first arisen - is better understood in terms of the three-dimensional structure of a molecule. In this regard, RNA serves as an interesting example. Evidence for information storage in RNA can be seen in a number of ways. One of these is that, instead of experiencing pressure to maintain a particular sequence of nucleotides, RNA experiences evolutionary pressure to maintain a particular three-dimensional structure through mechanisms such as covariation. Additionally, covariation allows for the relative fluidity of a nucleotide sequence while motifs provide a specific set of building block that RNA can use, both of which would increase the likelihood of producing an informational molecule. Conceptualizing biological information in terms of a functional three-dimensional structure, rather than a linear code, provides a useful paradigm for understanding how biological information could have arisen. Because the three-dimensional structure is less dependent on primary sequence, the likelihood of producing an informational molecule through random chance would increase in prebiotic conditions. Although I am proposing RNA as the best candidate for the origin of biological information, it is possible that other molecules may also be plausible candidates for the origin of information as long as they share these qualities with RNA. This method of understanding biological information also has implications for alternative models of the origin of information, including those proposed by the Intelligent Design community, which tend to focus primarily on DNA.

Biological Information and Its Origins: The Curious Case of RNA

Life is characterized and sustained by a number of information rich biological processes that govern cellular functions, and greatly contribute to its overall complexity. Because of this, any study of the origins of life must address the origin of biological information as well. Although it is fundamental to life today, the origin of biological information remains largely neglected. However, the following three assumptions seem to provide a reasonable starting point in the endeavor to provide an explanation for the beginnings of information. First, it is important to keep in mind that life and information need not have arisen simultaneously. In fact, it seems far more likely that information would have arisen prior to the onset of life, as cellular life is completely dependent on the flow of information. Second, information in biological systems today is largely understood in terms of the Central Dogma of Biology, in which information passes from DNA to RNA to protein, but prebiotic information, or information at the cusp of life, would undoubtedly have been simpler and looked much different. For example, one explanation of the origin of life, described in the RNA world hypothesis, posits that the earliest cells were so different from modern cells that they did not contain DNA or protein at all, but were run entirely by RNA. Finally, although our focus on the Central Dogma predisposes us to understand biological information in terms of a linear string of nucleotides or amino acids, similar to a digitized code, it is most fully realized as information in three-dimensional forms, when it can enact a specific change or reaction in a cell. These interactions are accomplished through molecules possessing specific shapes, orientations, or arrangements that communicate with other molecules that have different molecular affinities. A particular reaction or function resides in the shape and (chemical environment afforded by the shape) of a molecule, suggesting that the information residing in the three-dimensional shape may be equally, if not more, important

than the information encoded in the linear arrangement of monomers. In light of these three assumptions, I would propose that RNA provides an informative model for conceptualizing the origin of biological information. However, it is important to note that other molecules may also make plausible candidates for the first informational molecule so long as they share the characteristics of RNA that I describe throughout this paper.

Before exploring the possibility of RNA as the first informational molecule, it is first necessary to understand what information is and how it functions. Information is closely related to, although not interchangeable with, the entropy in a given system, and, in many cases, the difference between the possible and observed entropy of that system (Shannon, 1949). When considering this in terms of biological information, one might think of all the possible combinations a string of nucleotides in a particular gene could adopt as the potential entropy of the system and the combination of nucleotides that the gene *actually* adopts as the observed entropy of the system. The difference between the two entropies would then correspond to the amount of information that a linear string of nucleotides might convey. To increase the total possible entropy, one could either increase the length of a sequence or increase the number of possible variables for each unit. By increasing the total possible entropy, the potential informational capacity for a molecule would also increase due to the possibility of a larger difference between the possible and actual entropy of a molecule. However, this difference alone does not amount to information. In order for the information in a system to be fully realized, it must be *about* something. Information is found, not in a sequence or shape, but in the relationship between that sequence and the system it describes. (Adami, 2004). A string of nucleotides or a protein cannot carry information unless the sequence or shape correlates to a particular function, such as turning off a gene or catalyzing a reaction.

Until recently, RNA has been seen as merely a messenger between DNA and protein, but it is now also known to play numerous cellular roles including catalyzing reactions and regulating genes, earning it the title of the "dark matter" of the cell (Ridihough, 2005). These functions are rooted in the ability of RNA to adopt a variety of three-dimensional shapes, and it is in the configuration of these shapes that information is stored. By its nature, information can only convey meaning through its relationship to its environment (Deacon, 2011). Unlike DNA, which consistently forms into the well-known double helix, RNA possesses the ability to self-fold through Watson-Crick and noncanonical base-pairing, thus producing a variety of three dimensional structures, which enables RNA to perform a variety of functions. Because the shapes themselves are what have biological significance, it seems that the informational content of RNA is found more in the three-dimensional structure of the molecule than the linear sequence of its nucleotides. Evidence of this idea can be found in the way particular shapes are favored and evolutionarily conserved (Grabow et al., 2013). The informational content of RNA being found more in its shape than its linear sequence makes it an interesting case study for the origin of biological information. We will see that these shapes are not only more likely than a particular sequence to arise by chance, but, since they also convey a biological function, they meet the criteria for being considered informational molecules. This ability to function would also be more relevant to the emergence of life than the relatively inert DNA.

The need for an informational molecule that directly relates to its environment is an excellent starting point in making the case for RNA as the first informational molecule, particularly when compared to DNA. Information is efficiently stored in the linear string of DNA nucleotides due to its ability to form stable and predictable base pairs, and this information is even stored redundantly due to the ability of DNA to bind its own complementary strand.

However, it needs a complex system of non-DNA molecules to process the information and actualize its message within a cell. On its own, DNA is relatively inert, and the complex interpreting system required for DNA makes it ill-suited as a candidate for the origin of information. RNA, on the other hand, seems to be a more plausible candidate due to its ability to spontaneously fold into a variety of three-dimensional structures that are capable of performing a variety of dynamic functions, including catalysis and gene regulation. Through its three-dimensional structure, RNA would be able to communicate information about its environment independent of an interpreting system, since its shape would both hold and communicate the information on its own.

Although DNA and RNA are chemically very similar, RNA differs from DNA in several substantial ways that make the formation of these structures possible. Perhaps the most notable difference is that RNA, unlike DNA, does not interact with its own complementary strand. Instead, it acquires its three-dimensional structure from the intramolecular interactions of nucleotides within a single strand of RNA. This allows RNA to base-pair to itself, creating a number of unique three-dimensional structures and making RNA more structurally similar to protein and, by extension, more functionally similar as well. Another noteworthy difference is the expanded repertoire of RNA base-pairing. While DNA is much more limited to the iconic Watson-Crick canonical base pairs, where guanine binds exclusively with cytosine and adenine with thymine, RNA is able to form several thermodynamically stable noncanonical base pairs (Figure 1A), which expand the possible structures it can adopt. These noncanonical base pairs are characteristic of the interactions forming the tertiary structures of RNA, while Watson-Crick base pairs are used to form the hairpins and stem loops of the secondary structures (Figure 1B). Because information, by its nature, must be *about* something, the order of nucleotides in a strand

of RNA is less important than the shape which results from that sequence. It is through its shape that an RNA molecule interacts with its environment through its own particular function. The various three-dimensional structures of RNA are able to interact with other molecules, including protein, other RNA molecules, or even other parts of the same RNA strand. The storage of information in a three-dimensional structure makes RNA an excellent candidate for the origin of information as its variety of functions would be useful for the onset of life.

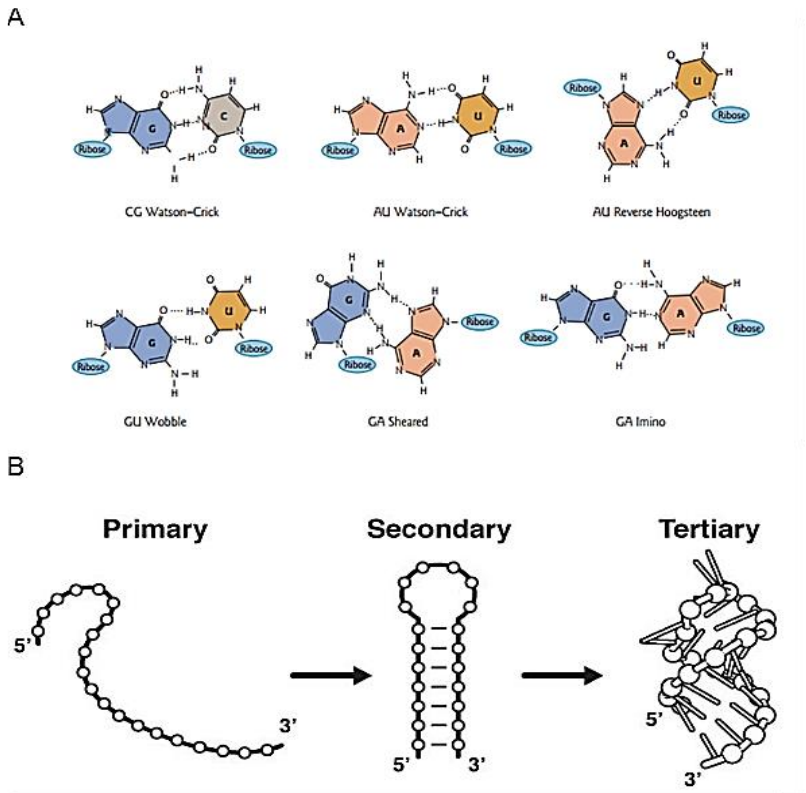

Figure 1. (A) Both Watson-Crick and noncanonical base pairs are pictured (Allison, 2012). (B) RNA exhibits primary, secondary, and tertiary levels of organization. Secondary structure contains Watson Crick base pairs, while tertiary structure can have both Watson-Crick and noncanonical base pairs (Hecht and Huc, 2007).

Another notable difference between DNA and RNA is that, for DNA, there is evolutionary pressure to maintain individual nucleotides with high informational content, but, for RNA, the pressure to maintain information results in the preservation of secondary and tertiary structure rather than the primary sequence. This tendency can be observed in the covariation of nucleotides that often occurs in RNA in an attempt to preserve structural features throughout the course of evolution (Figure 2). In covariation, if a mutation occurs to cause a change in a base,

then the corresponding base that it pairs with can also be observed to mutate in order to retain a certain secondary structure (Parsch et al, 2000). In this way, a Watson-Crick GC base pair may change to an AU pair without disrupting the overall structure. Covariation allows a certain amount of fluidity to a sequence so that not every change in base will result in the loss of function. This makes an RNA molecule less dependent on its primary sequence and also increases the likelihood of randomly producing an RNA of a particular shape, since multiple sequences can produce the same structure. Although the primary sequence is important, it appears that the identity of each nucleotide is less important than the shape that is produced from that sequence. This further demonstrates how RNA primarily stores its information in its tertiary structure.

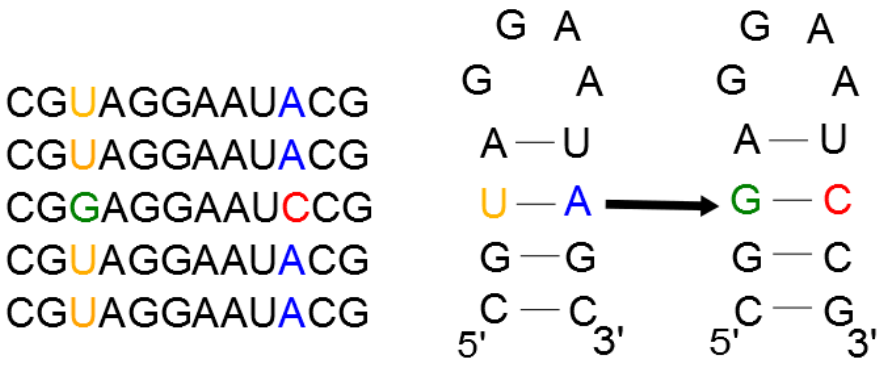

Figure 2. On the left, a sequence alignment is shown with positions that contain covarying nucleotides in color. On the right, the role these nucleotides play in the formation of secondary structure is shown. The secondary structure is maintained even though the nucleotides change.

Another example of the importance of three-dimensional structure in information storage is the structural motif. An RNA motif is a semi-conserved sequence that is found in remarkably high abundance and produces a characteristic three-dimensional structure through the conservation of specific hydrogen-bonding contacts. These recurring motifs function

independently of the context in which they are found (Moore, 1999). This results in the conservation of identifiable three-dimensional shapes held together by hydrogen-bonding networks between RNA nucleobases. These nucleobases are conserved only to the degree that they can maintain those contacts that define and hold the motif together. Many different motifs exist in RNA including the 11-nucleotide motif, T-loop, kink turn, GNRA tetraloop, right angle motif (RA) (Figure 3), and GA minor motif (Moore, 1999). These motifs are often fairly small, but they play a pivotal role in helping to direct local and long-range RNA folding as well as providing the basic building blocks through which more complex secondary and tertiary structures are formed (Grabow and Jaeger 2014). By directing the folding of RNA sequences into particular structures, motifs help to ensure that a sequence will fold into a functionally useful shape. In addition to demonstrating the importance of three-dimensional structure, the high abundance of motifs in RNA and the way in which they facilitate the folding and function of RNA provide the groundwork for understanding how functional structures could have arisen prebiotically.

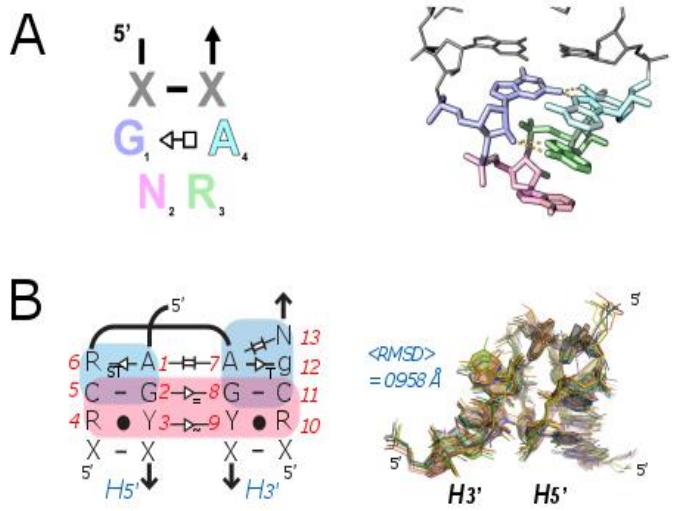

Figure 3. (A) GNRA tetraloop (B) RA turn (Grabow et al., 2012). Positions marked with an N can be any base, while an R can be either A or G, and Y can be either C or U.

The pressure placed on an RNA molecule to maintain its shape, and therefore function, throughout the course of evolution can also be seen as evidence for the importance of three-dimensional structure. When considering the informational content of a single nucleotide in a strand of RNA, one might think of the informational content of that nucleotide in terms of bits. For RNA, where there are four possible nucleotides - adenine, guanine, cytosine, and uracil (A, G, C, and U) - the most bits of information a single nucleotide can contain is two according to the Shannon equation for the determination of maximal entropy (Adami, 2004):

$$H = -\sum_{i=G,C,A,T} p(i) log_2 p(i) = log_2(4) = 2 \text{ bits}$$

If *p(i)* is ¼ for each nucleotide at maximal entropy, then the number of possible bits of information is two. Although we can determine the maximal entropy of a string of nucleotides, it is not possible to determine the actual entropy of a nucleotide based on a single sequence, making it impossible to estimate the informational content of that nucleotide. However, by comparing sequences of nucleotides that are evolutionarily related and functionally equivalent, one can more accurately determine the bits of information contained by each nucleotide in a given position. This is accomplished by calculating the probability that a particular nucleotide will appear at a certain position in the sequence. For example, if a number of evolutionarily related sequences are aligned and every nucleotide at position sixty-five is a G, then that nucleotide contains two bits of information. However, if only 50% of the nucleotides at position sixty-five in the sequence are G's then there is only one bit of information at that location. Nucleotides that contain a large amount of information are generally those with evolutionary pressure to remain the same due to their functional importance. They may be a necessary component of a binding site or essential for maintaining a particular structure.

To determine whether or not nucleotides that were essential for tertiary structure formation would be high in informational content, the bits of information for each nucleotide in a tetrahydrofolate (THF) riboswitch (Figure 4) were calculated via the method previously described and using an alignment of fifty-seven different sequences from a variety of organisms (Figure 4A). A riboswitch is a structural element associated with a portion of an mRNA sequence that is not translated into a protein, but can regulate either transcription or translation through the binding of a particular molecule, which in this case is tetrahydrofolate (Ames et al., 2010). The primary three-dimensional contacts that take place within the molecule occur as an interaction between nucleotides 65-71 and 185-191 (Figure 4D). The informational content of each nucleotide in these regions shows a great deal of variation, with each nucleotide containing

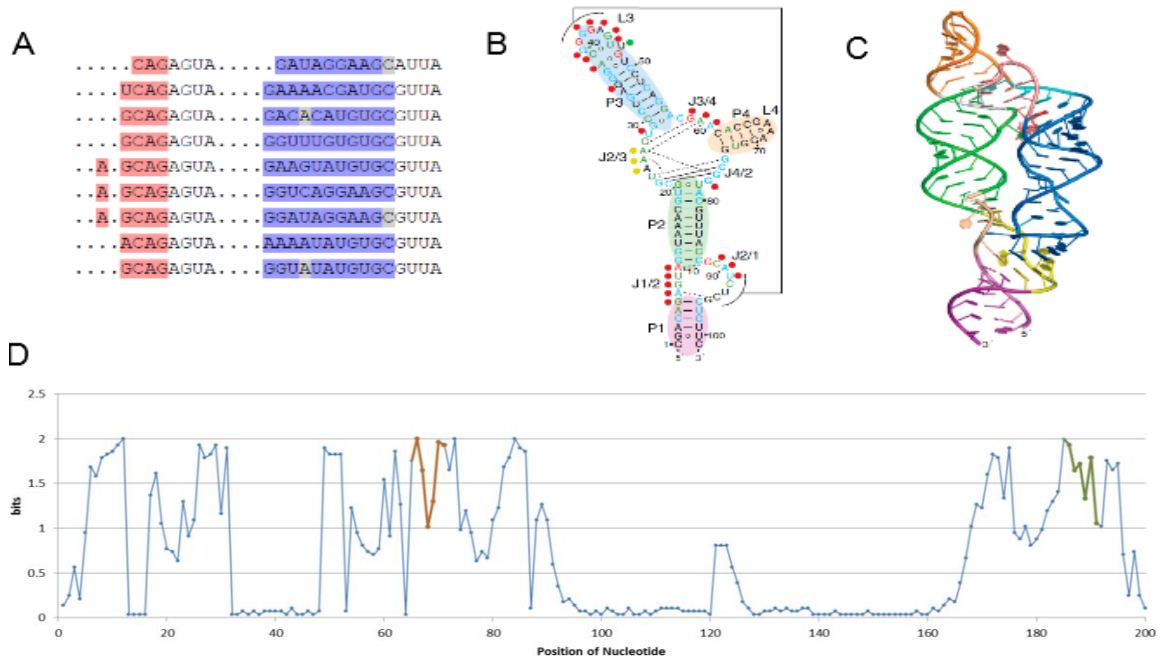

Figure 4. (A) A section of the sequence alignment for the THF riboswitch (Ames et al., 2010). (B) Secondary structure of a THF riboswitch with a line drawn to indicate the long range tertiary contact (Huang et al., 2011). (C) Tertiary structure of the THF riboswitch (Huang et al., 2011). (D) Bits of information for a nucleotide at a given position. Nucleotides that participate in the tertiary contact shown in (B) are colored orange and green.

anywhere between one and two bits of information. Although the reasonably high informational content of nucleotides in these regions suggests that they convey an important function, the high degree of variation may cause one to question how important these intramolecular contacts actually are. If the three-dimensional structure of the molecule is the most important element of the riboswitch, and this structure is maintained by the contact of these two regions, then one might expect the informational content of the nucleotides in these regions to be uniformly high. While this does not seem to be the case, it does not necessarily mean that the structure is not important or maintained through all 57 organisms. The necessity of a particular contact or base-pair being made can also be analyzed by observing covariation in an RNA molecule.

When aligning sequences, as was done for the analysis of informational content, covariation will decrease the occurrence of the dominant nucleotide, thus reducing the calculated informational content for the nucleotide at that position. By checking for covariation in the sequence alignment, we see that some of the variation in informational content is explained by this phenomenon. Because covariation is evidence of the importance of tertiary structure, the observed variation in informational content may not imply insignificance for the structure. Rather, the covariation of these nucleotides supports the concept that it is the shape of RNA that makes it meaningful and functional within its context rather than the primary sequence of nucleotides. Furthermore, when considering the origin of biological information, the variation in informational content of the nucleotides in RNA may have actually been beneficial for the first informational molecule. If a nucleotide is less constrained in terms of its identity, it will not contain as many bits of information, but it will also be easier to produce an RNA molecule of a particular shape. Having a variety of primary sequences that will yield a certain structure increases the likelihood of that structure being formed. Because of this, the efficiency with which

information is stored decreases, but the likelihood of producing a functional shape increases. The decreased informational content required of each nucleotide would then make it easier for RNA to cross what I will call the "information barrier" at the origin of information. Similar to the way a decreased energy of activation makes a chemical reaction easier to initiate, the fewer bits the first informational molecule needs, the easier it will be for the first informational molecule to be produced. In this way, decreased information storage in RNA makes it a well-suited candidate for the origin of information.

The decreased informational capacity of RNA is related to the fact that many different primary sequences can yield the same secondary or tertiary structure. In the case of DNA, because its capacity to store information is directly dependent on the primary sequence, the probability of randomly producing a specific string of nucleotides that is relevant to its environment is quite low. To calculate the probability of generating a particular sequence, one simply needs to find the probability of a specific nucleotide occurring and raise it to the power of however long the sequence of interest is. The probability of a nucleotide occurring at a particular position will always be ¼ since the identity of a nucleotide is not affected by those adjacent to it. For

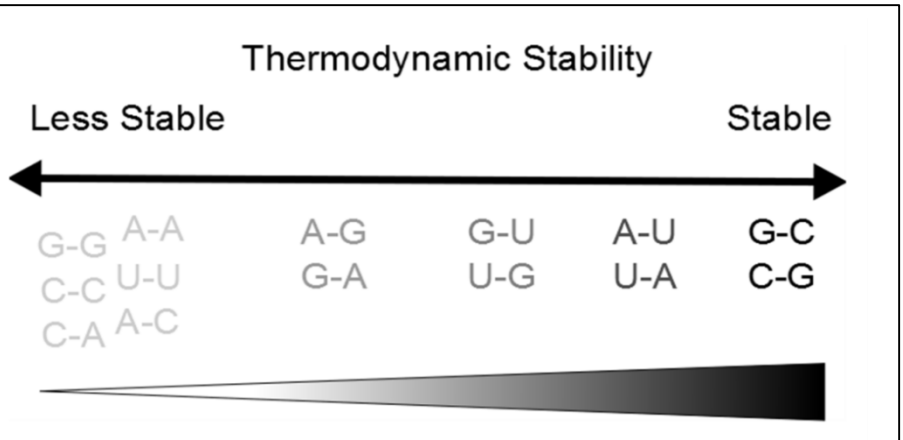

Figure 5. Relative stabilities of each possible Watson-Crick base pair. GC and AU pairs are the most stable while bases pairing to themselves and AC pairs are the least stable.

example, a specific 12 nucleotide sequence would have a $(¼)^{12}$, or 1/16,777,216 chance of occurring. Although the same would be true for producing a specific 12 nucleotide sequence of RNA, RNA is less tied to the specificity of the primary sequence in terms of its informational content. This is beneficial for the production of information in RNA as the probability of generating some kind of three-dimensional structure is relatively high. In a typical 25 nucleotide sequence, where every nucleotide is represented, the percentage of nucleotides involved in base-pairing is 40-50% (Yarus, 2010). This would virtually guarantee that a few thermodynamically stable structures would result for any random sequence of 25 nucleotides. This large degree of base-pairing can be attributed to the relatively high thermodynamic stability of 8 of the 16 possible base-pairs (Figure 5), with the AG pairs and, to a lesser extent, the GU pairs stability being dependent on the stability of adjacent base pairs. Because about half of the possible base pairs are relatively stable, it is reasonable to assume that, in a random sequence of nucleotides, about half of the bases would form stable base pairs. The abundance of stable base pairs makes the possibility of forming these contacts fairly high, which also increases the probability of forming a functional RNA molecule. As was alluded to in the case of covariation, a number of different primary sequences can even produce the *same* tertiary structure. There still may be 16,777,216 different sequences, but many combinations would now exist that could produce the same shape and therefore convey the same information. For example, 11,880 of the 16,777,216 possible sequences for a 12 nucleotide RNA would be capable of forming four base pairs (See Appendix A). Because it is the tertiary structure that allows the information of RNA to be fully actualized, the probability of randomly producing an informationally potent molecule from a string of nucleotides is much higher for RNA than it is for DNA.

In addition to the increased statistical likelihood of a particular fold being produced, certain folds would be favored based on their inherent thermodynamic stability and rate of polymerization. Although the specific functions of different RNA or protein molecules are often thought of as the result of evolutionary selection for biologically useful purposes, Michael Denton makes the case for there being a relatively small number of possible RNA folds which are achieved by adopting a structure corresponding to the minimum free energy possible for a given sequence (Denton et al., 2002; Denton et al., 2003). This means that, even though there may be several ways in which an RNA molecule can fold, it will favor the configuration that makes it the most stable. The most stable structures may be achieved through a number of different paths and would be selected for based on thermodynamic stability. In this way, we can see why commonly recurring structures, such as RNA motifs, would be found in such high abundance. Additionally, further selection can take place in prebiotic systems on the basis of varying rates of polymerization. Small increases in rates of polymerization can lead to an increase in the abundance of certain molecules, and the ability of a molecule, or system of molecules, to replicate would greatly accentuate the potential increase in abundance (Chen and Nowak, 2012). Selection for certain molecules, either through thermodynamic stability or rates of polymerization, would lead to an increased likelihood of informationally useful RNA molecules being formed from a random sequence of nucleotides. In addition to this, a "fold-first, function-second" approach to RNA folding would suggest that early life had perhaps taken advantage of the already structurally useful RNA and simply enhanced its functional properties through natural selection. In this scenario, prebiotic RNA would have existed, ready-made, with relevant information.

Although the case for RNA as the first informational molecule is compelling, it is not without its problems, many of which are shared with the RNA World Hypothesis, which is still largely debated (Bernhardt, 2012). One such problem is the question of whether it is plausible that ribonucleotides could have been generated in prebiotic conditions. RNA is a considerably complex molecule and determining probable synthetic pathways in prebiotic conditions has proven to be quite challenging. However, despite the difficulty, scientists continue to make progress in solving this dilemma. A plausible pathway has recently been developed for the synthesis of pyrimidine ribonucleotides in pre-biotic conditions (Powner et al., 2009), and, while this does not completely answer the problem, it does seem to be a promising step towards providing a mechanism by which RNA could have existed in prelife conditions.

Another common argument against the RNA World Hypothesis is the notion that catalysis is a function associated only with relatively long RNA molecules, and it is unlikely that such long strands of RNA would be forming in any degree of abundance. This is problematic since the need for a functional RNA molecule to be long would pose a problem for understanding the origin of information in terms of RNA, as it would decrease the likelihood of forming functional RNA molecules. However, just as with the prebiotic synthesis of ribonucleotides, promising research shows that exceedingly short RNA can possess catalytic functions. There have been several cases of very short sequences being capable of performing reactions, including the self-cleavage of a seven nucleotide duplex (Vlassov et al., 2005) and the aminoacylation of an RNA substrate with a ribozyme truncated to only five nucleotides (Yarus, 2011). These “mini-ribozymes” provide support for the functionality of ribozymes that are considerably shorter in length than the ribozymes we see in nature today. In addition, the probability of producing such small ribozymes is considerably greater than the production of

their more extensive counterparts. This is because it is not only easier to polymerize a shorter RNA strand, but it also requires a smaller pool of those polymerized RNA strands to cover the sequence space necessary to promote a functional ribozyme. Although these mini-ribozymes are less efficient than longer ribozymes (Vlassov et al., 2005), they do provide a simpler possible starting point for catalytic RNA molecules. The simplification of ribozymes would allow for a prebiotic system in which the production of a functioning RNA molecule, capable of carrying information, is possible. This would make the participation of RNA in the origin of information more plausible.

Although it seems possible that RNA could have developed as an informational molecule in prebiotic conditions, in order to be significant for the origin of life, it would also need to develop the ability to replicate. To be considered life, a system must have more bits of heritable information than the bits required for its initiation (Joyce, 2012), which would require RNA to have the ability to replicate if it were to be a viable candidate for the origin of information. While there has been some progress in finding a stand-alone self-replicating RNA (Ma and Yu, 2006), cooperative networks of RNA molecules, working to replicate the entire system, present a promising new perspective on the origin of RNA replication and the origin of information. These cooperative networks of RNA replicators have been demonstrated to exhibit favorable growth dynamics and are actually able to out-compete single molecule replicators. Not only are they more efficient, but, through cooperation, RNA networks are also able to evolve greater complexity over time (Vaidya et al., 2012). To take this a step further, it is also possible that, after these potential replicating complexes were formed, the most efficient replicating system could have been selected for through a prebiotic selection process based on polymerization and replication rates and stability (Nowak and Ohtsuki, 2008; Chen and Nowak,

2011). This would mean that selective pressures could have been at work even before the existence of the cell, and would have been capable of selecting for the most efficient replicator. Cooperative replication not only helps to answer the question of how prebiotic information eventually shifted to life, but it also follows the proposed model of information as relatively sequence independent. Because replication would be dependent on the interaction of different molecules, the sequences could be somewhat fluid due to the fact that individual nucleotides could change as long as these interactions were maintained.

While RNA provides a compelling model for understanding the origin of information, many of the arguments that have been made here in favor of RNA could possibly be applied to other molecules, including protein. For instance, one of the primary reasons that RNA is such an excellent case study for considering the origin of information is that its three-dimensional structure is an essential part of this information. However, a similar argument could be made for protein since it is also highly dependent on its tertiary/quaternary structure and uses it to interact with its environment. In addition to this, another way in which the case for RNA and protein overlap is that Denton's argument regarding the favoring of certain RNA folds is just as applicable to protein as it is to RNA. But despite their similarities, one of the ways in which RNA would make a better candidate than protein for the first informational molecule is its reduced number of monomers. Proteins are made using 20 amino acids, but, in comparison, RNA is only comprised of four nucleotides. Although the nucleotides are more complex than the amino acids, once they were synthesized it would be easier to produce a functional RNA from a small sequence, as was seen in the case of the mini-ribozymes. A second advantage of RNA is that it would seem to produce a more straightforward path to information in the way we observe it today. RNA is capable of retrosynthesis, which is the process by which RNA is used as the

template for the synthesis of DNA instead of the other way around (Goff, 1990). Information has never been seen to flow from protein to RNA or DNA, but information can flow in either direction through RNA. This seems to put RNA in a central position for the flow of information and makes it a more plausible candidate for the origin of information than protein. Finally, because the folding rules of RNA are simpler and more limited, it would be easier to produce something that would fold predictably, which would be advantageous for the origin of information.

The connection between information and the three-dimensional structure of a molecule may have a wide range of implications for studies of the origin of information and life, particularly in the Intelligent Design community. In his book *Signature in the Cell: DNA and the Evidence for Intelligent Design*, a leading figure of the Intelligent Design community, Stephen Meyer, has proposed that the information we see in today's cells, carried primarily by specific strings of nucleotides in DNA, could not have arisen by random processes. Here I would like to address the way in which Meyer's fixation on DNA as the first informational molecule is problematic and neglects other viable mechanisms for the origin of biological information.

One of Meyer's arguments for information being a product of intelligent design is that the probability of generating a sequence of DNA that relates to a biological function, as well as the molecules needed to make this information biologically relevant, is so low as to be essentially impossible. But this idea makes the assumption that early life would have had to look fairly similar to the way it does today, and disregards the possibility that information could have arisen in a form different than DNA. He suggests that "any minimally complex protocell resembling cells we have today would have required not only genetic information, but a sizeable preexisting suite of proteins for processing that information" (2009). But, in this statement, Meyer neglects

the possibility that the earliest forms of life may not have needed such a complex informational processing system. He seems to be implying that biological information must have always looked like the genetic code, where a three-nucleotide codon in DNA codes for a specific amino acid. In this scenario, DNA, RNA, and protein would all have had to come together at just the right time and in just the right way to produce the necessary components of informational processing, but this is not necessarily the case. The RNA World Hypothesis posits an alternative view on this subject by proposing that the earliest cells may not have needed DNA and protein in order to function because RNA is capable of both storing genetic information and performing a variety of functions within a cell (Higgs and Lehman, 2014). The idea that all the components of biology's Central Dogma would have to come together at once ignores the possibility that the earliest forms of life may have looked quite different than they do today and would have undoubtedly been much simpler.

Part of the reason Meyer gives for suggesting that functional DNA could not have arisen through random chance is that, because the bases of adjacent nucleotides do not interact with each other, there are no forces influencing the order in which nucleotides are assembled. A guanine preceding a uracil is just as probable as an adenine preceding a cytosine since there is no interaction between adjacent nucleotides in either RNA or DNA. Meyer suggests that this is a problem, not only for DNA, but for RNA as well since "for strands of RNA to perform catalytic functions (including self-replication), they, like proteins, must display specific arrangements of their constituent building blocks" (2009). But this makes the assumption that it is only the primary sequence of nucleotides that dictates the informational content of RNA. This would be a valid argument if the origin of information were approached from the perspective of DNA since the primary sequence *is* the most important aspect of its informational content and any natural

influences on the order of nucleotides in DNA would actually diminish its informational storage capacity (Polanyi, 1969). But while it is true that there are no interactions between adjacent bases that influence the order of nucleotides in DNA and RNA, and that this is actually a good thing for DNA, it is not true that the primary sequence is the main conveyor of information in RNA or that extremely specific sequences are needed for RNA to be functionally relevant. As I have attempted to show, the tertiary structure of RNA is what carries the bulk of its information because this is what allows it to interact with its environment, and the same tertiary structure can be produced by a variety of different primary sequences. This lowered dependence on primary sequence is due in part to the pressure to adopt the structure with the lowest free energy, as was suggested by Denton (2002). Therefore, if information can be stored in the three-dimensional structure of a molecule, there actually would be natural laws that would favor the formation of certain informational shapes over others. Additionally, by influencing which three-dimensional structures are selected for, factors such as covariation and thermodynamic stability can indirectly exert pressure on the primary sequence by favoring those sequences which provide the most stable structures. This provides an elementary example as to how natural laws can in fact influence the order of nucleotides.

Finally, Meyer dismisses much of the RNA World Hypothesis and instead focuses exclusively on DNA as an informational molecule. He says that "the theory did not solve the problem of biological information – it merely displaced it" (2009). This argument assumes that information in RNA works in a similar fashion to that of DNA. However, as has been demonstrated throughout this paper, the two are quite different in terms of the ways in which they store their information and the ability of each to facilitate the rise of information and be selected for in prebiotic conditions. The problem is not merely shifted from one place to another,

but it is understood in a completely different fashion when considered in terms of RNA. The ability of RNA to store information in its three-dimensional structure and the natural laws that favor the formation of certain structures over others change the way in which we can think about the origin of information. Difficulties intrinsic to the nature of DNA, such as the formation of specific sequences accompanied by processing tools, are less problematic in regard to RNA. With this understanding, the problem is not simply displaced onto RNA; rather, a deeper understanding of biological information from the standpoint of three-dimensional structures actually solves many of the problems that Meyer associates with DNA's perceived role in the origin of information. For reasons already stated, very few scientists studying the origin of life would view DNA as a realistic candidate for the origin of information.

RNA seems remarkably well-suited for the onset of biological information. While it may be possible that a different biomolecule is responsible for the origin of information, this biomolecule would have to share certain characteristics with RNA. RNA's three-dimensional structure would allow for the earliest informational molecules to be functionally relevant to their surroundings, which is an essential component of information storage. The ability of the first informational molecules to function would also allow the most efficient of these to be selected for via natural law and thereby increase in abundance. In addition, the dependence of information in RNA on secondary and tertiary structure, rather than primary sequence, allows for the sequence of nucleotides to be more fluid, which would decrease the initial hurdle associated with the emergence of information. Because a variety of different sequences can make the same secondary and tertiary contacts, the random production of a functional RNA molecule becomes far more likely. This way of understanding biological information also has implications for alternative models of the origin of information, including those proposed by the Intelligent

Design community, which tend to focus primarily on DNA. In light of the way that RNA functions and stores information, it provides a compelling new way of understanding the origin of information.

Appendix A

This appendix outlines the calculations used to determine that, in a 12-nucleotide sequence, 11,880 of the 16,777,216 possible sequences would base-pair four times.

This calculation is best conceptualized in terms of dot-bracket notation for RNA where each nucleotide is either a dot, meaning it does not base-pair, or a left or right bracket. Matching brackets represent base-pairing nucleotides. This means that the number of right and left brackets must be even. For example, the dot-bracket notation for the 12 nucleotide hairpin loop in figure 2 is ((((….)))). There are 12! ways of arranging the dots and brackets in this 12 nucleotide sequence, but, because there are four each of three identical symbols, the actual number of different arrangements is $\frac{12!}{4!4!4!}$ which equals 831,600 possible different arrangements. However, the sequence is so short that it is reasonable to assume that all the right brackets must precede left brackets in order to produce a stable structure. When the dots are removed, the total of possible arrangements of brackets in relation to each other is $\frac{8!}{4!4!}$ which equals 70. Of these 70, only one will have a structure where all the right brackets precede the left brackets. The number of possible sequences in a 12 nucleotide strand is then $\frac{831,600}{70}$ which is 11,880. However, the number of realized structures may actually be less than this since not each of these sequences will yield a thermodynamically stable structure.

Literature Cited

Adami, C. Information Theory in Molecular Biology. *Physics of Life Reviews*. **2004,** *1*, 3-22.

Allison, L. A. *Fundamental Molecular Biology, 2e*; Wiley: Hoboken, NJ, 2012.

Ames, T. D.; Rodionov, D. A.; Weinberg, Z.; Breaker, R. R. A Eubacterial Riboswitch Class That Senses the Coenzyme Tetrahydrofolate. *Chemistry & Biology*. **2010**, *17*, 681–685.

Bernhardt, H. S. The RNA World Hypothesis: the Worst Theory of the Early Evolution of Life (except for All the Others)a. *Biology Direct Biol Direct*. **2012**, *7*, 23.

Chen, I. A.; Nowak, M. A. From Prelife To Life: How Chemical Kinetics Become Evolutionary Dynamics. *Accounts of Chemical Research Acc. Chem. Res.* **2012**, *45*, 2088–2096.

Deacon, T. W. *Incomplete Nature: How Mind Emerged from Matter*; W.W. Norton & Co.: New York, 2012.

Denton, M. J.; Marshall, C. J.; Legge, M. The Protein Folds As Platonic Forms: New Support for the Pre-Darwinian Conception of Evolution by Natural Law. *Journal of Theoretical Biology*. **2002**, *219*, 325–342.

Denton, M. J.; Dearden, P. K.; Sowerby, S. J. Physical Law Not Natural Selection as the Major Determinant of Biological Complexity in the Subcellular Realm: New Support for the Pre-Darwinian Conception of Evolution by Natural Law. *Biosystems*. **2003**, *71*, 297–303.

Goff, S.P.; Retroviral Reverse Transcriptase: Synthesis, Structure, and Function. *Journal of Acquired Immune Deficiency Syndromes*, **1990**, *3*, 817-831.

Grabow, W. W.; Zhuang, Z.; Swank, Z. N.; Shea, J.-E.; Jaeger, L. The Right Angle (RA) Motif: A Prevalent Ribosomal RNA Structural Pattern Found in Group I Introns. *Journal of Molecular Biology*. **2012**, *424*, 54–67.

Grabow, W. W.; Zhuang, Z.; Shea, J.-E.; Jaeger, L. The GA-Minor Submotif as a Case Study of RNA Modularity, Prediction, and Design. *Wiley Interdisciplinary Reviews: RNA WIREs RNA*. **2013**, *4*, 181–203.

Grabow, W. W.; Jaeger, L. RNA Self-Assembly And RNA Nanotechnology. *Accounts of Chemical Research Acc. Chem. Res.* **2014**, *47*, 1871–1880.

Hecht, S.; Huc, I. *Foldamers: Structure, Properties, and Applications*; Wiley-VCH: Weinheim, 2007.

Higgs, P. G.; Lehman, N. The RNA World: Molecular Cooperation at the Origins of Life. *Nat Rev Genet Nature Reviews Genetics*. **2014**, *16*, 7–17.

Huang, L.; Ishibe-Murakami, S.; Patel, D. J.; Serganov, A. Long-Range Pseudoknot Interactions Dictate the Regulatory Response in the Tetrahydrofolate Riboswitch. *Proceedings of the National Academy of Sciences*. **2011**, *108*, 14801–14806.

Joyce, G. F. Bit By Bit: The Darwinian Basis of Life. *PLoS Biology PLoS Biol*. **2012**, *10*.

Ma, W.; Yu, C. Intramolecular RNA Replicase: Possibly the First Self-Replicating Molecule in the RNA World. *Origins of Life and Evolution of Biospheres Orig Life Evol Biosph*. **2006**, *36*, 413–420.

Meyer, S. C. *Signature In the Cell: DNA and the Evidence for Intelligent Design*; HarperOne: New York, 2009.

Moore, P. B. Structural Motifs In RNA. *Annual Review of Biochemistry*. **1999**, *68*, 287-300.

Nowak, M. A.; Ohtsuki, H. Prevolutionary Dynamics and the Origin of Evolution. *Proceedings of the National Academy of Sciences*. **2008**, *105*, 14924–14927.

Parsch, J.; Braverman, J. M.; Stephan, W. Comparative Sequence Analysis and Patterns of Covariation in RNA Secondary Structures. *Genetics*. **2000**, *154*, 909-921.

Polanyi, M.; Grene, M. G. *Knowing And Being*; Routledge & Kegan Paul: London, 1969.

Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis Of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature*. **2009**, *459*, 239–242.

Riddihough, G. In The Forests of RNA Dark Matter. *Science*. **2005**, *309*, 1507–1507.

Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, 1949.

Vaidya, N.; Manapat, M. L.; Chen, I. A.; Xulvi-Brunet, R.; Hayden, E. J.; Lehman, N. Spontaneous Network Formation among Cooperative RNA Replicators. *Nature*. **2012**, *491*, 72–77.

Vlassov, A. V.; Kazakov, S. A.; Johnston, B. H.; Landweber, L. F. The RNA World On Ice: A New Scenario for the Emergence of RNA Information. *J Mol Evol Journal of Molecular Evolution*. **2005**, *61*, 264–273.

Yarus, M. *Life From an RNA World: the Ancestor Within*; Harvard University Press: Cambridge, MA, 2010.

Yarus, M. The Meaning of A Minuscule Ribozyme. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **2011**, *366*, 2902–2909.

Appendix B: Integration of Faith and Learning

Both faith and science approach the world in different ways, but each one is useful for providing a more complete understanding of the reality in which we live. Science seeks to understand how the natural world works, while faith seeks to understand God and, by extension, what humanity's relationship to the natural and supernatural ought to look like. Science achieves its goal by studying repeating patterns in nature, and faith pursues understanding in a variety of ways, including experience, reason, and natural and divine revelation. However, problems arise when these two different ways of understanding the world are seen as mutually exclusive, or an attempt is made to answer a question with science that can only be answered by faith or vice versa. Using scientific epistemology to understand ideas relating to faith or the supernatural proves rather fruitless as science is only equipped to deal with the natural world. This does not invalidate the existence of the supernatural, but it simply means that science is not able to detect its existence since science depends on the observation of repeating patterns. In a similar way, using faith to understand the patterns we see in the world around us can ultimately keep us from analyzing these patterns and learning useful information about the underlying mechanisms of our environment. When considering the origin of information, the Intelligent Design community seems to make the mistake of answering a scientific question by introducing a designer who could not be subject to scientific investigation. This can lead to a block in the scientific endeavor to understand a particular question.

Using faith to answer an unanswered scientific question can be considered a "God of the gaps" approach and seems to be employed by Stephen Meyer in his attempt to explain the origin of information. In a "God of the gaps" approach, a gap in scientific knowledge is explained by means of the supernatural. Meyer argues against this claim, saying that "when we observe effects

that we know only agents can produce, we rightly infer the presence of a prior intelligence even if we did not observe the action of the particular agent responsible" (Dembski and Kushiner, 2001). However, the assumption made in this statement is that we know only intelligent agents can produce information, but, based on all the research done in the area, this is not an assumption common to the scientific community as a whole. The claim that we "know" only intelligent agents can produce information disregards the consensus of a large part of the scientific community. This "God of the gaps" approach is not only a scientifically invalid means of approaching the origin of information, but, for people of faith, it can also undermine faith in God when scientific evidence is produced that suggests God may not be responsible for a particular gap in knowledge. In one of his letters from prison, Dietrich Bonhoeffer explains why he finds the "God of the gaps" approach problematic:

> If in fact the frontiers of knowledge are being pushed further and further back (and that is bound to be the case), then God is being pushed back with them, and is therefore continually in retreat. We are to find God in what we know not in what we do not know; God wants us to realize his presence, not in unsolved problems but in those that are solved. That is true of the relationship between God and scientific knowledge. (Bonhoeffer, 1953)

Using God to fill in the gaps of scientific knowledge will only serve to diminish him as those gaps begin to be closed. We must try to fill in the gaps of scientific knowledge through scientific means lest we find God reduced when we discover that we do not need him in order to understand these questions after all. This attempt to overlap the methodologies of faith and science is not only harmful to the scientific pursuit of knowledge, but it can be harmful to the pursuit of faith as well.

In the same way that faith cannot answer scientific questions, science cannot be used to answer the foundational questions of faith. Unlike science, faith does not depend on determining patterns in the natural world, but rather on divine revelation through, scripture, experience, and nature. In *Personal Knowledge; towards a Post-Critical Philosophy,* Michael Polanyi says of Christianity that "Christian faith does not express the assertion of observable facts and consequently you cannot prove or disprove Christianity by experiments or factual records" (1958). Because faith seeks to answer questions that are not defined by the natural world, scientific study of the natural world can neither prove nor disprove faith. However, the evangelical Christian community has often felt threatened by advances in scientific knowledge when they feel as if science *is* trying to disprove faith in some way. This can be seen in the evangelical church's reaction to evolution, which was perceived as negating the creative nature of God. Unfortunately, the perception that science could in some way disprove faith has caused a rejection in the evangelical community of science altogether. This can have serious ramifications for these communities. In some branches of the evangelical Christian community, children are often brought up believing that certain elements of faith and science are mutually exclusive or that science poses a threat to faith in some way. This sets up a false dichotomy between the two and forces people to ultimately choose between them. But when forced to choose, a person will miss all the knowledge that the other side has to offer.

Although the methodologies that faith and science employ do not overlap, both can still provide meaningful information for the other. Because God is understood to be the creator of the natural world, gaining more knowledge about this world can allow Christians to understand their creator in a more complete way. Additionally, new scientific discoveries can help Christians discern which biblical interpretations are the most consistent with the world in which we live.

This allows for a more robust understanding of divine revelation. Polanyi suggests that new scientific discoveries "may engender conceptual reforms which will renew and clarify, on the grounds of modern extra-religious experience, man's relation to God. An era of great religious discovery may lie before us" (1958). New scientific discovery need not be perceived as a threat against faith, but rather as an aid in reforming some of the ways that we understand faith, which would result in a deeper understanding of Christianity.

So it seems that faith is enriched by scientific pursuits, but science can also benefit from faith. Faith provides a guide for what we expect the world to look like. For example, because Christianity believes in a relational God, it would not make sense to have a universe with no possibility of life and, by extension, relationship. When confronted with equally valid and scientifically robust models, having some basis for understanding what we should expect from the world can help direct the way in which we choose to take our scientific exploration. By allowing faith and science to interact, members of the scientific community are provided with some idea of what the world ought to look like, and those in the religious community gain access to knowledge, which can cause them to more thoughtfully engage with their own faith and gain a deeper understanding of their Creator.

Because some branches of the evangelical church have been historically wary of science, an important component of scientific scholarship for people of faith today is an attempt to reconcile the strained relationship between the two. While both can benefit from the knowledge that the other has to offer, it is up to those intimately connected with each one to enact this reconciliation. Christian members of the scientific community are in a unique position to initiate dialogue between the two seemingly opposed camps. Although the methodologies applied in science must remain robust and consistent with the methodologies of the scientific community as

a whole, Christian scientists are able bring a different perspective to science and maintain that there are alternate avenues by which we can understand the world in a meaningful way. In addition to this, they can also be ambassadors of science to the Christian community and encourage the church to not dismiss the knowledge that science has to offer. Because of this unique position, reconciliation becomes an important aspect of what it means to work in the sciences as a person of faith, and it is an ideal that I hope to incorporate into my own work in the scientific field.

Literature Cited

Bonhoeffer, D. Letters And Papers From Prison. Edited by Eberhard Bethge; Macmillan C: New York, 1953.

Dembski, W. A.; Kushiner, J. M. Signs Of Intelligence: Understanding Intelligent Design; Brazos Press: Grand Rapids, MI, 2001.

Polanyi, M. Personal Knowledge; towards a Post-Critical Philosophy; University of Chicago Press: Chicago, 1958.