The Mystery of Life's Origin

The Continuing Controversy

CHARLES B. THAXTON,
WALTER L. BRADLEY,
ROGER L. OLSEN, JAMES TOUR,
STEPHEN MEYER, JONATHAN WELLS,
GUILLERMO GONZALEZ,
BRIAN MILLER, DAVID KLINGHOFFER

Seattle Discovery Institute Press 2020

Description

The origin of life from non-life remains one of the most enduring mysteries of modern science. *The Mystery of Life's Origin: The Continuing Controversy* investigates how close scientists are to solving that mystery and explores what we are learning about the origin of life from current research in chemistry, physics, astrobiology, biochemistry, and more. The book includes an updated version of the classic text *The Mystery of Life's Origin* by Charles Thaxton, Walter Bradley, and Roger Olsen, and new chapters on the current state of the debate by chemist James Tour, physicist Brian Miller, astronomer Guillermo Gonzalez, biologist Jonathan Wells, and philosopher of science Stephen C. Meyer.

Library Cataloging Data

The Mystery of Life's Origin: The Continuing Controversy by Charles B. Thaxton, Walter L. Bradley, Roger L, Olsen, James Tour, Stephen Meyer, Jonathan Wells, Guillermo Gonzalez, Brian Miller, and David Klinghoffer

486 pages, 6 x 9 x 1.0 inches & 1.4 lb, 229 x 152 x 25 mm. & 0.65 kg

Library of Congress Control Number: 9781936599745

ISBN-13: 978-1-936599-74-5 (paperback), 978-1-936599-75-2 (Kindle), 978-1-936599-76-9 (EPUB)

BISAC: SCI013040 SCIENCE / Chemistry / Organic

BISAC: SCI013030 SCIENCE / Chemistry / Inorganic

BISAC: SCI007000 SCIENCE / Life Sciences / Biochemistry

BISAC: SCI075000 SCIENCE / Philosophy & Social Aspects

Publisher Information

Discovery Institute Press, 208 Columbia Street, Seattle, WA 98104

Internet: http://www.discoveryinstitutepress.com/

Published in the United States of America on acid-free paper.

Second Edition, First Printing, January 2020.

Contents

13. We're Still Clueless about the Origin of Life

James M. Tour

Organisms have well-defined molecular assemblies, redox potentials across membranes, and metabolic pathways—all operating in exquisite states that we call "life."

Chemistry, by contrast, is utterly indifferent to whether anything is alive or not. Without a biologically derived entity acting upon them, molecules have never been shown to "evolve" toward life. Never.[1]

While organisms exploit chemistry for their own ends, chemicals have never been seen to assemble themselves into an organism. Origin-of-life research keeps attempting to make the chemicals needed for life, and then to have those assemble toward something to which they are inherently indifferent. But try as they might, without preexisting life no researchers have ever seen molecules assemble into a living cell, or anything even remotely resembling a living cell. Contrary to the hyperbole of press reports, any synthetic molecularly derived structures that have been touted as being cell-like are in reality far from it. This situation might change in the future, but it is unlikely to change under the current course of research. Scientists have no data to support molecular "evolution" leading to life. The research community remains clueless.

Many scientists and professors who are outside boutique origin-of-life circles have been led astray by researchers' claims and the subsequent press, thinking that far more is known about life's origin than really is

known. This has affected the highest seats in the academy where even some science professors confuse origin of life with biological evolution. Like a muddy prebiotic cesspool, confusion abounds in the academy.

Two-thirds of a century since the 1952 Miller-Urey experiment, where some racemic amino acids were formed from small molecules and an electrical discharge, the world is no closer to generating life from small molecules—or any molecules for that matter—than it was in 1952.[2] One could argue that origin-of-life research is even more befuddled now than it was in 1952 since more questions have evolved than answers, and the voluminous new data regarding the complexity within a cell makes the target much more daunting than it used to be.[3]

Consider what has occurred in other fields in the past sixty-seven years since Miller-Urey performed their experiments: human space travel, satellite interconnectivity, unlocking DNA's code and its precise genetic manipulation, biomedical imaging, automated peptide and nucleotide synthesis, molecular structure determination, silicon device fabrication, integrated circuits, and the internet, to name just a few.

By comparison, origin-of-life research has not made any progress whatsoever in addressing the fundamental questions of life's origin. Two-thirds of a century and all that has been generated are more suggestions on how life *might* have formed—suggestions that really show how life *probably did not* form. Nothing even resembling a synthetic cellular structure has arisen from its independent components, let alone a living cell. Not even close.

In 1775, the French Academy in Paris refused to entertain any further proposals for perpetual motion machines; the devices just did not work as advertised.[4] No one knew why not—the mature science of thermodynamics, which gave us a theoretical account for *why* the *perpetuum mobile* schemes failed, lay nearly one hundred years in the future—but the machines clearly failed. Today we need a French Academy-like directive toward origin-of-life proposals; for, like perpetual motion machines, such proposals just do not work as advertised. Instead we should explore why scientists have failed to produce life. Clearly life can exist—unlike

perpetual motion machines, we have the ubiquity of life surrounding us on this planet. But there needs to be a wholly different scientific approach to reveal life's origin.

This is an appeal to the origin-of-life research community: Step back and consider the claims within the research, the true state of the field, the retarded state of the science relative to other research areas, and the confusion or delusion of the public regarding life's origin. Many researchers in origin-of-life organic synthesis are superb scientists. However, overly confident assertions, exaggerated and spread by the over-zealous press, have led to gross public misconceptions regarding what is and is not known concerning the beginning of life.

We will now turn to an exploration of the two main classes of origin-of-life science: chemical synthesis and molecular assembly. After a brief summary of each, the two classes of experiments will be considered separately in depth.

First, the chemical synthesis of the four molecule types for life: nucleotides, carbohydrates, proteins, and lipids. Nucleotides are composed of a trimeric nucleobase-carbohydrate-phosphate combination, and once polymerized, constitute DNA and RNA. Five different nucleobases comprise the primary alphabet for DNA and RNA. The nucleotides and their subsequent DNA and RNA structures are homochiral, meaning that they exist as one mirror image form and not the other, yielding one of two possible enantiomers.

Amino acids are most often homochiral. When amino acids are polymerized, they form proteins and enzymes, the latter being nature's nanomachines that build the biological system. Like DNA and RNA, proteins and enzymes also have a tertiary homochirality based upon their coiling and folding.

Lipids are dipolar molecules having a polar water-soluble head and a non-polar water-insoluble tail. They too are most often homochiral.

Carbohydrates, in addition to being part of the backbone of DNA and RNA using their 5-carbon containing versions, also use 6-carbon containing structures. Cells live on carbohydrates for energy, and carbo-

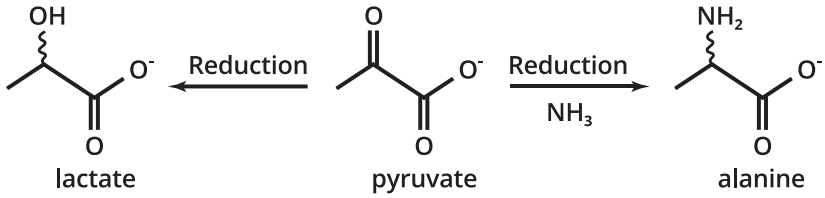

Figure 13-1. The reduction and reductive amination of pyruvate to lactate and alanine.
This was described in the NASA press release as "NASA Study Reproduces Origins of Life on Ocean Floor."

hydrates, along with proteins, are identification-receptors wherein some regulation on and within cells is controlled. Carbohydrates are also homochiral, and their polymeric forms take on tertiary homochiral shapes.

Origin-of-life efforts have spent much time trying to make these four classes of molecules and their polymers, starting from simple chemicals that were presumed to be available on the prebiotic earth, such as formaldehyde, hydrogen cyanide, and carbonates.

The second class of experiments that are performed in origin-of-life research deal with the assembly of the molecules. For example, when lipids are added to water and subjected to shear forces, they can form spherical bilayer vesicles. These vesicles have the lipid polar ends pointing inward and outward toward the water on the inside and outside of the vesicle, respectively, and the nonpolar tails pointing toward each other and away from the water phases. Sometimes researchers will add other compounds to the water that become engulfed when the bilayer vesicles form. In order to obtain a cell, molecules must precisely assemble into many higher-order structures.

Chemical Synthesis Experiments

Chemical synthesis experiments in origin of life can be summed up by a protocol analogous to this:

- Purchase some chemicals, generally in high purity, from a chemical company.

- Mix those chemicals together in water in high concentrations and a specific order under some set of carefully devised conditions in a modern laboratory—sets of conditions that often would be difficult to replicate in a non-laboratory environment on early earth.
- Obtain a mixture of compounds that have a resemblance to one or more of the basic four classes of chemicals needed for life: carbohydrates, nucleotides, amino acids, or lipids. Most of the time they are synthesized in racemic (both mirror images) or near racemic form, not in homochiral form.
- Identify the desired compound in a mixture of many other isomers and products. Then buy (or make, using modern non-abiotic methods) a purified version of that desired compound and proceed to the next step.
- Publish a paper making bold extrapolations about origin of life from these functionless crude mixtures of stereochemically scrambled intermediates.
- Engage with the often over-zealous press to dial up the knob of unjustified origin-of-life projections.
- Watch the misled and mesmerized layperson exclaim, "You see, scientists understand how life formed!"
- Accept a generation of science textbooks yielding colorful, deceptive cartoons of raw chemicals assembling into cells, which then emerge as slithering creatures from a prehistoric pond.

Even professors have been misled by this. Nor do the scientists themselves understand anything more about life's origin than they did before they performed their experiments, because their experiments offer no solution to the fundamental questions needed for a path to life.

How can the results be published if there is nothing new regarding life's origin? Because this becomes the norm in the field—there are no expectations of addressing grander questions. Reviewers are of the same mind, believing this is the best that can be done. Journal editors

have been numbed to believe the same, and to ignore unjustified claims regarding the origin-of-life implications. Some published work contains chemistry which is pedestrian, while other papers show remarkably ingenious routes to these molecular classes starting from simple chemicals—but in every case, fundamental questions of life are not addressed. Thus the field stagnates for two-thirds of a century while other areas of research make quantum leaps that advance humankind.

Here is a recent example of such a scenario of simple chemistry and the hyperbole that follows. In 2019, Laura Barge and coworkers at NASA Jet Propulsion Laboratory, the California Institute of Technology, and the Oak Crest Institute of Science simulated an undersea hydrothermal vent. Heating an aqueous solution of pyruvate to 70°C and introducing ammonia and iron hydroxides while limiting oxygen, they observed simple reduction and reductive amination to stereo-scrambled lactate and alanine, respectively.

Those are such simple reducing reactions that the chemistry is certain and therefore wholly unremarkable. Yet the authors write, "This shows that aqueous, partially reducing iron mineral systems (which would have been common in early-Earth seafloor/vent environments) could have facilitated synthesis and concentration of prebiotic organic molecules relevant for the emergence of life." [5]

The NASA press office then had a field day with this result, titling their article "NASA Study Reproduces Origins of Life on Ocean Floor" and writing, further, "Scientists have reproduced in the lab how the ingredients for life could have formed deep in the ocean 4 billion years ago. The results of the new study offer clues to how life started on Earth and where else in the cosmos we might find it."[6] The press cut and pasted from the NASA press release, resulting in a blitzkrieg of deeply misleading news.

Although the chemistry of this experiment is less complex and less interesting than that of the 1952 Miller-Urey experiment, it was published in the *Proceedings of the National Academy of Sciences*, a superb scientific journal. This underscores that journals themselves are com-

plicit in continuing this sort of rudimentary experiment claiming to be suggestive of life's origin. Unlike the far more sophisticated synthetic chemistry of origin-of-life leaders like John D. Sutherland, the work by NASA is nonsensically simple—so in that sense, much like the prebiotic earth would have been.

Unlike the artless 2019 NASA experiment, most origin-of-life researchers today put far more precision into their protocols to make more elaborate arrays of stereo-scrambled intermediates. One could easily argue, therefore, that the researchers are moving further from the heart of abiogenesis since they are filling the protocols with the best of their intellectual training to coax molecules into the form that the researcher desires. Yet even with all that intellectual input, the origin-of-life researchers overcome few if any of the hurdles noted below that need to be considered when dealing with chemical synthesis experiments common to all origin-of-life protocols that are being published.

Hurdle 1: Homochirality

Molecules that compose living systems almost always show homochirality. So one particular enantiomer, selected from the many possible stereoisomers, needs synthesizing. Generally there are 2^n possible stereoisomers where n is the number of stereocenters in the molecule. If discussing carbohydrates, there are eight possibilities among the abundant 5-carbon carbohydrate and sixteen possibilities from the 6-carbon-long carbohydrates. Claims that these structures could be prepared under prebiotic conditions in high enantiomeric purity using inorganic templates, or any presumed templates, have never been realized even with the advanced designs of the origin-of-life researchers. How much less could homochiral compounds have been obtained in a mindless prebiotic environment?

In addition, this would have to happen repeatedly for all the varying carbohydrates. Nobody has ever offered a demonstrative solution.

Moreover, each class of compound, the carbohydrates, the amino acids, and the lipids, and further each compound within each of those

classes, would require its own separate methodology to control its specific regiochemistry and stereochemistry. To merely say that all the diverse stereochemistries form and the required one preferentially reacts (kinetic resolution) repeatedly over its enantiomer or diastereomers negates what is known about the difficulties of selective synthesis, especially when envisioned in a mindless prebiotic system where no enzymes yet exist. The differences in reaction rates often require chiral systems acting upon chiral molecules.

If this can be done sufficiently well in a mindless prebiotic puddle, why cannot the experts in research repeatedly replicate it in sixty-seven years of trying while using their sophisticated modes of synthetic ingenuity?

Hurdle 2: Pre-DNA and -RNA

Abiogenesis starts long before DNA and RNA are formed. So en route to those compounds, one would have to select the 5-carbon carbohydrate for its backbone over the 6-carbon structure, and all this in homochiral form. Further, for DNA, it has to be one hydroxyl group deficient, or deoxyribose. If it is not, then it will be suitable for RNA, but far less stable. Prebiotic systems never knew any of this.

Hurdle 3: Selectors

In choosing the molecule types to go forward, there are no chemical selectors yet formed in a prebiotic system, or if there are selectors, they generally need to be more complex than the molecule that they are selecting. What is the origin of the selector in a prebiotic system?

Hurdle 4: Redesigns

When building molecular systems, constant redesigns are needed which take the synthesis back to step one. It is often impossible to remove a moiety once it has been added to a molecule. So if a prebiotic and mindless reaction makes one small mistake, the synthesis has to go back to the beginning—but that could mean sending it back a hundred million years, and it will likely make the same mistake again since it has no mem-

ory to prevent its repeated mistake. Plus it has no impetus to start over, because chemistry is indifferent to moving toward life. It is chaos.

Hurdle 5: Stopping Point

The synthetic reactions do not know how to stop their current course of progression, or why to stop. The prebiotic system will continue to make derivatives. Time, although claimed to be the great savior of abiogenesis, can actually be the enemy. Time works against obtaining desired chemicals, particularly when the needed target is a kinetic product. For example, carbohydrate prebiotic synthesis is generally conducted through the formose reaction, but then one gets aldol reactions in equilibrium with retro-aldol reactions, and Cannizzaro reactions, which, taken together over time, favor the branched and "caramelized" polymeric products. How does the system know when or how to stop if reaction times can be in the thousands of years or longer? Routes to carbohydrates from presumed prebiotic molecules are an all-around mess.

Hurdle 6: Purification

A prebiotic system does not have the ability to easily purify the structures. Sometimes selective crystallization can occur with the designed input of a synthetic chemist, but most often not. And the impurities contaminate and inhibit subsequent steps. Separations have to be done repeatedly across broad arrays of the four classes of compounds or else the impurities withdraw the resources from the chemical pools. Most origin-of-life researchers do not even purify the desired products. They simply identify the desired product in a morass of other isomers or related molecules, and then purchase a pure sample for the next step. That's cheating when it comes to total synthesis, but it's a cheat rarely acknowledged by the researchers.

Hurdle 7: Order

Reagent addition-order is essential. One cannot add the icing to a cake at the stage of mixing the flour and eggs. Chemistry is even more demanding with its sequences throughout multiple steps, each requiring their

own reaction conditions. To claim that compound A spilled in from one pool, and then another pool dumped its contents of compound B, seems far-fetched when such sequences are repeatedly required with specific timings.

Hurdle 8: Activation Steps

The making of the amino acid monomers is hard enough, but the synthesis of a single dipeptide bond generally requires activation steps that are complex if they are to be performed cleanly and repeatedly. Automated systems today require multiple individual steps to cleanly prepare a single amide bond. Likewise, nucleotide polymerization can be terribly messy unless proper activators (leaving groups) and blocking chemistry is exploited. No general solution to this problem has been offered.

Hurdle 9: Environmental Factors

The parameters of temperature, pressure, solvent, light, pH, and atmospheric gases have to be carefully controlled in order to build complex molecular structures. Ultraviolet light in particular is highly degrading to organic compounds. Some origin-of-life researchers use these wavelengths of light to make their compounds, and as soon as those compounds are synthesized, the lights are removed to prevent further rapid degradation. That is convenient in a lab, but how is that done outside the laboratory, such as at the edge of a volcano, and repeatedly? The ultraviolet light that is present in the atmosphere will severely degrade the molecules if left even for days or months.

Hurdle 10: Molecular Characterization

Molecular characterization at each step is essential. If the chemist doesn't know the molecular structure or at least the gross composition of the intermediates, the process is doomed for failure. So how might this be done in a prebiotic milieu? A prebiotic system knows nothing of molecular structure. It is mindless.

Hurdle 11: Isolation

Each organic reaction needs a carefully controlled work-up (isolation) protocol to prevent decomposition of the product. For example, nucleotides are sensitive compounds, and chemists today, even origin-of-life chemists, take great pains to work up these reactions very carefully.

Hurdle 12: Mass Transfer

The mass transfer problem will be the killer of all routes. How does one bring sufficient material through a complex multistep synthesis? If the route runs out of material after, say, 300 million years of progression, how does it go back to make more when nature has never kept a laboratory notebook of its former path?

In addition to origin-of-life researchers leaving this problem unaddressed, they exacerbate it. One origin-of-life research team will publish a paper where they make a trace amount of a stereochemically impure target, like a particular carbohydrate. And then the next researcher will use that formerly published carbohydrate as their starting point for the next synthetic step, claiming a protocol called "relay synthesis." But the new researcher will either buy the intermediate in large amounts and pure homochiral form or make it using purely advanced synthesis, separation, and characterization means. They will not use that former cumbersome proposed prebiotic route. So there is no accountability of mass transfer when going from one published work to the next; a prebiotic world would never have such a luxury.

And how many chemical steps are needed to make all the chemicals that compose a simple cell, and in sufficient quantities to build the higher-order structures within a cell? Nobody knows, but the number of steps must be enormous, regardless of whether the compounds are made by linear or convergent routes. Any synthetic chemist knows that the mass transfer would be daunting and impenetrable in their advanced laboratories. A typical thirty-step synthesis, using our most advanced methods, can often afford less than 1% overall yield of the final product in an optimized sequence.

How this could have been done with thousands of necessary steps to the thousands of requisite compounds in a prebiotic world is presently beyond our comprehension. So most researchers just bypass this difficulty by not mentioning it.

Furthermore, claims of "it only has to happen once" are incorrect. The chemistry would have to occur repeatedly, en masse, to produce the quantities needed to progress through a mindless and structurally blinded synthesis. Some might argue that higher molecular concentrations might have accumulated locally next to volcanoes where there is a heat source, or in gels, but how could this happen repeatedly through broad arrays of chemical classes? There is no reasonable explanation.

Elsewhere I have considered and discussed these deficiencies in synthesis in greater detail, presenting several examples from the recent literature.[7] When the obvious glaring problems are unaddressed, might this explain the arrested state of origin-of-life research when compared to the progress of other fields?

Molecular Assembly Experiments

In addition to chemical synthesis experiments that do not traverse the hurdles, there are origin-of-life experiments that deal with the assembly of chemicals into what researchers refer to as a "protocell," that is, "a self-organized, endogenously ordered, spherical collection of lipids proposed as a stepping-stone to the origin of life."[8]

Basically, if one takes a few drops of a lipid, adds them to water, and shakes, lamellae can form, which are lipid bilayer films. A small amount of spherical bilayer vesicles can break off from these lamellae, but much higher yields are realized if the lamellae are put through shear forces such as obtained during sonication. While origin-of-life researchers call the results "protocells," no life or pre-life exists. It remains lipid bilayer vesicles in water.

Most so-called protocell assembly experiments in origin-of-life research can be summed up by a protocol analogous to this:

- Purchase one homochiral lipid type from a chemical company or synthesize stereochemically scrambled lipids from smaller molecules. Add those lipids to water and observe the simple and expected thermodynamically driven assembly of those lipids into synthetic bilayer vesicles upon agitation. Sometimes the researchers will add other molecules, like nucleotides, that are engulfed by the vesicle as it forms.
- Publish a paper claiming that the synthetic vesicles are protocells and suggestive of early forms of cellular life.[9]
- Engage with the media to ramp up the hype.
- Watch the layperson being misled.

Here is one of many recent examples, published in 2017, of standard chemistry being portrayed as having something to do with the construction of a living cell.[10] A team from the *Origins of Life Initiative* at Harvard University performed a known type of polymerization reaction in water, called Reversible Addition-Fragmentation Chain Transfer (RAFT). This reaction type is not seen in nature—it is a purely synthetic process. The monomers that the research team chose are all synthetic and unnatural. This is standard chemistry used to make polymers wherein there is a controlled radical polymerization reaction that can afford a polymer chain bearing a hydrophobic block attached to a hydrophilic block when two different monomer types are used sequentially. The researchers observed these to form polymeric vesicles during the polymerization, which is interesting but surely not extraordinary.

While they kept the radical chain growing through ultraviolet light activation (a typical activating source) the vesicles grew, consuming monomer within the vesicles, to the point where the vesicles would burst. Again, nothing surprising; a critical vesicle size is reached and then the forces between the growing vesicle and the surrounding water dictate a critical growth volume before the vesicle ruptures. The vesicles move toward the ultraviolet light, likely by heating gradients induced by the light source or reaction thermodynamics.

Chemists like myself find this type of polymerization reaction to be interesting. It was a fine job by the researchers and well-worth publishing. The claims should have ended there. But here is how the work was portrayed in the published article:

> The observed net oscillatory vesicle population grows in a manner that reminds one of some elementary modes of sustainable (while there is available "food"!) population growth seen among living systems. The data supports an interpretation in terms of a micron scale self-assembled molecular system capable of embodying and mimicking some aspects of "simple" extant life, including self-assembly from a homogenous but active chemical medium, membrane formation, metabolism, a primitive form of self-replication, and hints of elementary system selection due to a spontaneous light triggered Marangoni instability [surface tension gradients].[11]

Was that statement justified? Just because A "reminds" me of B, it does not make A an "embodying" form of B—it is just my imagination. If the disc-shaped vesicle "reminds one" of a flying saucer, is it a "simple extant" flying saucer? No extant life, not even simple extant life, was demonstrated.

Following those excessive extrapolations by the authors, the claims were then rephrased and projected to the lay public by the *Harvard Gazette* and other news outlets: "A Harvard researcher seeking a model for the earliest cells has created a system that self-assembles from a chemical soup into cell-like structures that grow, move in response to light, replicate, and exhibit signs of rudimentary evolutionary selection."[12] Is that an accurate representation of the article? Surely not.

Here is a listing of a few of the challenges that need to be considered when dealing with lipid bilayer vesicle experiments common to most origin-of-life protocols that are being published.[13]

Challenge 1: Heterogeneity

Researchers have identified thousands of different lipid structures in modern cell membranes. These include glycerolipids, sphingolipids, ste-

rols, prenols, saccharolipids, and polyketides. All are homochiral or sp^2-stereo-defined.

For this reason, the origin-of-life researchers' selection of simple one-component synthetic vesicle lipid bilayers is far from realistic. When making synthetic vesicles—synthetic lipid bilayer membranes—mixtures with monoacyl lipids can destabilize the system, so researchers conveniently avoid these mixtures, while a prebiotic earth would not have that option. The heterogeneity of lipid bilayer structure is essential for cellular function, yet very hard for the researcher to reproduce.

Challenge 2: Varying Lipid Composition

Lipid bilayers surround subcellular organelles, such as nuclei and mitochondria, which are themselves microsystem assemblies. Each of these has their own lipid composition, different from the host vesicle.

Challenge 3: Symmetry

Lipid bilayers have a non-symmetric distribution. The outer and inner faces of the lipid bilayer are chemically inequivalent and cannot be interchanged without flippase enzymes, yet origin-of-life bilayer membranes are homogeneous across the bilayer; hence, they do not resemble the lipid bilayer of a living cell.

Challenge 4: Gatekeepers

Protein–lipid complexes and ionophores are the required passive transport sites and active pumps for the passage of molecules and ions through bilayer membranes, often with high specificity. Some allow passage for substrates into the compartment, and others their exit—they are highly specialized gatekeepers composed of very intricate structures. These complexes are rarely addressed by researchers working on their so-called protocell assemblies, yet they are essential for cell function.

Challenge 5: Glycans

Most cellular lipid bilayers have vast numbers of polycarbohydrate appendages, known as glycans. These are essential for cell regulation. For example, just six repeat units of the carbohydrate D-pyranose can form

more than one trillion different hexamers through branching (constitutional) and glycosidic (stereochemical) diversity. The diversity in branching patterns can store more information about the state of the cell than both DNA and RNA combined.[14]

Every cell membrane is coated with a complex array of glycans, and all cell-to-cell interactions take place through carbohydrate participation on the lipid bilayer membrane surface. Eliminating any class of carbohydrates from an organism results in its death, and almost every known cellular dysfunction involves carbohydrates.

Furthermore, in nature, these glycans are not made using a direct genetic template but result from the activity of several hundreds of enzymes organized in complex pathways—these are super-hard to construct and their structures selectively morph throughout cellular life changes.

So how do the origin-of-life researchers address the prebiotic synthesis of these complex lipid bilayers? They do not. Yet they claim a protocell through merely the formation of a homogeneous lipid bilayer vesicle. Might this retard the field?

Another example: Lipid bilayer assembly experiments were conducted by teams from the University of California at Santa Cruz and the University of New South Wales in Australia, and they disclosed a summary of the work in 2017.[15] These teams combined nucleotides and lipids in water to form lamellae with the nucleotides sandwiched between the layers. Recall that nucleotides are trimers of nucleobase-carbohydrate-phosphate, and in this case they were purchased in pure homochiral form—so already in a well-developed state. The lipids were also purchased in pure homochiral form.

The researchers showed that a condensation polymerization of the nucleotides via the pre-loaded phosphate with the purchased stereo-defined alcohol moiety on a neighboring nucleotide can take place within the lamella upon dehydration. They further demonstrated that similar reactions can occur at the edges of hydrothermal fields associated with volcanic landmasses to provide the heat needed for the reactions. The

chemistry is indifferent to the heat source, whether a volcano, a Bunsen burner, or a laboratory heating oven; the nucleotide will polymerize upon reaching a critical concentration and temperature.

The chemistry is unremarkable since it is preloaded through the purchased derivatives. This work addresses the essential concentration needs by removing the water and driving the intermolecular reactions to form oligomers that resemble the nucleic acids. The problem with a condensation (step growth) polymerization is that any alcohol can compete for the reactive electrophilic site, but in the researchers' case, they conveniently added only nucleotides and no other alcohols. In other words, the system is stacked to work through its purity. Condensation polymerization reactions need to be very pure, free of competing nucleophilic and electrophilic components, as explained by the Carothers equation defining degrees of polymerization based upon monomer purity.[16] If there happen to be amino acids or carbohydrates with the nucleotides, these would terminate or interrupt the growth of the oligonucleotides.

Moreover, the researchers did not confirm the detailed integrity of the claimed structures, which, if carefully analyzed, would likely show attacks from unintended hydroxyl sites. Nonetheless, even when short oligonucleotides form, they are not a usable form of RNA, since they have no useful sequences. It would be like a book of random letters, or in this actual case a small book of all the same letters.

The authors suggest that the lamellae sandwiching oligonucleotides eventually break off to form lipid bilayer vesicles containing the oligonucleotide-within-vesicle constructs, which they call protocells. The conversion of planar lamellae into multilamellar vesicles (onion-like structures) as they hydrate is well-established, but these generally need shearing (extrusion-type mechanical) forces, sonication, or peptides in order to form the requisite lipid bilayer vesicle, so the researchers' yields of the desired vesicles were sure to be very low.[17]

The conditions used in this experiment are hard to fathom being found in the prebiotic earth: homochiral nucleotides in high chemical purity, trapped in a lamella composed of homochiral stereo- and regio-

chemically pure lipids. Even accepting that improbability, those obtained vesicle structures have almost no resemblance to cellular lipid bilayers that have a vastly more complex constitution. The authors are merely forming lipid bilayer balls made from purchased homochiral lipids containing some randomly sequenced oligonucleotides from purchased homochiral nucleotides.

While exciting chemistry to the origin-of-life researcher, nothing here is chemically remarkable and it has almost no resemblance to a real cell. Nonetheless, behold the claims in the published paper:

- "Then, in the gel phase, protocells pack together in a system called a progenote and exchange sets of polymers, selecting those that enhance survival during many cycles."[18] But chemicals know nothing of survival since they are indifferent to "survival." There is no mechanism shown for how their protocells would bear different sets of polymers or exchange their sets of polymers between them, or make a "selection" process. The researchers misappropriate terms from biology and use them in a prebiotic world in a manner that makes no chemical sense.
- "The best-adapted protocells spread to other pools or streams, moving by wind and water, and some develop the ability to use carbon dioxide for photosynthesis."[19] However, there is no suggestion regarding the meaning of "best-adapted." It is again a misuse of terminology. Photosynthesis is a highly precise process requiring many enzymes, a well-ordered electron pathway, and precisely defined distances between photon receptors and electron ejectors, with electron transfers traveling down defined homochiral polypeptide channels. The authors' statement not only blurs the line of realism, but is fallacious.
- "After much trial and error, one protocell assembles the complicated molecular machinery that enables it to divide into daughter cells. This paves the way for the first living microbial community."[20] However, there is neither a demonstration of

how "molecular machinery" is made, nor even a proposal. The mechanisms needed for cellular division are highly complex, requiring cascades of enzymes functioning in precise and timed manner. This is utterly inconceivable based on the demonstrated results, and nothing proposed, let alone demonstrated, "paves the way for the first living microbial community."

- And these "ultimately evolve into a primitive metabolism required by the earliest forms of life."[21] It seems to be commonplace for origin-of-life researchers to co-opt terms from biological evolution and move them into the prebiotic vocabulary. This is unhelpful. Molecules are indifferent to moving toward life. Furthermore, what is a "primitive metabolism"? There is nothing being metabolized. There is only a condensation polymerization, a simple chemical reaction based upon the addition of nucleophiles to electrophiles. Such a reaction is never referred to as a metabolism within synthetic chemistry.

Those origin-of-life assembly claims are akin to buying twenty pounds of sliced turkey meat, adding a gallon of turkey broth, warming, sticking in a few feathers and suggesting that a "prototurkey," "primitive turkey," or "extant turkey" had just been synthesized.

A book by the famous science writer Ed Regis, entitled *What is Life?: Investigating the Nature of Life in the Age of Synthetic Biology*, attempts to describe life's origin from molecules: "Life began with little bags of garbage, random assortments of molecules doing some crude kind of metabolism. That is stage one. The garbage bags grow and occasionally split in two, and the ones that grow and split fastest win."[22] Few origin-of-life researchers would state it so shamelessly; nonetheless, "little bags of garbage" are precisely what origin-of-life researchers have been making. Those "little bags of garbage" have no more resemblance to living cells than a big bag of garbage resembles a horse.

There is a highly complex non-covalent interactive connectivity within a functioning cell—just like the parts of a machine need to be

fitted together—but with far more complexity in a biological organism. Nobody knows how a viable cell emerges from the massive combinatorial complexity of its molecular components. Of course, nobody has ever synthetically mimicked it either.

To begin to grasp the complexity involved, consider the interactome. An "interactome" is the whole set of molecular interactions in a particular cell.[23] Just as one sees the precise overlap and interconnectivity in human anatomical structures, in molecular biology the interconnectivity effects (through van der Waals interactions) are displayed trillions of times more abundantly than in gross human anatomy. The interactome can be protein-protein, gene-gene, or molecule-molecule interactions, and these greatly affect the function of the cellular system. It is through the molecular interactions that information is transferred. Electrostatic potentials permit information to flow through non-covalent molecular arrays, but these molecules need specific orientations relative to each other.[24] The interactome defines the intermolecular orientations, alignments that are unattainable through random mixing.

Peter Tompa of the University of Brussels and George Rose from Johns Hopkins University calculated that if one merely considers all protein-protein interactome combinations in just a single yeast cell, the result is an estimated $10^{79,000,000,000}$ combinations.[25] That is the number 1 followed by 79 billion zeros, a whoppingly large number. To put that in perspective, the number of elemental particles in the universe is estimated to be $\sim 10^{90}$. These numbers are beyond the realm of human appreciation.

The authors understate the ramifications, writing that "the numbers preclude formation of a functional interactome by trial and error complex formation within any meaningful span of time." Thus, "a complicated cellular sorting/trafficking and assembly system, made up of membranous organelles, receptors, membrane translocation devices, cytoskeletal tracks, motor proteins, and accessory chaperones guides the proper compartmentalization, localization, and assembly of proteins in the cell." But even with all that sophisticated biochemical guidance

and scaffolding, "in the absence of energy even this well developed infrastructure would be insufficient to account for the generation of the interactome, which requires a continuous expenditure of energy to maintain steady state." They conclude:

> The inability of the interactome to self-assemble de novo imposes limits on efforts to create artificial cells and organisms, that is, synthetic biology. In particular, the stunning experiment of 'creating' a viable bacterial cell by transplanting a synthetic chromosome into a host stripped of its own genetic material has been heralded as the generation of a synthetic cell (although not by the paper's authors). Such an interpretation is a misnomer, rather like stuffing a foreign engine into a Ford and declaring it to be a novel design. The success of the synthetic biology experiment relies on having a recipient interactome… that has high compatibility with donor genetic material. The ability to synthesize an actual artificial cell using designed components that can self-assemble spontaneously still remains a distant challenge.[26]

Regarding the ability to effect reactions through successive dehydration and rehydration steps as proposed by some researchers, Tompa and Rose write that "it is implausible that a completely 'denatured' cell could be reversibly renatured spontaneously, like a protein. Instead, new cells are generated by the division of pre-existing cells, an unbroken chain of renewal tracking back through contingent conditions and evolving responses to the origin of life on the prebiotic earth." Indeed, "all extant cells are generated by the division of preexisting cells that provide the necessary template for perpetuation of the interactome."[27]

Therefore, even if one were to try to simplify the problem with network connectivity theory, interactomes add a massive layer of complexity to all cellular structures. That further underscores the difference between a real cell and the so-called protocells or extant cells made by origin-of-life researchers. In fact, terms such as "protocells" or "extant cells" are misnomers that exacerbate the confusions.

So how close have researchers come to creating an artificial cell? In 2010, Craig Venter's team made a copy of a known bacterial genome and transplanted it into another cell.[28] In 2016 the Venter team did some-

thing related. They removed all but 473 genes from a natural genome, transplanting it into another cell.[29]

These are indeed exciting experiments, but the cells were already made, naturally, and alternate genomes were inserted. This is analogous to buying two Corvettes, removing one of the electronic engine control modules (ECUs) from the first Corvette, and swapping it as a substitute into the second Corvette; or copying the ECU in a fabrication facility and inserting the copied version into a car. One could not rightly claim the building of a Corvette; it is an exchange of parts, while the cars already existed.

More recently Henrike Niederholtmeyer, Cynthia Chaggan, and Neal K. Devaraj of the University of California at San Diego have made what they term "mimics of eukaryotic cells"[30]; the journal *Science* declared these "the most lifelike artificial cells yet."[31] In this experiment, semi-porous microcapsules made of *plastic* (from acrylate polymerization) containing *clay* were prepared using modern microfluidics techniques that are done within fabrication devices. Due to their inherent charges, these clays have a high affinity for binding DNA, so when DNA was then added to the solution, it diffused through the semi-porous plastic microcapsules and bound to the clay. The requisite RNA polymerases for mRNA transcription, ribosomes for polypeptide translation, tRNA, amino acids, enzymatic cofactors, energy sources, and cellular components essential for proper protein folding were similarly purchased or extracted from living systems, added to the medium, and permitted to diffuse into the plastic capsules.

The expected chemical reactions ensued, resulting in protein synthesis. The newly formed proteins could diffuse out of the plastic microcapsules to other nearby semi-porous plastic microcapsules that had been similarly prepared, and the nearer the neighboring plastic microcapsule was to the original microcapsule, the more exchange of reagents between them took place. Those neighboring plastic microcapsules could then similarly become production sites for proteins. This diffusion between nearby plastic microcapsules was termed "quorum sensing," relying on

standard local concentration gradients where the nearer neighbors received more of the leached materials. The chemistry of the exogenously added reagents will work regardless of the container, whether it be a plastic semi-porous microcapsule, a test tube, or a large-scale industrial production tank.

While the experimental design is clever and exciting, the actual chemical synthesis is unremarkable, and it is—as expected—based upon the purchased bio-extracted chemicals that were added. Such use of known and commercially available cellular components to synthesize new proteins is done every day in laboratories around the world, and one can buy commercial kits to do this.[32]

So it is far from the embarrassing press-hyped claim of "gene expression and communication rivaling that of living cells."[33] There is no rivalry here. All of the active chemical components for the synthesis were extracted from living systems. Further, one might arguably agree that these are indeed "the most lifelike artificial cells yet,"[34] but that only serves to underscore the point: Nobody has ever yet come close to generating the workings of life.

There are further demonstrations of such over-extrapolations. In a 2018 article entitled "How Did Life Begin?" in the top-ranked scientific venue in the world, *Nature*, Nobel laureate Jack Szostak wrote a synopsis for the process of life's origin. (The article appeared in the journal's special report, "Innovations In: The Biggest Questions in Science.") Directing his message to the non-expert, Szostak explained:

> …iron-cyanide compounds accumulated over time, building up into a concentrated stew of reactive chemicals. Life as we know it requires RNA. Some scientists believe that RNA emerged directly from these reactive chemicals, nudged along by dynamic forces in the environment. Nucleotides, the building blocks of RNA, eventually formed, then joined together to make strands of RNA. Some stages in this process are still not well understood. Once RNA was made, some strands of it became enclosed within tiny vesicles formed by the spontaneous assembly of fatty acids (lipids) into membranes, creating the first protocells. As the membranes incorporated more fatty acids, they grew

and divided; at the same time, internal chemical reactions drove replication of the encapsulated RNA.[35]

The descriptions listed here were derived from Szostak's earlier article in *Scientific American*,[36] and the presentation in *Nature* of Szostak's synopsis elevates its credibility in the eyes of the scientific community. But let us examine Szostak's claims.

First, Szostak's statement that "some scientists believe that RNA emerged directly from these reactive chemicals, nudged along by dynamic forces" is painful to a synthetic chemist because a complex pathway of reactions would be needed, along with all the steps of purification and then assembly, polymerization, and sequencing. All that is reduced to a simple passing sentence. For example, how could RNA emerge directly from iron cyanide? Iron cyanide is highly stable, and the concentration of free cyanide is minuscule. Nothing "emerges directly," let alone something as complex as RNA.

Further, words like "nudged along by dynamic forces" have no meaning in the realm of synthetic chemistry, though they seem acceptable to the layperson. That "nucleotides… eventually formed and then joined together to make strands of RNA," is an incredible statement for which there is no basis. Nucleotides do not merely join together with any significant precision without complex protection and deprotection steps.

In sum, Szostak's remark that "some stages in this process are still not well understood" would be more accurately phrased as "in almost all stages we remain clueless when it comes to the chemistry needed on a prebiotic earth."

Accompanying Szostak's article is a figure that purports to summarize the chemical process leading to the formation of RNA nucleotides.[37] However, the compounds listed in this figure as "simple sugars" are not sugars; they are glycerol and ethylene glycol. There are known routes to convert those to very simple sugars,[38] but only in gross relative and absolute stereochemically mixed states, and as a mixture of several different polyols—so separation problems abound that remain poorly

delineated. Carbohydrate synthesis is a very difficult problem for a prebiotic earth.[39] Further, the carbohydrate, as shown, is devoid of stereochemistry, and therefore is not ribose. If it is not ribose, then it cannot be an RNA nucleotide as written. Moreover, the nucleotide as drawn is dehydrated, and the "cyanide derivatives" as shown in the figure are unrecognizable as cyanide derivatives.

In an act of grace, let us attribute these chemical structural errors to the faulty renderings of a staff artist. However, far more importantly, there is no way that heat and light can directly make a nucleotide, even if there were simple sugars and cyanide derivatives. The primary literature upon which this scheme is based shows the process as involving ten to twelve chemical steps. Many of those steps form vast and unusable mixtures of products. As has been mentioned previously, researchers do not then use the desired compounds formed in those various mixtures; instead they just identify the compounds' presence, and then buy pure versions of them from a chemical vendor or else make them using purely modern synthetic chemical methods.

Granted, it is difficult to explain origin-of-life chemistry to the layperson, but that is part of the problem. Its portrayal to those outside the field has been so oversimplified as to mislead even the academic community. Professors themselves are confused regarding the state of origin-of-life research.

Chasing Fool's Gold?

When origin-of-life researchers are confronted by skeptics regarding the weakness of the data on the fundamental questions of life's origin, they will sometimes quote the famous late origin-of-life researcher, Leslie Orgel: "Anybody who thinks they know the solution to this problem [of the origin of life] is deluded. But anybody who thinks this is an insoluble problem is also deluded."[40] The remark is a strawman—the skeptics would merely enjoy seeing some new results that move the field toward an explanation of life's origin. The direction of origin-of-life research is suspect and the petty dismissal of questioning is unhelpful to the field.

So is the current fixation on extraneous experimental results. Consider an analogy from history: Alchemists tried to convert inexpensive metals into gold. They discovered that metals could be treated with sulfur to make yellow solids, sometimes even with lustrous facets, like pyrite, "fool's gold." While it was clear to the alchemists that they had not formed gold, would not the alchemist community have viewed sulfur additives as "a step in the right direction"?[41] These are the dangers of building a field of study around minute experimental results that do not even attempt to tackle the fundamental questions; one might be chasing fool's gold.

One such fundamental question that must be addressed is the origin of the chemical code; this is likely the single-most significant hurdle in any approach to understanding life's origin. The information or coding within the DNA or RNA that corresponds to the sequence of the nucleotides is essential to the entire discussion of life's origin. Some would rightly argue that the information is even more fundamental than the matter (molecules) upon which it is encoded. Present origin-of-life research does not address this foundational issue, but rather merely demonstrates that the requisite molecules are unlikely to have occurred in the states and quantities needed, and any assembly into an organism is even more unlikely.

This is grossly insufficient. The sequence of the nucleotides is the blueprint upon which life is founded. It is that code that will be translated to the enzymes that build the organism. The code defines the operating system for cellular function. The code vs. the molecules is analogous to the difference between the Library of Congress and a box of alphabetic letters—the library (DNA or RNA) has a huge amount of embedded information while the random box of letters (molecules) has little. We know from computer science that one needs complex non-regular patterns for complex computation and processing. Accordingly, complex patterns constitute the molecular assemblies seen in all living systems, even in the simplest bacterium. The simple regular pattern of thermodynamically driven crystallization or self-assembly is actually antithetical

to what is needed for organism function, even when considering a cellular lipid bilayer.

To demonstrate how far humankind is from generating life, if origin-of-life researchers were given all the molecules and their polymeric forms that they desperately seek, and all in 100% homochirality, and their advanced laboratories, and all the chemical literature, and the DNA and RNA in any sequence (code) that they wish, could they assemble even a simple cell? The answer is a resounding *No!* Moreover, there is not an origin-of-life researcher on earth that would claim differently. As with perpetual motion machines, the pieces just do not come together as advertised.

When all else fails for explanations, some call upon Father Time, suggesting that hundreds of millions of years solve their mysteries. No other field of chemistry would accept such a proposition. In chemical synthesis, as we have seen, time is often the enemy, especially when making kinetic products that constitute the requisite organic chemicals of life.

Interestingly, Edward Steele and his thirty-two co-authors, spread over eleven countries, in 2018 in *Progress in Biophysics and Molecular Biology* conceded the following:

> The transformation of an ensemble of appropriately chosen biological monomers (e.g., amino acids, nucleotides) into a primitive living cell capable of further evolution appears to require overcoming an information hurdle of superastronomical proportions, an event that could not have happened within the time frame of the Earth except, we believe, as a miracle. All laboratory experiments attempting to simulate such an event have so far led to dismal failure.[42]

Further, they add, "At this stage of our scientific understanding we need to *place on hold the issue of life's actual biochemical origins*—where, when and how may be too difficult to solve on the current evidence." [Italics added]

However, Steele and his co-workers then merely push back the problem by fancifully increasing the reaction space: "It would thus seem rea-

sonable to go to the biggest available 'venue' in relation to space and time. A cosmological origin of life thus appears plausible and overwhelmingly likely to us...." They write: "It is many orders of magnitude more likely that it emerged in one of the trillions of comet-like incubators or water-bearing planets (cosmic-wide versions of Darwin's 'warm little ponds') at a very early time in the growth of this Universe, perhaps 12 billion years ago which then went on to infect via knock-on effects other life-favourable sites (planets, moons, comets) throughout that Galaxy and then in an interconnected and interactive way throughout the Cosmos as the Universe expanded."[43]

In other words, while conceding that origin-of-life research has been a "dismal failure" and the community should "place a hold" on it, Steele and his colleagues reveal their own cluelessness regarding any of the details in life generation by hoping for a gigantic reaction space to overcome the vanishingly small probabilities of life originating from anything observable through "current evidence."

This too would require its own miracle.

In Praise of Humility

I have had cordial discussions with biologist proponents of origin-of-life research on these issues, and I am amazed that they fail to appreciate the magnitude of the problem in building molecules. These biologists see little difficulty in accepting a chemical synthesis where a desired product is mixed with a large array of closely related yet undesired compounds—mixtures from which separations would be enormously complex, and subsequent reactions unavailing.

But chemists see the inherent problems, even in their own research. John Sutherland of the University of Cambridge, one of origin-of-life's giants and the most skilled synthetic chemist to engage in origin-of-life research, has recently proposed that "chemical determinism can no longer be relied on as a source of innovation, and further improvements have to be chanced upon instead."[44]

"Chanced upon"? Why? Could it be due to chemistry's indifference to life and the cluelessness of the researchers?

It appears that Sutherland is grappling with the perplexity of the origin-of-life problem. The befuddlement is greatest for the synthetic chemist because he appreciates what molecules will and will not do, whereas to the biologist, all seems possible because he is used to using biology's constructs, while glossing over the requirements of the chemistry.

Another example: In 2017, Ramanarayanan Krishnamurthy of the Scripps Research Institute and his team cleverly showed that diamidophosphate can phosphorylate nucleosides, nucleotides, and stereoscrambled lipid precursors. These can further result in the formation of random oligonucleotides and oligopeptides. The fundamental challenges noted above for the synthesis and assembly experiments remain unaddressed, so Krishnamurthy was rightly measured in the claims within his publication, writing that "any comparison must be viewed with caution given the pitfalls of extrapolating extant biochemical pathways backwards all the way to prebiotic chemistry and vice versa."[45] More of these realistic conclusions are needed from the origin-of-life community.

Further refreshing comments are making their way into the primary literature. In a 2018 article in *Nature Communications*, Clemens Richert describes prebiotic chemistry versus human intervention. He explains that "the ideal experiment does not involve any human intervention."[46] Further, he even reflects upon the pure chemicals used by the researchers as being unrealistically available but prebiotically necessary for the syntheses to have ensued.

Thus, there is a glimmer of hope. The origin-of-life community is taking heed of their own unrealistic protocols that have supposedly been simulating prebiotic conditions.

And none too soon. Claims that mislead the all-too-patient taxpayer are not only dishonest, but unhelpful; the public will eventually realize that they have been taken for fools, and their ensuing distrust of scientific claims will carry over into other fields of scientific endeavor. Uncorrected or unfounded assertions jeopardize science beyond a singular

field, especially since there is mounting distrust of higher education in general.[47]

Going Forward

Bearing all this in mind, should origin-of-life research continue in the same vein as it has been practiced for the past two-thirds of a century? Does not the field's stagnation suggest that a dramatic change should be instituted?

The Defense Advanced Research Projects Agency (DARPA) presents challenges to shake the research and engineering community out of their stagnancy on topics related to technology, putting before them contests that demand proposals that are wholly unlike the status quo. DARPA mandates new fundamental ways to address problems, often embracing young nonconformist researchers who would not normally be funded by the seasoned research community of peers.

Origin-of-life research needs some such shake-up to do something beyond the making of yet another small chemical intermediate, *ad nauseam*, or forming suspensions of lipid bilayers, protocells as they call them, which have little resemblance to true cellular bylayer vesicles. Researchers must be challenged to address hurdles such as the origin of life's code, the complex assembly and interactomes that are essential to cellular functioning, and the mass-throughput in synthesis to provide the requisite quantities of molecules in their homochiral form. Alternatively, researchers must offer some conjectures, underpinned by experiments, to show that perhaps these features, such as the code or the interactomes, are irrelevant to life's origin from prebiotic chemicals.

Any moratorium needs to be initiated by the funding agencies and directed by the program managers. This starts with a thoughtful evaluation that compares origin-of-life progress to the progress in other fields of research over the past sixty-seven years. Are these current origin-of-life experiments taking us closer, or do the newer findings on cellular complexity drive the target further out of reach with each passing year?

Formulation of new programmatic goals should ensue, with those outside the mainstream origin-of-life community being encouraged to offer divergent thoughts. A moratorium is something that the scientific community might be obliged to request, because origin-of-life research uses taxpayer dollars, and its overexpressed assertions jeopardize trust in scientific claims in general.

As in any field, it is important to maintain engagement with the press so that the scientific message reaches the masses. The press has an essential role in the ecosystem of technical dissemination, and most science reporters will heed advice from the scientists whom they interview. We cannot continue to let them run unchecked. Their over-the-top claims jeopardize scientific credibility.

There also needs to be a cessation of the gross extrapolations and hyperbole within scientific publications themselves that give the impression that scientists are near to creating life. The field has migrated outside of the bounds of scientific credibility. Thus, journal editors should be held accountable to restrict grossly exaggerated claims and even terminology that is misleading. For example, when simply referring to a lipid bilayer vesicle, cavalier use of the term "protocell" should be discouraged; "lipid bilayer vesicle" or "liposome" is sufficient.

Therefore, I appeal to the research community and funding agencies to consider whether a moratorium on origin-of-life research is warranted.

This starts with a redefinition of targets that will address the fundamental questions: mass transfer of starting materials to the requisite four compound classes in high chemical and stereochemical purity, the origin of life's code, the massive combinatorial complexities present in any living system, and the precise non-regular assembly of required cellular components.

Without deliberate and widespread changes, origin-of-life progress will likely remain retarded.

Acknowledgment: The author thanks Paul Nelson for helpful insights. Walt Shaw and Steve Burgess of Avanti Polar Lipids kindly provided information on lipid bilayer assemblies.

Endnotes

1. Some portions of this chapter have been adapted or copied from the author's previous article: James Tour, "Time Out," *Inference: International Review of Science* 4, no. 4 (July 2019), https://inference-review.com/article/time-out.
2. "Wikipedia: Miller-Urey Experiment," Wikimedia Foundation, last modified July 29, 2019, 11:44, https://en.wikipedia.org/wiki/Miller-Urey_experiment. Stanley L. Miller, "A Production of Amino Acids Under Possible Primitive Earth Conditions," *Science* 117, no 3046 (May 15, 1953): 528–529, doi:10.1126/science.117.3046.528. Stanley L. Miller and Harold C. Urey, "Organic Compound Synthesis on the Primitive Earth," *Science* 130, no 3370 (July 31, 1959): 245–251, doi:10.1126/science.130.3370.245.
3. A. G. Cairns-Smith, *Genetic Takeover* (New York: Cambridge University Press, 1982).
4. "Wikipedia: History of Perpetual Motion Machines," Wikimedia Foundation, last modified July 8, 2019, 08:40, https://en.wikipedia.org/wiki/History_of_perpetual_motion_machines.
5. Laura M. Barge, Erika Flores, Marc M. Baum, David G. VanderVelde, and Michael J. Russell, "Redox and pH Gradients Drive Amino Acid Synthesis in Iron Oxyhydroxide Mineral Systems," *PNAS* 116, no. 11 (March 12, 2019) 4828-4833, doi.org/10.1073/pnas.1812098116.
6. Arielle Samuelson, "NASA Study Reproduces Origins of Life on Ocean Floor" (February 25, 2019), NASA Science: Solar System Exploration, last modified February 27, 2019, https://solarsystem.nasa.gov/news/863/nasa-study-reproduces-origins-of-life-on-ocean-floor/.
7. James Tour, "Animadversions of a Synthetic Chemist," *Inference: International Review of Science* 2, no. 2 (May 2016), https://inference-review.com/article/animadversions-of-a-synthetic-chemist.
8. "Wikipedia: Protocell," Wikimedia Foundation, last modified July 24, 2019, 12:57, https://en.wikipedia.org/wiki/Protocell.
9. Dimitri Fayolle, Emiliano Altamura, Alice D'Onofrio, Warren Madanamothoo, Bernard Fenet, Fabio Mavelli, René Buchet, Pasquale Stano, Michele Fiore and Peter Strazewski, "Crude Phosphorylation Mixtures Containing Racemic Lipid Amphiphiles Self-Assemble to Give Stable Primitive Compartments," *Scientific Reports* 7 (December 22, 2017), doi:10.1038/s41598-017-18053-y.
10. Anders N. Albertsen, Jan K. Szymański, and Juan Pérez-Mercader, "Emergent Properties of Giant Vesicles Formed by a Polymerization-Induced Self-Assembly (PISA) Reaction," *Scientific Reports* 7 (January 27, 2017), doi:10.1038/srep41534.
11. Albertsen, "Emergent Properties of Giant Vesicles."
12. Alvin Powell, "Mimicking Life in a Chemical Soup," *The Harvard Gazette*, March 31, 2017, https://news.harvard.edu/gazette/story/2017/03/harvard-researcher-creates-chemical-system-that-mimics-early-cell-behavior/. Harvard University, "Researcher Creates Chemical System That Mimics Early Cell Behavior," Science X (website), April 3, 2017, https://phys.org/news/2017–04-chemical-mimics-early-cell-behavior.html. Ava Jones, "Harvard Scientist Discovers Phoenix Vesicles in Quest to Mimic Life," *University Herald*, April 4, 2017, https://www.universityherald.com/articles/71576/20170404/harvard-scientist-discovers-phoenix-vesicles-quest-mimic-life.htm. "Researcher Creates Chemical Complement That Mimics Early Dungeon Behavior, Health Medicine Network (website),

no date, http://healthmedicinet.com/researcher-creates-chemical-system-that-mimics-early-cell-behavior/.

13. James Tour, "An Open Letter to My Colleagues," *Inference: International Review of Science* 3, no. 2 (August 2017), https://inference-review.com/article/an-open-letter-to-my-colleagues.

14. Roger A. Laine, "Invited Commentary: A Calculation of All Possible Oligosaccharide Isomers Both Branched and Linear Yields 1.05 × 1012 Structures for a Reducing Hexasaccharide: the Isomer Barrier to Development of Single-Method Saccharide Sequencing or Synthesis Systems," *Glycobiology* 4, no. 6 (December 1994): 759–767, doi:10.1093/glycob/4.6.759.

15. Martin J Van Kranendonk, Tara Djokic, and David Deamer, "Life Springs," *Scientific American* 317, no. 2 (July 2017): 28–35, doi:10.1038/scientificamerican0817-28.

16. "Wikipedia: Carothers Equation," Wikimedia Foundation, last modified December 7, 2017, 00:28, https://en.wikipedia.org/wiki/Carothers_equation.

17. Avanti Polar Lipids (website), https://avantilipids.com/tech-support/liposome-preparation/. Barbara Bianca Gerbelli, Emerson Rodrigo da Silva, Bruna Miranda Soares, Wendel Andrade Alves, and Elisabeth Andreoli de Oliveira, "Multilamellar-to-Unilamellar Transition Induced by Diphenylalanine in Lipid Vesicles," *Langmuir* 34, no. 5 (December 28, 2017): 2171–2179, doi:10.1021/acs.langmuir.7b03869.

18. Van Kranendonk,"Life Springs."

19. Van Kranendonk,"Life Springs."

20. Van Kranendonk,"Life Springs."

21. Van Kranendonk,"Life Springs."

22. Ed Regis, *What is Life?: Investigating the Nature of Life in the Age of Synthetic Biology* (Oxford: Oxford University Press, 2009), 103.

23. "Wikipedia: Interactome," Wikimedia Foundation, last modified June 3, 2019, 15:10, https://en.wikipedia.org/wiki/Interactome.

24. James M. Tour, Masatoshi Kozaki, and Jorge M. Seminario, "Molecular Scale Electronics: A Synthetic/Computational Approach to Digital Computing," *Journal of the American Chemical Society* 120, no. 33 (1998), 8486–8493, doi:10.1021/ja9808090.

25. Peter Tompa and George D. Rose, "The Levinthal Paradox of the Interactome," *Protein Science* 20, no. 12 (October 10, 2011): 2074–2079, https://doi.org/10.1002/pro.747.

26. Tompa and Rose, "The Levinthal Paradox."

27. Tompa and Rose, "The Levinthal Paradox."

28. Daniel G. Gibson, John I. Glass, Carole Lartigue, Vladimir N. Noskov, Ray-Yuan Chuang, Mikkel A. Algire, Gwynedd A. Benders, Michael G. Montague, Li Ma, Monzia M. Moodie, Chuck Merryman, Sanjay Vashee, Radha Krishnakumar, Nacyra Assad-Garcia, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Lei Young, Zhi-Qing Qi, Thomas H. Segall-Shapiro, Christopher H. Calvey, Prashanth P. Parmar, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter, "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome," *Science* 329, no. 5987 (July 2, 2010): 52–56, doi:10.1126/science.1190719.

29. Clyde A. Hutchison III, Ray-Yuan Chuang, Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogumil J. Karas, Li Ma, James F. Pelletier, Zhi-Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski,

Lijie Sun, Yo Suzuki, Billyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman, Daniel G. Gibson, and J. Craig Venter, "Design and Synthesis of a Minimal Bacterial Genome," *Science* 351, no. 6280 (March 25, 2016): aad6253, doi 10.1126/science.aad6253.

30. Henrike Niederholtmeyer, Cynthia Chaggan, and Neal K. Devaraj, "Communication and Quorum Sensing in Non-Living Mimics of Eukaryotic Cells," *Nature Communications* 9 (November 28, 2018), https://doi.org/10.1038/s41467-018-07473-7.
31. Mitch Leslie, "Biologists Create the Most Lifelike Artificial Cells Yet," *Science* (November 19, 2018), doi:10.1126/science.aaw1173.
32. "Cell Free Protein Expression," ThermoFisher Scientific (website), https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/cell-free-protein-expression.html.
33. University of California, "Copycat Cells Command New Powers of Communication," Science X (website) December 7, 2018, https://phys.org/news/2018-12-copycat-cells-powers.html.
34. Leslie, "Biologists Create the Most Lifelike Artificial Cells Yet."
35. Jack Szostak, "How Did Life Begin? Untangling the Origins of Organisms Will Require Experiments at the Tiniest Scales and Observations at the Vastest," *Nature* (May 9, 2018), https://www.nature.com/articles/d41586-018-05098-w.
36. Jack Szostak, "How Did Life Begin? Untangling the Origins of Organisms Will Require Experiments at the Tiniest Scales and Observations at the Vastest," *Scientific American* 318, no. 6 (June 2018): 65–67, doi:10.1038/scientificamerican0618-65.
37. The figure can be viewed at https://www.nature.com/articles/d41586-018-05098-w.
38. Dougal Ritson and John D. Sutherland, "Prebiotic Synthesis of Simple Sugars by Photoredox Systems Chemistry," *Nature Chemistry* 4 (2012): 895–899, doi:10.1038/nchem.1467.
39. James Tour, "Animadversions of a Synthetic Chemist," *Inference: International Review of Science* 2, no. 2 (May 2016), https://inference-review.com/article/animadversions-of-a-synthetic-chemist.
40. Jason Socrates Bardi, "Study Suggests Component of Volcanic Gas May Have Played a Significant Role in the Origins of Life on Earth," *Scripps Research Institute News & Views* 3, no. 30 (October 11, 2004), https://www.scripps.edu/newsandviews/e_20041011/ghadiri.html.
41. A. G. Cairns-Smith, *Genetic Takeover* (New York: Cambridge University Press, 1982).
42. Edward J. Steele, Shirwan Al-Mufti, Kenneth A. Augustyn, Rohana Chandrajith, John P. Coghlan, S. G. Coulson, Sudipto Ghosh, Mark Gillman, Reginald M. Gorczynski, Brig Klyce, Godfrey Louis, Kithsiri Mahanama, Keith R. Oliver, Julio Padron, Jiangwen Qu, John A. Schuster, W. E. Smith, Duane P. Snyder, Julian A. Steele, Brent J. Stewart, Robert Temple, Gensuke Tokoro, Christopher A. Tout, Alexander Unzicker, Milton Wainwright, Jamie Wallis, Daryl H. Wallis, Max K. Wallis, John Wetherall, D. T. Wickramasinghe, J. T. Wickramasinghe, N. Chandra Wickramasinghe, Yongsheng Liu, "Cause of Cambrian Explosion—Terrestrial or Cosmic?" *Progress in Biophysics and Molecular Biology* 136 (August 2018): 3–23, https://doi.org/10.1016/j.pbiomolbio.2018.03.004.
43. Steele, "Cause of Cambrian Explosion."

44. John D. Sutherland, "Opinion: Studies on the Origin of Life—the End of the Beginning," *Nature Reviews Chemistry* 1 (January 18, 2017), https://doi.org/10.1038/s41570-016-0012.

45. Clémentine Gibard, Subhendu Bhowmik, Megha Karki, Eun-Kyong Kim, and Ramanarayanan Krishnamurthy, "Phosphorylation, Oligomerization and Self-Assembly in Water Under Potential Prebiotic Conditions," *Nature Chemistry* 10 (2018): 212–217, https://www.nature.com/articles/nchem.2878.

46. Clemens Richert, "Prebiotic Chemistry and Human Intervention," *Nature Communications* 9 (December 12, 2018), https://doi.org/10.1038/s41467-018-07219-5.

47. Shiloh Rea, "Uncertainty Perception Drives Public's Trust, Mistrust of Science," Carnegie Mellon University, Dietrich College of Humanities and Social Sciences (website), February 22, 2017, https://www.cmu.edu/dietrich/news/news-stories/2017/february/broomell-scientific-uncertainty.html; Tania Lombrozo, "Science Isn't Partisan, But Public Perception Of Science Often Is," NPR: Cosmos & Culture (website), April 24, 2017, https://www.npr.org/sections/13.7/2017/04/24/525360310/science-isn-t-partisan-but-public-perception-of-science-often-is; Benjamin Wermund, "University Presidents: We've Been Blindsided," *Politico*, December 19, 2017, https://www.politico.com/story/2017/12/19/college-university-backlash-elitism-296898.

17. Evidence of Intelligent Design in the Origin of Life

Stephen C. Meyer

Theories about the origin of life necessarily presuppose knowledge of the attributes of living cells. As historian of biology Harmke Kamminga has observed, "At the heart of the problem of the origin of life lies a fundamental question: What is it exactly that we are trying to explain the origin of?"[1] Or as the pioneering chemical evolutionary theorist Alexander Oparin put it, "The problem of the nature of life and the problem of its origin have become inseparable."[2] Origin-of-life researchers want to explain the origin of the first and presumably simplest—or, at least, minimally complex—living cell. As a result, developments in fields that explicate the nature of unicellular life have historically defined the questions that origin-of-life scenarios must answer.

Since the late 1950s and 1960s, origin-of-life researchers have increasingly recognized the complex and specific nature of unicellular life and the biomacromolecules on which such systems depend. Further, molecular biologists and origin-of-life researchers have characterized this complexity and specificity in informational terms. Molecular biologists routinely refer to DNA, RNA, and proteins as carriers or repositories of "information."[3] Many origin-of-life researchers now regard the origin of the information in these biomacromolecules as the central question facing their research. As Bernd-Olaf Kuppers has stated, "The

problem of the origin of life is clearly basically equivalent to the problem of the origin of biological information."[4]

This chapter will evaluate competing explanations for the origin of the information necessary to build the first living cell. To do so will require determining what biologists have meant by the term information as it has been applied to biomacromolecules. As many have noted, "information" can denote several theoretically distinct concepts. This chapter will attempt to eliminate this ambiguity and to determine precisely what type of information origin-of-life researchers must explain "the origin of." What follows will first seek to characterize the information in DNA, RNA, and proteins as a fact in need of explanation; and, second, to evaluate the efficacy of competing classes of explanation for the origin of biological information.

Part I will seek to show that molecular biologists have used the term "information" consistently to refer to the joint properties of complexity and functional specificity or specification. Biological usage of the term will be contrasted with its classical information-theoretic usage to show that "biological information" entails a richer sense of information than the classical mathematical theory of Shannon and Wiener. Part I will also argue against attempts to treat biological "information" as a metaphor lacking empirical content and/or ontological status.[5] It will show that the term biological information refers to two real features of living systems, complexity and specificity, features that jointly do require explanation.

Part II will evaluate competing types of explanation for the origin of the specified biological information necessary to produce the first living system. From the 1920s to the mid-1960s, origin-of-life researchers relied heavily on theories emphasizing the creative role of random events—"chance"—often in tandem with some form of prebiotic natural selection. Since the late 1960s, theorists have instead emphasized deterministic self-organizational laws or properties—that is, physical-chemical "necessity." Part II will show the causal inadequacy of explanations involving "chance," "necessity," and the combination of the two.

Part III will suggest that the origin of biological information requires a radically different explanatory approach. It will argue that our present knowledge of causal powers suggests intelligent design as a better, more causally adequate explanation for the origin of the specified complexity (the information so defined) present in large biomolecules such as DNA, RNA, and proteins.

I.

A. The Growing Recognition of the Complexity of the Cell

After Darwin published the *Origin of Species* in 1859, many scientists began to think about a problem that Darwin had not addressed.[6] Although Darwin's theory purported to explain how life could have grown gradually more complex starting from "one or a few simple forms," it did not explain, or attempt to explain, how life had first originated. Yet in the 1870s and 1880s, evolutionary biologists like Ernst Haeckel and Thomas Huxley assumed that devising an explanation for the origin of life would be fairly easy, based on their assumption that life was, in essence, a chemically simple substance called "protoplasm" that could easily be constructed by combining and recombining simple chemicals such as carbon dioxide, oxygen, and nitrogen.

Over the next sixty years, biologists and biochemists gradually revised their view of the nature of life. During the 1860s and 1870s, biologists tended to see the cell, in Haeckel's words, as an undifferentiated and "homogeneous globule of plasm." By the 1930s, however, most biologists had come to see the cell as a complex metabolic system.[7] Origin-of-life theories reflected this increasing appreciation of cellular complexity. Whereas nineteenth-century theories of abiogenesis envisioned life arising almost instantaneously via a one- or two-step process of chemical "autogeny," early twentieth-century theories, such as Oparin's theory of evolutionary abiogenesis, envisioned a multibillion-year process of transformation from simple chemicals to a complex metabolic system.[8] Even

so, most scientists during the 1920s and 1930s still vastly underestimated the complexity and specificity of the cell and its key functional components—as developments in molecular biology would soon make clear.

B. The Complexity and Specificity of Proteins

During the first half of the twentieth century, biochemists had come to recognize the centrality of proteins to the maintenance of life. However, they repeatedly underestimated the complexity of proteins. Beginning in the 1950s a series of discoveries caused this simplistic view of proteins to change. Researchers ultimately found that proteins exhibit an extraordinarily complex and irregular three-dimensional shape: a twisting, turning, tangle of amino acids. As John Kendrew explained in 1958, "The big surprise was that it was so irregular… the arrangement seems to be almost totally lacking in the kind of regularity one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure."[9]

By the mid-1950s, biochemists recognized that proteins possess another remarkable property. In addition to their complexity, proteins also exhibit *specificity*. Whereas proteins are built from chemically rather simple amino acid "building blocks," their function (whether as enzymes, signal transducers, or structural components in the cell) depends crucially on a specific arrangement of those building blocks.[10] In particular, the specific sequence of amino acids in a chain and the resultant chemical interactions between amino acids largely determine the specific three-dimensional structure that the chain as a whole will adopt. Those structures or shapes in turn determine what function, if any, the amino acid chain can perform in the cell.

For a functioning protein, its three-dimensional shape gives it a hand-in-glove fit with other molecules, enabling it to catalyze specific chemical reactions or to build specific structures within the cell. Because of its three-dimensional specificity, one protein can usually no more substitute for another than one tool can substitute for another. A topoisomerase can no more perform the job of a polymerase than a hatchet

can perform the function of a soldering iron. Instead, proteins perform functions only by virtue of their three-dimensional specificity of fit, either with other equally specified and complex molecules or with simpler substrates within the cell. Moreover, the three-dimensional specificity derives in large part from the one-dimensional sequence specificity in the arrangement of the amino acids that form proteins. Even slight alterations in sequence often result in the loss of protein function.

C. The Complexity and Sequence Specificity of DNA

During the early part of the twentieth century, researchers also vastly underestimated the complexity (and significance) of nucleic acids such as DNA and RNA. By then, scientists knew the chemical composition of DNA. Biologists and chemists knew that in addition to sugars (and later phosphates), DNA was composed of four different nucleotide bases, called adenine, thymine, cytosine, and guanine. In 1909, chemist P. A. Levene thought he had shown that the four different nucleotide bases always occurred in equal quantities within the DNA molecule.[11] He conjectured that the four nucleotide bases in DNA linked together in repeating sequences of the same four chemicals in the same sequential order. Yet if those sequential arrangements of nucleotides were repetitive and invariant, their potential for expressing any genetic diversity seemed inherently limited. To account for the heritable differences between species, biologists needed to discover some source of variable or irregular specificity, some source of information, within the germ lines of different organisms. Yet insofar as DNA was seen as an uninterestingly repetitive molecule, many biologists assumed that DNA could play little if any role in the transmission of heredity.

That view began to change in the mid-1940s for several reasons. Crucially, work by Erwin Chargaff of Columbia University in the late 1940s undermined Levene's "tetranucleotide hypothesis." Chargaff showed that nucleotide frequencies actually do differ between species, even if they often hold constant within the same species or within the same organs or tissues of a single organism.[12] More important, Char-

gaff recognized that even for nucleic acids of exactly "the same analytical composition"—meaning those with the same relative proportions of the four bases (abbreviated A, T, C, and G)—"enormous" numbers of variations in sequence were possible.[13] Thus, Chargaff showed that base sequencing in DNA might well display the high degree of variability and aperiodicity required by any potential carrier of heredity.

Eventually, elucidation of the three-dimensional structure of DNA by Watson and Crick in 1953 made clear that DNA could function as a carrier of hereditary information.[14] The model proposed by Watson and Crick envisioned a double-helix structure. According to the now well-known Watson and Crick model, the two strands of the helix were made of sugar and phosphate molecules linked by phosphodiester bonds. Nucleotide bases were linked horizontally to the sugars on each strand of the helix and to a complementary base on the other strand to form an internal "rung" on a twisting "ladder."

The Watson-Crick model made clear that DNA might possess an impressive chemical and structural complexity. The double-helix structure for DNA presupposed an extremely long and high-molecular-weight structure, possessing an impressive potential for variability and complexity in sequence. As Watson and Crick explained, "The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information."[15]

The notion of a "code" was important. Discovery of the complexity and specificity of proteins had led researchers to suspect a functionally specific role for DNA. Molecular biologists assumed that proteins were much too complex to arise by chance *in vivo*. Moreover, given their irregularity, it seemed unlikely that a general chemical law or regularity could explain their assembly. Instead, molecular biologists had begun to look for some source of information or "specificity" within the cell that could direct the construction of such highly specific and complex

structures. To explain the presence of the specificity and complexity in the protein, Monod would later insist, "you absolutely needed a code."[16]

The structure of DNA as elucidated by Watson and Crick suggested a means by which information or "specificity" might be encoded along the spine of DNA's sugar-phosphate backbone.[17] Their model suggested that variations in sequence of the nucleotide bases might find expression in the sequence of the amino acids that form proteins. In 1955, Crick proposed this idea as the so-called sequence hypothesis. According to Crick's hypothesis, the specificity of arrangement of amino acids in proteins derives from the specific arrangement of the nucleotide bases on the DNA molecule.[18] The sequence hypothesis suggested that the nucleotide bases in DNA functioned like letters in an alphabet or characters in a machine code. Just as alphabetic letters in a written language may perform a communication function depending on their sequence, so, too, might the nucleotide bases in DNA result in the production of a functional protein molecule depending on their precise sequential arrangement. In both cases, function depends crucially on sequence. The sequence hypothesis implied not only the complexity but also the functional specificity of DNA base sequences.

By the early 1960s, a series of experiments had confirmed that DNA base sequences play a critical role in determining amino acid sequence during protein synthesis.[19] By that time, the processes and mechanisms by which DNA sequences determine key stages of the process were known (at least in outline). Protein synthesis or "gene expression" proceeds as long chains of nucleotide bases are first copied during a process known as transcription. The resulting copy, a "transcript" made of single-stranded "messenger RNA," now contains a sequence of RNA bases precisely reflecting the sequence of bases on the original DNA strand. The transcript is then transported to a complex organelle called a ribosome. At the ribosome, the transcript is "translated" with the aid of highly specific adaptor molecules (called transfer-RNAs) and specific enzymes (called amino-acyl tRNA synthetases) to produce a growing amino acid chain.[20]

Whereas the function of the protein molecule derives from the specific arrangement of twenty different types of amino acids, the function of DNA depends on the arrangement of just four kinds of bases. This lack of a one-to-one correspondence means that a group of three DNA nucleotides (a triplet) is needed to specify a single amino acid. In any case, the sequential arrangement of the nucleotide bases determines (in large part) the one-dimensional sequential arrangement of amino acids during protein synthesis.[21] Since protein function depends critically on amino acid sequence and amino acid sequence depends critically on DNA base sequence, the sequences in the coding regions of DNA themselves possess a high degree of specificity relative to the requirements of protein (and cellular) function.

D. Information Theory and Molecular Biology

From the beginning of the molecular biological revolution, biologists have ascribed information-bearing properties to DNA, RNA, and proteins. In the parlance of molecular biology, DNA base sequences contain the "genetic information" or the "assembly instructions" necessary to direct protein synthesis. Yet the term "information" can denote several theoretically distinct concepts. Thus, one must ask which sense of "information" applies to these large biomacromolecules. In fact, molecular biologists employ a concept of information stronger than that of mathematicians and information theorists, but slightly weaker conception than that of linguists and ordinary users.

During the 1940s, Claude Shannon at Bell Laboratories developed a mathematical theory of information.[22] His theory equated the amount of information transmitted with the amount of uncertainty reduced or eliminated by a series of symbols or characters.[23] For example, before one rolls a six-sided die, there are six possible outcomes. Before one flips a coin, there are two. Rolling a die will thus eliminate more uncertainty and, on Shannon's theory, will convey more information than flipping a coin. Equating information with the reduction of uncertainty implied a mathematical relationship between information and probability (or its

inverse, complexity). Note that for a die each possible outcome has only a one in six chance of occurring, compared to a one in two chance for each side of the coin. Thus, in Shannon's theory the occurrence of the more improbable event conveys more information. Shannon generalized this relationship by stating that the amount of information conveyed by an event is inversely proportional to the prior probability of its occurrence. The greater the number of possibilities, the greater the improbability of any one being actualized, and thus more information is transmitted when a particular possibility occurs.

Moreover, information increases as improbabilities multiply. The probability of getting four heads in a row when flipping a fair coin is ½ × ½ × ½ × ½, or $(½)^4$. Thus, the probability of obtaining a specific sequence of heads and/or tails decreases exponentially as the number of trials increases. The quantity of information increases correspondingly. Even so, information theorists found it convenient to measure information additively rather than multiplicatively. Thus, the common mathematical expression ($I = -\log_2 p$) for calculating information converts probability values into informational measures through a negative logarithmic function, where the negative sign expresses an inverse relationship between information and probability.[24]

Shannon's theory applies most easily to sequences of alphabetic symbols or characters that function as such. Within any given alphabet of x possible characters, the placement of a specific character eliminates x-1 other possibilities and thus a corresponding amount of uncertainty. Or put differently, within any given alphabet or ensemble of x possible characters (where each character has an equi-probable chance of occurring), the probability of any one character occurring is $1/x$. The larger the value of x, the greater the amount of information that is conveyed by the occurrence of a specific character in a sequence. In systems where the value of x can be known (or estimated), as in a code or language, mathematicians can easily generate quantitative estimates of information-carrying capacity. The greater the number of possible characters at each site and the longer the sequence of characters, the greater is the

information-carrying capacity—or Shannon information—associated with the sequence.

The essentially digital character of the nucleotide bases in DNA and of the amino acid residues in proteins enabled molecular biologists to calculate the information-carrying capacity (or syntactic information) of those molecules using the new formalism of Shannon's theory. Because at every site in a growing amino acid chain, for example, the chain may receive any one of twenty amino acids, placement of a single amino acid in the chain eliminates a quantifiable amount of uncertainty and increases the Shannon or syntactic information of a polypeptide by a corresponding amount. Similarly, since at any given site along the DNA backbone any one of four nucleotide bases may occur with equal probability, the p value for the occurrence of a specific nucleotide at that site equals 1/4, or .25.[25] The information-carrying capacity of a sequence of a specific length n can then be calculated using Shannon's familiar expression ($I = -\log_2 p$) once one computes a p value for the occurrence of a particular sequence n nucleotides long where $p = (1/4)^n$. The p value thus yields a corresponding measure of information-carrying capacity or syntactic information for a sequence of n nucleotide bases.[26]

E. Complexity, Specificity, and Biological Information

Though Shannon's theory and equations provided a powerful way to measure the amount of information that could be transmitted across a communication channel, it had important limits. In particular, it did not and could not distinguish merely improbable sequences of symbols from those that conveyed a message. As Warren Weaver made clear in 1949, "The word information in this theory is used in a special mathematical sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning."[27] Information theory could measure the information-carrying capacity or the syntactic information of a given sequence of symbols but could not distinguish the presence of a meaningful or functional arrangement of symbols from a random sequence (for example, "we hold these truths to be self-evident"

versus "ntnyhiznlhteqkhgdsjh"). Thus, Shannon information theory could quantify the amount of functional or meaningful information that *might be present* in a given sequence of symbols or characters, but it could not distinguish the status of a functional or message-bearing text from gibberish. Thus, paradoxically, random sequences of letters often have more syntactic information (or information-carrying capacity), as measured by classical information theory, than do meaningful or functional sequences that happen to contain a certain amount of intentional redundancy or repetition. Thus, Shannon's theory remains silent on the important question of whether a sequence of symbols is functionally specific or meaningful.

In its application to molecular biology, Shannon information theory did succeed in rendering rough quantitative measures of the information-carrying capacity or syntactic information (where those terms correspond to measures of brute complexity),[28] establishing that DNA and proteins were highly complex, and quantifiably so; yet it could not establish whether base sequences in DNA or amino acid sequences in proteins possessed the property of functional specificity. Information theory helped establish that DNA and proteins *could* carry large amounts of functional information; it could not establish whether they *did*.

The ease with which information theory applied to molecular biology (to measure information-carrying capacity) has created considerable confusion about the sense in which DNA and proteins contain "information." Since as early as 1958, leading molecular biologists have defined biological information so as to incorporate the notion of specificity of function (as well as complexity).[29] Molecular biologists such as Monod and Crick recognized that sequences of nucleotides and amino acids in functioning biomacromolecules possessed a high degree of specificity relative to the maintenance of cellular function. As Crick explained in 1958, "By information I mean the specification of the amino acid sequence of the protein… Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein."[30] Crick's "precise determination of sequence" is

now equated with the extra-information-theoretic property of specificity or specification. Biologists have defined specificity tacitly as "necessary to achieve or maintain function." They have determined that DNA base sequences, for example, are specified not by applying information theory but by making experimental assessments of the function of those sequences within the overall apparatus of gene expression.[31] Similar experimental considerations have established the functional specificity of proteins.

Further, developments in complexity theory have now made possible a fully general theoretical account of specification, one that applies readily to biological systems. According to mathematician William Dembski, specification involves a match or correspondence between a physical system or sequence and an independently recognizable pattern or set of functional requirements.[32]

To illustrate Dembski's notion of specification, consider these two strings of characters:

"iuinsdysk]idfawqnzkl,mfdifhs"

"Time and tide wait for no man."

Given the number of possible ways of arranging the letters and punctuation marks of the English language for sequences of this length, both of these two sequences constitute highly improbable arrangements of characters. Thus, both have a considerable and quantifiable information-carrying capacity. Nevertheless, only the second of the two sequences exhibits a specification on Dembski's account.

The reason for this is that English has many functional requirements. For example, to convey meaning in English one must employ existing conventions of vocabulary (associations of symbol sequences with particular objects, concepts, or ideas) and existing conventions of syntax and grammar. When symbol arrangements "match" existing vocabulary and grammatical conventions (i.e., functional requirements), communication can occur. Such arrangements exhibit "specification." The sequence "Time and tide wait for no man" clearly exhibits such a match, and thus performs a communication function.

Biological organisms also exhibit specifications, though not necessarily semantic or subjectively "meaningful" ones. The nucleotide base sequences in the coding regions of DNA are highly specific relative to the independent functional requirements of protein function, protein synthesis, and cellular life. To maintain viability, the cell must regulate its metabolism, pass materials back and forth across its membranes, destroy waste materials, and do many other specific tasks. Each of these functional requirements in turn necessitates specific molecular constituents, machines, or systems (usually made of proteins) to accomplish these tasks. Building these proteins with their specific three-dimensional shapes requires specific arrangements of nucleotide bases on the DNA molecule.

Since the chemical properties of DNA allow a vast ensemble of combinatorially possible arrangements of nucleotide bases, any particular sequence will necessarily be highly improbable and rich in Shannon information or information-carrying capacity. Yet within that set of possible sequences a very few will, given the multimolecular system of gene expression within the cell, produce functional proteins.[33] Those that do are thus not only improbable but also functionally "specified" or "specific," as molecular biologists use the terms. Thus, the nucleotide sequences in the coding regions of DNA possess both syntactic information and "specified" information.

A note of definitional clarity must be offered about the relationship between "specified" information and "semantic" information. Though natural languages and DNA base sequences are both specified, only natural language conveys meaning. If one defines "semantic information" as "subjectively meaningful information that is conveyed syntactically (as a string of phonemes or characters) and is understood by a conscious agent," then clearly the information in DNA does not qualify as semantic. Rather, the coding regions of DNA function in much the same way as a software program or machine code, directing operations within a complex material system via highly complex yet specified sequences of characters. As Richard Dawkins has noted, "The machine code of the

genes is uncannily computer-like."[34] Or as software developer Bill Gates has noted, "DNA is like a computer program, but far, far more advanced than any software we've ever created."[35] Just as the specific arrangement of two symbols (0 and 1) in a software program can perform a function within a machine environment, so, too, can the precise sequencing of the four nucleotide bases in DNA perform a function within the cell.

Since the late 1950s, the concept of information as employed by molecular biologists has conjoined the notions of complexity (or improbability) and specificity of function. The crucial biomolecular constituents of living organisms possess not only Shannon or syntactic information but also "specified information" or "specified complexity."[36] Biological information so defined, therefore, constitutes a salient feature of living systems that any origin-of-life scenario must explain "the origin of." Further, as we will see below, all naturalistic chemical evolutionary theories have encountered difficulty explaining the origin of such functionally "specified" biological information.

F. Information as Metaphor: Nothing to Explain?

Though most molecular biologists would see nothing controversial in characterizing DNA and proteins as "information-bearing" molecules, some historians and philosophers of biology have challenged that description. Before evaluating competing types of explanation for the origin of biological information, this challenge must be addressed. In 2000, the late historian of science Lily Kay characterized the application of information theory to biology as a failure, in particular because classical information theory could not capture the idea of meaning. She suggests, therefore, that the term information as used in biology constitutes nothing more than a metaphor. Since, in Kay's view, the term does not designate anything real, it follows that the origin of "biological information" does not require explanation. Instead, only the origin of the use of the term "information" within biology requires explanation. As a social constructivist, Kay explained this usage as the result of various social forces operating within the "Cold War Technoculture."[37] In a different but re-

lated vein, Sarkar has argued that the concept of information has little theoretical significance in biology because it lacks predictive or explanatory power.[38] He, like Kay, seems to regard the concept of information as a superfluous metaphor lacking empirical reference and ontological status.

Of course, insofar as the term "information" connotes semantic meaning, it does function as a metaphor within biology. That does not mean, however, that the term functions only metaphorically or that origin-of-life biologists have nothing to explain. Though information theory has a limited application in describing biological systems, it has succeeded in rendering quantitative assessments of the complexity of biomacromolecules. Further, experimental work established the functional specificity of the sequences of monomers in DNA and proteins. Thus, the term "information" as used in biology does refer to two real and contingent properties of living systems: complexity and specificity. Indeed, since scientists began to think seriously about what would be required to explain the phenomenon of heredity, they have recognized the need for some feature or substance in living organisms possessing precisely these two properties together. Thus, Schrödinger envisioned an "aperiodic crystal"; Chargaff perceived DNA's capacity for "complex sequencing"; Watson and Crick equated complex sequences with "information," which Crick in turn equated with "specificity"; Monod equated irregular specificity in proteins with the need for "a code"; and Orgel characterized life as a "specified complexity."[39] Further, Davies has recently argued that the "specific randomness" of DNA base sequences constitutes the central mystery surrounding the origin of life.[40] Whatever the terminology, scientists have recognized the need for, and now know the location of, a source of complex specificity in the cell to transmit heredity and maintain biological function. The incorrigibility of these descriptive concepts suggests that complexity and specificity constitute real properties of biomacromolecules—indeed, properties that could be otherwise, but only to the detriment of cellular life. As Orgel notes: "Living organisms are distinguished by their specified complexity. Crystals… fail to qualify as

living because they lack complexity; mixtures of random polymers fail to qualify because they lack specificity."[41]

The origin of specificity and complexity (in combination), to which the term "information" in biology commonly refers, therefore does require explanation, even if the concept of information connotes only complexity in classical information theory and even if it has no explanatory or predictive value in itself. Instead, as a descriptive (rather than as an explanatory or predictive) concept, the term "information" helps to define (either in conjunction with the notion of "specificity" or by subsuming it) the effect that origin-of-life researchers must explain "the origin of." Thus, only where "information" connotes subjective meaning does it function as a metaphor in biology. Where it refers to an analog of meaning, namely, *functional* specificity and complexity, it defines an essential feature of living systems.

II.

A. Naturalistic Explanations for the Origin of Specified Biological Information

The discoveries of molecular biologists during the 1950s and 1960s raised the question of the ultimate origin of the specified complexity or specified information in both DNA and proteins. Since at least the mid-1960s, many scientists have regarded the origin of information (so defined) as the central question facing origin-of-life biology.[42] Accordingly, origin-of-life researchers have proposed three broad types of naturalistic explanation to explain the origin of specified genetic information: those emphasizing chance, necessity, or the combination of the two.

B. Beyond the Reach of Chance

Perhaps the most common popular naturalistic view about the origin of life is that it happened exclusively by chance. A few serious scientists have also voiced support for this view, at least, at various points in their careers. In 1954, biochemist George Wald, for example, argued for the causal efficacy of chance in conjunction with vast expanses of time. As

he explained, "Time is in fact the hero of the plot… Given so much time, the impossible becomes possible, the possible probable, and the probable virtually certain."[43] Later, in 1968, Francis Crick would suggest that the origin of the genetic code—that is, the translation system—might be a "frozen accident."[44] Other theories have invoked chance as an explanation for the origin of genetic information, though often in conjunction with prebiotic natural selection (see part C below).

Almost all serious origin-of-life researchers now consider "chance" an inadequate causal explanation for the origin of biological information.[45] Since molecular biologists began to appreciate the sequence specificity of proteins and nucleic acids in the 1950s and 1960s, many calculations have been made to determine the probability of formulating functional proteins and nucleic acids at random. Various methods of calculating probabilities have been offered by Morowitz, Hoyle and Wickramasinghe, Cairns-Smith, Prigogine, Yockey, and, more recently, Robert Sauer.[46] For the sake of argument, these calculations have often assumed extremely favorable prebiotic conditions (whether realistic or not), much more time than was actually available on the early earth, and theoretically maximal reaction rates among constituent monomers (that is, the constituent parts of proteins, DNA, or RNA). Such calculations have invariably shown that the probability of obtaining functionally sequenced biomacromolecules at random is, in Prigogine's words, "vanishingly small… even on the scale of… billions of years."[47] As Cairns-Smith wrote in 1971: "Blind chance… is very limited. Low levels of cooperation [it] can produce exceedingly easily (the equivalent of letters and small words), but [it] becomes very quickly incompetent as the amount of organization increases. Very soon indeed long waiting periods and massive material resources become irrelevant."[48]

Functioning proteins require amino acids that link up in functionally specified sequential arrangements, like the arrangements required in meaningful sentences. In some cases, changing even one amino acid at a given site results in the loss of protein function. Moreover, because there are twenty biologically occurring amino acids, the probability of

getting a specific amino acid at a given site is small—1/20. (Actually the probability is even lower because, in nature, there are also many non-protein-forming amino acids.) On the assumption that each site in a protein chain requires a particular amino acid, the probability of attaining a particular protein 150 amino acids long would be $(1/20)^{150}$ or roughly 1 chance in 10^{195}.

Molecular biologists have known for a while that most sites along the chain can tolerate several of the different twenty amino acids commonly found in proteins without destroying the function of the protein, though some cannot. This raised an important question: How rare, or common, are the *functional* sequences of amino acids among all the possible sequences of amino acids in a chain of any given length? In the late 1980s, several important studies were conducted in the laboratory of MIT biochemist, Robert Sauer, in order to investigate this question. His research team used a sampling technique known as "cassette mutagenesis" to determine how much variance among amino acids can be tolerated at any given site in several proteins. So what did they find? Their most clear-cut experiments[49] seemed to indicate that, even taking the possibility of variance into account, the probability of achieving a functional sequence of amino acids in several known (roughly 100-residue) proteins at random is still "exceedingly small," about 1 chance in 10^{63} (to put this in perspective, there are 10^{65} atoms in our galaxy).[50] Using a variety of mutagenesis techniques, they and other scientists showed that proteins (and thus the genes that produce them) are highly specified relative to biological function.[51] Earlier studies had shown that amino acid residues at many sites cannot vary without functional loss.[52] Now Sauer and others had shown that even for sites that do admit some variance, not just any amino acid will do. Instead, they showed that functional requirements place significant constraints on sequencing at sites where some variance is allowed. By quantifying that allowable variance, they made it possible to calculate the probability of finding a protein with a functional sequence among the larger ensemble of combinatorial possibilities.

Further work in this area has been done by Douglas Axe. He asked a question similar to that which had motivated Sauer: "How rare, or common, are the amino acid sequences that produce the stable folds that make it possible for proteins to perform their biological functions?" The results of his work were published in a series of papers between 1996 and 2004.

The results of a 2004 paper were particularly telling.[53] Axe performed a mutagenesis experiment, using his refined method, on a functionally significant 150-amino-acid section of a protein called beta-lactamase, an enzyme that confers antibiotic resistance upon bacteria. On the basis of his experiments, Axe was able to make a careful estimate of the ratio of (a) the number of 150-amino-acid sequences that could perform that function to (b) the whole set of possible amino acid sequences of this length. Based on his experiments, Axe estimated this ratio to be $1/10^{77}$.

This was a staggering number, and it suggested that a random process would have great difficulty generating a protein with that particular function by chance. But origin-of-life researchers didn't just want to know the likelihood of finding a protein with a particular function within a space of combinatorial possibilities. They wanted to know the odds of finding *any* functional protein whatsoever within such a space. That number would make it possible to evaluate chance-based origin-of-life scenarios, by assessing the probability that a single protein—*any working protein*—would have arisen by chance on the early Earth.

Fortunately, Axe's work provided this number as well. Axe knew that in nature proteins perform many specific functions. He also knew that in order to perform these functions their amino acid chains must first fold into stable three-dimensional structures. Thus, before he estimated the frequency of sequences performing a specific (beta-lactamase) function, he first performed experiments that enabled him to estimate the frequency of sequences that will produce stable folds. On the basis of his experimental results, he calculated the ratio of (a) the number of 150-amino-acid sequences capable of folding into stable "function-ready" structures to (b) the whole set of possible amino acid sequences of

that length. He determined that ratio to be 1 in 10^{74}. Axe's ratio of 1 in 10^{74} implied that a random process producing amino acid chains of this length would stumble onto a functional protein only about once in every 10^{74} attempts.

Axe's improved estimate of how rare functional proteins are within "sequence space" has now made it possible to calculate the probability that a 150-amino-acid compound assembled by random interactions in a prebiotic soup would be a functional protein. This calculation can be made by multiplying three independent probabilities by one another: the probability of incorporating only peptide bonds (1 in 10^{45}), the probability of incorporating only left-handed amino acids (1 in 10^{45}) and the probability of achieving correct amino acid sequencing (using Axe's 1 in 10^{74} estimate). Making that calculation (multiplying the separate probabilities by adding their exponents: $10^{45+45+74}$) gives a dramatic answer. The odds of getting a functional protein of modest length (150 amino acids) by drawing a compound of that size from a prebiotic soup is no better than 1 chance in 10^{164}. In other words, the probability of constructing a rather short functional protein at random becomes so small (no more than 1 chance in 10^{164}) as to appear absurd on the chance hypothesis.

Yet the probabilities, as small as they are, are not by themselves conclusive. One also has to consider the number of opportunities that the event in question might have had to occur. That is, one has to take into account what William Dembski calls the *probabilistic resources*.

But what were those resources—how many opportunities did the necessary proteins or genes have to arise by chance? The advocates of the chance hypothesis envisioned amino acids, or nucleotide bases, phosphates and sugars, knocking into each other in an ocean-sized soup until the correct arrangements of these building blocks arose by chance somewhere. Surely, they think, such an environment would have generated many opportunities for the assembly of functional proteins and DNA molecules. But how many? And were there enough such opportunities to render these otherwise exceedingly improbable events probable?

In order to establish an upper bound on the probabilistic resources that might be available to produce functional proteins and DNA by chance,[54] Dembski calculated the maximum number of events that could actually have taken place during the history of the observable universe.[55] His calculation was elegantly simple and yet made a powerful point.

He noted that there were about 10^{80} elementary particles[56] in the observable universe.[57] He also noted that there had been roughly 10^{16} seconds since the Big Bang. He then introduced another parameter: the shortest time in which any physical event can occur. This unit of time is the Planck time of 10^{-43} seconds. Since elementary particles can only interact with each other so many times per second (at most 10^{43} times), and since there are a limited number (10^{80}) of elementary particles, and since there has been a limited amount of time since the Big Bang (10^{16} seconds), Dembski was able to calculate the total number of events that could have taken place in the observable universe since the origin of the universe. He obtained this number by simply multiplying the three relevant factors together: the number of elementary particles (10^{80}) times the number of seconds since the Big Bang (10^{16}) times the number of possible interactions per second (10^{43}). The product, i.e., 10^{139}, provided a measure of the probabilistic resources of the entire observable universe.[58] Other mathematicians and scientists have made similar calculations.[59]

Recall Axe's calculation that the probability of producing a single 150-amino acid functional protein by chance stands at about 1 in 10^{164}. Thus, for each functional sequence of 150 amino acids, there are 10^{164} other non-functional sequences of the same length. Therefore, to have a good (i.e., better than 50/50) chance of producing a single functional protein of this length by chance, a random process would have to generate (or sample) more than half of the 10^{164} non-functional sequences corresponding to each functional sequence of that length. Unfortunately, as we see from Dembski's calculation, that number vastly exceeds the most optimistic estimate of the probabilistic resources of the universe, i.e., 10^{139}.

It seems, then, that what Mora said in 1963 still holds: "Statistical considerations, probability, complexity, etc., followed to their logical implications suggest that the origin and continuance of life is not controlled by such principles. An admission of this is the use of a period of practically infinite time to obtain the derived result. Using such logic, however, we can prove anything."[60]

C. Prebiotic Natural Selection: A Contradiction in Terms

Of course, even many early theories of chemical evolution did not rely exclusively on chance as a causal mechanism. For example, Oparin's original theory of evolutionary abiogenesis, first published in the 1920s and 1930s, invoked prebiotic natural selection as a complement to chance interactions. Oparin's theory envisioned a series of chemical reactions that he thought would enable a complex cell to assemble itself gradually and naturalistically from simple chemical precursors.

Developments in molecular biology during the 1950s cast doubt on Oparin's scenario. Oparin originally invoked natural selection to explain how cells refined primitive metabolism once it had arisen. His scenario relied heavily on chance to explain the initial formation of the constituent biomacromolecules on which even primitive cellular metabolism would depend. Discovery during the 1950s of the extreme complexity and specificity of such molecules undermined the plausibility of his claim. For that and other reasons, Oparin published a revised version of his theory in 1968 that envisioned a role for natural selection earlier in the process of abiogenesis. His new theory claimed that natural selection acted on random polymers as they formed and changed within his coacervate protocells.[61] As more complex and efficient molecules accumulated, they would have survived and reproduced more prolifically.

Even so, Oparin's concept of prebiotic natural selection acting on initially unspecified biomacromolecules remained problematic. For one thing, it seemed to presuppose a preexisting mechanism of self-replication. Yet self-replication in all extant cells depends on functional and,

therefore, (to a high degree) sequence-specific proteins and nucleic acids. Yet the origin of specificity in these molecules is precisely what Oparin needed to explain. As Christian de Duve has stated, theories of prebiotic natural selection "need information which implies they have to presuppose what is to be explained in the first place."[62] Oparin attempted to circumvent the problem by claiming that the first polymers need not have been highly sequence-specific. But that claim raised doubts about whether an accurate mechanism of self-replication (and thus natural selection) could have functioned at all.

Thus, the need to explain the origin of specified information created an intractable dilemma for Oparin. On the one hand, if he invoked natural selection late in his scenario, he would need to rely on chance alone to produce the highly complex and specified biomolecules necessary to self-replication. On the other hand, if Oparin invoked natural selection earlier in the process of chemical evolution, before functional specificity in biomacromolecules would have arisen, he could give no account of how such prebiotic natural selection could even function. Thus, Dobzhansky would insist that "prebiological natural selection is a contradiction in terms."[63]

Nevertheless, during the 1980s, Richard Dawkins and Bernd-Olaf Kuppers attempted to resuscitate prebiotic natural selection as an explanation for the origin of biological information.[64] Both accepted the futility of naked appeals to chance and invoke what Kuppers calls a "Darwinian optimization principle." Both used computers to demonstrate the efficacy of prebiotic natural selection. In these computer simulations, a target sequence is selected, to represent a desired functional polymer. After creating a crop of randomly constructed sequences and generating variations among them at random, the computers select those sequences that match the target sequence most closely. The computers then amplify the production of those sequences, eliminate the others (to simulate differential reproduction), and repeat the process. As Kuppers puts it, "Every mutant sequence that agrees one bit better with the meaningful or reference sequence… will be allowed to reproduce more rapidly."[65] In

his case, after a mere thirty-five generations, his computer succeeded in spelling his target sequence, "NATURAL SELECTION."

Despite superficially impressive results, such "simulations" conceal an obvious flaw: Molecules *in situ* do not have a target sequence "in mind." Nor will they confer any selective advantage on a cell, and thus differentially reproduce, until they combine in a functionally advantageous arrangement. Thus, nothing in nature corresponds to the role that the computer plays in selecting functionally non-advantageous sequences that happen to agree "one bit better" than others with a target sequence. The sequence NORMAL ELECTION may agree more with NATURAL SELECTION than does the sequence MISTRESS DEFECTION, but neither of the two yields any advantage over the other in trying to communicate something about NATURAL SELECTION. If that is the goal, both are equally ineffectual. Even more to the point, a completely nonfunctional polypeptide would confer no selective advantage on a hypothetical protocell, even if its sequence happened to agree "one bit better" with an unrealized target protein than some other nonfunctional polypeptide.

Both Kuppers's and Dawkins's published results of their simulations show the early generations of variant phrases awash in nonfunctional gibberish.[66] In Dawkins's simulation, not a single functional English word appears until after the tenth iteration (unlike the more generous example above that starts with actual, albeit incorrect, words). To make distinctions on the basis of function among sequences that have no function is entirely unrealistic. Such determinations can be made only if considerations of *proximity to possible future function* are allowed, but that requires foresight, which natural selection does not have. A computer, programmed by a human being, can perform such functions. To imply that molecules can do so as well illicitly personifies nature. Thus, if these computer simulations demonstrate anything, they subtly demonstrate the need for intelligent agents to elect some options and exclude others; that is, to create information. In *Signature in the Cell*, I show that other,

more recent genetic algorithms such as *Ev* and *Avida* demonstrate this same need.[67]

D. Self-Organizational Scenarios

Because of the difficulties with chance-based theories, including those relying on prebiotic natural selection, most origin-of-life theorists after the mid-1960s attempted to address the problem of the origin of biological information in a completely different way. Researchers began to look for self-organizational laws and properties of chemical attraction that might explain the origin of the specified information in DNA and proteins. Rather than invoking chance, such theories invoked necessity. Given a limited number of broad explanatory categories, the inadequacy of chance (with or without prebiotic natural selection) has, in the minds of many researchers, left only one option. Christian de Duve articulates the logic: "A string of improbable events—drawing the same lottery number twice, or the same bridge hand twice in a row—does not happen naturally. All of which lead me to conclude that life is an obligatory manifestation of matter, bound to arise where conditions are appropriate."[68]

When origin-of-life biologists began considering the self-organizational perspective that de Duve describes, several researchers proposed that deterministic forces (stereochemical "necessity") made the origin of life not just probable but inevitable. Some suggested that simple chemicals possessed "self-ordering properties" capable of organizing the constituent parts of proteins, DNA, and RNA into the specific arrangements they now possess.[69] Steinman and Cole, for example, suggested that differential bonding affinities or forces of chemical attraction between certain amino acids might account for the origin of the sequence specificity of proteins.[70] Just as electrostatic forces draw sodium (Na^+) and chloride (Cl^-) ions together into highly ordered patterns within a crystal of salt (NaCl), so, too, might amino acids with special affinities for each other arrange themselves to form proteins. A discussion of other self-organization scenarios can be found in my book *Signature in the Cell*.[71]

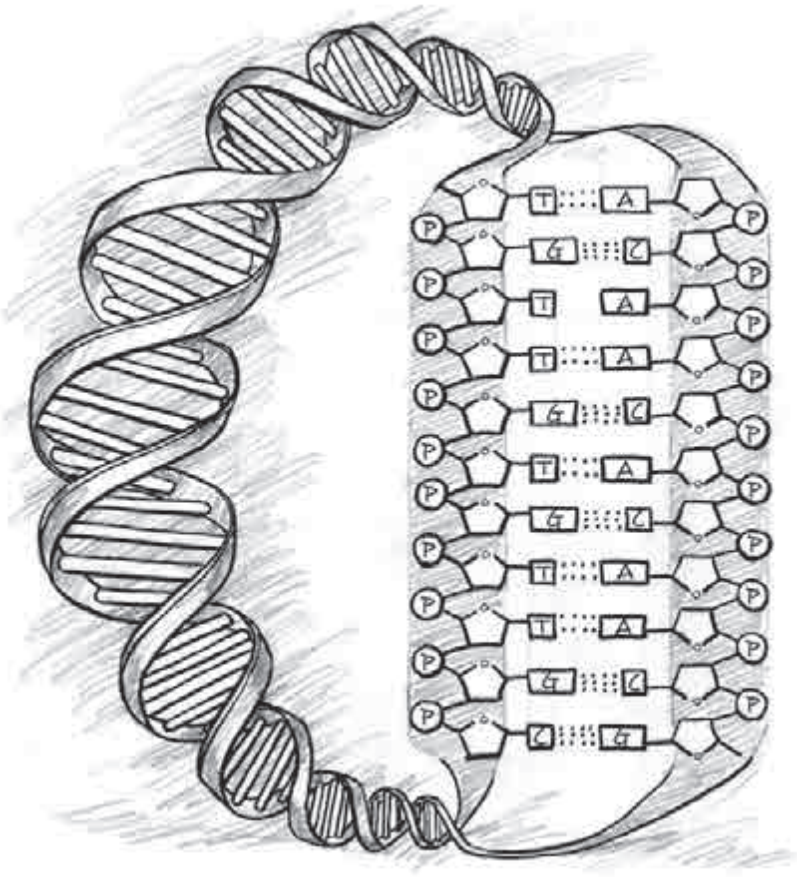

Figure 17-1. The bonding relationship between the chemical constituents of the DNA molecule.
Sugars (designated by the pentagons) and phosphates (designated by the circled Ps) are linked chemically. Nucleotide bases (A's, T's, G's and C's) are bonded to the sugar-phosphate backbones. Nucleotide bases are linked by hydrogen bonds (designated by dotted double or triple lines) across the double helix. But no chemical bonds exist between the nucleotide bases along the message-bearing spine of the helix.

For many current origin-of-life scientists, self-organizational models now seem to offer the most promising approach to explaining the origin of specified biological information. Nevertheless, critics have called into question both the plausibility and the relevance of self-organizational models. Ironically, a prominent early advocate of self-organization, Dean Kenyon, later explicitly repudiated such theories as both incompatible with empirical findings and theoretically incoherent.[72] Kenyon voiced his doubts in his Foreword to *The Mystery of Life's Origin*, reprinted earlier in this volume.

It is true that empirical studies have shown that some differential affinities do exist between various amino acids; that is, certain amino acids do form linkages more readily with some amino acids than with others.[73] Nevertheless, such differences do not correlate to actual sequences in large classes of known proteins.[74] In short, differing chemical affinities do not explain the multiplicity of amino acid sequences existing in naturally occurring proteins or the sequential arrangement of amino acids in any particular protein.

In the case of DNA, this point can be made more dramatically. Figure 17-1 shows that the structure of DNA depends on several chemical bonds. There are bonds, for example, between the sugar and the phosphate molecules forming the two twisting backbones of the DNA molecule. There are bonds fixing individual (nucleotide) bases to the sugar-phosphate backbones on each side of the molecule. There are also hydrogen bonds stretching horizontally across the molecule between nucleotide bases, making so-called complementary pairs. The individually weak hydrogen bonds, which in concert hold two complementary copies of the DNA message text together, make replication of the genetic instructions possible. It is important to note, however, that there are no chemical bonds between the bases along the longitudinal axis in the center of the helix. Yet it is precisely along this axis of the DNA molecule that the genetic information is stored.

Just as magnetic letters can be combined and recombined in any way to form various sequences on a metal surface, so, too, can each of the four

bases (A, T, G, and C) attach to any site on the DNA backbone with equal facility, making all sequences equally probable (or improbable). Indeed, there are no significant differential affinities between any of the four bases and the binding sites along the sugar-phosphate backbone. The same type of N-glycosidic bond occurs between the base and the backbone regardless of which base attaches. All four bases are acceptable; none is chemically favored. As Kuppers has noted, "The properties of nucleic acids indicate that all the combinatorially possible nucleotide patterns of a DNA are, from a chemical point of view, equivalent."[75]

Thus, "self-organizing" bonding affinities cannot explain the sequentially specific arrangement of nucleotide bases in DNA because (1) there are no bonds between bases along the information-bearing axis of the molecule, and (2) there are no differential affinities between the backbone and the specific bases that could account for variations in sequence. And because the same holds for RNA molecules, researchers who speculate that life began in an RNA world have also failed to solve the sequence specificity problem—that is, the problem of explaining how information in functioning RNA molecules could have arisen in the first place.

For those who want to explain the origin of life as the result of self-organizing properties intrinsic to the material constituents of living systems, these rather elementary facts of molecular biology have decisive implications. The most obvious place to look for self-organizing properties to explain the origin of genetic information is in the constituent parts of the molecules that carry that information. But biochemistry and molecular biology make clear that forces of attraction between the constituents in DNA, RNA, and proteins do not explain the sequence specificity of these large, information-bearing biomolecules.

The properties of the monomers constituting nucleic acids and proteins simply do not make a particular gene, let alone life as we know it, inevitable. Imagine a pool of all four DNA bases and all necessary sugars and phosphates; would any particular genetic sequence inevitably arise? Given all necessary monomers, would any particular functional

protein or gene, let alone a specific genetic code, replication system, or signal transduction circuitry, inevitably arise? Clearly not. Yet de Duve has claimed that "the processes that generated life" were "highly deterministic," making life as we know it "inevitable" given "the conditions that existed on the prebiotic earth."[76]

In the parlance of origin-of-life research, monomers are "building blocks," and building blocks can be arranged and rearranged in innumerable ways. The properties of stone blocks do not determine their own arrangement in the construction of buildings. Similarly, the properties of biological building blocks do not determine the arrangement of functional polymers. Instead, the chemical properties of the monomers allow a vast ensemble of possible configurations, the overwhelming majority of which have no biological function whatsoever. Functional genes or proteins are no more inevitable, given the properties of their "building blocks," than, for example, the Palace of Versailles was inevitable, given the properties of the stone blocks that were used to construct it.

Significantly, information theory makes clear that there is a good reason for this. If chemical affinities between the constituents in the DNA determined the arrangement of the bases, such affinities would dramatically diminish the capacity of DNA to carry information. Recall that classical information theory equates the reduction of uncertainty with the transmission of information, whether specified or unspecified. The transmission of information, therefore, requires physical-chemical contingency. As Robert Stalnaker has noted, "[information] content requires contingency."[77] If, therefore, forces of chemical necessity completely determine the arrangement of constituents in a system, that arrangement will not exhibit complexity or convey information.

Consider, for example, what would happen if the individual nucleotide bases (A, C, G, and T) in the DNA molecule did interact by chemical necessity (along the information-bearing axis of DNA). Suppose that every time adenine (A) occurred in a growing genetic sequence, it attracted cytosine (C) to it.[78] Suppose every time guanine (G) appeared, thymine (T) followed. If this were the case, the longitudinal axis of

DNA would be peppered with repetitive sequences in which C followed A and T followed G. Rather than a genetic molecule capable of virtually unlimited novelty and characterized by unpredictable and aperiodic sequences, DNA would contain sequences awash in repetition or redundancy—much like the arrangement of atoms in crystals. In a crystal, the forces of mutual chemical attraction do determine, to a very considerable extent, the sequential arrangement of its constituent parts. Hence, sequencing in crystals is highly ordered and repetitive but neither complex nor informative. In DNA, however, where any nucleotide can follow any other, a vast array of novel sequences is possible, corresponding to a multiplicity of possible amino acid sequences and protein functions.

The forces of chemical necessity produce redundancy (roughly, law- or rule-generated repetition) or monotonous order but reduce the capacity to convey information and express novelty. Thus, as chemist Michael Polanyi noted:

> Suppose that the actual structure of a DNA molecule were due to the fact that the bindings of its bases were much stronger than the bindings would be for any other distribution of bases, then such a DNA molecule would have no information content. Its code-like character would be effaced by an overwhelming redundancy… Whatever may be the origin of a DNA configuration, it can function as a code only if its order is not due to the forces of potential energy. It *must be* as physically indeterminate as the sequence of words is on a printed page [emphasis added].[79]

Bonding affinities, to the extent they exist, inhibit the maximization of information because they determine that specific outcomes will follow specific conditions with high probability.[80] Yet information-carrying capacity is maximized when just the opposite situation obtains, namely, when antecedent conditions allow many improbable outcomes. Chemical affinities do not generate complex sequences. Thus, they cannot be invoked to explain the origin of information, whether specified or otherwise.

A tendency to conflate the qualitative distinctions between "order" and "complexity" has characterized self-organizational scenarios—whether those that invoke internal properties of chemical attraction or an external organizing force or source of energy. That tendency calls into question the relevance of these scenarios of the origin of life. What needs explaining in biology is not the origin of order (defined as symmetry or repetition) but of specified information—the highly complex, aperiodic, and specified sequences that make biological function possible. As Yockey warns: "Attempts to relate the idea of order… with biological organization or specificity must be regarded as a play on words that cannot stand careful scrutiny."[81]

In the face of these difficulties, some self-organizational theorists have claimed that we must await the discovery of new natural laws to explain the origin of biological information. As Manfred Eigen has argued, "our task is to find an algorithm, a natural law, that leads to the origin of information."[82] Such a suggestion betrays confusion on two counts. First, scientific laws don't generally produce or cause natural phenomena, they describe them. For example, Newton's law of gravitation described, but did not cause or explain, the attraction between planetary bodies. Second, laws necessarily describe highly deterministic or predictable relationships between antecedent conditions and consequent events. Laws describe highly repetitive patterns in which the probability of each successive event (given the previous event) approaches unity. Yet information sequences are complex, not repetitive—information mounts as improbabilities multiply. Thus, to say that scientific laws can produce information is essentially a contradiction in terms. Instead, scientific laws describe (almost by definition) highly predictable and regular phenomena—that is, redundant order, not complexity (whether specified or otherwise).

One could argue that we might someday discover a very particular configuration of initial conditions that routinely generates high informational states. Yet the statement of this hypothetical seems itself to beg the question of the ultimate origin of information, since "a very particu-

lar set of initial conditions" sounds precisely like an information-rich—a highly complex and specified—state. In any case, everything we know experientially suggests that the amount of specified information present in a set of antecedent conditions necessarily equals or exceeds that of any system produced from those conditions.

E. The RNA World Scenario and the Displacement of the Information Problem

In addition to the general categories of explanation already examined, origin-of-life researchers have proposed many more specific scenarios, each emphasizing random variations (chance), self-organizational laws (necessity), or both. Some of those scenarios purport to address the information problem; others attempt to bypass it altogether. Yet on closer examination, even scenarios that appear to alleviate the problem of the origin of specified biological information merely shift the problem elsewhere. Genetic algorithms can "solve" the information problem, but only if programmers provide informative target sequences and selection criteria. Simulation experiments can produce biologically relevant precursors and sequences, but only if experimentalists manipulate initial conditions or select and guide outcomes—that is, only if they add information themselves. As discussed in detail in my book *Signature in the Cell*, origin-of-life theories can leapfrog the problem altogether, but only by presupposing the presence of information in some other preexisting form.[83]

For example, some have claimed that the RNA-world scenario offers a promising approach to the origin-of-life problem and with it, presumably, the problem of the origin of the first genetic information. The RNA world was proposed as an explanation for the origin of the interdependence of nucleic acids and proteins in the cell's information-processing system. In extant cells, building proteins requires genetic information from DNA, but information in DNA cannot be processed without many specific proteins and protein complexes. This poses a chicken-or-egg problem. The discovery that RNA (a nucleic acid) possesses some

limited catalytic properties similar to those of proteins suggested a way to solve that problem. "RNA-first" advocates proposed an early state in which RNA performed both the enzymatic functions of modern proteins and the information-storage function of modern DNA, thus allegedly making the interdependence of DNA and proteins unnecessary in the earliest living system.

Nevertheless, many fundamental difficulties with the RNA-world scenario have emerged. First, synthesizing (and/or maintaining) many essential building blocks of the RNA molecules under realistic conditions has proven either difficult or impossible.[84] Further, the chemical conditions required for the synthesis of ribose sugars are decidedly incompatible with the conditions required for synthesizing nucleotide bases.[85] Yet both are necessary constituents of RNA. Second, naturally occurring RNA possesses very few of the specific enzymatic properties of proteins necessary to extant cells. In fact, RNA catalysts do not function as true enzyme catalysts. Enzymes are capable of coupling energetically favorable and unfavorable reactions together. RNA catalysts, so-called "ribozymes," are not. Third, RNA-world advocates offer no plausible explanation for the transitions from (1) RNA-based RNA synthesis to (2) RNA-based protein synthesis to (3) the modern DNA, RNA *and* protein-based protein synthesis translation system used in cells today.[86] Fourth, attempts to enhance the limited catalytic properties of RNA molecules in so-called ribozyme engineering experiments have inevitably required extensive investigator manipulation, thus simulating, if anything, the need for intelligent design, not the efficacy of an undirected chemical evolutionary process.[87]

Most importantly for our present considerations, the RNA-world hypothesis presupposes, but does not explain, the origin of sequence specificity or information in the original functional RNA molecules. As noted, the RNA-world scenario was proposed as an explanation for the functional interdependence problem, not the information problem. Even so, some RNA-world advocates seem to envision leapfrogging the sequence-specificity problem. They imagine oligomers of RNA aris-

ing by chance on the prebiotic earth and then later acquiring an ability to polymerize copies of themselves—that is, to self-replicate. In such a scenario, the capacity to self-replicate would favor the survival of those RNA molecules that could do so and would thus favor the specific sequences that the first self-replicating molecules happened to have. Thus, sequences that originally arose by chance would subsequently acquire a functional significance as "an accidental choice remembered."

This suggestion, however, merely shifts the information problem out of view. To date, scientists have been able to design RNA catalysts that will copy only about 10% of themselves.[88] For strands of RNA to perform even this limited replicase (self-replication) function, they must, like proteins, have very specific arrangements of constituent building blocks (nucleotides in the RNA case). Further, the strands must be long enough to fold into complex three-dimensional shapes (to form so-called tertiary structures). Thus, any RNA molecule capable of even limited replicase function must have possessed considerable (specified) information[89]—information that, in the case of actual (partial) RNA replicators was produced by intelligent "ribozyme engineers."

Indeed, explaining how the building blocks of RNA arranged themselves into functionally specified sequences in a prebiotic environment has proven no easier than explaining how the constituent parts of DNA might have done so, especially given the high probability of destructive cross-reactions between desirable and undesirable molecules in any realistic pre-biotic soup. As de Duve noted in a critique of the RNA-world hypothesis, "hitching the components together in the right manner raises additional problems of such magnitude that no one has yet attempted to do so in a prebiotic context."[90]

Recently some have claimed that a scientific study by chemists Matthew Powner, Béatrice Gerland, and John Sutherland of the University of Manchester[91] has rendered the RNA scenario "eminently plausible,"[92] as Stephen Fletcher, a chemist from the University of Loughborough, has put it. Starting with several simple chemical compounds, Powner and his colleagues successfully synthesized a pyrimidine ribonucleotide,

one of the two types of the four bases of the RNA molecule. (Of the four information-carrying nucleotide bases in DNA and RNA, chemists classify two as "pyrimidines" and two as "purines" due to differences in chemical structure.)

Nevertheless, this work did nothing to address the much more acute problem of explaining how the nucleotide bases in DNA or RNA acquired their specific information-rich arrangements. In effect, the Powner study putatively explains the origin of two of the "letters" in the genetic text, but not the specific arrangements of the four different "letters" into functional genetic "words" or "sentences."

Moreover, Powner and his colleagues only partially addressed the problem of generating the constituent building blocks of RNA under plausible pre-biotic conditions. The weakness in their demonstration, ironically, was their own skillful intervention. To ensure a biologically relevant outcome, they had to intervene—repeatedly and intelligently—in their experiment: first, by selecting only the "right-handed" versions of sugar that life requires (sugars, like amino acids, come in two mirror-image chemical structures called isomers); second, by purifying their reaction products at each step to prevent interfering cross-reactions; and third, by following a precise procedure in which they carefully selected chemically purified reagents and then choreographed the order in which those reagents were introduced into the reaction series. As my colleague David Berlinski pointed out, "They began with what they needed and purified what they got until they got what they wanted."

Thus, not only did this study *not* address the problem of getting nucleotide bases to arrange themselves into functionally specified sequences, but the extent to which it did succeed in producing biologically relevant chemical constituents of RNA actually illustrates the indispensable role of *intelligence* in generating such chemistry.

Proponents of chemical evolution have also cited the more recent work of Tracey Lincoln and Gerald Joyce,[93] who have ostensibly established the capacity of RNA to self-replicate as a way of demonstrating the plausibility of the RNA World. Nevertheless, their "self-replicating"

RNA molecules could not copy a template of genetic information from free-standing nucleotides as protein machines (called polymerases) do in actual cells. Instead, in the experiment, a pre-synthesized *specifically sequenced* RNA molecule merely catalyzed a single chemical bond, fusing together two other pre-synthesized partial RNA chains. Their version of "self-replication," therefore, amounted to nothing more than joining two sequence-specific pre-made halves together.

More significantly, Lincoln and Joyce themselves *intelligently arranged* the base sequences in these RNA chains. They generated the sequence-specific functional information that made even this limited form of "self-replication" possible. Thus, the experiment not only demonstrated that even a limited capacity for RNA self-replication depends upon information-rich RNA molecules, it also lent inadvertent support to the idea that intelligence is necessary to produce such functionally specified information. The Lincoln and Joyce experiment illustrates a well-known problem in origin-of-life research known as "investigator interference," wherein the "success" of the experiment invariably and crucially depends on the intervention, guidance, or choreography of *intelligent* chemists doing the organic synthesis experiments.

III.

A. The Return of the Design Hypothesis

If attempts to solve the information problem only relocate it, and if neither chance nor physical-chemical necessity, nor the two acting in combination, explains the ultimate origin of specified biological information, what does? Do we know of any entity that has the causal powers to create large amounts of specified information? We do. As Henry Quastler recognized, "creation of new information is habitually associated with conscious activity."[94]

Indeed, experience affirms that functionally specified information routinely arises from the activity of intelligent agents. A computer user who traces the information on a screen back to its source invariably

comes to a mind, that of a software engineer or programmer. Similarly, the information in a book or newspaper column ultimately derives from a writer—from a mental, rather than a strictly material, cause.

But could this intuitive connection between information and the prior activity of a designing intelligence justify a rigorous scientific argument for intelligent design? I first began to consider this possibility during my PhD research at Cambridge University in the late 1980s, after reading *The Mystery of Life's Origin* and extensive discussions with Charles Thaxton during my last year in Dallas before leaving for England. During my PhD work, I began to examine how scientists investigating origins events developed and evaluated their hypotheses and arguments. Specifically, I examined the method of reasoning that historical scientists use to identify causes responsible for events in the remote past.

I discovered that historical scientists often make inferences with a distinctive logical form (known technically as *abductive inferences*).[95] Paleontologists, evolutionary biologists, and other historical scientists reason like detectives and infer *past* conditions or causes from *present* clues. As Stephen Jay Gould notes, historical scientists typically "infer history from its results."[96]

Nevertheless, as many philosophers have noted, there is a problem with this kind of historical reasoning, namely, there is often more than one cause that can explain the same effect. This makes reasoning from present clues (circumstantial evidence) tricky because the evidence can point to more than one causal explanation or hypothesis. To address this problem in geology, the nineteenth-century geologist Thomas Chamberlain delineated a method of reasoning he called "the method of multiple working hypotheses."[97]

Contemporary philosophers of science such as Peter Lipton have called this the method of "inference to the best explanation."[98] That is, when trying to explain the origin of an event or structure from the past, scientists often compare various hypotheses to see which would, if true, best explain it. They then provisionally affirm the hypothesis that best

explains the data as the one that is most likely to be true. But that raises an important question: Exactly what makes an explanation best?

As it happens, historical scientists have developed criteria for deciding which cause, among a group of competing possible causes, provides the best explanation for some event in the remote past. The most important of these criteria is called "causal adequacy." This criterion requires that historical scientists, as a condition of a successful explanation, identify causes that are known to have the power to produce the kind of effect, feature or event that requires explanation. In making these determinations, historical scientists evaluate hypotheses against their present knowledge of cause and effect. Causes that are known to produce the effect in question are judged to be better candidates than those that are not. For instance, a volcanic eruption provides a better explanation for an ash layer in the earth than an earthquake because eruptions have been observed to produce ash layers, whereas earthquakes have not.

One of the first scientists to develop this principle was the geologist Charles Lyell who also influenced Charles Darwin. Darwin read Lyell's *magnum opus, The Principles of Geology,* on the voyage of the *Beagle* and employed its principles of reasoning in *The Origin of Species*. The subtitle of Lyell's *Principles* summarized the geologist's central methodological principle: *Being an Attempt to Explain the Former Changes of the Earth's Surface, by Reference to Causes Now in Operation*.[99] Lyell argued that when scientists seek to explain events in the past, they should not invoke unknown or exotic causes, the effects of which we do not know. Instead they should cite causes that are known from our uniform experience to have the power to produce the effect in question. Historical scientists should cite "causes now in operation" or presently acting causes. This was the idea behind his uniformitarian principle and the dictum: "The present is the key to the past." According to Lyell, our *present* experience of cause and effect should guide our reasoning about the causes of *past* events. Darwin himself adopted this methodological principle as he sought to demonstrate that natural selection qualified as a *vera causa*, that is, a true, known, or actual cause of significant biological change. He

sought to show that natural selection was "causally adequate" to produce the effects he was trying to explain.[100]

Both philosophers of science and leading historical scientists have emphasized causal adequacy as the key criterion by which competing hypotheses are adjudicated. But philosophers of science also have noted that assessments of explanatory power lead to conclusive inferences only when it can be shown that there is *only one known cause* for the effect or evidence in question. Philosophers of science Michael Scriven and Elliot Sober, for example, have pointed out that historical scientists can make inferences about the past with confidence when they discover evidence or artifacts for which there is only one cause known to be capable of producing them.[101] Indeed, when scientists can infer a *uniquely* plausible cause, they can avoid the fallacy of affirming the consequent and the error of ignoring other possible causes with the power to produce the same effect.[102]

B. Intelligent Design as the Best Explanation?

What did all this have to do with the origin of the information necessary to produce the first life? As a PhD student I wondered if a case for an intelligent cause could be formulated and justified in the same way that historical scientists would justify any other causal claim about an event in the past. My study of historical scientific reasoning and origin-of-life research suggested to me that it was possible to formulate a rigorous scientific case for intelligent design as an inference to the best explanation, specifically, as the best explanation for the origin of biological information. The action of a conscious and intelligent agent clearly represents a known (presently acting) and adequate cause for the origin of information. Uniform and repeated experience affirms that intelligent agents produce information-rich systems, whether software programs, ancient inscriptions, or Shakespearean sonnets. Minds are clearly capable of generating functionally specified information.

Further, the functionally specified information in the cell also points to intelligent design as the *best* explanation for the ultimate origin of

biological information. Why? Experience shows that large amounts[103] of such information (especially when digitally or alphabetically encoded) *invariably* originate from an intelligent source—from a mind or a personal agent. In other words, intelligent activity is *the only known cause* of the origin of functionally specified information (at least, starting from a non-living source, that is, from purely physical or chemical antecedents).[104] Since intelligence is the only known cause of specified information in such a context, the presence of functionally specified information sequences in even the simplest living systems points definitely to the past existence and activity of a designing intelligence.

Notice also that one can detect (or retrodict) the past action of a designing intelligence from an information-rich effect even if the cause itself cannot be directly observed.[105] For example, the information-rich inscriptions in the famed Rosetta Stone clearly allow archeologists to infer the activity of intelligent scribes even if they did not see such agents chisel the letters and hierogylphs into the stone. Similarly, the specified and complex arrangements of nucleotide bases in DNA imply the past action of intelligence, even if such activity cannot be directly observed.

Ironically, the generalization that intelligence is the only known cause of specified complexity or information (at least, starting from a nonbiological source) has received support from origin-of-life research itself. During the last fifty years, every naturalistic model proposed has failed to explain the origin of the specified genetic information required to build a living cell.[106] Instead, attempts to solve the origin-of-life problem with pre-biotic simulation experiments and computer simulations have invariably required inputs of functional information from intelligent agents, further confirming intelligence as the only known or "presently acting" cause of the origin of functionally specified information.

When I first noticed the subtitle of Lyell's book, referring us to "causes now in operation," a light came on for me. I immediately asked myself a question: "What causes 'now in operation' produce digital code or specified information?" Is there a known cause—a *vera causa*—of the origin of such information? What does our uniform experience tell us?

As I thought about this further, it occurred to me that by Lyell's and Darwin's own rule of reasoning and test of a sound scientific explanation, intelligent design must qualify as the currently best scientific explanation for the origin of biological information. Why? Because we have independent evidence—"uniform experience"—that intelligent agents are capable of producing specified information and, as origin-of-life research itself has helped to demonstrate, we know of no other cause capable of producing functional or specified information starting from a purely physical or chemical state.

Scientists in many fields recognize the connection between intelligence and specified information and make inferences accordingly. Anthropologists establish the intelligence of early hominids from chipped flints that are too improbably specified in form (and function) to have been produced by natural causes; NASA's search for extraterrestrial intelligence (SETI) presupposes that any information embedded in electromagnetic signals coming from space would indicate an intelligent source.[107] Astronomers have not found such information-rich signals coming from space, but closer to home, molecular biologists have identified information-rich sequences and systems in the cell, suggesting, by the same logic, an intelligent cause for those effects.

Indeed, our uniform experience affirms that specified information—whether inscribed in hieroglyphs, written in a book, encoded in a terrestrial radio signal, or produced in an RNA-world "ribozyme engineering" experiment—*always* arises from an intelligent source, from a mind and not a strictly material process. So the discovery of the functionally specified digital information in DNA and RNA provides strong grounds for inferring that intelligence played a role in the origin of these molecules. Whenever we find specified information and we know the causal story of how that information arose, we always find that it arose from an intelligent source. It follows that the best, most likely explanation for the origin of the specified, digitally encoded information in DNA and RNA is that it too had an intelligent source. Intelligent design

best explains the specified genetic information necessary to produce the first living cell.

C. Argument from Ignorance? Or an Inference to the Best Explanation?

Objectors charge that this design argument constitutes an argument from ignorance. They say that design advocates use our present ignorance of any sufficient materialistic cause of specified information as the sole basis for inferring an intelligent cause of the information present in the cell. Since we don't yet know how specified biological information could have arisen, we invoke the mysterious notion of intelligent design. On this view, intelligent design functions not as an explanation but as a placeholder for ignorance.

My response is that arguments from ignorance occur when evidence against a proposition X is offered as the sole (and conclusive) grounds for accepting some alternative proposition Y. The inference to design as sketched above (see part III, sections A and B) does not commit this fallacy.

True, the previous part of this chapter (see part II, sections A–E) argued that at present all types of natural causes and mechanisms fail to account for the origin of biological information from a prebiotic state. And clearly, this lack of knowledge of any adequate natural cause does provide *part* of the grounds for inferring design from information in the cell; but our "ignorance" of any sufficient natural cause is only part of the basis for inferring design. We also *know* that intelligent agents can and do produce information-rich systems: we have positive experience-based knowledge of an alternative cause that is sufficient, namely, intelligence or "conscious activity."

For this reason, the design inference defended here does not constitute an argument from ignorance but an inference to the best explanation.[108] Inferences to the best explanation do not assert the adequacy of one causal explanation merely on the basis of the inadequacy of some other causal explanation. Instead, they compare the explanatory power

of many competing hypotheses to determine which hypothesis would, if true, provide the best explanation for some set of relevant data based upon our *knowledge* of the causal powers of competing explanatory entities.[109]

This chapter has followed precisely this method to make a case for intelligent design as the best explanation for the origin of biological information. It has evaluated and compared the causal efficacy of four broad categories of explanation—chance, necessity, the combination of those two, and intelligent design—with respect to their ability to produce large amounts of specified complexity or information. As we have seen, neither scenarios based on chance nor those based on necessity (nor those that combine the two) can explain the origin of specified biological information in a prebiotic context. That result comports with our uniform human experience. Natural processes do not produce information-rich structures starting from purely physical or chemical antecedents. Nor does matter, whether acting at random or under the force of physical-chemical necessity, arrange itself into complex, information-rich sequences.

On the other hand, we know from experience that conscious intelligent agents can create informational sequences and systems. To quote Quastler, "creation of new information is habitually associated with conscious activity."[110] Further, experience teaches that whenever large amounts of specified complexity or information are present in an artifact or entity whose causal story is known, invariably creative intelligence—intelligent design—played a causal role in the origin of that entity. Thus, when we encounter such information in the biomacromolecules necessary to life, we may infer—based on our *knowledge* (not our ignorance) of established cause-effect relationships—that an intelligent cause operated in the past to produce the specified complexity or information necessary to the origin of life.

Insofar as the inference to design depends on present knowledge of the demonstrated causal powers of natural entities and intelligent agency, it no more constitutes an argument from ignorance than any

other well-grounded inference in geology, archaeology, or paleontology—where present knowledge of cause-effect relationships guides the inferences that scientists make about the causal past.

Some objectors would characterize the design inference presented here as invalid or unscientific because it depends on a negative generalization—i.e., "purely physical and chemical causes do not generate large amounts of specified information"—which future discoveries may later falsify. We should "never say never," they say.

Yet science often says "never," even if it can't say so for sure. Negative or proscriptive generalizations often play an important role in science. As many scientists and philosophers of science have pointed out, scientific laws often tell us not only what does happen but also what does not happen.[111] The conservation laws in thermodynamics, for example, proscribe certain outcomes. The first law tells us that energy is never created or destroyed. The second tells us that the entropy of a closed system will never decrease over time. Those who claim that such "proscriptive laws" do not constitute knowledge, because they are based on past but not future experience, will not get very far if they try to use their skepticism to justify funding for research on, say, perpetual motion machines.

Further, without proscriptive generalizations, without knowledge about what various possible causes cannot or do not produce, historical scientists could not make determinations about the past. Reconstructing the past requires making abductive inferences from present effects back to past causal events.[112] Making such inferences requires a progressive elimination of competing causal hypotheses. Deciding which causes can be eliminated from consideration requires knowing what effects a given cause can—and cannot—produce. If historical scientists could never say that particular entities lack particular causal powers, they could never eliminate them, even provisionally, from consideration. Thus, they could never infer that a specific cause had acted in the past. Yet historical and forensic scientists make such inferences all the time, without worrying about committing fallacious arguments from ignorance. And for good reason. A vast amount of human experience shows that intelligent agents

have unique causal powers that matter (especially nonliving matter) does not. When we observe features or effects that we know from experience only agents produce, we rightly infer the prior activity of intelligence.

To determine the best explanation, scientists do not need to say "never" with absolute certainty. They need only say that a postulated cause is best, given what we know at present about the demonstrated causal powers of competing entities or agencies. That cause C can produce effect E makes it a better explanation of E than some cause D that has never produced E (especially if D seems incapable of doing so on theoretical grounds), even if D might later demonstrate causal powers of which we are presently ignorant.[113]

Thus, the objection that the design inference constitutes an argument from ignorance reduces in essence to a restatement of the problem of induction. Yet one could make the same objection against any scientific law or explanation or against any historical inference that takes present, but not future, knowledge of natural laws and causal powers into account. Our knowledge of what can and cannot produce large amounts of specified information may later have to be revised, but so might the laws of thermodynamics. Inferences to design may later prove incorrect, as may other inferences implicating various natural causes. Such possibilities do not stop scientists from making generalizations about the causal powers of various entities or from using those generalizations to identify probable or most plausible causes in particular cases.

D. But Is It Science?

Of course, many simply refuse to consider the design hypothesis on grounds that it does not qualify as "scientific." Such critics affirm an extra-evidential principle known as methodological naturalism.[114] Methodological naturalism asserts that, as a matter of definition, for a hypothesis, theory, or explanation to qualify as "scientific," it must invoke only naturalistic or materialistic entities. On that definition, critics say, the intelligent design hypothesis does not qualify. Yet, even if one grants this definition, it does not follow that some nonscientific (as defined by

methodological naturalism) or metaphysical hypothesis may not constitute a better, more causally adequate, explanation. This chapter has argued that, whatever its classification, the design hypothesis does constitute a better explanation than its materialistic or naturalistic rivals for the origin of specified biological information. Surely, simply classifying an argument as metaphysical does not refute it.

In any case, methodological naturalism now lacks justification as a normative definition of science. First, attempts to justify methodological naturalism by reference to metaphysically neutral (that is, non-question-begging) demarcation criteria have failed.[115] Second, to assert methodological naturalism as a normative principle for all of science has a negative effect on the practice of certain scientific disciplines, especially the historical sciences. In origin-of-life research, for example, methodological naturalism artificially restricts inquiry and prevents scientists from seeking some hypotheses that might provide the best, most causally adequate explanations. To be a truth-seeking endeavor, the question that origin-of-life research must address is not "Which materialistic scenario seems most adequate?" but rather "What actually caused life to arise on Earth?" Clearly, one possible answer to that latter question is this one: "Life was designed by an intelligent agent that existed before the advent of humans." If one accepts methodological naturalism as normative, however, scientists are not allowed to consider the design hypothesis as possibly true. Such an exclusionary logic diminishes the significance of any claim of theoretical superiority for non-design hypotheses and raises the possibility that the best "scientific" explanation (as defined by methodological naturalism) may not be the best in fact.

As many historians and philosophers of science now recognize, theory-evaluation is an inherently comparative enterprise. Theories that gain acceptance in artificially constrained competitions can claim to be neither "most probably true" nor "most empirically adequate." At best, such theories can be considered "the most probably true or adequate among an artificially limited set of options." Openness to the design hypothesis would seem necessary, therefore, to any fully rational histori-

cal biology—that is, to one that seeks the truth, "no holds barred."[116] A historical biology committed to following the evidence wherever it leads will not exclude hypotheses *a priori* on metaphysical grounds. Instead, it will employ only metaphysically neutral criteria—such as explanatory power and causal adequacy—to evaluate competing hypotheses. Yet this more open (and seemingly rational) approach to scientific theory evaluation would now suggest the theory of intelligent design as the best, most causally adequate explanation for the origin of the information necessary to build the first living organism.

Endnotes

1. Harmke Kamminga, "Protoplasm and the Gene," in *Clay Minerals and the Origin of Life*, ed. A. G. Cairns-Smith and H. Hartman (Cambridge: Cambridge University Press, 1986), 1.
2. Alexander Oparin, *Genesis and Evolutionary Development of Life* (New York: Academic Press, 1968), 7.
3. F. Crick and J. Watson, "A Structure for Deoxyribose Nucleic Acid," *Nature* 171 (1953): 737–38; F. Crick and J. Watson, "Genetical Implications of the Structure of Deoxyribose Nucleic Acid," *Nature* 171 (1953): 964–67, esp. 964; T. D. Schneider, "Information Content of Individual Genetic Sequences," *Journal of Theoretical Biology* 189 (1997): 427–41; W. R. Loewenstein, *The Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life* (New York: Oxford, 1999).
4. Bernd-Olaf Kuppers, *Information and the Origin of Life* (Cambridge: MIT Press, 1990), 170–72.
5. L. E. Kay, "Who Wrote the Book of Life? Information and the Transformation of Molecular Biology," *Science in Context* 8 (1994): 601–34; L. E. Kay, "Cybernetics, Information, Life: The Emergence of Scriptural Representations of Heredity," *Configurations* 5 (1999): 23–91; L. E. Kay, *Who Wrote the Book of Life?* (Stanford, CA: Stanford University Press, 2000), xv–xix.
6. Darwin's only speculation on the origin of life is found in an unpublished letter to Joseph Hooker. In it, he sketched the outlines of the chemical evolutionary idea, namely, that life could have first evolved from a series of chemical reactions. He wrote: "… if (& oh what a big if) we could conceive in some warm little pond with all sorts of ammonia & phosphoric salts,—light, heat, electricity &c present, that a protein compound was chemically formed, ready to undergo still more complex changes…" Darwin to Hooker, February 1, 1871, Darwin Correspondence Project, https://www.darwinproject.ac.uk/letter/DCP-LETT-7471.xml. Darwin's original punctuation and abbreviations have been preserved.
7. E. Haeckel, *The Wonders of Life*, trans. J. McCabe (London: Watts, 1905), 111; T. H. Huxley, "On the Physical Basis of Life," *Fortnightly Review* 5 (1869): 129–45.
8. A. I. Oparin, *The Origin of Life*, trans. S. Morgulis (New York: Macmillan, 1938); S. C. Meyer, "Of Clues and Causes: A Methodological Interpretation of Origin of Life Studies" (PhD dissertation, Cambridge University, 1991).

9. J. C. Kendrew, in H. F. Judson, *The Eighth Day of Creation* (NY: Simon and Schuster, 1979), 562–63; J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis," *Nature* 181 (1958): 662–66, esp. 664.
10. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell* (New York: Garland, 1983), 111–12, 127–31.
11. Judson, *Eighth Day*, 30.
12. Judson, *Eighth Day*, 95–96; E. Chargaff, *Essays on Nucleic Acids* (Amsterdam: Elsevier, 1963), 21.
13. Chargaff, *Essays*, 21.
14. Crick and Watson, "Structure."
15. Crick and Watson, "Genetical Implications," 965.
16. Judson, *Eighth Day*, 611.
17. Crick and Watson, "Structure"; Crick and Watson, "Genetical Implications."
18. Judson, *Eighth Day*, 245–46, 335–36.
19. *Ibid*., 470–89; J. H. Matthei and M. W. Nirenberg, "Characteristics and Stabilization of DNAase-Sensitive Protein Synthesis in *E. coli* Extracts," *Proceedings of the National Academy of Sciences, USA* 47 (1961): 1580–88; J. H. Matthei and M. W. Nirenberg, "The Dependence of Cell-Free Protein Synthesis in *E. coli* upon Naturally Occurring or Synthetic Polyribonucleotides," *Proceedings of the National Academy of Sciences, USA* 47 (1961): 1588–1602.
20. Alberts et al., *Molecular Biology*, 106–8; S. L. Wolfe, *Molecular and Cellular Biology* (Belmont, CA: Wadsworth, 1993), 639–48.
21. We now know, of course, that in addition to the process of gene expression, specific enzymes must often modify amino acid chains after translation in order to achieve the precise sequencing necessary to allow correct folding into a functional protein. The amino acid chains produced by gene expression may also undergo further modification in sequence at the endoplasmic reticulum. Finally, even well-modified amino acid chains may require preexisting protein "chaperones" to help them fold into a functional three-dimensional configuration. All these factors make it impossible to predict a protein's final sequence from its corresponding gene sequence alone. See S. Sarkar, "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology," in *The Philosophy and History of Molecular Biology: New Perspectives*, ed. S. Sarkar (Dordrecht, Netherlands: Boston Studies in Philosophy of Science, 1996), 196, 199–202. Nevertheless, this unpredictability in no way undermines the claim that DNA exhibits the property of "sequence specificity," or the isomorphic claim that it contains "specified information" as argued here in part I, section E. Sarkar argues, for example, that the absence of such predictability renders the concept of information theoretically superfluous for molecular biology. Instead, this unpredictability shows that the sequence specificity of DNA base sequences constitutes a necessary, though not sufficient, condition of attaining proper protein folding; that is, DNA does contain specified information (part I, section E), but not enough to determine protein folding by itself. Instead, the presence of both post-translation processes of modification and pretranscriptional genomic editing (through exonucleases, endonucleases, spliceosomes, and other editing enzymes) only underscores the need for other preexisting, information-rich biomolecules in order to process genomic information in the cell. The presence of a complex and functionally integrated information-processing system does suggest that the informa-

tion on the DNA molecule is insufficient to produce proteins. It does not show that such information is unnecessary to produce proteins, nor does it invalidate the claim that DNA, therefore, stores and transmits specified genetic information.

22. C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27 (1948): 379–423, 623–56.

23. F. Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981), 6–10.

24. *Ibid*.; Shannon, "A Mathematical Theory."

25. B. Kuppers, "On the Prior Probability of the Existence of Life," in *The Probabilistic Revolution*, ed. Lorenz Kruger et al. (Cambridge: MIT Press, 1987), 355–69.

26. Schneider, "Information Content"; see also H. P. Yockey, *Information Theory and Molecular Biology* (Cambridge: Cambridge University Press, 1992), 246–58, for important refinements in the method of calculating the information-carrying capacity of proteins and DNA.

27. C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949), 8.

28. Schneider, "Information Content," 58–177; Yockey, *Information Theory*, 58–177.

29. Schneider, "Information Content"; Yockey, *Information Theory*; Sarkar, "Biological Information," 199–202, esp. 196; F. Crick, "On Protein Synthesis," *Symposium for the Society of Experimental Biology* 12 (1958): 138–63, esp. 144, 153.

30. Crick, "On Protein Synthesis," 144, 153.

31. Recall that the determination of the genetic code depended, for example, on observed correlations between changes in nucleotide base sequences and amino acid production in "cell-free systems." See Judson, *Eighth Day*, 470–87.

32. For a much more detailed discussion of specification, see W. A. Dembski, *The Design Inference: Eliminating Chance through Small Probabilities* (Cambridge: Cambridge University Press, 1998), 1–35, 136–74. The simplified discussion here draws in part on Stephen Meyer, "Yes, Intelligent Design Is Detectable by Science," *Evolution News*, April 24, 2018, https://evolutionnews.org/2018/04/yes-intelligent-design-is-detectable-by-science/.

33. J. Bowie and R. Sauer, "Identifying Determinants of Folding and Activity for a Protein of Unknown Sequences: Tolerance to Amino Acid Substitution," *Proceedings of the National Academy of Sciences, USA* 86 (1989): 2152–56; J. Reidhaar-Olson and R. Sauer, "Functionally Acceptable Solutions in Two Alpha-Helical Regions of Lambda Repressor," *Proteins, Structure, Function, and Genetics* 7 (1990): 306–10.

34. R. Dawkins, *River out of Eden* (New York: Basic Books, 1995), 11.

35. B. Gates, *The Road Ahead* (Boulder, CO: Blue Penguin, 1996), 228.

36. L. E. Orgel, *The Origins of Life* (New York: John Wiley, 1973), 189.

37. Kay, "Who Wrote the Book of Life?" 611–12, 629; Kay, "Cybernetics"; Kay, *Who Wrote the Book of Life?* (For full details for Kay references, see note 5 above.)

38. Sarkar, "Biological Information," 199–202.

39. E. Schrödinger, *What Is Life?* and *Mind and Matter* (Cambridge: Cambridge University Press, 1967), 82; Alberts et al., *Molecular Biology*, 21; Crick and Watson, "A Structure"; Crick and Watson, "Genetical Implications"; Crick, "On Protein"; Judson, *Eighth Day*, 611; Orgel, *Origins of Life*, 189.

40. P. Davies, *The Fifth Miracle* (New York: Simon and Schuster, 1998), 120.

41. Orgel, *Origins of Life*, 189.

42. Loewenstein, *Touchstone*; Davies, *Fifth Miracle*; Schneider, "Information Content"; C. Thaxton and W. Bradley, "Information and the Origin of Life," in The *Creation Hypothesis: Scientific Evidence for an Intelligent Designer*, ed. J. P. Moreland (Downers Grove, IL: InterVarsity Press, 1994), 173–210, esp. 190; S. Kauffman, *The Origins of Order* (Oxford: Oxford University Press, 1993), 287–340; Yockey, *Information Theory*, 178–293; Kuppers, *Information and Origin*, 170–72; F. Crick, *Life Itself* (New York: Simon and Schuster, 1981), 59–60, 88; J. Monod, *Chance and Necessity* (New York: Vintage Books, 1971), 97–98, 143; Orgel, *Origins*, 189; D. Kenyon and G. Steinman, *Biochemical Predestination* (New York: McGraw-Hill, 1969), 199–211, 263–66; Oparin, *Genesis*, 146–47; H. Quastler, *The Emergence of Biological Organization* (New Haven, CT: Yale University Press, 1964).

43. G. Wald, "The Origin of Life," *Scientific American* 191 (August 1954): 44–53; R. Shapiro, *Origins: A Skeptic's Guide to the Creation of Life on Earth* (New York: Summit Books, 1986), 121.

44. F. Crick, "The Origin of the Genetic Code," *Journal of Molecular Biology* 38 (1968): 367–79; H. Kamminga, "Studies in the History of Ideas on the Origin of Life from 1860" (PhD dissertation, Chelsea College, University of London, 1980), 303–4.

45. C. de Duve, "The Constraints of Chance," *Scientific American*, Jan. 1996, 112; Crick, *Life Itself*, 89–93; Quastler, *Emergence*, 7.

46. H. J. Morowitz, *Energy Flow in Biology* (New York: Academic Press, 1968), 5–12; F. Hoyle and C. Wickramasinghe, *Evolution from Space* (London: J. M. Dent, 1981), 24–27; A. G. Cairns-Smith, *The Life Puzzle* (Edinburgh: Oliver and Boyd, 1971), 91–96; I. Prigogine, G. Nicolis, and A. Babloyantz, "Thermodynamics of Evolution," *Physics Today* 23 (Nov. 1972); Yockey, *Information Theory*, 246–58; Yockey, "Self-Organization, Origin of Life Scenarios and Information Theory," *Journal of Theoretical Biology* 91 (1981): 13–31; Bowie and Sauer, "Identifying Determinants"; Reidhaar-Olson et al., *Proteins*; Shapiro, *Origins*, 117–31.

47. Prigogine, "Thermodynamics of Evolution."

48. Cairns-Smith, *The Life Puzzle*, 95.

49. John Reidhaar-Olson and Robert Sauer, "Functionally Acceptable Substitutions in Two Alpha-Helical Regions of Lambda Repressor," in *Proteins: Structure, Function, and Bioinformatics* 7, no. 4 (1990): 306–316; James Bowie and Robert Sauer, "Identifying the Determinants of Folding and Activity for a Protein of Unknown Structure," *Proceedings of the National Academy of Sciences USA* 86 (1989): 2152–2156.

50. Interestingly, their descriptions of their own results often downplay the rarity of functional sequences within sequence space. Instead, they often emphasize the tolerance for different amino acids that is allowable at each site. For example, the abstract of the paper reporting the figure of 1 in 10^{63} makes no mention of that figure or its potential significance, stating instead that their results "reveal the high level of degeneracy in the information that specifies a particular protein fold." Reidhaar-Olson and Sauer, "Functionally Acceptable Substitutions."

51. Bowie and Sauer, "Identifying the Determinants of Folding"; Reidhaar-Olson and Sauer, "Functionally Acceptable Substitutions"; Cyrus Chothia, Israel Gelfand, and Alexander Kister, "Structural Determinants in the Sequences of Immunoglobulin Variable *Acid* Domain," *Journal of Molecular Biology* 278 (1998), 457–479; Douglas Axe, "Extreme Functional Sensitivity to Conservative Amino Acid Changes on Enzyme Exteriors" *J. Mol. Biol.* 301 (2000), 585–595; Taylor, Sean V., Kai U. Walter, Peter Kast, and Donald Hilvert,

"Searching Sequence Space for Protein Catalysts," *Proceedings of the National Academy of Sciences USA* 98 (2001), 10596–10601.

52. See, for example, Max F. Perutz and Hermann Lehmann, "Molecular Pathology of Human Hemoglobin," *Nature* 219 (1968), 902–909.

53. Douglas Axe, "Estimating the Prevalence of Protein Sequences Adopting Functional Enzyme Folds" *J. Mol. Biol*. 341 (2004), 1295–1315.

54. For Dembski's treatment of probabilistic resources at the scale of the known universe, see Dembski's above-cited *Design Inference*, ch. 6.

55. The number of possible ways to combine elementary particles (and thus the number of combinatorial possible events) is actually much greater than the number of different events that could have taken place in the history of the universe. Why? Because the occurrence of each individual event precludes the occurrence of many other possible events within the larger combinatorial space. The number of combinatorial possible events represents the number of different events that might have occurred before the universe actually unfolded in the way that it did. Dembski correctly identifies the maximum number of events that could *actually* occur in any given history of the universe as the number that determines the probabilistic resources of the universe. This smaller number determines how many opportunities the universe has to produce a particular outcome by chance. As Dembski explains, it is not the total number of combinatorial possible events (or elementary particles) in the universe which determines the available probabilistic resources, but how many opportunities there are to "individuate" actual events. See *The Design Inference*, ch. 6, and in that chapter p. 209 n.15.

56. The elementary particles enumerated in this calculation include only protons, neutrons, and electrons (fermions), because only these particles have what physicists call "half-integral spin" which allows them to form material structures. This calculation does not count bosons, which cannot form material structures, but instead only transmit energy. Nor does this calculation count the quarks out of which protons and neutrons are made, because quarks are necessarily bound together within these particles. Even if quarks were counted, however, the total number of elementary particles would change by less than one order of magnitude, because there are only three quarks per proton or neutron.

57. Because there is an upper limit on the speed of light, only those parts of the universe that are observable to us can affect events on Earth. Thus, the observable universe is the only part of the universe with probabilistic resources relevant to explaining events on Earth.

58. To be safe, Dembski rounded the number that he had calculated up a few orders of magnitude to 10^{150}, though without any physical or mathematical justification. Since he didn't need to do this, I decided to use his more accurate, if less round, number as the actual measure of the probabilistic resources of the universe in my evaluations of the chance hypothesis.

59. See discussion and references in *Signature in the Cell*, 217.

60. P. T. Mora, "Urge and Molecular Biology," *Nature* 199 (1963): 212–19.

61. Oparin, *Genesis*, 146–47.

62. C. de Duve, *Blueprint for a Cell: The Nature and Origin of Life* (Burlington, NC: Neil Patterson, 1991), 187.

63. T. Dobzhansky, "Discussion of G. Schramm's Paper," in *The Origins of Prebiological Systems and of Their Molecular Matrices*, ed. S. W. Fox (New York: Academic Press, 1965), 310; H. H. Pattee, "The Problem of Biological Hierarchy," in C. H. Waddington, ed.,

Organization, Stability, & Process [vol. 3 of *Toward a Theoretical Biology*] (Piscataway, NJ: Transaction, 1970), 123.

64. Richard Dawkins, The *Blind Watchmaker: Why the Evidence Reveals a Universe Without Design* (New York: Norton, 1987), 47–49; Kuppers, "On the Prior Probability."

65. Kuppers, "On the Prior Probability," 366.

66. Dawkins, *Blind Watchmaker*, 47–49; P. Nelson, "Anatomy of a Still-Born Analogy," *Origins and Design* 17, no. 3 (1996): 12.

67. Stephen C. Meyer, *Signature in the Cell: DNA and the Evidence for Intelligent Design* (San Francisco: HarperOne, 2009), 283–291.

68. C. de Duve, "The Beginnings of Life on Earth," *American Scientist* 83 (1995), 437.

69. Morowitz, *Energy Flow*, 5–12.

70. G. Steinman and M. N. Cole, "Synthesis of Biologically Pertinent Peptides Under Possible Primordial Conditions," *Proceedings of the National Academy of Sciences USA* 58 (1967): 735–41; G. Steinman, "Sequence Generation in Prebiological Peptide Synthesis," *Archives of Biochemistry and Biophysics* 121 (1967): 533–39; R. A. Kok, J. A. Taylor, and W. L. Bradley, "A Statistical Examination of Self-Ordering of Amino Acids in Proteins," *Origins of Life and Evolution of the Biosphere* 18 (1988): 135–42.

71. Meyer, *Signature in the Cell*, 229–270.

72. C. Thaxton, W. Bradley, and R. Olsen, *The Mystery of Life's Origin: Reassessing Current Theories* (Dallas: Lewis and Stanley, 1992), v–viii; D. Kenyon and G. Mills, "The RNA World: A Critique," *Origins and Design* 17, no. 1 (1996): 9–16; D. Kenyon and P. W. Davis, *Of Pandas and People: The Central Question of Biological Origins* (Dallas: Haughton, 1993); S. C. Meyer, "A Scopes Trial for the '90s," *Wall Street Journal*, Dec. 6, 1993, A14; Kok et al., "Statistical Examination."

73. Steinman and Cole, "Synthesis"; Steinman, "Sequence Generation."

74. Kok et al., "Statistical Examination"; B. J. Strait and G. T. Dewey, "The Shannon Information Entropy of Biologically Pertinent Peptides," *Biophysical Journal* 71 (1996): 148–155.

75. Kuppers, "On the Prior Probability," 64.

76. C. de Duve, "Beginnings of Life," 437.

77. R. Stalnaker, *Inquiry* (Cambridge: MIT Press, 1984), 85.

78. This, in fact, happens where adenine and thymine do interact chemically in the complementary base-pairing across the information-bearing axis of the DNA molecule. Along the message-bearing axis, however, there are no chemical bonds or differential bonding affinities that determine sequencing.

79. M. Polanyi, "Life's Irreducible Structure," *Science* 160 (1968): 1308–12, esp. 1309.

80. Yockey, "Self-Organization," 18.

81. H. P. Yockey, "A Calculation of the Probability of Spontaneous Biogenesis by Information Theory," *Journal of Theoretical Biology* 67 (1977): 377–98, esp. 380.

82. M. Eigen, *Steps Toward Life* (Oxford: Oxford University Press, 1992), 12.

83. Meyer, *Signature in the Cell*, 267–268, 284–286, 288–290.

84. R. Shapiro, "Prebiotic Cytosine Synthesis: A Critical Analysis and Implications for the Origin of Life," *Proceedings of the National Academy of Sciences, USA* 96 (1999): 4396–

4401; M. M. Waldrop, "Did Life Really Start Out in an RNA World?" *Science* 246 (1989): 1248–49.

85. R. Shapiro, "Prebiotic Ribose Synthesis: A Critical Analysis," *Origins of Life and Evolution of the Biosphere* 18 (1988): 71–85; Kenyon and Mills, "RNA World."

86. G. F. Joyce, "RNA Evolution and the Origins of Life," *Nature* 338 (1989): 217–24. Yuri I. Wolf and Eugene V. Koonin, "On the Origin of the Translation System and the Genetic Code in the RNA World by means of Natural Selection, Exaptation, and Subfunctionalization," *Biology Direct* 2 (2007): 1–25.

87. A. J. Hager, J. D. Polland Jr., and J. W. Szostak, "Ribozymes: Aiming at RNA Replication and Protein Synthesis," *Chemistry and Biology* 3 (1996): 717–25.

88. Johnston et al, "RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension," *Science* 292 (2001): 1319–25.

89. In addition, for a single-stranded RNA catalyst to self-replicate (the only function that could be selected in a prebiotic environment), it must find another catalytic RNA molecule in close vicinity to function as a template, since a single-stranded RNA cannot function as both enzyme and template. Thus, even if an originally unspecified RNA sequence might later acquire functional significance by chance, it could perform a function only if another RNA molecule—that is, one with a highly specific sequence relative to the original—arose in close vicinity to it. Thus, the attempt to bypass the need for specific sequencing in an original catalytic RNA only shifts the specificity problem elsewhere, namely, to a second and necessarily highly specific RNA sequence. Put differently, in addition to the specificity required to give the first RNA molecule self-replicating capability, a second RNA molecule with an extremely specific sequence—one with essentially the same sequence as the original—would also have to arise. Yet RNA-world theorists do not explain the origin of the requisite specificity in either the original molecule or its twin. Joyce and Orgel have calculated that to have a reasonable chance of finding two identical RNA molecules of a length sufficient to perform enzymatic functions would require an RNA library of some 10^{54} RNA molecules. The mass of such a library vastly exceeds the mass of the earth, suggesting the extreme implausibility of the chance origin of a primitive replicator system. Yet one cannot invoke natural selection to explain the origin of such primitive replicators, since natural selection only ensues once self-replication has arisen. Further, RNA bases, like DNA bases, do not manifest self-organizational bonding affinities that could explain their specific sequencing. In short, the same kind of evidentiary and theoretical problems emerge whether one proposes that genetic information arose first in RNA or DNA molecules. The attempt to leapfrog the sequencing problem by starting with RNA replicators only shifts the problem to the specific sequences that would make such replication possible.

90. Christian de Duve, *Vital Dust: The Origin and Evolution of Life on Earth* (New York: Basic Books, 1995), 23.

91. Matthew W. Powner, Béatrice Gerland, and John D. Sutherland, "Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions," *Nature* 459 (2009), 239–42.

92. Stephen Fletcher, *Times Literary Supplement*, Letters, 3 February, 2010.

93. Tracey A. Lincoln and Gerald F. Joyce, "Self-Sustained Replication of an RNA Enzyme," *Science* 323 (2009), 1229–32.

94. Quastler, *Emergence*, 16.

95. Charles Sanders Peirce, *Collected Papers*, vol. II, edited by Charles Hartshorne and Paul Weiss (Cambridge, MA: Harvard University Press, 1932), 372–78. Abductive reason-

ing was first described by the American philosopher and logician C. S. Peirce. He noted that, unlike inductive reasoning, in which a universal law or principle is established from repeated observations of the same phenomena, and unlike deductive reasoning, in which a particular fact is deduced by applying a general law or rule to another particular fact or case, scientists use abductive reasoning to infer unseen facts, events, or causes in the past from clues or facts in the present. As Peirce himself showed, however, there is a problem with abductive reasoning. Consider the following syllogism:

If it rains, the streets will get wet.

The streets are wet.

Therefore, it rained.

This syllogism affirms a past condition (i.e., that it rained) but it commits a logical fallacy known as *affirming the consequent*. Given that the street is wet (and without additional evidence to decide the matter), one can only conclude that *perhaps* it rained. Why? Because there are many other possible ways by which the street may have gotten wet. Rain may have caused the streets to get wet; a street cleaning machine might have caused them to get wet; or an uncapped fire hydrant might have done so. It can be difficult to infer the past from the present because there are many possible causes of a given effect.

Given that abductive inferences affirm the consequent, Peirce wondered how we can nevertheless frequently make reliable inferences about the past? He noted, for example, that no one doubts the existence of Napoleon. Yet we use abductive reasoning to infer Napoleon's existence. That is, we must infer his past existence from present effects. But despite our dependence on abductive reasoning to make this inference, no sane or educated person would doubt that Napoleon Bonaparte actually lived. How could this be if the problem of affirming the consequent bedevils our attempts to reason abductively? Peirce's answer was revealing: "Though we have not seen the man [Napoleon], yet we cannot explain what we have seen without [the hypothesis of his existence]." For Peirce, a particular abductive hypothesis can be reasonably believed (in practice) if it explains in a way that no other hypotheses do. In other words, an abductive inference is a strong one if it represents the best or the only adequate explanation of the effects in question.

96. Stephen Jay Gould, "Evolution and the Triumph of Homology: or, Why History Matters," *American Scientist* 74 (1986): 61.

97. Thomas Chamberlain, "The Method of Multiple Working Hypotheses," *Science* (old series) 15 (1890): 92–96. Reprinted in *Science* 148 (1965): 754–759.

98. P. Lipton, *Inference to the Best Explanation* (New York: Routledge, 1991).

99. Charles Lyell, *Principles of Geology: Being an Attempt to Explain the Former Changes of the Earth's Surface, by Reference to Causes Now in Operation,* three volumes (London: John Murray, 1830–1833).

100. V. Kavalovski, "The Vera Causa Principle: A Historico-Philosophical Study of a Meta-Theoretical Concept from Newton through Darwin" (PhD dissertation, University of Chicago, 1974), 78–103.

101. E. Sober, *Reconstructing the Past* (Cambridge, MA: MIT Press, 1988), 4–5; M. Scriven, "Causes, Connections, and Conditions in History," in *Philosophical Analysis and History*, ed. W. Dray (New York: Harper and Row, 1966), 238–64, esp. 249–50; Michael Scriven, "Explanation and Prediction in Evolutionary Theory," *Science* 130 (1959): 477–82, especially 480.

102. Meyer, "Of Clues," 96–108.

103. Of course, the phrase "large amounts of specified information" again begs a quantitative question, namely, "How much specified information or complexity would the minimally complex cell have to have before it implied design?" Recall that Dembski has calculated a universal probability bound of $1/10^{150}$ corresponding to the probabilistic/specificational resources of the known universe. Recall further that probability is inversely related to information by a logarithmic function. Thus, the universal small probability bound of $1/10^{150}$ translates into roughly 500 bits of information. Chance alone, therefore, does not constitute a sufficient explanation for the *de novo* origin of any specified sequence or system containing more than 500 bits of (specified) information. Further, since systems characterized by complexity (a lack of redundant order) defy explanation by self-organizational laws and since appeals to prebiotic natural selection presuppose but do not explain the origin of the specified information necessary to a minimally complex self-replicating system, intelligent design best explains the origin of the more than 500 bits of specified information required to produce the first minimally complex living system. Thus, assuming a nonbiological starting point, the *de novo* emergence of 500 or more bits of specified information will reliably indicate design. One gene capable of coding for one protein of average length easily exceeds this threshold.

104. A possible exception to this generalization might occur in biological evolution. If the Darwinian mechanism of natural selection acting on random variation can account for the emergence of all complex life, then a mechanism does exist that can produce large amounts of information—assuming, of course, a large amount of preexisting biological information in a self-replicating living system. Thus, even if one assumes that the selection/variation mechanism can produce all the information required for the macroevolution of complex life from simpler life, that mechanism will not suffice to account for the origin of the information necessary to produce life from nonliving chemicals. As we have seen, appeals to prebiotic natural selection only beg the question of the origin of specified information. Thus, based on experience, we can affirm the following generalization: "for all nonbiological systems, large amounts of specified complexity or information originate only from mental agency, conscious activity, or intelligent design." Strictly speaking, experience may even affirm a less qualified generalization (such as "large amounts of specified complexity invariably originate from an intelligent source"), since the claim that natural selection acting on random mutations can produce large amounts of novel genetic information depends on debatable theoretical arguments and extrapolation from observations of small scale microevolutionary changes that do not themselves manifest large gains in biological information. I have argued elsewhere (see Stephen. C. Meyer, "The Origin of Biological Information and the Higher Taxonomic Categories," in *Proceedings of the Biological Society of Washington* 117 (2004): 213–239; Stephen C. Meyer, *Darwin's Doubt: The Explosive Origin of Animal Life and the Case for Intelligent Design*, San Francisco: HarperOne, 2013), that neither the neo-Darwinian mechanism nor any other current naturalistic mechanism adequately accounts for the origin of the information required to build the novel protein folds and body plans that arise in the Cambrian explosion. In any case, the more qualified empirical generalization (stated above in this note) is sufficient to support the argument presented here, since this chapter seeks only to establish intelligent design as the best explanation for origin of the specified information necessary to the origin of the first life.

105. Meyer, "Of Clues," 77–140.

106. K. Dose, "The Origin of Life: More Questions Than Answers," *Interdisciplinary Science Reviews* 13 (1988): 348–56; Yockey, *Information Theory*, 259–93; Thaxton et al., *Mystery*, 42–172; Thaxton and Bradley, "Information and the Origin," 193–97; Shapiro, *Origins*.

107. Less exotic (and more successful) design detection occurs routinely in both science and industry. Fraud-detection, forensic science, and cryptography all depend on the application of probabilistic or information-theoretic criteria of intelligent design. See Dembski, *Design Inference*, 1–35. Many would admit that we may justifiably infer a past human intelligence operating (within the scope of human history) from an information-rich artifact or event, but only because we already know that human minds exist. But, they argue, since we do not know whether an intelligent agent or agents existed prior to humans, inferring the action of a designing agent that antedates humans cannot be justified, even if we observe an information-rich effect. Note, however, that SETI scientists do not already know whether an extraterrestrial intelligence exists. Yet they assume that the presence of a large amount of specified information (such as the first 100 prime numbers in sequence) would definitively establish the existence of one. Indeed, SETI seeks precisely to establish the existence of other intelligences in an unknown domain. Similarly, anthropologists have often revised their estimates for the beginning of human history or civilization because they discovered information-rich artifacts dating from times that antedate their previous estimates. Most inferences to design establish the existence or activity of a mental agent operating in a time or place where the presence of such agency was previously unknown. Thus, to infer the activity of a designing intelligence from a time prior to the advent of humans on Earth does not have a qualitatively different epistemological status than other design inferences that critics already accept as legitimate. See T. R. McDonough, *The Search for Extraterrestrial Intelligence: Listening for Life in the Cosmos* (New York: Wiley, 1987).

108. P. Lipton, *Inference to the Best Explanation*, 32–88.

109. *Ibid*.; S. C. Meyer, "The Scientific Status of Intelligent Design: The Methodological Equivalence of Naturalistic and Non-Naturalistic Origins Theories," in *Science and Evidence for Design in the Universe*, The Proceedings of the Wethersfield Institute, vol. 9 (San Francisco: Ignatius Press, 2000), 151–212; Meyer, "The Demarcation of Science and Religion," in *The History of Science and Religion in the Western Tradition: An Encyclopedia*, ed. G. B. Ferngren (New York: Garland, 2000), 17–23; E. Sober, *The Philosophy of Biology* (San Francisco: Westview Press, 1993); Meyer, "Of Clues," 77–140.

110. Quastler, *Emergence*, 16.

111. Oparin, *Origin of Life*, 28; M. Rothman, *The Science Gap* (Buffalo, NY: Prometheus, 1992), 65–92; K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Routledge and Kegan Paul, 1962), 35–37.

112. Meyer, "Of Clues," 77–140; Sober, *Reconstructing the Past*, 4–5; de Duve, "Beginnings of Life," 249–50.

113. R. Harré and E. H. Madden, *Causal Powers* (London: Basil Blackwell, 1975).

114. M. Ruse, "McLean v. Arkansas: Witness Testimony Sheet," in *But Is It Science?* ed. M. Ruse (Amherst, NY: Prometheus Books, 1988), 103; Meyer, "Scientific Status"; Meyer, "Demarcation."

115. Meyer, "Scientific Status"; Meyer, "Demarcation"; L. Laudan, "The Demise of the Demarcation Problem," in Ruse, *But Is It Science?* 337–50; L. Laudan, "Science at the Bar—Causes for Concern," in Ruse, *But Is It Science?* 351–55; A. Plantinga, "Methodological Naturalism?" *Origins and Design* 18, no. 1 (1986): 18–26; A. Plantinga, "Methodological Naturalism?" *Origins and Design* 18, no. 2 (1986): 22–34.

116. Bridgman, *Reflections of a Physicist*, 2nd ed. (New York: Philosophical Library, 1955), 535.