Deep salty RNA world

"Deep polymers" easily-testable (and already confirmed) hypothesis about the origin of life; and how to measure a lifeform's information content and estimate genesis rates

Warren D. Smith, warren.wds@gmail.com. This is v3, Nov.2025. http://vixra.org/abs/2510.0088

Short Abstract:

I. HYPOTHESIS: Aqueous RNA becomes stable against hydrolysis under high hydrostatic pressure, 1-4000 atm. Weaker also-sufficient hypothesis: it works for some exponentially-large *subset* of RNA-*like* molecules; also *salinity* has most of the same beneficial effects as *pressure*, and both can work together. This and related hypotheses can resolve the gaping holes in L.E.Orgel's "RNA world" paradigm about life's origin. It deserves direct experimental tests, but even without them we present 4 independent lines of evidence for it apparently yielding confidence>99.9999%.

II. I explain how to measure the "true vital information content" of a lifeform. It is hugely wrong to claim an N-base pair DNA genome has information content 2N bits. The simplest estimate instead should be **$\le -N\log_2(1-3\varepsilon/4) \sim 3\varepsilon N/\ln 16 \approx 1.082\varepsilon N$** bits, where ε is the chance that a random single-base mutation would kill the organism and experimentally for bacteria ε≈1%, and the "≤" is because of further refinements I'll describe, which can reduce the bit-count further. I explain how to get more precise estimates with more experimental and computational work. Then calculating numbers shows that the genesis of life was not necessarily absurdly improbable; predicted and observed genesis rates plausibly can roughly agree.

Together, ideas I (if correct) and II overcome the two top obstacles blocking understanding how genesis could have happened, and suggest many experiments. Unfortunately version 3 adds a new section "(Pessimistic) Post Mortem" which discusses another class experiments (on "apparent molar volumes") whose results suggest I is false.

Long Abstract:

I. HYPOTHESIS: Aqueous RNA becomes stable against hydrolysis under high hydrostatic pressure (HHP). I have in mind the pressure range 1-4000 atm, corresponding to up to 10 km below seafloor. Weaker also-sufficient hypothesis: it works for some exponentially-large *subset* of RNA-*like* molecules. (Also, as the paper explains, *salinity* has most of the same beneficial effects as *pressure*; and both can work together.) If correct, this hypothesis resolves gaping holes a, b, and c in L.E.Orgel's "RNA world" paradigm about life's origin: (a) unlike today's robust highly-evolved bacteria which in good conditions can double in only 10 minutes, the earliest life presumably was extremely "delicate" and "inefficient," with doubling times likely 10-2000 kyears; (b) RNA is thermodynamically unstable to hydrolysis in water, destroying any nascent "RNA life" on far faster time scales than that; (c) Although Miller-Urey style experiments can abiotically generate many basic biochemical *monomers* such as amino acids, sugars, and nucleobases, they have not been observed to *polymerize* them into proteins, polysaccharides, and RNA, all of which are thermodynamically unstable in water. They also have not even been able to *dimerize* them, or to dimerize nucleo-bases such as adenine and cytosine with pentoses such as ribose (or deoxyribose) to create nucleo*s*ides such as adenosine (or deoxyadenosine) and cytidine, nor to dimerize *those* by adding a phosphate group to create nucleo*t*ides such as adenosine monophosphate (AMP). All those dimerizations and polymerizations happen by removal of water, and are reversed by hydrolysis, i.e. adding water.

The *only* environment on Earth stable enough to permit survival of delicate life for long enough for (a), is deep underground (or beneath ocean floor). And since those are *exactly* the places with HHP (often also with a lot of salt in the water), with the aid of the Hypothesis they overcome (b) too. If more polymers than RNA (and more dimers) become thermodynamically stabler than monomers at high enough pressures, then (c) also is overcome.

This all remains capable of compatibility with some versions of G.Wächtershäuser and A.G.Cairns-Smith's ideas about minerals and "clay templating," "pre-RNA-world" notions including "PAH world," and P.L.Luisi, D.W.Deamer and

C.M.Dobson about "lipid vesicle self-assembly," and with the abiotic generation of basic biochemical life-ingredients as in Miller-Urey experiments. However, if correct, our hypothesis indicates that the 70+ years of Milley-Urey experimentation have been misleading, because they did not include high pressure components. That is why I (primarily a mathematician) have fruitlessly urged approximately all the biochemists I ever met for 30 years to do such experiments. I've now given up on them and instead now simply publish my hypothesis, reasoning supporting it, and outline how to do the (sometimes fairly easy!) experiments.

Why is the Hypothesis likely? In 1954, in one of the earliest and most spectacular successes of "anthropic reasoning," Fred Hoyle made the first-ever prediction of an excited state of any atomic nucleus: carbon-12 had to contain an excited nuclear state ≈7.7 MeV above ground. Ridiculously, this prediction was based solely on the fact of Hoyle's own existence – which he couldn't see any reasonable way to explain without his hypothesis. Result: 3 years later experiments by Cook et al proved Hoyle correct. I predict my chemical Hypothesis for (A) that same reason, albeit unlike Hoyle I also have non-anthropic reasons: (B) crude estimates show RNA, DNA, peptide, and glycogen hydrolysis energies should have the *same order of magnitude* as $\Delta E=P\Delta V$ energy adjustments arising from P≈4000 atm; (C) both general and specific experimental evidence, and (D) simple mathematical models and theorems, indicate that the *sign* of the effect should be correct, i.e. for all of these, HHP should thermodynamically favor dimerization, polymerization, and biomolecule stability while disfavoring hydrolysis. (E) Statistical-miracle clues inside the genetic code suggest deep/hot origin. (F) phylogenetic back-deducers believe today's lifeform most-resembling the last common ancestor is a hot deep methanogen halophile that cannot tolerate O_2. (G) Observational proofs I'll discuss show stability of proteins and either RNA or DNA (or both) for at least 1000-year timespans in water 2 miles underground.

II. I explain for the first time how to measure the "true vital information content" of a lifeform. It is hugely wrong to claim an N-base pair DNA genome has information content 2N bits. The simplest estimate instead should be **$\leq -N\log_2(1-3\varepsilon/4) \sim 3\varepsilon N/\ln 16 \approx 1.082\varepsilon N$** bits, where ε is the chance that a random point mutation would kill the organism and experimentally for bacteria ε≈1%, and the "≤" is because of further refinements I'll describe, which can reduce the bit-count further. I explain how to get more precise estimates with more experimental and computational work. Then calculating numbers shows that the genesis of life was *not* necessarily absurdly improbable; predicted and observed genesis rates plausibly can roughly *agree*.

Together, ideas I (if correct) and II overcome the two top obstacles blocking understanding how genesis could have happened (indeed seem capable of overcoming all 5 of Benner 2014's "paradoxes"), and suggest many experiments. Unfortunately version 3 adds a new section "(Pessimistic) Post Mortem" which discusses another class of "apparent molar volume" experiments whose results suggest I is false and severely diminish my confidence in it.

Table of contents

Leisurely reading plan: start at beginning and read to end. §6-7 discuss pressure (I) while §4 discusses valid versus invalid information-content calculations (II) and compares them versus observation-based estimates of genesis rates. Impatient readers: read summary first; it has hyperlinks back to things it summarizes.

Definition of "life"

For us **"life"** shall mean: an entity which uses *energy and materials* found in its environment, and a large amount of internal "genetic" *information*, to build *more copies* of itself; and those copies can be inexact ("mutations"), thus enabling "Darwinian *evolution*" to "improve" that life every generation. I shall not add to this definition (but some people want to): the further demand that life *self-repair* many common kinds of damage.

The clearest **examples** are cyanobacteria capable of survival alone in aqueous environments on today's or the primordial Earth, or, say, ferns on suitable land. Note that this definition of "alive" depends heavily on the **environment** – algae would be dead on the moon. Is a virus "alive"? Yes, *provided* its "environment" consists of "suitable host cells." But alone on the primordial Earth, viruses would be dead. Are "prions" alive? Arguably yes, but only assuming even more special "environments"; and prions can be criticized as not being mutable. Also, neither viruses nor prions can self-repair. Call life which can survive in natural environments in the *absence* of all other life, "**standalone**" life. Prions and viruses are not standalone. Although monkeys are normally considered alive, they could not survive on Earth if all its other life were obliterated, so they too are not standalone. Is a hammer "alive"? In the present-day world, arguably yes, in the sense that it inspires humans in its "environment" to use their industrial infrastructure to make more hammers. However, if all humans were killed, hammers would be dead. Also, hammers (and perhaps prions) could be criticized as not containing their "genetic" information *internally*.

"Machine life." If a collection of smart robots cooperated to mine materials and design and build more robots, then I would consider that collection both alive and standalone; and indeed I suspect this completely different form of "life" will be what matters in the future galaxy, while humans who (roughly speaking) cannot survive off Earth, will become unimportant. (Machine life will be capable of far longer lifespan and far greater physical and mental capabilities, e.g. survival in outer space and "Lamarckian" evolution, than old-style life.)

"Szostak's dream molecule (SDM)." Jack W. Szostak has long tried (so far unsuccessfully, but with partial success) to create a "lifeform" consisting of a single molecule of RNA, with the property that it can catalyze polymerization to create new copies of itself. If he succeeded, his molecule would qualify as "alive" in an aqueous "environment" continually artificially replenished with adenosine triphosphate (ATP), UTP, GTP, and CTP supplies as polymerase feedstocks. Since a single molecule, this would be approximately the simplest possible lifeform. However, in any realistic (not artificially replenished) environment, an SDM would be dead. I believe SDMs must exist, and speculate the reason Szostak has been unable to discover one is simply that:

i. very few RNA molecules work;
ii. SDMs likely also would replicate plenty of other "parasite" RNAs otherwise incapable of reproducing;
iii. the simplest SDMs likely have doubling times ≥1 year since the SDM's catalytic rate-enhancement effect is small. That would prevent Szostak from observing the SDM working, and even an ultra-patient, ultra-long-lived improved-Szostak still would be incapable of seeing the SDM working because it would self-destruct via RNA hydrolysis before that year ended.

J.H.Conway's mathematical game, which I here will call "**conlife**," is the following three "laws of physics" in an infinite 2D square-grid mathematical universe with discrete integer time:

A. Each grid point has two possible states: 0 and 1; and eight nearest-neighbor gridpoints.
B. Any 1 with fewer than two or more than three 1-neighbors converts to 0 next timestep, but if it has either two or three 1-neighbors, stays 1.
C. Any 0 with exactly three live neighbors becomes 1 next timestep, otherwise stays 0.

Conlife generates recreational time-evolving pixel patterns on a black/white computer screen. A community of enthusiasts investigated it for years, devising all sorts of local "gadgets" in the conlife universe enjoying interesting properties, such as "oscillators," "still patterns," "gliders and spaceships" which move, "logic gates" which input glider-carried "information signals" then output others, "eaters," "glider guns," recirculating "memories," etc. Using those gadgets, Conway proved mathematically that the conlife universe supports universal computers (i.e. "Turing machines" can be built inside it). Consequently (Conway showed) it is a Turing-undecidable question whether a given life-pattern will eventually "vanish," i.e. evolve into the all-0 state. Conway also showed how to design "assemblers" universal-enough so that such a computer could be programmed to "build" copies of itself located at arbitrary

locations arbitrarily far away. (Conway's proof was not fully explicit, but later conlifers constructed fully explicit versions.) That would qualify as "alive" under my definition, within those laws of physics, except that the conlife laws of "physics" does not include any conserved notions of "matter" and "energy"; and since everything is wholy deterministic, arguably no random "mutations" can occur. But Conway's computer could be programmed to employ arbitrarily good *pseudo*random bit generators to produce *pseudo*random "mutations," so the latter objection does not bother me. Also, if it bothers you, then we could modify Conway's underlying conlife laws of physics to cause random gridpoint 0↔1 state changes with some tiny probability p. Provided p^{-1} was, say, at least 10× greater than the total number of gridpoint state changes during a conlife life-replication, life could then still occur and prolifierate, at least for some substantial number of doublings. I.e: "life" can exist inside mathematical "environments" implementable inside computers. The reason to care about that is: we can *prove theorems* about this kind of life, reaching a level of understanding of it not attainable for less-simple kinds of life.

However, I must criticize the following aspect of Conway's "proof of life": his replicants work in a pristine conlife universe initially entirely filled with 0s (aside from the initial replicator). What if the environment is polluted with random crud? What if other replicators are out there to "interfere with" or "compete" with ours? And those phenomena presumably eventually would happen with randomly-mutating replicators. In that case, it remains unclear whether any Conway "life" can exist that can repair and protect itself to keep surviving. Maybe somehow Darwinian evolution could improve the lifeforms fast enough to overcome the obstacles and attacks they would eventually face. Or, maybe that is not possible and they would ultimately be doomed. The answer is not known, and it isn't clear to me even how to formulate the question.

Given that (1) lifeforms by repeatedly doubling, grow *exponentially*, but information in conlife cannot travel faster than the "speed of light" (one gridpoint nearest-neighbor L^{∞}-distance per timestep), but (2) the area in the Euclidean plane enclosed by an expanding "lightwave" grows only *quadratically* with time, we conclude that eventually, any life that doubles after a constant number of timesteps, must experience "overcrowding" from a "population explosion"; i.e. exponential growth cannot continue forever. (Permanent exponential growth would not necessarily be forbidden in the *non*Euclidian "hyperbolic plane," however.) But I claim that it is possible in conlife to overcome that limitation by adding a binary counter to each Turing machine to keep track of time; and to make each lifeform double not after a *constant* time interval, but rather to keep increasing those time-intervals unboundedly; and make each lifeform locate its "child" not a constant distance away, but rather a distance that grows unboundedly with time sufficiently quickly (although much more slowly than the interdoubling times grow); and finally to make lifeforms "commit suicide" and vanish once their binary time-counters overflow. (Later lifeforms would have more bits in their binary counters, permitting longer lifespans.) In this way an infinity of conlife creatures could be made to arise from one initial one, with their lifespans known ahead of time and increasing unboundedly.

Back-deductions from today's lifeforms; nature of early earth

Today's lifeforms are based almost wholy on DNA to encode their genetic information, proteins to catalyze metabolic reactions, lipids and polysaccharides for their cell membranes, sugars and ATP for energy, and RNA for information transfer from DNA to proteins. (Call that the "**orthodox plan**.") The key players are macromolecules, and almost everything is made almost wholy of the six **atomic elements** H,O,C,N (96-98% of the mass), P (1%), and S (0.2%), all swimming in a solvent bath consisting mostly of liquid water. (Even in arid deserts, all active cellular life is mostly water by mass, although some spores and dormant lifeforms can survive with much-reduced water fractions.) Small amounts of the three elements K, Mg, Zn also all seem essential. The following 8 more elements, while perhaps not essential for *all* microbial life, are essential for *most* microbes in all three of {Archaea, Bacteria, Eukaryotes}: Ca, Mo, Mn, Fe, Co, Ni, Cu, Se. Apparently it is possible for a substantial number of microbes to live even if deprived of all further elements, although it is infeasible to prove that with complete certainty by removing *all* atoms of some element from each cell, since those purity levels are not technically attainable. But additional elements play useful roles in many organisms (e.g. humans certainly need Na, Cl, Br, and I, while plants need B).

The following is a list of **elements** found in the solar system in decreasing order of atom-count: H, He, O, C, N, Ne, Mg, Si, Fe, S, Ar, Al, Ca, Na, Ni, Cr, P, Mn, Cl, K, Ti, Co, Zn, F, Cu, V, Ge, Se,... The rarest element essential for most microbial life is Mo, whose abundance is about 10^{-10} of the count of H atoms. The elements in the human body, ordered by decreasing atom-count: H, O, C, N, Ca, P, S, Na, K, Cl, Mg, ... In the Earth's crust, the order is O, Si, Al, Fe, Ca, Na, Mg, K, Ti, H, P, C, N, Sr, ... The most common elements in the Earth's crust *not* essential for most

microbial life (besides the inert gases) are Si, Al, Na, Ti.

The present paper cares about "the **primordial Earth**" as the "environment"; and wants to understand Earth's *first* lifeform.

As far as has been back-inferred by examining DNA sequences and biochemical idiosyncrasies (Carl R. Woese, Lynn Margulis, W.Ford Doolittle, Madeline C. Weiss) the oldest lifeform that can be dimly perceived is some common ancestor of both the bacteria and archaea, which probably resembled an archaean more than a bacterium. Within today's archaeans, it most resembled high-temperature methanogens, and within today's bacteria most-resembled clostridia. To review that: originally the mainstream view was that there were *two* kingdoms of microbes: Eukaryotes and Prokaryotes. Euks have much more complicated cell structure, containing "cells within cells": e.g. nuclei and mitochondria. Proks are much simpler with only one (outer) cell membrane, and since simpler were presumed older. Carl Woese decided to classify microbes *not* by their naive visual appearance, but rather by assessing similarities and dissimilarities between crucial parts of their DNA. Very wisely, Woese selected the "16S RNA component of ribosomes" as the part of the genomic sequence to use for this purpose. Every known cell has ribosomes. Every known ribosome contains a 16S (or, for some Euks, 18S) RNA subunit. Due to its importance and role, 16S presumably changes very slowly with evolutionary time – slower than almost everything else. Note that genes like 16S that are present in *all* known life are very **rare**: Charlebois & Doolittle 2004 claim there are fewer than 50 (they name 34 common to the 147 prokaryotic genomes they surveyed), and probably only about 30. Every one encodes a ribosome component, helps synthesize tRNA, or is transcription related, with the sole exception of O-sialoglycoprotein endopeptidase which presumably plays an important role in manipulating cell walls. If Orgel's "RNA world" indeed existed before "orthodox plan" life supplanted and eliminated it, then the best hope for understanding what RNA-based life was like, presumably are today's highly-conserved "RNA enzymes" like 16S. These (and perhaps also "ribo-switches") hopefully are "surviving relics" of RNA world, still present in every cell today. The sequence lengths (≈1550) for 16S RNAs happen to be long enough to be useful for this purpose, yet still small enough for it to be technologically feasible (even back in Woese's era) to sequence them.

Cells reproducing solely by fission genuinely would form a "binary tree" whose "root node" would be the first live cell, and all other nodes are its descendant cells. Call that the "**one true tree.**" But unfortunately for simplicity, rare "horizontal gene transfers" (**HGT**) can also occur, and when they do the lifeforms *stop* forming a "tree" (mathematically defined as a "connected graph with no cycles"). Rather, we have a "blue" tree defined by cell fissions (the one true tree), *plus* extra "red" cycle-forming edges representing HGTs. Helpfully for Woese's goal of finding the one true tree, something like the 16S RNA among the *least* likely genes to spread via HGT, and therefore perhaps the single *best* gene to use to try to reconstruct the one true tree. There are many single genes encoding single proteins which also can be very crucial, so much that most known living cells have a version of that gene. However, most such genes could easily spread by HGT – if an organism luckily acquires an improved gene by HGT, it will happily reproduce from then on, dominating its ecological niche due to its improved fitness. To make clear the magnitude of the problem, Weiss, Preiner et al 2018 estimate that "**97%** of the [genes] present in bacteria and archaea apparently underwent some transdomain HGT." In contrast, a component like **16S** of a multi-macromolecule machine like a ribosome, would find it very ***difficult* to spread via HGT**, because, if some other organism had an improved ribosome containing an improved 16S, it probably would be unable to transfer its 16S gene to you because it probably would be incompatible, as a machine part, with the rest of your (unaltered) ribosome. The only way it could transfer would be to transfer a *full many-gene package* encoding *all* components of its ribosome, to you, while hoping that somehow your old ribosome machine parts would not try to inter-assemble with the new parts, ruining them both. Such a huge coordinated HGT is asking much more, hence would be expected to be a much rarer lucky event. So hopefully 16S is a member of the 3%.

Mechanics – sensible assessments – in contrast to the utter garbage Fox, Woese et al actually used: Woese's original assessments of "**sequence similarity**" were completely subjective, but still (in his mind) good enough to deduce the existence of the archaea, a whole new "third kingdom" of life. Hopefully some mathematically-defined metric would be better. It is algorithmically possible to compute the "edit distance" S between two arbitrary input strings of N and M characters, N≤M (where in the DNA case, all strings are made of letters from the 4-character alphabet {A,C,T,G}) in polynomial time, namely O(NM) steps. Improved algorithms are known that do it in O(SN) and $O(S^2+M)$ steps. Here changing a single letter ("point mutation") costs 1, while inserting or deleting a letter costs some amount X>0 (most simply X=1, giving "Levenshtein distance"); and the "edit distance" is the least summed operation-cost among all possible sequences of edit-operations converting one string into the other. We also could charge

differing costs for different point-mutations and different insertions and deletions. For biochemical realism, it would be better to add "cut-and-paste" (aka "substring move") to the suite of allowed editing operations: delete an arbitary-length substring and optionally re-insert it someplace else, costing Y>0. We also could consider "*copy*-and-paste" and substring "flipping" aka "reversal with complementation" (where "complementation" means A↔T and G↔C). Unfortunately, edit-distance with substring-moves is no longer a polynomial-time task (unless P=NP) since proven NP-complete (Shapira & Storer 2002). However, if we demand that all "wide operations" occur on *non-overlapping* substrings, then it turns out that edit distance is computable in polynomial time even with all those extra operations added (Alicioglu & Alkan 2024). That's great, but unfortunately all the computations by the likes of Woese and Fox did not use these better kinds of distance – nor even the less-good kinds – and therefore optimally *all their work on the "tree of life" should be* ***completely redone****.*

Instead, atrociously, the Fox et al 1980 paper based everything on a mysterious "association coefficient S_{AB}" between two genomes A and B, *without even defining* it! (Incredibly, subsequent papers by Woese were even *more* opaque about his methodology!) As a mathematician speaking to biochemists, let me give you some free advice: If you plan to rest your entire life's work upon a mathematical definition, then *provide that definition* to your reader, stated in a mathematically unambiguous clear way. They failed that test.

Apparently the closest approach to a definition is found on pages 45 & 52 of Fox, Pechman, Woese 1977. But it would be instantly rejected as completely inadequate unreadable meaningless garbage by every mathematics journal I ever saw (also probably get an "F" if submitted as a homework assignment in a freshman math class). Nevertheless I tried to understand what (I think) it was trying to say. The result was

Counter-examples: To keep this simple and short, I'll treat this for a 2-symbol alphabet {0,1}. Consider the 69-bit genome

```
G = 10010100110000001011100100001110101011001111011011111100011010000 10010.
```

You might imagine (or hope) that no other sequence could be as similar to G, as G is to itself ("**distance zero**"); and also that the 69-bit genome *least* similar to G would be its bitwise complement

```
Ḡ = 01101011001111110100011011110001010100110000100100000011100101111 01101.
```

But, as far as I can tell, Fox, Pechman, Woese 1977 violently disagree. They contend that the number of 69-bit genomes at "distance zero" to G is ≥4294967296=2^{32}, and $\overline{G}$ is one of them! Three (random) examples, which everybody besides FP&W would claim hugely differ from both G and $\overline{G}$, are:

```
01110101011110000001000111001001010000110001011010011011001111111 01110,
00101010000001000110011011111100001110010011110110001011010111101 00101,
01000000110000100111111000111010101111011100100010100101100110111 01000.
```

(For combinatoricists: all these are order-6 noncyclic De Bruijn sequences, which I got by selecting cyclic 64-bit circular De Bruijns uniformly at random from the 2^{32} of them, then "unwrapping" each of them by 5 bits, shown after a space, to create a noncyclic one 69 bits long. Any additional 69-bit De Bruijns created by some other method would also provide more counterexamples, beyond my 2^{32}.) Their S_{AB} is not, and cannot be converted into, a "metric," e.g. every nonconstant function of S_{AB} disobeys either the demand dist(G,G)=0, the demand dist(G,H)≥0 for all H≠G, and/or the triangle inequality. It therefore also has no connection to |log likelihood| of genome A↔B interconversion, given that the latter should *be* a metric. And had I instead employed the full 4-character alphabet {A,C,T,G}, the number of, e.g, 1028-character noncircular genomes at "distance zero" to any one particular among a huge set of such beasts, using FP&W's 5-mer-based similarity measure (I have constructed examples, but will spare the reader) is ≥24^{256}, far exceeding the number of particles in the observable universe. This criticism holds for *every* S_{AB} definition based solely on the counts of the 5-long substrings within the genome. This is simply absurd. And

Question: Quantitatively, how **confident** should we be about deductions of the "tree of life" from such methods? Fox et al 1980 stated no confidence numbers whatsoever! Unfortunately, everybody kowtowed to the "Great And Brilliant Pioneers" Fox, Woese et al from then on (actual quote: "the most important paper in biology"), never questioning their work. Had I been their referee, I would have rejected it until they redid it with sensible mathematics. Despite this sorry

story, it may well be (and I hope) that Fox et al's biological conclusions are valid, since I have no doubt that they fooled around for a while devising ad hoc "similarity" measures to try to find one that empirically seemed to "work well" (Woese probably stared at 16S RNA sequences longer than anybody else, acquiring great intuition about them). But the bottom line is: their work was atrocious.

We now return to the task of explaining an actually sensible method. After you choose X and your basic mutation-costs and then compute all **distances** between all pairs of lifeforms, we then may use simple known polynomial-time algorithms to compute their minimum spanning tree (**MST**). Indeed, with L lifeforms, and hence (L-1)L/2 pairwise distances, "Prim's algorithm" computes the MST in $O(L^2)$ steps. The MST is the minimum-summed-distance way to interconnect all the lifeforms via a tree made of lifeform-pair edges; and for *generic real* distances this tree by theorem is *unique*.

Then each MST edge (A,B) models a *path* in the actual tree-of-cell-fissions, which goes *both* forward and backward in time. That is, known lifeform A changes, going "backward in time," until it reaches the unknown common ancestor of A and B; then further changes ("forward in time") to convert that ancestor into known lifeform B. The MST hopefully represents the set of paths with *shortest* length-sum that each are subgraphs of the one true tree and suffice to connect all your lifeforms. The *second*-least-costly spanning tree (which I'll call M_2ST) may also be found in $O(L^2)$ steps, by finding the shortest tree-edge in the MST-path from A to B (for all A,B not directly connected by an MST edge), and considering removing its longest edge and then inserting the direct AB edge. The least-cost such nontrivial edge-replacement converts MST→M_2ST. More generally, for any desired K with 1≤K≤the number of spanning trees, it also is algorithmically possible to find the K least-cost spanning trees in $O(KL^2 \log L)$ steps (Gabow 1977). We then may assess "confidence" our MST is the "right" one, by computing the cost-differences between MST and the M_KST, converting those cost differences to "relative likelihood factors," and normalizing the resulting probability distribution on trees. My point: Each genome-change, e.g. point-mutation, is an unlikely event. The more such events are needed to explain your set of lifeforms, the more unlikely that explanation becomes. This can be made quantitative. To best accomplish that, the basic operation "costs" should be chosen to be proportional to |log likelihood| values. Then the MST will correspond to the "most likely explanation," and M_2ST to the second-most likely explanation. One also could assess confidences that any particular (sub-forest) part of the MST is correct, by finding the cost of the least-cost spanning tree not containing it. All this is doable in polynomial time, in fact $\tilde{O}(L^2)$.

A flaw in that MST-based methodology is that inter-lifeform paths could *overlap*. For example, suppose the MST of three known lifeforms A, B, C consists of the two edges A-B and A-C. In reality there is some fourth unknown lifeform D such that the one true tree involves nonoverlapping paths A-D, B-D, and C-D. The MST's paths A-B and A-C are really A-D-B and A-D-C which overlap along the sub-path A-D, which therefore is counted twice. For this reason, the MST does not necessarily really represent the "most parsimonious" (i.e. max-likelihood) explanation for your lifeforms. To find the truly max-likelihood explanation, we want the min-cost "**Steiner tree**," i.e: extra unknown lifeforms (such as D) are added, then the least-cost MST is found among the thus-expanded set of lifeforms, and with minimization not only over all possible trees, but also over all possible sets of extra lifeforms as extra nodes in those trees. The trouble with min-cost Steiner trees is that finding them is known to be NP-hard, i.e. algorithmically infeasible. A known theorem: The MST can never cost more than a factor 2 more than the min-cost Steiner tree (in any metric space), so it constitutes a "factor≤2 approximation algorithm." Better approximation guarantees than "2" are achievable via slower polynomial-time algorithms, for example factor≤11/6≈1.84 is achievable by Zelikovsky's $O(L^3)$-time "greedy" algorithm.

New lemma: Given *three* "genome" character-strings A,B,C, there is a polynomial-time algorithm to compute a string D minimizing dist(D,A)+dist(D,B)+dist(D,C).

> **Proof sketch:** First, to convince yourself (as I'd already claimed) that edit-distance algorithms running in polynomial time are possible for all the distance-definition-flavors I mentioned: assume all distances between all strict *sub*strings of the M-long and N-long strings have been previously computed and tabulated [a polynomial(N,M) amount of information], then realize that from that tabulated information we can compute the distance between the full strings in polynomial(N,M) additional steps. Second, to solve my new three-string problem, the same kind of "dynamic programming" approach is possible: solve the problem for all substrings, then use that table of data to solve the problem for the full strings (by

considering all allowed possible ways to glue together substring solutions; there are only a polynomially bounded number of gluing choices to worry about). **Q.E.D.**

Again: this new lemma suggests all prior tree-of-life deductions should be completely redone; and this lemma could actually lead to a deduced guess for LUCA's *full genome.*

Horizontal gene transfers (HGT) are different than MST edges. First, they always are non-tree edges that create cycles. Second, HGTs always have known time-direction arrows, whereas MST edges are (virtually always) time-*bi*directional and their "predominant" time-direction arrow (if and when they can be said to have one) is often unknown and is not deducible from lifeform-pair distances alone. (It perhaps sometimes could be deducible from some very nontrivial human assessment that A is an "improved" version of B and therefore came later, or that B is "more adapted to earlier-earth conditions" therefore came earlier.) The more HGTs happen, the more confused any attempt to determine the "one true tree" of cell-fissions becomes. The only way I can think of to avoid that confusion is to try to deduce *both* the one true tree *and* the set of HGTs simultaneously, with maximum combined likelihood. That is probably algorithmically very difficult and nobody has tried; but since this is the correct goal, it should at least be mentioned.

Result of computations by Fox, Woese, et al: Woese and his student George E. Fox applied their absurdly flawed methodology to their collection of 16S sequences to discover there are actually **three kingdoms** of microbes: Eukaryotes, and Bacteria and Archaea now being two kingdoms replacing the one prokaryote kingdom. Fox, Woese et al were able to produce "trees of life" connecting different members of these families, allegedly in the correct way. Probably the **Archaea came first**, before the Bacteria, because they include the "methanogens" which happily grow in a CO_2 and H_2 atmosphere – excreting CH_4 – i.e. what some people think was the primordial Earth's atmosphere (or the H_2 getting produced by hot lava reducing, or radioactivity splitting, H_2O). The fact that many methanogens are killed if they contact O_2 also suggests their early-earth origins. The Archaea have "funnier lipids" in their cell membranes, and tend to be capable of survival in a more-diverse range of more-extreme conditions than the other kingdoms, such as high salt, high temperatures, methanogens, sulfur diet, acid lovers, alkaliphiles, and microbes which grow deep underground at high pressure. That greater diversity also is indicative of greater age, for the same reason that Africa, the original source of *Homo Sapiens*, features greater human genetic diversity than all the other continents combined (and in general, the most genetic diversity for almost every species, is found in the geographic location where that species first originated). The third main set of microbes, the Eukaryotes (which later developed into *multi*cellular life), arose as a bacteria-archaea obligate-symbiont, e.g. today's mitochrondria inside eukaryotes arose as bacteria that lived *inside* archeans. Perhaps these bacteria originally were parasites, but after losing some genes became incapable of external survival while the outer cells also grew dependent on their internal symbionts. This "capture" hypothesis was proven by sequencing means. Similarly chloroplasts arose as former bacteria which also became obligate symbionts and now are a permanent internal component of plant cells.

Inadequate look at much work by others following up on Fox & Woese 1980. Ciccarelli, Doerks, et al (2006) started with the *whole genome* sequences of 191 microbes. They began by surgically *removing* everything involved in any HGT their computer detected, from all genomes, resulting in 31 genes common to all their organisms (a tiny subset of their original genomes) – apparently all never involved in any HGT. They claimed that without such an HGT detect/remove first step, tree-of-life deduction attempts are garbage. It then turned out that these 31 genes *all* were ribosome genes. That confirms Woese's (and my) intuitions about the especial comparative HGT-immunity of ribosome genes and about their consequent especial usefulness for tree-of-life deduction purposes – only now it is not mere "intuition"; it is "proven."

Prof. Erik Winfree, in email, told me that I was being obnoxious by criticizing Fox & Woese 1980, because that was merely a "quick and dirty" attempt which during the following 45 years was later redone "ad nauseam" by many others who used much better techniques to deduce trees-of-life. (I agree about "dirty," but not "quick"!) What were those numerous much better redos? Unfortunately, I am not confident that I know the answer. But I have looked, and consequently have some further critical remarks – now not about Fox & Woese 1980, but in fact about this entire area. It is diseased. And I will now explain why. Ciccarelli et al used their 31-gene set plus some (poor) kind of distance-notion to deduce a hopefully-good phylogenetic tree for their 191 organisms (pictured in their paper using tiny font); then claimed their tree resolved a number of puzzles. Unfortunately, there are absolutely no confidence numbers associated with any part of their tree. A more recent much huger computation (3840 hours on a supercomputer!) is Hug, Baker, et al 2016. Although they end up with beautiful-looking graphic depicting their tree of

life, there again are *no confidence numbers anywhere in the paper*. What benefit, if any, was obtained by such an enormous expenditure of computer time? Nobody knows. Without confidence numbers for any conclusion, none of these claims about the "tree of life" are assessible. And this unfortunately is typical. So far in my (probably quite partial) tour though this literature, I don't see confidence numbers attached to any claim by anybody. If two papers yield differing conclusions, who is right? Nobody knows. I prefer *low-order polynomial-time algorithms* that come with *theorems* about the properties of their output, so that we can use a normal computer (so that a normal person without millions of dollars in funding can reproduce the results without expending such huge computer resources) and so that we can have at least *some* certainty about what the output actually tells us.

Theorems: The minimum spanning tree is the absolute minimum-cost among all possible spanning trees, it generically is unique, and it is findable very efficiently, and you then can also find the K least-cost spanning trees and use them to quantify confidences. The MST is not the right model, but it by theorem is within a factor 2, and within that model we have great understanding of what we are doing.

Their huge computation: these papers gave no theorem whatsoever about their output-tree being "optimal" in any way, nor any confidence of anything. And if you *are* going to expend huge money to employ exponential-time algorithms, then you should do so in *selected* places and ways that yield *quantifiable* confidence-payoffs. They don't. It is feasible to find the unique absolutely best Steiner tree among a *small-enough* number of lifeforms (say 10), and also all near-best-tree rivals, by *exhaustive search.* If you did that, then you'd be able to find out what you know and what you do not know with numerical confidence claims attached, and generally know what you are doing. They don't. "Branch and bound" algorithmic approaches also might be able to output confidence statements as side effects of the "bounds"; and another idea that could be helpful is to construct *many* trees of life based on *disjoint* sets of genes; whenever such trees disagree, then that disagreement is some fact about the Tree that we are insufficiently confident about. That has rarely been done, e.g. Ciccarelli et al did not despite finding the key 31-gene set that they could have used to do this. (I also find Hug et al 2016 largely unreadable with massive use of acronyms, but presume if I worked hard enough I could eventually decode it.) And sure enough, here is Moody, Mahendrarajah et al 2022's *opening sentence*:

> "Core gene phylogenies provide a window into early evolution, but different gene sets and analytical methods *have yielded substantially different* views of the tree of life."

That is a huge indictment of this entire area, even 42 years after Fox & Woese. (And M,M et al might actually be disputing Fox & Woese's entire discovery of the archea "kingdom," although I must confess I don't understand exactly what they are claiming re that.) And if Alicioglu & Alkan 2024 are indeed the first to realise how to compute edit distance in polynomial time with substring "cut & paste" and "flip" as editing primitives, then that means *every* tree of life computation before 2024 is at best suspect, at worst total hogwash; and they all need to be redone.

The "**last universal common ancestor**" (LUCA) of the archaea, bacteria, and eukaryotes actually cannot currently be claimed to be any single organism because the tree structure of the tree of life becomes confused and largely irrelevant in the far past due to HGTs. Madeline C. Weiss et al 2016, have inferred from genetic backtracing of overlapping gene families that the LUCA was "anaerobic, [fixed carbon from] CO_2, H_2-dependent with a Wood-Ljungdahl pathway, N_2-fixing, and thermophilic," while Woese 1998 claimed it "was a prototroph with a complete [Krebs-like] tricarboxylic acid cycle, polysaccharide metabolism, both sulfur oxidation and reduction, and nitrogen fixation; it was motile by means of flagella; and had [regulation]." Moody, Alvarez-Carretero, et al 2024 infer from "molecular clocks" that LUCA lived between 4.09 and 4.33 Gyr ago (albeit if their "clocks ticked faster" inside ancient life than they do inside today's life, then this age is an overestimate; Feng, Cho, Doolittle 1998 instead claimed LUCA lived 3.2 to 3.8 Gyr ago), had ATP, a genome with $(2.49\text{-}2.99)\times10^6$ base pairs encoding 2451 to 2855 proteins, synthesized nucleotides via the "pentose phosphate pathway," and had an early immune system indicating viruses were already important at that time.

> Slesarev, Mezhevaya et al 2002 (and several phylogenetics papers they cite) claimed that a good candidate for today's lifeform most resembling LUCA is the deep hot methanogen Methanopyrus Kandleri. Takai et al 2008 discovered it on the wall of a 2000m deep ocean "black smoker" off California. This organism is killed if exposed to O_2; and sulfur might also be toxic for it (although less so). It needs the water to be very salty (≥1 molar ion concentrations) in which case one strain can grow and reproduce at 122°C. *M.kandleri* proteins have unusually many negatively-charged amino acids (Asp & Glu), possibly an

adaption to high intracellular salinity. *M.kandleri* has unusually few proteins involved in signaling and regulation of gene expression. and appears to have fewer HGT-acquired genes than other archaea, . It lives in an environment rich in hydrogen (H_2) and carbon dioxide (CO_2) and like other methanogens reduces the latter to methane. Its genome is a single loop of double-stranded DNA containing 1694969 nucleotides, with 62.1% GC fraction, encoding 1691 proteins and 39 RNAs. *Kandleri* possesses many genes lacking homologous counterparts in any other known species, e.g. is the only species known to have topoisomerase V. However it shares the gene-set implicated in methanogenesis (and, in part, its operon organization) with *Methanococcus jannaschii* and *Methanothermobacter thermoautotrophicum*.

It seems widely agreed that LUCA

- excreted acetic acid as a waste product ("vinegar world"?),
- had "methanogenic roots,"
- methylated its RNA and/or DNA,
- had seleno-proteins,
- used chemiosmosis to produce energy,
- had an outer cell membrane,
- fixed carbon from CO_2 using the Wood-Ljungdahl exergonic-reaction pathway, but had not yet developed photosynthesis,
- fixed N_2,
- performed sulfate→sulfide and sulfite→sulfide reduction,
- Neither LUCA, nor any prior lifeform, ever had sexual reproduction,
- May have had reduced abilities compared to present-day life in synthesizing amino acids and lipids,
- Was "thermophilic" and was probably involved with FeS and H_2 and likely inhabited deep-sea "hot vents."
- had the usual "orthodox plan" roles for DNA, proteins, and lipids, including the genetic code, ribosomes and messenger RNA – although perhaps it did not replicate its DNA the way today's life does, but instead replicated its DNA via RNA, with DNA merely serving as a long-term storage "backup copy" of the (actually primary) RNA genome.

What life was *before* the "orthodox plan" took over, is very difficult or impossible to deduce *genetically*. It is like trying to back-deduce the nature of the earliest computers before the Von Neumann architecture took over, from current computers – almost impossible. Presumably the last step before the orthodox plan took over was "RNA world," and this paper's main goal is to go toward removing RNA world's main obstacles. What happened *before* RNA world life is even more mysterious. Conceivably it had no living predecessor. If in this paper I make it clear that genesis of "RNA world" life was acceptably probable, and if some "pre-RNA world" would have made it even *more* probable, then so much the better.

Nature of early earth. The age of the earth-moon system, as determined by the oldest solid minerals yet found, is 4.46 Gyr, based on minerals in plagiocase moon rocks collected by Apollo-16 from the Descartes highlands in 1972, and there are 4.40 Gyr old zircon crystals inside rocks from Jack Hills, W.Australia. The early earth was like an autoclave, only more violent – too hot for any liquid surface water. Presumably life only become possible once it had cooled enough for liquid surface water to exist, whereupon the oceans formed. That occurred 4.2-4.3 Gyr ago according to Zahnle, Arndt et al 2007 and Arndt & Nisbet 2012 citing Mojzsis, Harrison, Pidgeon 2001. Therefore, the cooling from molten rock to liquid water temperatures took 100-260 Myr. Because the enthalpy of vaporization of water is about 2260 J/gram, whereas the specific heat of water is only 4.184 J/gram/°K (and 2260/4.184≈540), presumably once the oceans had formed, further cooling to about the present average Earth surface temperature ≈15°C took only ≈40 more Myr (albeit perhaps it was considerably warmer than 15°C due to comparatively immense amounts of greenhouse gases like CO_2

Energy Budgets

For present-day and 4 Gyr older ancient Earth mantle (mostly from D.L.Turcotte & G.Schubert: *Geodynamics*, Cambridge Univ. Press 2002). Power units: 10^{-15} watts/kg. The ratio 25739/7385≈3.49. The sun presently supplies 1.7×10^{17} watts to the Earth, a factor ≈10^4 times more than its radioisotopes.

Isotope	Power now	Halflife (Gyr)	Power Then
Th-232	3270	14.0	3986
U-238	2910	4.47	5411
K-40	1080	1.25	9925
U-235	125	0.704	6417
(total)	7385		25739

then in the atmosphere – or colder since the sun was less luminous then). It is unknown when ice first appeared, but quite possibly very soon after liquid water did. Plate tectonics (subduction of ocean crust) also was occurring already over 4 Gyr ago based on zircon crystals found in Australia. The internal heat from radioisotopes inside the earth is now 10^4× smaller than the power supplied by sunlight; for the early earth, with 3.5× more radioisotope power, that ratio instead was about 3000. Consequently (at least naively) the early earth's crust would have been 3.5× thinner, and the thermal gradient within that crust 3.5× today's. Meteor hits also would have been much more common than now. Wikipedia claims typical thermal gradients today are 25°C/km, which would have been 87°C/km on the early Earth. But these vary considerably with location. For example (Onstott 2017), near the world's deepest mine, the Mponeng gold mine in S.Africa (>4 km deep), today's thermal gradient is only 9.5°C/km (and actually the main reason the deep levels of that mine are hot inside is *not* the crust's thermal gradient, but rather adiabatic heating of ventilation air pumped down into the mine which gets compressed to half its volume by twice the usual atmospheric pressure; the miners counteract that by also pumping chilled *water*). So on the 4 Gyr old earth, Mponeng's thermal gradient instead would have been 34°C/km. The original oceans are believed to have had about twice the water, and twice today's average depth of 3688 meters. (The "lost" half of the water later got absorbed/subducted into the crust and mantle according to Hirschmann 2006 and Fei, Yamazaki et al 2017. Indeed, the fact that there is water even very deep inside the earth was shown by the discovery of diamonds containing trapped water inclusions; these diamonds formed 150-200 km deep.) The increased water and greater temperatures than today both suggest the planet continually had ultra-violent storms, with greater rock-water chemical interaction than now. Without photosynthetic life, there was no free oxygen in the atmosphere, whose chemical nature instead was reducing. Since ≤3% of Earth's surface area today is land ≥3000 meters above sea level, there probably was very little dry land. At then-typical ocean-bottom depths of 7000m, the pressure would have been about 700 atm (and more if huge amounts of CO_2 were dissolved in the water or present in the atmosphere above) and more at deeper places; and if we consider going as much as 10 km deeper within sediments *underneath* ocean bottoms, then add an extra 2600 atm. Because hot water rises (and, which is less-well known, water very near the freezing point *also* rises, as a consequence of ice's density being only 92% of the density of liquid water; water's density is maximized at 4°C) the oceans get colder as we go deeper and are coldest at their bottoms, but slightly above the freezing point. However, within all *solid* parts of the Earth, temperatures rise as we go deeper, today by about 25°K per km. By 3.95 Gyr ago microbial cellular life was prevalent (Tashiro, Ishida et al 2017). Therefore, the genesis, early evolutionary improvement, and exponential explosion and worldwide spread of life all had occurred within 150-250 Myr after the first day life became possible.

Estimating the slowness and fragility of early life

Humans' earliest **computing** devices and methods were over 10^{10} times slower than year-2025 desktop silicon computers. Also, some early electronic computers such as Harwell's CADET (1955), Brookhaven national lab's MERLIN (1960), and Australia's SILLIAC (1955) exhibited mean time between failures of ≤6 hours, i.e. ≤10^9 cycle times. That contrasts with multiyear uptimes for high-reliability computer systems today, e.g. ≥10^{17} cycle times. (The computer inside NASA's Voyager 1 spacecraft has been operating continuously since August 1977, i.e. over 48 years as of year 2025, i.e. 4×10^{14} cycles of its rather slow 250 kHz clock.) On top of that, the earliest chess-playing computer programs, in order to produce chessplaying strength comparable to the best year-2025 software (which uses better algorithms and improved techniques which have enjoyed many years of refinements), would need to run for perhaps 10^{10} times as many CPU cycles per chess move. These **improvement factors** 10^{10}, 10^8, 10^{10} would probably become much larger if the earliest computer hardware, or chess-playing software, had *not* been intelligently designed, but rather created by random luck.

Humanity's earliest successful powered **airplane**, the Wright brothers' "flyer" (1903), flew 4 times, carrying one human for a maximum distance of 260 meters, and never exceeding about 4 meters altitude, then self destructed. It had a maximum speed of Mach 0.04. In contrast, Airbus's A380 (first manufactured exactly 100 years later) can carry 853 passengers plus 20 crew up to 14800 km, has maximum design speed Mach 0.96, and has flown over 800000 flights with no losses. That's a factor 873 improvement in #people, factor 24 speedup, factor 57000 range increase, and factor≥200000 reliability improvement. Airbus's Zephyr unmanned solar powered airplane flew continuously for 67 days in 2025, mostly at about 18 km altitude at Mach 0.2. That's a factor≈300000 flight-duration increase, factor≈6000 altitude improvement, and very large fuel-efficiency improvement (namely: no fuel needed). Another highly evolved flyer, the bar-tailed godwit, in some cases migrates between New Zealand and Alaska via one 13000km 11-day

nonstop flight. These improvement factors again would probably be even larger had the earliest airplane been the product of random luck rather than intelligent design.

One may continue in this way to explore other **improvement factors caused by technological refinement or Darwinian evolution**:

- Man-made **magnets**: the earliest (steel needles rubbed against "lodestones" – very rare rocks, made of impure Fe_3O_4, that had been magnetized by lightning strikes) probably had fields of around 0.05 Tesla, whereas today's strongest have 45 Tesla fields, a factor 900 increase.
- **Tall buildings**: The earliest were 1 story, while the Burj Khalifa has 163 floors and stands 830 meters high. Also notable is the tallest known tree, a coast redwood 116 meters tall; in contrast the earliest land plants presumably were ≤10 cm tall, a factor>1000 increase.
- **Temperature**: When hominids discovered how to make fire, they achieved temperatures 900-1200°K. The highest temperatures yet achieved, in collisions of heavy ions in particle accelerators, are above 5×10^{12}°K; and on a macroscopic scale inside H-bombs (3×10^{8}°K). The lowest are below 10^{-10}°K. Presumably no pre-hominid life ever got above 500°K or below 150°K.
- **Chemical synthesis atom-count**: The earliest chemists did reactions involving only a few atoms, but the syn3.0 artificial bacterial genome is a precisely designed molecule containing about 32 million atoms; and genome syntheses in vivo can be 3 orders of magnitude larger. Also Lim et al 2014 synthesized a "dendrimer" with molecular weight≈8.4×10^{6} made of H,C,O, and N atoms.
- **Mathematical proofs**: Some of the earliest theorems still in use today (over 2000 years old), such as the formula for the roots of a quadratic, "there are infinitely many primes," the irrationality of √2, and the Pythagorean right-triangle theorem, can be proven "formally" with machine-verifiable proofs 49-659 lines long, verifiable in $<10^{10}$ compute cycles. In contrast, some of the most impressive recent "formal proofs" include the "flyspeck project" by Hales et al 2015 showing Kepler's conjecture: "no packing of congruent balls in Euclidean 3-space has density greater than that of the face-centered cubic packing." Their Kepler proof is verifiable in about 4×10^{13} computer cycles. The proof that Tibor Rado's "busy beaver function" S(n) = the maximum number of steps that an n-state 2-symbol Turing machine can perform from the all-zero tape before halting in finite time for n=1,2,3,4,5 assumes the values 1, 6, 21, 107, 47176870 was shown with a "coq" machine-checkable proof verifiable in about 10^{14} computer cycles. The proof that 8×8 American checkers is a perfect-play draw is verifiable in about 2×10^{18} computer cycles. M.J.H.Heule's 2017 proof that the first 5 "Schur numbers" are 2, 5, 14, 45, 161 occupies over 2 petabytes and can be verified in about 3×10^{18} compute cycles. Rokicki et al in 2009 and 2010 proved that every "Rubik's cube" configuration is solvable in ≤20 face-turning moves, and also in ≤29 face-quarter-turn moves, where 20 and 29 are least possible; this required about 3×10^{19} computer cycles. In short, the best mathematicians today can prove theorems 10^{10} times "harder" (reckoned in terms of proof-verification effort) than the earliest still-important theorems.
- **Explosives**: The first explosive discovered, "black powder," achieves pressures≈4400 atm, albeit that figure is for a modern powder, produced by improved (purified ingredients, wet grinding, "corning") processes versus the original kind, which perhaps achieved only 1000 atm. The later nitroglycerin has detonation pressures≈200000 atm, i.e. about 45 (or 200) times greater.
- **Light sources**: Compare burning branches as sources of light, versus modern LEDs. The lumens/watt efficiency improvement factor≈500.

A different analogy is industrial chemical syntheses. Consider the history of **methods for "fixing" atmospheric nitrogen** N_2, arguably the single most important industrial synthetic reaction. The first four such methods were

1. The 1775-1785 observation that electrical discharges in air (N_2 & O_2 mixture; can use either artificial electrical sparks, or natural lightning) produce nitrogen oxides, which then can be dissolved in water and/or cryogenically liquified. This consumes over a gigajoule of electrical energy per kg N fixed.
2. The 1857-1862 discovery that magnesium (Mg) metal would react with N_2 at 800-850°C to yield Mg_3N_2. One then could react $Mg_3N_2+6H_2O\rightarrow3Mg(OH)_2+2NH_3$ whereupon the $Mg(OH)_2$ could be used to regenerate Mg. [Also extremely energy-consuming. A similar process could be based on $6Li+N_2\rightarrow2Li_3N$, $3H_2O+2Li_3N\rightarrow6LiOH+2NH_3$, then regenerating Li from the LiOH.]

3. The "cyanamide process" (first invented 1895-1899, but there was a related earlier process based on barium invented in the 1850s), consisting of the cycle of 4 reactions
$CaC_2+N_2+heat(1000°C)\rightarrow CaCN_2+C$, $CaCN_2+3H_2O\rightarrow 2NH_3+CaCO_3$, $CaCO_3+heat(900°C)\rightarrow CaO+CO_2$,
$CaO+3C+heat(2000\text{-}2200°C)\rightarrow CaC_2+CO$.
4. Today's main process is Haber-Bosch $N_2+3H_2\rightarrow 2NH_3$ (pressures 100-400 atm, temperatures 400-500°C, iron-based catalyst, invented 1909).

Since the Haber-Bosch reaction theoretically is exothermic (3294 kJ released per kg of N_2), it ought to occur even at room temperature. But if you try that, you will find the reaction rate to be negligibly small. In principle perhaps some wonderful catalyst might exist which could enable high reaction rates even at room temperature, but nobody ever invented any catalyst that good in over 100 years of trying. When we take inefficiencies into account and include the cost of producing the H_2 from CH_4, in net today's Haber-Bosch processes consume 25-50 megajoules per kg N fixed. At present about 3-5% of the economy's natural gas production is (unsustainably) consumed by Haber-Bosch. Meanwhile,

5. Bacteria that fix nitrogen do it at room temperature and pressure (actually releasing H_2 gotten from H_2O as a side-effect!) and are thought to consume 16 ATP→ADP energy units per N_2, which if so would cost about 23.4 megajoules per kg of N fixed. This reckoning did not take into account energy needed to run "proton pumps" to provide electrons and protons needed by the reaction (which might effectively cost 4 additional ATP→ADP energy units), nor of inefficiencies in producing ATP – and therefore 23.4 should be increased to perhaps 30. But a large fraction of those energy costs could be recouped by burning the H_2 in O_2 to release 20.4 megajoules of heat per kg N fixed.

So the net **efficiency improvement** from "evolving" from process 1 to 5 was a factor 50-300.

Now consider **catalysis** e.g. aqueous fructose↔glucose interconversion, catalyzed by xylose isomerase; and peroxidases catalyzing $2H_2O_2\rightarrow 2H_2O+O_2$. The factors by which highly-evolved enzymes speed up the **rates** of reactions (such as those) that they catalyze are usually 10^6-10^9. And vastly larger rate-speedup factors occur for the net reactions catalyzed by *multi*enzyme systems such as yeast converting glucose to ethanol+CO_2, or by bacteria converting food and oxygen into an extra bacterium.

One could continue – it is a matter of common experience that for any complicated system, the first version is improvable by large factors. This might approach the status of a "law of nature."

That collection of historical improvement-factors tell me, by analogy, that the earliest life probably was **far more fragile, slower, less "competitive," and less "efficient"** than today's highly-evolved improved life. Under good conditions today's fastest-known bacterium, the halophile Vibrio natriegens, if placed in a nutrient-rich, oxygenated environment, can double in 10 minutes. If the earliest life doubled 10^9 times slower, its "doubling time" would be 19000 years. Furthermore, today's *deep underground* bacteria, which live in nutrient- and energy-*poor* conditions, appear to have doubling times ≳1000 years (Phelps, Murphy et al 1994; Whitman, Coleman, Wiebe 1998; Jørgensen & D'Hondt 2006; Lomstein,Langerhuus, et al 2012; p.28 of Lever, Rogers, et al 2015; Barras 2012; Bird, Tague, Zinke, et al 2019; Vuillemin, Vargas et al 2020; Onstott 2017's note 26 to ch.2 on p.408-9 derives an estimate that doubling time≥3640 years). That suggests the earliest life's doubling times might have been much *longer* than 19000 years.

I find it inconceivable that the earliest life could double in any timespan shorter than 10^5× the 3-hour doubling time of the syn3.0 bacterium, i.e. 3×10^5 hours=34 years, and would conjecture **≥20 kyears**.

However, the fact that Earth's life (apparently) took only 150-250 Myr to appear, self-improve, and spread all over the Earth, proves that the doubling time of most of that life could not have been longer than **2500 kyr**. (The estimated total number of prokaryotes on Earth today is $\approx 2^{102}$, and of viruses $\approx 2^{103}$; and Earth's total biomass is equivalent to about 1/300000 times the mass of Earth's oceans.) So, although both are attackable as not based on very solid evidence, we possess *both* rough upper and lower bounds on the "**speed**" of early life.

Consequences of that slowness & fragility. I deduce from the observed speedup factors from technological

improvement, aka in the biological context "evolutionary development," from those analogies that Darwin's "warm little pond" idea is nonsense: *No pond-type environment could be safe and protected enough to permit early fragile slow life to survive for 1 doubling time*. There's weather, seasons, temperature changes, hail, rain, storms, lightning, waves, day/night, earthquakes, sunlight including heavy UV (no protective ozone layer back then), meteors, and volcanos, all happening on timescales 1000s of times faster than earliest-life doubling times – and the earliest life had to withstand that despite essentially zero self-repair capability. That all is unacceptable. Ditto for the oceans. The whole idea (often associated with Stanley Miller) that the earliest life occurred on the surface of surface rocks, or near air-water surfaces, is absurd. Ideas associated with Jack B. Corliss, John A. Baross, and Sarah E. Hoffman, and later with G.Wächtershäuser, that the first life arose near or in deep sea hydrothermal "hot vents" similarly are instantly ruled out. I have no objection to contentions that those environments provided some unique advantages for early life *once evolved enough to stop being ultra-fragile and ultra-slow*. But they are unacceptable settings for the *earliest* life.

The *only* possible common-enough and safe-enough for long-enough aqueous environments would have been water inside rock cracks and interstices between particles deep **underground** (or underneath seafloor). Also, these places, aside from being far **safer** over long time durations, also simply contain far greater total *volume* and *mass*, making the unlikely first genesis ≋10000× more likely than if you stupidly insisted on places with 10000× smaller total volume and mass. (Also Life Deep Underground Is Twice the Volume of the Oceans to quote the title of an 11 Dec. 2018 "the-scientist.com" article about subterranean organisms.) Also, it is not just a matter of more *volume* – there also is the factor of *number of chemical reactions*, i.e. pseudo-*time*. A commonly used rule of thumb (e.g. often used in food industries) for the effect of temperature on biochemical reaction rates is approximately: one rate-doubling for each 10°C temperature increase. By going deep underground, getting more pressure, the boiling point of water rises, permitting life to exist at higher temperatures. That hotter life can perform far more chemical reactions, e.g. according to that rule of thumb, proto-life at 120°C will explore 2^{10}=1024 times as many chemical reactions per second, versus proto-life at 20°C, even without accounting for the effect of increased pressure. (Increased pressure generally also increases chemical reaction rates because it causes more frequent collisions between reactants.) So if you stupidly insist initial life be located on the surface at 20°C, then you are stupidly giving up not only a 10000× volume & mass factor, but *also* a 1000× pseudo-time factor, making your scenario **10^7** times less likely. And actually the factor should be a lot bigger than that. Essentially every chemical engineer tries to *optimize their reaction conditions*, e.g. in our case by adjusting pressure, temperature, concentrations, and rock type – a high-dimensional space of adjustable parameters – to try to maximize "performance" (Taylor, Pomberger, et al 2023), which in our case means "probability of genesis." This optimization happened automatically on the early earth by simply choosing the right location, time, and depth. If you stupidly sacrifice most of that parameter-optimization by insisting the location must be on the surface (or stupidly insisted it be "in the oceans," giving you pretty-uniform conditions *everywhere* and thus sacrificing almost all your adjustability) then the performance penalty you will suffer will far exceed merely the volume×time factor. I.e. *non*uniform conditions (i.e. inside rocks) always give you a greater genesis-probability boost than merely the volume×time factor when compared with much more uniform conditions such as ocean. [Today's oceans almost everywhere have boringly uniform composition and surface temperature between 292(1±0.065)°K, and with *sub*surface temperatures a lot more uniform than ±6.5%.]

If and when the first life somehow managed to arise, it then would multiply exponentially, while also experiencing rapid Darwinian adaption making it faster, more efficient, and less delicate. That *eventually* would enable it to survive in less-protected environments, for example ocean hot vents, where it enjoyed comparatively rich energy sources; and near and above Earth's surface, even exposed to (highly destructive) sunlight, where it could spread much faster. Life then proliferated and varied so widely and enormously that it became unstoppable.

Estimating the unlikeliness and rarity of life-geneses, and the information content of life

Life exists, i.e. chemical configurations that are "alive" exist, and therefore the probability is positive that life will arise; and if the universe is infinite that probability becomes 1. Mathematically the only difficulty is that the genesis probability per meter^{3}second seems, at least **naively**, to be very *small*. I'll now examine the numbers and find some conclusions quite different from every prior analysis I ever saw.

Initial guesstimate: Nonlife→life "genesis" events happen at a rate of about 1 genesis per 150-250 Myr elapsed time throughout the entirety of the top 10 km of the solid part (plus oceans) of planet Earth. I base this estimate on the observed fact that after Earth's environment stopped being far too hostile for life, life on Earth, in the form of microbes

with outer membranes, apparently took 150-250 Myr to appear and spread widely.

Some of the clearest/earliest life evidence is based on carbon "isotope-lightening" found in tiny blobs of carbon-rich material (presumably former bacteria) inside rocks up to 3.95 Gyr old. That is: all cellular life imports carbon-containing molecules through its cell membrane. Any such process favors molecules containing more carbon-12 and fewer carbon-13 isotopes, since lighter molecules diffuse faster. Consequently all life features a slightly lighter carbon 12+13 isotope mix than found in non-biological carbon sources such as carbonate rocks. (This effect is larger inside *photosynthetic* cells, because their main carbon source is CO_2, which nearly-maximizes isotope-fractionating effect because it is an especially light molecule containing only a single C. Especially large carbon-lightening also should occur in bacteria whose main carbon source is CO and/or CH_4; and these effects should increase as we go up the food chain to herbivores then carnivores.) The same phenomenon also should occur for sulfur-32, 33, 34 and 36 isotopes, but that approach currently seems more difficult.

"Stromatolites" – fossilized mounds of marine bacteria – have been found, and make the existence of 3.5 Gyr old life very clear.

I now suggest a different, extremely simple, method of proving the existence of very old life – which as far as I can tell, origin-of-life biochemists, life-on-Mars investigators, etc, surprisingly (and unfortunately?) never tried:

1. Dissolve organic compounds trapped inside very old rocks in a high purity solvent with mirror-symmetric molecular structure, such as H_2O, ethanol (C_2H_5OH), cyclohexane (C_6H_{12}), toluene ($C_6H_5CH_3$), or dimethyl sulfoxide $(CH_3)_2S{=}O$.
2. Shine a plane-polarized laser beam through it.
3. If the plane of polarization of the laser gets rotated (whereas control fluids do not cause any such rotation), that proves those organics were chiral, hence biogenic.

However it also is quite possible that all chiral organics would have racemized after 1 Gyr, in which case this method would be useless for life that old. This only can work if some class of chiral biomolecules exist with chirality halflife≳1 Gyr.

Earth's geological history therefore suggests the **lower bound**

$$\text{Earth-based genesis rate lower bound} \approx$$
$$1 \text{ genesis per } (250 \text{ Myr}) \times (\text{volume of Earth's biosphere}) \approx 2.4\times10^{-35} \text{meter}^{-3}\text{second}^{-1}.$$

(For this purpose I am regarding the "biosphere" as the planet's top 10km.) From the point of view of atomic chemistry it is most natural to express this number not in terms of meters and seconds, but rather in terms of "atomic units." The natural atomic-physics **length** unit is $\hbar/(\alpha m_e c)\approx 52.91772105$ picometers, the smallest orbit-radius in Bohr's simplest and first quantum mechanical model of the hydrogen atom (with "immovable" proton). The unit of **time** is the Bohr-electron's orbital period divided by 2π (also equals Planck's constant $\hbar$ divided by twice the ionization energy of that hydrogen atom model), i.e. $16\pi^2 (m_e)^{-1}\hbar^3 e^{-4} (\varepsilon_0)^2 \hbar^3 e^{-4} \approx 2.418884327\times10^{-17}$ second. (The resulting unit of speed=length/time is 2.188×10^6 meter/sec≈0.0073c≈αc.) Re-expressed in **atomic units**, then, that genesis rate lower bound estimate is

$$\text{Earth based genesis rate lower bound} \approx$$
$$\mathbf{9.0\times10^{-83}}\ \hbar^{-6}\ \alpha^3\ (m_e)^4\ c^3\ e^4\ (\varepsilon_0)^{-2}\ 2^{-4}\ \pi^{-2}.$$

Also note that $\log_2(9.0\times10^{-83})\approx -272.6$. In other words, this genesis rate lower bound corresponds, roughly, to geneses being events of probability comparable to, if repeatedly tossing a fair coin, getting 272.6 consecutive "heads."

Actually, this could be criticized since, e.g, the speed of sound in hydrogen at room temperature is about 1 km/sec, whereas the speed of a hydrogen atom with kinetic energy equal to its ionization energy is about 50 km/sec. Therefore, perhaps we need to change our frequency of "coin tossing" by factor≈50. This would additively alter the coin-toss head-count by $\log_2(50)\approx 5.6$, which is going to be too small to matter in view of the sizes of the errors in our

estimates.

Warning: That was a very unreliable and noisy lower bound, because based on only a *single*, quite random, event (Earth's genesis), and subject to probably-enormous observer bias since human observations have been almost entirely about our one planet and one solar system. That observer bias acts in an "exaggerate rate" direction. If we postulate that a goodly fraction of geneses within our galaxy would yield life that eventually would evolve to a tool-using intelligent stage that would create galaxy-colonizing "machine life," then we conclude from the *lack* of any observation of such machine life all over our galaxy and solar system (as in Fermi's quote "where is everybody?") that genesis events are much rarer than would be wrongly concluded by examining Earth alone. To try to avoid the Fermi paradox and Earth-centric observer bias: I expect the true genesis rate is ≲10 geneses per Milky Way type galaxy throughout the entire active timespan of that galaxy (which I expect to be about 10 times its present age). Exoplanet observations provide the rough estimate that the Milky Way contains 5-50 billion earthlike planets, suggesting in view of the Fermi paradox the upper bound

$$\text{Milky Way based genesis rate upper bound} \approx$$
$$1 \text{ genesis per (100 Myr)} \times (\text{volume of Earth's biosphere}) / (5\times10^{10}) \approx 1.2\times10^{-45}\text{meter}^{-3}\text{second}^{-1}$$

within the upper 10 km of earthlike planets in our galaxy. In atomic units this is 4.5×10^{-93}, whose binary logarithm is ≈-306.8.

The contradiction between these upper and lower bounds is presumably is due to earth-centric observer bias. One could even postulate *maximal* observer/sampling bias: the Earth contains the only life in the entire observable universe! If Earth life is the product of that much extreme luck, we must cut the rate by another factor of $10^{10}\approx2^{33}$ or more.

Also, life apparently never appeared on Mars, despite the fact it *was* an Earthlike planet during a ≈1 Gyr long timespan and has about 1/4 of the surface area of Earth. That suggests

$$\text{Mars-based genesis rate upper bound} \approx$$
$$1 \text{ genesis per (1 Gyr)} \times (\text{volume of Earth's biosphere}) \times (1/4) \approx 2.2\times10^{-35}\text{meter}^{-3}\text{second}^{-1}$$

which in atomic units is 7.9×10^{-83}, whose binary logarithm is ≈-272.7. Mars and Earth thus provide roughly-matching upper *and* lower bounds on genesis rate, but only if we foolishly ignore the rest of our galaxy.

Consequence of that rarity. I expect that no human-done lab experiment will ever be able to cause/observe a realistic genesis, because the total time×volume inside labs run by humans, is necessarily at least $10^{20}\times$ tinier than $10^{35}\text{meter}^3\text{second}$. And even if such a genesis *did* occur in a lab experiment, in view of the postulated $\geq10^4$-year doubling time of the resulting life, the humans running that experiment wouldn't even notice it!

Nevertheless, what might be feasible is a **sequence** of experiments, each showing that (artificially created) conditions N can create scenarios featuring (more genesis-friendly) conditions N+1, which then in turn could be more-artificially created, and so on, starting from N=0 conditions mimicking the primordial Earth, and ending with some moderately large stage-number N where one could argue genesis was likely.

There also are many more **theoretical** ways to estimate genesis rarity that naively suggest such events are far-*more*-extremely rare and lucky. I'll now explain some. The main problem genesis-explainers face is how to make those theoretical estimates large enough to match the above observational estimates. I shall largely solve that problem, simply by thinking about it more clearly.

The bacterium Nasuia deltocephalinicola has only N=112091 DNA base pairs encoding G=137 genes (the smallest bacterial genome known in 2013). But it is thought that *Nasuia* cannot survive alone, but only as a symbiont that lives inside certain insects. That objection hopefully does not matter because Earth's earliest life presumably analogously benefitted from a pool of abiotically pre-created bio-ingredient chemicals, which it could use as food and did not need to synthesize itself. Only later life needed to perform such syntheses, after the initial pool was consumed. (In particular, we can be essentially certain Earth's earliest life could *not* fix nitrogen.) The "JCV-syn3.0" bacterium, the

first "synthetic bacterium," was created (at cost of about $40M) in a successful intentional effort to produce a near-simplest life form capable of *free living* survival (Hutchison et al 2016). It has N=531490 base pairs encoding G=473 genes and doubles in about 3 hours in optimal conditions. The most **naive estimate** (which we shall disparage!) would be that the chance of creating such a genome at random is around 4^{-N}. Robert, Ollion, et al 2018 claim that a random point mutation (i.e. altering a single base-pair) in a bacterial genome is fatal only with 1% chance. So in some sense, then ("**supercrude estimate**"), only 1% of *Nasuia*'s DNA base pairs "matter."

> To make an analogy: suppose you saw a 500×700 black and white pixel image of an ASCII character. I claim you probably could still recognize that character even if 98% of the pixels were replaced by random coin toss bits, or equivalently if the pixels were XORed with 51:49 biased-coin random bits. Therefore, it is wrong to regard such an image as having information content 350000=500×700 bits. A better estimate is 2% of that: 7000 bits. An even better estimate would be $[1+0.51\log_2(0.51)+0.49\log_2(0.49)]\times 350000\approx 101$ bits. In fact, since there are ≤128 ASCII characters, it is really only 7 bits.

Presumably, you could randomly alter about 100 of *Nasuia* DNA's quats and still enjoy a good chance$\approx 0.99^{100}\approx 0.37$ the resulting bacterium would remain alive. (If not, then "backtrack" by resetting to unaltered state, then try a different such random alteration.) Here a "quat" is a 4-way, rather than binary, bit, (i.e. 1 quat=2 bits) corresponding to the amount of information in each DNA base. But then, presumably you could keep on doing those "safeguarded random 100-quat alterations" to *Nasuia*'s genome for a very long time, say only stopping once 100 consecutive deaths occur. The result presumably would be a still-living version of *Nasuia* with genome extremely different from the original. The unanswered question is: **How many** such altered genomes still correspond to living Nasuia-variants? If the answer is H, then *Nasuia*'s information content is $2\times 112091-\log_2(H)$ bits, or more generally, for an N-base-pair-DNA bacterium, $2N-\log_2(H)$ bits.

So the question reduces to: How to estimate H? Answer: Use computer-science estimates of the **size of "backtrack trees"** – a topic initiated by Knuth 1975 then further studied by e.g. Purdom 1978; Chen 1992; Cornuejols, Karamanov, Li 2006; and Kilby, Slaney, Thiebaux, Walsh 2006. These estimate H from which we derive the **"true vital information content" of a bacterium**. Unfortunately, as far as I know no experimenter ever tried this kind of backtracking and subsequent estimation. I urge them to do it. But for now, that embarrassing non-investigation by the biologists forces me to estimate H crudely – but I can do better than the "supercrude estimate."

My simple formula to estimate H and true information content of bacterium: Let ε denote Robert, Ollion, et al 2018's "kill chance." I.e, randomly changing one random among the N base-pairs in the genome will kill the organism with chance=ε. Experimentally ε≈0.01. Regard that experimental ε estimate as a backtracking tree investigation that went only along one single root→leaf path in the tree of all possible N-base-pair genomes. (This is the complete 4-ary tree with depth N.) Then assume the entire backtrack-tree, which has 4^N potential leaves and is N ply deep (each node having 4 children), has independent failure probabilities 3ε/4 along each parent→child edge. I.e. we are assuming (**"homogeneous tree" model**) the *same* single-edge failure probabilities as Robert, Ollion et al observe along the root→leaf path corresponding to the bacterial genome, work for *every* parent→child edge in the tree. (The backtracking experiments I am suggesting for the future would improve that by taking some account of non-homogeneity.) In that case we estimate the number of full-depth tree-leaves to be $H=(4-3\varepsilon)^N$ (each a "live" genome), whereupon our final estimate of the bacterium's true vital information content is $\mathbf{-N\log_2(1-3\varepsilon/4)}$ **bits**, which for small ε approximately equals $\mathbf{3\varepsilon N/\ln 16\approx 1.082\varepsilon N}$ due to the identity $\log_2(X)=\ln(X)/\ln(2)$ and the Maclaurin series $\ln(1-X)=-X+X^2/2-X^3/3+...$ when $|X|<1$. Meanwhile the (larger) supercrude estimate was 2εN.

In the case of *Nasuia* with N=112091 and ε=0.01, we find that this less-crude estimate of *Nasuia*'s true vital information content is 1217.4 bits. The supercrude estimate was 2241.8 bits. Now actually, to create bacterial life of today's ilk you require *both* its DNA genome, *and* the machinery (RNA transcriptases, ribosomes to synthesize proteins, cell membrane, enyzmes to synthesize building block chemicals, etc) needed to use it, both of which have the same information content. Therefore, the unlikelihood of genesis is **squared** versus the unlikelihood of the genome alone. However, if we instead were considering Orgel-style hypothetical early "RNA world" life, which did not need transcription or ribosomes or proteins because its RNA genome *was* its enzymatic machinery, and assuming its cell membrane self-assembled for free ala Luisi and Deamer, then no squaring would be necessary.

So perhaps the "true vital information contents" of the Nasuia and syn3.0 bacteria "really" are about 1121 and 5315 "*quats*" respectively by the supercrude estimate (which is more conservative than my less crude estimate, e.g. already incorporates most of that "squaring"); which since each "quat" is two binary "bits" respectively are 2242 and 10630 bits. By the $-N\log_2(1-3\varepsilon/4)$ formula we instead find 1217 and 5773 bits. Therefore the supercrude estimate of the chance of creating viable life (anyhow Nasuia-level "life") in a "random experiment" (by creating a random genome of Nasuia's length 112091) is $2^{-2242}=4^{-1121}\approx 10^{-675}$.

We can do a little better by using more understanding about the nature of DNA genomes, in particular *Nasuia*'s genome. First assume (admittedly a bit too optimistically) that **permuting** the bacterium's G genes in all G! possible ways will not hurt the organism. We also could allow "**flipping**" each gene in any of 2^G possible ways. Finally, if we guess that J of the N base-pairs in the DNA genome are **"junk"** intergene spacers (the guess J=0.1N seems reasonable in view of data in Gil & Latorre 2012), i.e. say J=11209 such junk bases in all for *Nasuia*, then the genes can be **shifted** in (J+G)!/(J!G!) possible ways. Actually bacterial genomes are *circular* not linear, a fact that may be approximately accounted for by saying all N circular shifts of the genome will also work. Combining all of the above ideas yields

CLAIM: In the "homogeneous backtrack-tree model" (abandoning the "supercrude model"): an N-base-pair-long circular DNA genome, J of whose base pairs are "junk" intergene spacers, and whose non-junk consists of G genes (any promotor and repressor regions counted as "part of the gene" for this purpose; genes assumed arbitrarily permutable and direction-flippable), and with Robert-Ollion "kill chance" ε, has "true vital information content" (in binary bits)

$$\#\text{bits} = (-\tfrac{1}{2})\log_2(1-3\varepsilon/4)\left\{ 2N - G - \log_2([J+G]!\ N\ /\ J!\ )\right\}$$

which if J is large may be approximated [by using Stirling's approximation $\ln(X!)=(X+\tfrac{1}{2})\ln(X)-X+\ln(2\pi)/2\pm O(X^{-1})$ to get rid of the factorials of large numbers] accurate to better than ±0.2/J bits, by

$$\#\text{bits} = (-\tfrac{1}{2})\log_2(1-3\varepsilon/4)\left\{ 2N - (1-1/\ln 2)\ G + (J+\tfrac{1}{2})\log_2(J) - (G+J+\tfrac{1}{2})\log_2(G+J) - \log_2(N)\right\}$$

Example. For the syn3.0 bacterium (which was intentionally designed to have very little junk) if we use N=531490, G=473, and J=1900, we would find

$$\#\text{bits} = (-\tfrac{1}{2})\ \log_2(1-0.03/4)\left\{ 2\times 531490 - (1-1/\ln 2)\ 473 + (1900.5)\ \log_2(1900) - (473+1900.5)\ \log_2(G+J) - \log_2(531490)\right\} \approx 5741.$$

For Nasuia, if we use N=112091, G=137, and J=11209, we find #bits=1207.

Mismatch: We needed to match the observed $-\log_2$(genesis chance) = 272.6 to 306.8 *versus* our Nasuia and syn3.0 theoretical estimates 1207 and 5741. For 1207, this is a factor 3.9 to 4.4 mismatch. For 5741 the mismatch factor is 18-21. As far as I can see, the only hope to overcome this mismatch is this

> **"4× Postulate":** Although the information content of the Nasuia and syn3.0 genomes might be near minimum possible for life at *today's* levels of speed, efficiency, fragility, reliability, and competitiveness, the *earliest* life, which doubled vastly more slowly, and was hugely more fragile, less reliable, less efficient, and entirely uncompetitive, could be specified with 4× fewer bits of information.

Furthermore, as I warned you, *Nasuia*'s and syn3.0's information content estimates were based on a crude "homogeneous tree model" and also depended heavily on the precise value of Robert, Ollion, et al 2018's kill-chance "1%" which, note, they'd only estimated with poor (single significant figure) accuracy. That is why I outlined an experiment and math to make better estimates, and continue to urge biochemists to actually *do* those experiments and estimates. Don't you think the "information content of life" is worth trying to estimate right? Ever?

Now let me offer two little analogies suggesting something like the "4× postulate" should make sense, then I'll give further arguments supporting it.

1. Suppose you take some long piece of English text, e.g. the novel Jane Eyre by C.Brontë. I claim that *if* a team of people good at English, crossword puzzles, and cryptograms works hard at it, they probably will be able to mostly reconstruct the novel even if a random subset of 70-75% of its characters are blacked out by a censor. (A feat somewhat similar to that was accomplished, entirely manually, by allied codebreakers led by the USA's Joseph Rochefort and Agnes M. Driscoll, building on insights by Britain's John Tiltman; they largely "broke" some of Japan's secret codes, ultimately including Japan's primary World War II naval code family JN-25. This occurred over a period of many years of record-keeping with gradually-improving guesses. The JN-25 break prevented planned Japanese invasions of New Guinea and/or Australia, then led to the USA's tide-turning victory at the battle of Midway Islands.) This process will take much longer than just reading the uncensored novel, but will still almost-fully succeed in accomplishing Brontë's purpose: transferring the information well enough for you to answer most questions about it. I.e. *Jane Eyre* still almost-wholy "functions" if 70-75% censored – just at a much slower speed. Here is an exercise for the reader: a 220-character English quote, randomly 51% censored. Can you reconstruct it? (Answer at end of paper.)

```
In*t*e*beg*n*i****od***e**ed t*e **a**** and ****ea*t**
*o**the *arth*was**orml**s*a***e*pty* **r****s was*o**r *h***u*f*** ** t***d**p,
**d **e*S*i**t o* Go* wa**h*ve**n**ov****h**wa*e*s*
*** *o* *ai* "L*t*******be ***ht.*
```

2. Consider one of the simplest machines, a waterwheel, made of planks. Suppose 70% of the planks were removed. It might still function (poorly) for some time period.

Another way the "4× postulate" would automatically happen (in view of our earlier math) would be if Robert, Ollion, et al 2018's "1%" kill-chance simply were, instead, **(¼)%**. And maybe it is! Robert, Ollion et al considered changing one random base-pair in a bacterial genome and estimated that would kill the bacterium with chance ε=1%. But look deeper: *how* could they distinguish a "killed" versus "live" bacterium? They simply observed, and if it continued to divide, they judged it "alive." If, however, the random genome change had slowed the doubling time to (say) 1 week, then they probably would (falsely) have judged it "killed." The truth would have been "it remains alive, but much slower." This kind of error is *1-sided*; they never regard something dead as "alive." So, obviously, the Roberts-Ollion "ε=1%" claim really is merely an **upper bound** on the actual *truth*: 0<ε≤1%.

And suppose that many (perhaps even "every") enzyme in some bacterium that catalyses a thermodynamically-favored reaction, were simply *obliterated*. (And all the genes that specify those enzymes.) In that case, those reactions still would happen (uncatalyzed) – just much more slowly. (And perhaps it still *would* be catalyzed, in the early-life setting, by, e.g, minerals or interactions with random pollutants; much more poorly than by a modern enzyme, but faster than nothing.) Such obliteration would reduce the bit-count tremendously, but still allow a much slower, but perhaps still barely functioning, kind of "life."

You need to understand: the earliest life was just *barely* alive. Probably no biologist would recognize its aliveness even if it were placed right in front of them. It would be completely uncompetitive in today's ecologies. It was like a waterwheel with 70% of its planks missing – very poorly functioning. No miller, if given such a "waterwheel," would consider it acceptable. But nevertheless it might, just barely, "work."

In my opinion/guess, thanks to my information estimates combined with the "4× postulate," our universe is such that it makes genesis just barely likely enough to happen, *but* nevertheless so rare that it only will happen a few times per galaxy. (I have another theory, too crazy to discuss here, which arguably predicts exactly that rate!) The numbers I have presented look compatible with that opinion.

Is this new? Apparently. E.g. Yockey 2005 wrote an entire book on information theory vis-a-vis DNA genomes, which failed to mention my factor>100 effects, plus (therefore) giving yet another numerically extremely poor discussion of the notion that perhaps life had to be "intelligently designed" since "irreducibly complex." As prior examples, Fred Hoyle once did a naive calculation of this ilk and concluded

> "The likelihood of the formation of life from inanimate matter is [10^{-40000}] enough to bury Darwin and the whole theory of evolution. There was no primeval soup, neither on this planet nor any other, and [since] the beginnings of life were not random, they must therefore have been the product of purposeful intelligence."
> – Fred Hoyle: *The Intelligent Universe* (Holt Rinehart Winston 1984; QH331.H69). I saw some attempts by Darwinians to dispute Hoyle which nowhere mentioned my factor>100 effect. A different calculation of the same naive sort by Jean O'Micks [Journal of Creation Science 29,2 (2015) 110-118] "conservatively" got "approximately $10^{-167500}$," which Benner 2014 dubbed the [unsolved] "information paradox."

An even crazier number came from a "thought experiment" by H.J.Morowitz concerning a random re-assignment of chemical bonds, concluding

> "The probability for the chance of formation of the smallest, simplest form of living organism known is 1 to $10^{340000000}$... The size of this figure is truly staggering."
> – Harold Morowitz: *Energy Flow in Biology: Biological Organization as a Problem in Thermal Physics* (Academic Press 1968; QH505.M6).

Many "creationist" opponents of Darwinian evolution enjoy repeating those quotes. However, they all were wrong since they all rested on the most naive possible probability estimate, which I refuted. And Morowitz's calculation for "randomly bonded chemicals" disregards the observed fact that only certain chemicals appear in Miller-Urey-type simu-Earth experiments, so only *they* should get bonded – exactly as today's life's molecules are actually observed to be. (If today's life actually *did* seem to be made of remarkably brilliantly optimal molecules with "apparently random" bond structure, *that* really *would* have been evidence for a superintelligent creator. The contrasting fact that actual life's bonds form very *simple* linear polymers based on only a *few* building blocks, therefore actually is evidence favoring a Darwinian explanation.) And, e.g, the "creation scientist" prof. Micheal J. Behe argued in Nov.2019 that even bacteriophages (he mentions phage T4, with double-stranded DNA genome 168903 base-pairs long, encoding 289 proteins) are "irreducibly complex." Actually T4 is quite complex relative to typical bacteriophages. Much simpler is ΦX174 with a single-stranded 5386-base DNA genome encoding 11 genes; and exceptionally simple is the MS2 phage with a 3569-base single-stranded RNA genome encoding just four proteins. So if complicated phages evolved from simpler ones, Behe's T4 "counterexample" is just stupid. Further Behe refutations: Viruses presumably arose from bacteria as, initially, naked bits of DNA or RNA that accidentally got lopped off (and then the virus self-improved via Darwinian evolution, including getting much more complicated in cases where that increased fitness); or coevolved with the precursors of bacteria; so there is no need for viruses to have been created entirely de novo. And even if ΦX174 or MS2 foolishly *were* assumed to have been created de novo, their complexity would have posed no problem using our *non*naive calculation if their single-mutation kill-chances ε were ≤4% and ≤5% respectively (which they presumably indeed are since I expect the single-mutation kill chances for viruses like these featuring a lot of gene-*overlap* ought to be 2-3 times the kill chances for bacteria).

What about other universes? Well:

Question for the conlife enthusiasts and mathematicians: Find lower and upper bounds on the maximum (maximized over W,H and p with W≥H≥1 and 0<p<1) of the **"probability of life,"** for a random conlife initial state inside an H×W rectangle got by flipping HW coins, each coin having prob(1)=p.

Adam P. Goucher in 2009 designed and built a universal computer inside the conlife mathematical world, equipped with a "constructor arm" (by Paul Chapman & Dave Greene) capable of synthesizing anything having a "slow-salvo glider synthesis." (Nicolas Loizeau made a video showing *his* different computer in operation inside conlife.) Because Goucher's "hardware" consists exclusively of "still objects" with ≤7 one-bits each (while its "software" is "gliders" with 5 one-bits each), a sufficient program tape would allow Goucher's machine to self-replicate forever. Goucher estimates that it would take ≈10^{18} conlife-timesteps to completely reproduce. (Coincidentally, the fastest known bacteria reproduce in about 10 minutes, i.e. 2.5×10^{19} atomic time units, remarkably close to Goucher's timestep-estimate.) Goucher's construction has K=481672 one-bits within an H×W bounding box with H=84625, W=72374, which therefore could be described using about **B**=$\lfloor\log_2([H\times W-K]!K!/[H\times W]!)\rfloor\approx7.26\times10^6$ bits of information. (Coincidentally, this bit-count is about the same as the naive bit-count of many bacterial genomes. For example the genome of Deinococcus radiodurans R1 consists of 3284156 base pairs, corresponding to 6.6×10^6 bits.) Therefore, the

"probability of life" in conlife clearly is positive, and indeed should be **lower bounded** by about 2^{-2B}, where the 2 in the exponent is due to the aforementioned squaring effect. However, Goucher was merely trying to *design* life, not trying to produce a good lower bound on its probability. If he wanted to try for a good lower bound, he would have noted that his design is merely one instance of a enormous set of comparable-size designs (since he could move his components around in a large number of ways, plus add inert and/or quickly-vanishing junk in an enormous number of ways, while still making it work), then bounded how many, then determined the resulting lower bound on conlife's life-probability, which would hugely exceed 2^{-2B}. That's a task that, as far as I know, nobody ever worked on.

In the other direction, nobody ever observed "spontaneously generated life" during conlife simulations with "random soup" initial conditions. That non-observation perhaps indicates an **upper bound** of order 2^{-65}.

The importance of being first: Only Earth's *first* genesis mattered. Even assuming a new one occurs every 100 Myr, the resulting new, slow, fragile, and uncompetitive life would simply get eaten.

Quick review of some "origin of life" ideas by others

Charles Darwin in a Feb.1871 letter to Joseph Hooker suggested that the origin of life may have been in a "**warm little pond**, with all sorts of ammonia and phosphoric salts, light, heat, electricity &c present, [such] that a protein compound was chemically formed ready to undergo still more complex changes. At the present day such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures [existed]."

Space dust in the early pre-solar gas/dust disk tended to nucleate precipitation of hydrocarbons formed from things like CH_2 radicals produced by gases and sunlight, which is why many asteroids and Kuiper belt bodies contain organic materials including PAHs (polycyclic aromatic hydrocarbons), alkanes, and "tholins." For experiments suggesting that perhaps such chemicals can survive events as violent as a comet impact, see Zellner, McCaffrey, Butler 2020 and works they cite.

Miller-Urey. The first "Milley-Urey experiments" came in 1952.

Predecessors of Milley-Urey. Actually, F.Wöhler's 1828 abiotic sythesis of urea could be regarded as a far earlier "Miller-Urey" experiment; and so could today's industrial Bosch-Meiser urea-production process $2NH_3+CO_2 \rightarrow CO(NH_2)_2+H_2O$, patented in 1922. Walther Löb did experiments very similar to Miller-Urey in 1913. Adolph Strecker's 1850-1854 syntheses of amino acids from cyanides plus aldehydes or ketones, and Aleksandr Butlerov's 1861 "formose reactions" synthesizing sugars from formaldehyde $CH_2=O$, may also be regarded as "Miller-Urey" style reactions.

Miller & Urey's idea was to set up chemistry in vitro intended to resemble the early pre-biotic earth (water "ocean"; atmosphere: some subset of N_2, CO_2, CH_4, NH_3, H_2, CO, H_2S, and sulfur oxides – the latter three are known to be outgassed from volcanos; free oxygen O_2 was at least 10^5 times less available then than now; see Zahnle, Schaefer, Fegley 2010 about the prebiotic atmosphere:) complete with sources of heat, precipitation/evaporation cycle, electric discharges ("lightning"), possibly rocks, clays, minerals (which often can act as catalysts), dissolved salts, possibly "volcanism," possibly day-night light cycle, including UV light since there was no ozone layer protecting the early Earth, possibly "cosmic radiation," etc. After running their experiment for some weeks, Miller & Urey's "warm little pond" turned red. They and later investigators analysed "primordial soup" samples from it, finding such environments automagically self-generate many basic biochemical life-ingredients such as:

- the amino acid glycine $H_2N\text{-}CH_2\text{-}COOH$, alanine $H_3C\text{-}CH_2NH_2\text{-}COOH$, serine $HO\text{-}CH_2\text{-}H_2CNH_2\text{-}COOH$, isoserine $H_2N\text{-}CH_2\text{-}H_2COH\text{-}COOH$, valine, isovaline, norvaline, β-alanine $H_2N\text{-}CH_2\text{-}CH_2\text{-}COOH$, phenylalanine, methionine, isoleucine, leucine, S-methylcysteine, and glutamate.
- However, several of the fancier amino acids found in today's life, such as arginine, proline, cysteine, histidine, lysine, asparagine, tryptophan, threonine, and tyrosine, remain *not* found in Miller-Urey-type experiments.
- Also present in Miller-Urey experiments were aspartic acid $HOOC\text{-}CH_2\text{-}HCNH_2\text{-}COOH$, aminobutyric acid $CH_3\text{-}CH_2\text{-}HCNH_2\text{-}COOH$, and ethanolamine.
- Profusely produced were hydrogen cyanide HCN, methanol CH_3OH, acetylene $HC{\equiv}CH$, cyanoacetylene

H-C≡C-C≡N, and formaldehyde CH_2O.

HCN is highly toxic to humans (in fact was used by the Nazis to kill millions) – and CO also is toxic – so Miller-Urey experimenters need to be careful. Another reason for care is "lightning" in highly flammable H_2 or CH_4-rich atmospheres. Formaldehyde, with water, is known to synthesize sugars such as ribose via "Butlerov's formose reaction"; while aldehydes and cyanides react to form amino acids via the "Strecker synthesis."

MacNevin 1953 found electrically sparking CH_4 and H_2O produces "complicated resinous solids." Joan Oro in 1961 produced the nucleobase adenine and several amino acids from a concentrated aqueous solution of HCN and NH_3. Bar-Nun & Hartman 1978 found that UV-photolysis of gaseous H_2O and CO yields various alcohols, aldehydes, and organic acids.

Ferus et al 2017 were able to obtain the other three RNA nucleobases cytosine, uracil, and guanine from prebiotic Miller-Urey-like conditions.

Most of these ingredients (if dilute) of course then could become highly **concentrated** e.g. via evaporation of a puddle in the desert (or tidal pools), or "zone refining" occurring automatically during directional freezing of water (or at ice-crystal grain boundaries), or extraction of water into some geological dessicant such as silica gel. (Similarly, many natural processes can and do produce concentrated "ores" of various metals, e.g. copper and molybdenum.) Indeed, Miller himself accidentally(?) discovered that freezing dilute aqueous HCN for several years, synthesizes adenine (Levy, Miller, Brinton, Bada 2000).

The **Murchison meteorite**, a particularly uncontaminated carbonaeous chrondite, contained over 0.3% aromatic hydrocarbons, and 0.1% lipid-like organics. It also contained all straight chain 5-carbon sugars, urea, 15 amino acids, alcohols, other carboxylic acids, sulfonic and phosphonic acids, and purines and pyrimidines – all of it presumably abiotic. Also, many interesting chemicals were found in the 5.4 gram sample of the Ryugu near-earth asteroid collected by Japan's Hayabusa2 spaceprobe, and in the 122 gram sample of the Bennu near-earth asteroid collected by USA's OSIRIS-REx mission. E.g. 33 amino acids, including 14 of the usual 20 (albeit racemic), ammonia, and numerous PAHs and C1-C7 carboxylic acids, were found in Bennu; and both asteroids have water-formed minerals indicating they clearly had (briny) water far in the past. Both Bennu and the Murchison meteorite contained all 5 DNA & RNA nucleobases (albeit Ryugu has more pyrimidines than purines, and both Ryugu and Murchison violate Chargaff's rules). Ryugu had, among many other chemicals, uracil, niacin, imidazole, and pyrene. These facts about the Murchison meteorite and the Ryugu and Bennu asteroids prove that Miller & Urey and their followers were correct that conditions on the asteroid-progenitor planet (or planets) indeed generated these chemicals. They also prove that Zellner et al were correct: even the violent breakup of that planet (which produced the asteroid belt), and then the entry of the Murchison meteorite into the earth's atmosphere at over 11 km/sec, causing a daylight-visible fireball and explosion blasting it into fragments and smoke (maximum known fragment mass 7kg; total mass of all fragments collected>100 kg) and producing a "strong smell of methylated spirits" over an area about $(13km)^2$, did not destroy those chemicals.

Tommy Gold as part of his "deep biosphere" notion proposed that crude oil and natural gas could have formed abiotically deep underground, and indeed controversially suggested it *mostly* arose that way. It now seems clear that *some* oil is abiotic, but the *vast majority within commercially exploited oil fields*, has biogenic origin. All known oil fields arose in sedimentary rock after an impermeable "cap" layer formed on top of large amounts of organic former life.

Certainly Saturn's moon Titan has ethane, ethylene (C_2H_4), diacetylene, methylacetylene, acetylene, propane, propyne, propene, benzene, toluene, acrylonitrile (H_2C=CH-C≡N), cyanoacetylene (H-C≡C-C≡N), HCN, cyanogen (N≡C-C≡N), CO_2, CO, H_2O, NH_3, and complicated organic "tholins" – all presumably abiotic. Indeed, Titan has more liquid hydrocarbons than all known oil and gas reserves on Earth. The higher alkanes butane, pentane, etc have not been detected there; but since polyethylene can be got by polymerizing ethylene (exothermic reaction, even more favored under pressure) presumably also could also easily arise abiogenically.

Kolesnikov, Kutcherov, Goncherov 2009 showed that methane at pressures>2 GPa and temperatures 1000-1500°K (conditions like those 100 km deep inside the Earth's mantle) will abiotically form heavier alkanes, molecular H_2, and graphite. CO_2 can also get reduced inside magma to produce hydrocarbons. Gold hypothesized those then would

seep upward. Gold claimed the existence of abiotic oil on Earth was proven after he convinced the Swedish government and the company Dala Djupgas to spend $40 million to drill boreholes in a fractured *igneous granite* rock formation, the "Siljan ring" in 1989-1992; they found about 1 barrel of oily sludge 2833 meters down, plus also recovered about 12 tons of oil-magnetite black paste 6700 meters down (the oil was mainly C8-C16 alkanes; there also was gas consisting of about 80% methane, 10% He, and 10% H_2). Gold unconvincingly claimed the latter oil had to be abiotic, but e.g. failed to perform a carbon 12/13 isotope ratio analysis to prove that, plus in some of their boreholes confused matters by stupidly intentionally injecting diesel-containing drilling mud. More convincingly, Lollar et al 2002 analyzed abiotic natural gas from a 6800-6900 foot deep borehole in Kidd Creek, Ontario, Canada. The Kidd Creek gas contained methane, ethane, H_2 and N_2 along with minor amounts of helium, propane, and butane. Lollar's measurements of its carbon 12/13 isotope ratio patterns, she claimed, proved its abiotic origin. There also is abiotic methane emitted from volcanos and the East Pacific Rise (reviewed in Etiope & Lollar 2013).

Oil geologists have loudly objected that Gold's kind of oil is not commercially exploitable, that investors had lost lots of money trying to find it, and Gold-like thinking fails to predict the locations of commercial oil fields. Those *economic* objections, while correct, are *scientifically* irrelevant to origin of life questions and to the question of whether total abiotic oil amounts exceed biogenic oil amounts.

Known and claimed commercially already-extracted or extractible oil reserves amount to 2.65 trillion barrels. Those are (usefully for commercial drillers) concentrated in fields constituting a tiny fraction of the surface area of the Earth. If Gold's two Siljan boreholes were regarded as random samples from Earth's surface and the oil in them were genuine and abiotic, we would conclude that the *entire* surface area of the Earth contains small amounts of oil underneath. If we estimate (probably conservatively) that there is 1 barrel of oil per square-meter of earth surface (provided you are willing to drill as far as 7km down to find it), then that would add up to 510 trillion barrels, containing 6×10^{16} kg of carbon. Therefore, it is entirely possible that the total amount of abiotic oil and gas far exceeds the biogenic amount. Indeed, this 6×10^{16} kg estimate exceeds *all* Earth's present biomass (which is thought to contain only about 5.5×10^{14} kg of carbon) by a factor>100.

Abiogenic lipids, sugars, amino acids, and nucleobases provide a full palette of all main basic ingredients of known life. The problem then is to explain how monomer ingredients got linked together into polymers (proteins, polysaccharides, lipids, RNA, DNA) and somehow eventually yielded a living organism.

A.G.Cairns-Smith during 1968-1990 noted that **clay particles** contain atomic layers of alumino-silicates with impurities, and those impurities (and/or crystal grain boundaries and defects) can form a huge number of possible 2D "patterns" which could perhaps get copied by adjacent layers, perhaps similarly to information carried by the two strands of DNA and permitting reproduction, perhaps with some sort of natural-selection-driven evolution. Also, these patterns could serve catalytic purposes for organic chemistry in water near a clay particle. The clay layers have water between, and charged atoms in the clay layers can attract and concentrate chemicals in the water which otherwise would have been dilute. All that plus their ubiquity suggested that clays might have played an important role in the earliest life. I'll call this set of early-life ideas **clay world.** Some experiments have verified that some chemical phenomena of the right kinds (e.g. polypeptide formation from amino acids; aiding of amphiphile-lipid vesicle formation) actually can be made to occur in, or are aided by, clay particles. For the most part, though, clay world remains almost wholy bereft of experimental support even 50+ years later.

If clay world really was the earliest life, then it is strange that no trace of it seems to persist inside today's life. One might imagine that *some* clay-based chemical mechanism, hard won via extreme luck and optimized through 100 megayears of evolutionary development, would not be easily replaceable and discardable and therefore still would survive inside some life *somewhere*. [For example, all vertebrate eyes are incredibly stupidly designed "backwards" in a way no human camera designer today would ever duplicate, as opposed to the obviously-more-sensible eye designs of cephalopods such as octopuses, which indeed resemble commercial camera designs. All mammals control their larynxes via an incredibly stupid design involving a nerve that travels all the way down their neck – quite a long way, for a giraffe – then winds around inside the chest for a while, then finally comes all the way back up the neck. The reason for these clearly-stupid anatomical design mistakes is that they simply have been too difficult for evolution to repair without killing the lifeform by trying. If life arose from clay world there presumably would exist some clay-aided chemistry that was hard to replace even if sub-optimal; plus probably some clay-aided mechanisms actually still *would* be optimal, i.e. would outperform anything made of proteins.] But even in microbes that actually live deep

underground, apparently no trace of clay-biomechanisms remain. Indeed, the very common elements Si and Al, and Na and Ti, the first two of which are the most important clay components (and the second two also often are present), seem *entirely unnecessary* for most microbes in all three Woese-kingdoms – even though these all are among the most common elements in the Earth's crust – indicating an *incredibly total* divorce!

For that reason I cannot bring myself to believe Cairns-Smith's clay world notions. Probably clays and minerals played a significant role in the Milley-Urey-style "prologue" that set the stage for life by creating its basic building block (plus perhaps some not-so-simple) ingredient chemicals. I have no objection to that idea. I just do not believe clays actually were an important *part* of early life.

Simon Nicholas Platts proposed "PAH world" in his 2004 PhD thesis as an idea perhaps able to explain how DNA or RNA (or something like them) molecules appeared on the early earth. That is: presumably there were a lot of PAHs (polyaromatic hydrocarbons) on early earth that arrived from space. Platts in email also pointed out that "metamorphic aromatization of carbon at depth can produce beautifully-formed crystals of mineral graphite" and presumably also PAHs. Examples found in today's coal tars (and therefore obviously stable for many millions of years) include (3 rings) anthracene; (4 rings) chrysene, fluoranthene, pyrene, triphenylene, naphthacene, benzanthracene; (5 rings) picene, benzo[a]pyrene, benzo[e]pyrene, benzofluoranthenes, perylene; (6 rings) dibenzopyrenes, dibenzofluoranthenes, benzoperylenes; (7 rings) coronene. These are flat molecules resembling parts of 1 layer of graphite. Platts hypothesized that

1. if exposed to water and UV light, such PAHs could acquire hydrophilic groups on their edges, such as phenol's -OH group or the -COOH group in benzoic acid, thereby becoming water-soluble. Platts then further hypothesized
2. that, in water, such solubilized PAHs would naturally "discotically self-assemble" into **stacks** (like a stack of plates in your kitchen cabinet, or multiple layers of graphite) in order to keep their hydrophobic parts away from water, while their exterior hydrophilic groups would be in the water. This behavior in fact is observed in certain "liquid crystal" substances (reviewed by Wöhrle, Wurzbach, et al 2015).
3. The spacings between these layers should automatically nearly equal the 34Å spacing between bases inside the DNA molecule. Hence these PAH-stacks would greatly resemble RNA and DNA.
4. Various other chemicals could then try to chemically attach to some of the side-groups; and because of continual plate "horizontal sliding" within the vertical stacks, flat molecules like purine and pyrimidine bases would naturally be preferred, thus perhaps catalyzing and/or "scaffolding" self-assembly of purines and pyrimidines into a "genetic" molecular assembly resembling DNA or RNA.
5. That assembly might then get attached into a polymer by formation of a "backbone" of sugars or something.
6. Then the backbone-equipped genetic polymer could somehow detach from the PAH-stack to roam wild and free.
7. Then those in turn could cause a genesis.

While each of Platts' seven component-ideas in the above story sounds not too implausible by itself, the entire story combines many different implausible ideas together, resulting via correctness-chance-multiplication in a story just too implausible for me to believe without experimental support. (E.g. if each of Platts' 7 sub-hypotheses were valid with chance 0.6, the whole story would work with chance 0.6^7<3%.) And as far as I know, none of those 7 ideas have acquired any experimental support during the 20+ years since Platts cooked this up. But perhaps some will come. Note that PAHs mostly are **carcinogens** because they tend to insert themselves like knives between your DNA bases, then act as "decoys" to confuse DNA replication processes, causing mutations. Therefore, today's life tries to stay away from PAHs, hence little or no trace of them would be expected to remain in today's life even if Platts' speculation that is how genesis got started, is completely correct. Also notice that high pressure (as advocated in the present paper) should enhance hydrophobic/philic pseudoforces at small length scales in water, helping PAH world (and probably also Cairns-Smith's "clay world") to happen.

"RNA world." Alexander Rich, and then Francis H.C. Crick, Carl R. Woese, and (ultimately most prominently) Leslie E. Orgel, originated the "RNA world" hypothesis in 1962-1967, although not originally by that name. An improved "multiple overlapping redundant genome" version of it was suggested by Falk, Zhou, et al 2025, Ding, Zhou et al 2023, and Zhou, Ding et al 2021. The trouble with today's "orthodox plan" life involving DNA storing genetic information, while protein enzymes direct and control all metabolism, with messenger RNA carrying the information from DNA to direct the construction of proteins by ribosomes, is that it is just too complicated to have arisen abiotically. Orgel at al pointed out that **(1)** RNA is simpler to make than DNA and also capable of storing genetic

information and getting replicated. Indeed, polio, dengue, the common cold, and tobacco mosaic virus all have RNA, not DNA, genomes. The faster and cheaper manufacture and replication of RNA presumably outweighs its stable-storage and repairability fitness-deficiencies as far as those viruses are concerned.

Orgel also speculated **(2)** that RNA could be used (instead of proteins) to make enzymes. That speculation is now known to be correct! The discovery of the first RNA enzymes by Thomas R. Cech and Sidney Altman got them the 1989 Nobel prize in chemistry. (Amazingly enough, neither Cech nor Altman were motivated by Orgel's RNA world ideas, which indeed they originally were only very dimly aware of. They instead discovered their RNA enzymes by a combination of shear luck followed by perspiration, and only later were informed of the importance of their discoveries for the "RNA world" life-genesis hypothesis.) Altman discovered RNase P, an RNA enzyme that helps synthesizes tRNAs. Cech discovered RNA enzymes that can splice "introns." Many other RNA enzymes have since been discovered. They include the RNA components inside ribosomes which are present in and centrally crucial to all orthodox-plan cellular life; telomerase which plays a very crucial role in Eukaryote reproduction; and the hammerhead ribozyme. There also are so-called ribo-switches, RNA "thermometers," and RNA regulatory elements.

Although, generally speaking, proteins are better than RNA for making enzymes (better at folding, last longer, 20 amino instead of only 4 nucleobase components for much more chemical and hydrophobic/philic versatility to usually provide better catalysts) they only are better in *most*, not in *every* instance – or even if some hypothetical protein would be superior, it evidently was too extremely difficult for 4 Gyr of Darwinian evolution to find. That is why ① every known cellular lifeform, even today, still includes some RNA (non-protein) enzymes. There also are ② various other "living fossil" relics of RNA world floating around in today's cells, such as tRNAs, and ③ RNA-component cofactors in various very ancient and important metabolic reactions, including the Krebs citric acid cycle and glycogen synthesis. There also are many key cofactors that are not RNA components, but strikingly resemble them, such as ③ NADH, ④ acetyl-CoA, ⑤ FADH, ⑥ S-adenosylmethionine (SAM), and ⑦ F420. Presumably the reason these exist containing a very RNA-like part (when presumably other variants of the these molecules could have existed doing the same job but with something non-RNA-like replacing their RNA-like part) is because they are relics of ubiquitous RNA world technology for handling molecules with RNA-like parts. The ubiquity of ATP as the ⑧ cell's "energy currency" similarly is compatible with RNA being important – indeed with the idea that early life's *primary goal* was to use its energy to create RNA or DNA (since nucleotide triphosphates are exactly the construction material used to build them, and includes the extra energy needed to do so as part of the package). All those are evidence in favor of RNA world, *or* a remarkably deceptive miraculous combination of flukes. Also not hurting: the recent discovery ⑨ by Becker, Okamura et al 2019 of a "unified" pre-biotic synthesis pathway for *both* purines and pyrimidines (phosphates and diphosphates) simultaneously. Put those all together and I'm ≥98% convinced RNA world had to have happened. (E.g. if each of those 9 evidence items made "RNA world is wrong" 35% less likely, then "98%" arises since since 35+65=100 and $0.65^9 \approx 2\%$.)

A life-form based on RNA *both* carrying its genetic information *and* directly enzymatically controlling its metabolism does not suffer the squaring effect which makes genesis of orthodox-plan life vastly more improbable.

The net effect of both **(A)** that theoretical reasoning, **(B)** that observational evidence, and **(C)** experiments by the likes of S.Spiegelman, J.W.Szostak, and others on RNA "improvement" via "selective breeding," which has enabled the discovery by synthesis of RNA enzymes capable of numerous interesting tasks (including catalyzing the formation of peptide bonds) all leaves me even more convinced that "**RNA world**," and/or perhaps some analogous idea based on some RNA-analogue like GNA, TNA, or PNA, **really happened** and later evolved into "orthodox plan" life, which then outcompeted and therefore entirely supplanted it *except* for RNA viruses (if considered "alive") and the "biochemical relic/clues" of RNA world still present inside orthodox plan cells. If anybody wants to search to try to find RNA life cells that hypothetically still might survive somewhere, then I suggest that the only place they might be is in **very deep** underground borehole drilling-cores, in rare places geologists might suggest have been able to stay biologically isolated for many gigayears.

However, until now, RNA world has faced the major **objections** noted in the abstract: RNA is thermodynamically unstable, quickly hydrolyzed in water, and has not been synthesized in any Miller-Urey-type experiment so far. A main purpose of the present work is to propose that high pressure can overcome those objections.

"Lipid world." A long molecule hydrophobic at one end, but hydrophilic at the other end – the classic example is a

lipid consisting of a hydrocarbon chain attached at one end to an alcohol, -COOH, or phosphate group – is called **amphiphilic**. An important theoretical realization and experimental discovery is that many kinds of amphiphile molecules, if placed in water, will magically **self-assemble** into spherical vesicles, about the same size as bacteria, with bilayer walls. The point of "bilayers" is that the outer layer's hydrophilic ends point outward into the water, while the inner layer's hydrophilic ends point inward into the water, while both layers' hydrophobic ends point toward each other, avoiding water. For this reason, such vesicles are stable and minimize energy. Such automatic self-assembly can be disrupted or slowed by ions dissolved in (salty) water, but has been shown to be encouraged by clay particles. C.M.Dobson pointed out that they also should be able to form as tiny aerosol droplets sprayed into the air by ocean wave action – we then get a monolayer of amphiphiles on the droplet surface with hydrophilic ends pointing inward – and then when the droplet falls back into a lake whose surface has an amphiphile layer with hydrophilic ends pointing downward, the result again is spherical *bi*layer vesicles. Experimenters have also shown that such vesicles, if split apart, can re-assemble into two "daughter" vesicles (just like cell fission!) and can be induced to form *enclosing* selected molecules such as RNA. These ideas were pushed by P.L.Luisi as probably playing an important role in the initial start of life. Cells living in water need to have an outer **wall** to *confine* all their important chemicals which otherwise would simply dissolve in the water, diffuse far away, and never come back. [Instead of a wall, the lifeform could glue its important macromolecules to some, e.g. rock, surface. That also would prevent them drifting away, but such macromolecule gluing would *not* be applicable to the myriads of *tiny* molecules, e.g. sugars and amino acids, in cells. For them, only a wall could work.] Deamer & Pashley 1989 then put abiotic amphihile molecules from the Murchison meteorite in water and showed *they* auto-magically formed into bilayer vesicles. Another recent interesting class of ideas has been that some bio-chemical "transporter" mechanism developed that **concentrated chemicals inside vesicles**.

Deep hot biosphere. An important discovery about our world is that life does *not* occur only at or near the land surface, or in the ocean. There also are microscopic organisms living ≥3 meter underground or below the ocean floor, in tiny pores, rock cracks, or gaps between particles. This is called the "deep" biosphere. [And as we dig deeper, it gets hotter, hence "deep hot biosphere." Since pressures also get high, and life cell-doubling times get much slower, and life's mobility is greatly slowed, I suppose it really should have been called the "deep, hot, slow-doubling, slow-migrating, pressurized biosphere." We also might want to add "salty" to that list of adjectives, since, as Onstott 2017 (e.g. p.167) points out, old water found in deep mines and boreholes is often quite saline.] Presumably eventually it gets *too* hot (molten rock!), so that all life must stop existing below some depth. But if so, that depth has not yet been found! No matter how deep investigators have looked (inside deep mines and boreholes) as long as there is water there, they never have yet been able to go deep enough to avoid finding life.

> Onstott 2017 gives two scatterplots on his page 360 relevant to the question of the depth where life stops. Plot A shows data extracted from the literature up to year 2014 of microbe cells counted per gram of rock, finding (with a lot of "noise") this power law: cell count$\approx 10^8 H^{-5/6}$ where H is the depth below ground of that rock in meters and $1 \le H \le 5000$. (Plot B is the same except in "water" not "rock," and for it the noise is much larger.) If this law were (unwisely) extrapolated to arbitrarily large H, it would predict that life never stops, since the Earth's diameter is 1.3×10^7 meters, and even with that H this formula still would predict 120 cells/gram! Also note that with a flat earth, this formula would predict an *infinite* number of microbial cells underneath each square meter of earth-surface! Inagaki et al 2015 also find that there is life at least 2.5km beneath ocean floors.

All three families (eukaryotes, bacteria, and archaea) are represented, but predominantly the latter. (See Borgonie, Garcia-Moyano, et al 2011 for the eukaryotes.) Because it is dark down there, there is no photosynthesis. Deep life can and must power itself from energy sources other than the sun, such as non-ground states of chemicals found in rocks, and perhaps chemicals falling from above. In particular, radioactivity in rocks splits water, producing H_2 gas and H_2O_2, which microbes can use as food-energy, and e.g. to reduce carbonate in rocks to acquire carbon, while also acquiring nitrogen from ammonia inside rocks. Such fuel sources, although omnipresent and sufficient to power an ecology of microbes to live underground for billions of years without chemical energy from sunlight, provide much less available power than the sun provides to surface life. Hence deep life is much sparser: "Samples obtained from 400 meters below [ground surface] can contain as few as 100 to as many as 10^7 bacteria in each gram of rock" ("John R. Parkes and his colleagues at the University of Bristol have found somewhat higher concentrations of microorganisms living in sediments beneath the ocean floor.") while in contrast, "agricultural topsoil typically contains more than 10^9 bacteria in each gram of dirt." Those quotes were from Fredrickson & Onstott 1996, who also reported discoveries of

live prokaryotes up to 2.8km deep at temperatures of 75°C. Live cells and viruses have both been found in water seeping out of the walls of South African gold mines over 2 miles deep. But I doubt that 2 miles is the upper limit. Rather, it is merely lower bounds the depths that are survivable. Other cells are known to survive in 122°C pressurized water, but I similarly doubt 122°C is the upper limit. "Ellyn M. Murphy of Battelle, Pacific Northwest National Laboratory has determined, for example, that the groundwater now present deep beneath the USA's Savannah River facility has not been in contact with the surface for thousands of years. In the deepest sites we have examined, our measurements and computer modeling indicate that the groundwater has been isolated from the surface for millions of years... [so] some subsurface microbial communities must be at least several million years old." Also, deep cells reproduce far slower than surface life, with doubling times exceeding 1000 years!

What is the upper temperature limit for life? As of year 2025, the highest temperature at which microbes have been observed growing and reproducing is 122°C for one strain of Methanopyrus Kandleri. The chemical reactions/effects posing the most severe problems for any "orthodox plan" lifeform trying to live hot are:

- At 110°C Kawamura & Yukioka 2001 found halflives for racemization of Alanine, Leucine, and Phenylalanine of 32, 52, and 15 days. At 225°C these particular 3 halflives were 250-950 seconds (apparently these experiments were done at the vapor pressures of water at those temperatures); and a different author claimed that at 160°C most amino acids will racemize in 10-20 hours. (I would like experiments to measure the effects of pressure and salinity on those racemization rates, but apparently nobody ever did.) And
- Hydrolytic depurination of DNA happens at about those same time scales.
- DNA "melts" (i.e. the double helix breaks apart into two single-helices) at high temperatures.
- Water boils at 100°C, destroying all life relying on *liquid* water.

Modern life is capable of repairing amino-acid racemization and DNA depurination, *if* they happen slowly enough; and these effects indeed are very slow (30 kyr or longer time scales) at room temperature. (If they happen too quickly the repair mechanisms will be unable to keep up; and even if they could, this could amount to a severe energy-burden for the organism.) Let me just remark that **high pressure** (and high salinity) have the ability or potential to reduce those problems:

- Water's boiling point increases with both pressure and salinity, e.g. becomes ≈366°C at 200 atm pressure.
- DNA melts hotter at higher pressures and salinities: from Lah & Hadzi 2024 and Wu & Macgregor 1993/5 we see that T_{melt} increases by about 0.004°C per atm for CG-only, but remains near-constant for TA-only DNA; and T_{melt} also increases with salinity, with the increment being about 16.6°C times the change in the $\log_{10}$ of the salt concentration.
- Although the exploration of this has been very limited, I found two papers studying the effect of pressure on racemization rates of these two chemicals:
  a. the sodium salt of 6-nitro-2,2'-carboxybiphenyl [McCune, Cagle, Kistler 1960 fig.2],
  b. Tris(acetylacetonato)germanium(IV)perchlorate in 4 organic solvents [Ueno, Nagasaka, Saito 1981 fig.1].
  In case (a), 12000 atm pressure was found to slow racemization by a factor≈2 at 64°C. In case (b), pressures of 3000 atm slowed racemization in 3 solvents at 60°C (by as much as a factor≈4.5) but actually sped it up about 65% in the fourth solvent, DMF, at 50°C, which they suspect is because DMF actually serves as a catalyst. Neither of these chemicals was an amino acid, but the tentative general conclusion I reach from both these experiments, and from theoretical reasoning, is that (at least assuming the solvent cannot catalyse racemization) *high pressures usually slow racemization*. My "theoretical reasoning" is that if a molecule attempts to rearrange itself to racemize, it will get blocked by bumping into obstacles, e.g. solvent molecules, more often at higher pressures. [Indeed, at all high enough pressures liquids no longer exist – they all become solid – so that racemization could get "frozen" similarly to the effect of lowering the temperature. Of course this paper's pressures are not anywhere near that high; I say this merely to illustrate the principle.]
- Depurination should be thermodynamically disfavored at all high enough pressures.

For all those reasons I believe 122°C is *not* the upper temperature limit for life. I speculate **130°C** ought to be possible at, say 2000 atm with 2 mole/liter NaCl salinity. Temperature 130°C should occur about about 11.7km underground in the vicinity of S.Africa's Mponeng gold mine based on the 9.5°C/km thermal gradient there and the average surface temperature of 19°C. At the depth the pressure would be about 4440 atm assuming average rock density 2.6 gram/cm^3.

What is the upper pressure limit for life? As of year 2016, the highest pressure in which organisms are known to live in Nature on Earth does not exceed 150 MPa≈1500 atm. Wikipedia's phase diagram of water indicates that, at all temperatures≤300°C, water becomes *solid* if pressurized sufficiently high, where 80000 atm is always "sufficiently high," and only 20000 atm suffices at all temperatures≤82°C. (Bridgman 1912 discovered many high-pressure ice phases.) Therefore any life dependent on *liquid* water should be unable to live (it would freeze solid) at pressures higher than that. And in fact, a lot of common 1 atm microbes seem surprisingly unbothered by pressures not much below water-solidification pressures. Sharma, Scott, et al 2002 "observed physiological and metabolic activity of Shewanella oneidensis strain MR1 and E.coli MG1655 at pressures of 670-16580 atm in diamond anvil cells. We measured biological formate oxidation at 670-10460 atm. At pressures 11800-16000 atm, living bacteria resided in fluid inclusions in ice-VI crystals and continued to be viable upon subsequent release to 1 atm pressures. Evidence of microbial viability and activity at these extreme pressures expands by an order of magnitude the range of conditions representing the habitable zone." Both these bacteria are common in sewage treatment ponds at 1 atm hence would not a priori have been suspected to be especially pressure-resistant. Critics could complain that quote was deceptive in the sense that only a *small fraction* of Sharma et al's E.coli survived that pressurization. To riposte: Kish et al 2012 found that *nearly 100%* of Halobacterium salinarum NRC-1 (a bacterium that likes to eat salted pork at 1 atm) survived 1 hour at 400 MPa≈3950 atm. Furthermore, Vanlint et al 2011 and Hazael et al 2014 showed that even Sharma et al's original strains of E.coli and S.oneidensis can be *selected* for pressure resistance by repeated cycles of pressurization and culturing survivors, whereupon they can survive 20000-30000 atm. Probably their metabolisms are greatly slowed at these pressures, although that is not known for certain. Artemia cysts (the dried embryos of brine shrimp), and various plant spores and seeds, as well as "tardigrade" animals (aka "water bears") all survive pressurization to 74000 atm for timespans ranging from a few hours to several days (although dormant) then revive in 1 atm water. Those facts suggest the conjecture that microbes that (unlike all those) actually *are* pressure-adapted would have no problem surviving pressures up to nearly the solidification limit, albeit their metabolisms might be slowed at the highest pressures. Considering that all known organisms that live deep underground or deep sub-seafloor *already* have slow metabolic rate, such additional slowdowns might not be easy to detect and measure. Perhaps isotopic methods could do it. Girard, Prange et al 2007 showed using Xray crystallography that because the double-helix structure of DNA "acts like a spring" it is unbothered by pressures up to 2 GPa, and presciently remark "These features may have contributed to the emergence of [prebiotic] RNA World." In short, the pressure range up to 4000 atm that I'm considering in this paper, should present no inherent obstacle for life.

Although deep life has fewer cells/gram than surface soil, keep in mind that the 9997 meters below the upper 3 meters, is over 3333 times as many grams. As a result, it is quite possible that the total mass and cell-count of deep life is comparable to, or even far *exceeds*, the total mass of the sort of life humans are familar with (Whitman, Coleman, Wiebe 1998). In other words, almost all humans throughout history may well have had completely the wrong notion of what life is like.

With that prevalent observer-bias-based error about the nature of life now corrected, we may enquire: was deep life the *first* kind of life? For reasons this paper advances, I believe the answer clearly is "yes." And the observed fact that archaea dominate the deep, while gene-tree evidence suggests deep hot archaea are the oldest kind of life, is one kind of experimental support for that (perhaps not very convincing support, but support).

Two clues within the Genetic Code suggest Deep Origins. Every known form of life uses the same genetic code (up to slight variations) which is a known standard table (reprinted at the end of this paper) mapping the 4×4×4=64 DNA-base triplets to the 20 amino acids. Ribosomes continually consult this table when synthesizing proteins as directed by DNA via messenger RNA. Actually, it maps 61 triplets (with many duplicates of course) to the 20 aminos, while the 3 remaining triplets are "stop codons" indicating the protein's amino-sequence has ended.

Before biochemists deduced the genetic code, there was a brilliant but *incorrect* proposal (called "the cleverest wrong idea in biology"): Crick, Griffith, Orgel 1957's "comma-free code." CG&O pointed out that since 16<20, there exists no way to map 4×4=16 DNA-base 2-tuples to the 20 amino acids. But a "comma-free" map *does* exist from the 4×4×4=64 DNA-base triplets to the 20 amino acids. The advantage of their "comma free" property is that it makes "shifts of the reading frame" inherently impossible. (The actual way life does it is to *only* start reading genes at *correct* positions modulo 3. If wrong starts or ±1 shifts accidentally happen, then life *can* suffer reading-frame shifts, which whenever they occur are very damaging.) And furthermore, CG&O proved 20 was the greatest number for which this claim could be made!

Why did life stupidly not adopt CG&O's cleverer, better, safer genetic code? One big reason came when it was discovered that vertebrate mitochondria have their own DNA which uses a slightly *different* genetic code than their main cell chromosomal DNA. It does *not* map DNA base-triplets to 20 amino acids, but rather to **22** amino acids! Specifically, consider the three stop codons in the "usual" genetic code UAA, UGA, and UAG. In your mitochondria, UGA instead is mapped to the 21st amino acid selenocysteine, and UAG to the 22nd amino acid L-pyrolysine. Other such variations in the genetic code are surveyed by Yockey's §7.5 and wikipedia's list of genetic codes.

Another question is: Why is life so stupid that it encodes 20 aminos with 64 triplets? That's inefficient. And actually, the entropy of proteome amino-sequences is not $\log_2(20)\approx 4.32$ bits per amino, but rather [if we take advantage of observed non-uniform amino-acid frequencies (e.g. from Hormoz 2013) and also of the fact that each amino in a protein gives you useful statistical clues about what the next one will be] somewhat below 4 bits. Therefore in principle it would be possible to encode aminos with ≤2 DNA bases on average, not 3. Why doesn't life do that? The answer is that the "decoding machinery" needed to garner those savings would be so much more complicated that it isn't worth it.

OK, but why not encode consecutive-amino *pairs*? Even if we expanded the set of aminos from 20 up to 31, we still could encode the $961=31^2$ amino-pairs with 5-base DNA pentuplets, since there are $4^5=1024$ of them. This would allow bacteria to reduce the length of their DNA by $1/6\approx 16.67\%$. And prokaryotes definitely care about efficiency improvements smaller than that, which is why they typically have only 10% "junk" DNA according to Gil & Latorre 2012. Perhaps the reason life did not choose to save 16.67% in this way again merely is that this also would complicate ribosomes too much to be worth it. If that is the reason, then you can ignore the rest of my argument.

But perhaps such a 5-tuple-based code might be simple enough that the DNA savings it enables *are* worth the extra decoding complexity. In that case we must again ask: why was life so stupid? I speculate the answer is: it wasn't! The "wasteful" standard genetic code, by "foolishly" providing duplicate triplets actually grants life the boon of beneficial long-term *thermal-adaptive* and *pressure-adaptive* evolutionary flexibility. Specifically, the genetic code has this **amazing property**: *Every* "wasteful" multiple triplet codon option provides the lifeform the option of changing its DNA's CG:TA ratio. Example: The two triplets GAU and GAC both code for asparagine; one has GC:TA ratio 1:2, the other 2:1. Note the key word "every." *Every* amino besides the rarely-used tryptophan and methionine (because they each have unique codes: UGG and AUG) has multiple coding options, *always* with at least two unequal GC:TA ratios. This seems too amazing to be attributed to chance. **How amazing?** I produced 2^{30} "**random genetic codes**" by randomly permuting the 64 entries of the standard genetic code table. **99.9956%** of them failed to enjoy this property. (That same 99.9956% also is true for the vertebrate-mitochondria code table with 22 aminos.)

The GC:TA ratio governs the "melting temperature" of DNA at any given hydrostatic pressure and salinity. (DNA melts hotter at higher salinity, higher pressures, and higher GC:TA ratio.) A naively "more efficient" information-theoretic coding system would have sacrificed much of that ability of life to adapt, via changing its DNA, to temperature and/or pressure changes. So any species of deep life stuck with such a "more efficient" coding system might find itself trapped at one particular depth (pressure) and temperature range, unable to adapt its DNA to migrate out of it. Note: this "always can vary GC:TA" feature would not be useful for life growing in the ocean or in volcanic "hot vents." It *only* benefits life which originated *deep* underground whose descendants kept migrating upward and/or downward very *slowly*, adapting as they went, before eventually colonizing the surface. The key word is "slowly," meaning pressures and temperatures change at evolutionary timescales, much *longer* than cell-doubling times.

So that miraculous "GC:TA ratio can always be varied" property of the genetic code is our **first clue** that life originated deep below ground, and *not* in the ocean, in volcanic hot vents, or near the surface.

The **second clue** is the "mini-code" explained in my caption for the standard genetic code table at the end of this paper. It also is amazing: I generated 3×10^{11} "random genetic codes" as above, without finding a single one containing any GC-only 2-base→4-amino subcode in the first two positions of each triplet! Were we to allow *any* two positions, and also TA-only instead of GC-only, then this likelihood should become nearly a factor 6 greater, albeit still undetectably small. Because this brute force computational approach evidently was incapable of estimating how rare this property of the standard genetic code is, I resorted to manual combinatorial mathematics. Since there are 3 aminos (Leu, Arg, Ser) with 6 codes, 5 aminos (Val, Pro, Thr, Ala, Gly) with 4 codes, etc, the number of distinct ways to permute the 64 entries of the standard genetic code table is

$$X = 64! / (6!^3 4!^5 3!^2 2!^9) = 2316278295874198383354290667060208126761409719904245070233600000000000.$$

The number of permuted tables featuring a 16-entry GC-only "mini-code," coding for 4 aminos only, in which the amino depends only on the first two DNA bases in each triplet, then lies between

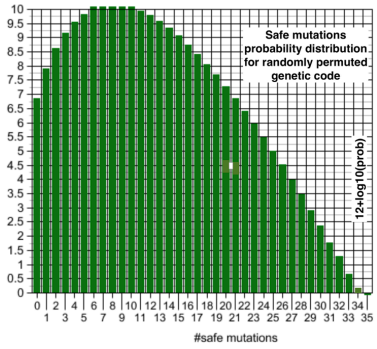
$$Y = (8!/4!)\ 48! / (6!^3 4!^1 3!^2 2!^9) = 126309750444226625719022501325866940609331200000000$$

and

$$Z = (8!/4!)\ 48! / (2!^3 4!^4 3!^2 2!^9) = 426295407749264861801700941974800924556492800000000000.$$

(Here the 8!/4!=1680 factor comes from the number of possible ordered ways to select the 4 aminos used in the mini-code from the 8=3+5 choices Leu, Arg, ..., Gly.) Consequently, the chance that my brute force Monte Carlo computer run was trying to estimate lies between $Y/X=1833807990062473425^{0-1}$ and $Z/X=5433505155740662^{-1}$, that is between 5.45×10^{-20} and 1.85×10^{-16}. With more effort, I could compute this chance exactly; but the bottom line is that >99.99999999999988% of random genetic codes have no 2-base→4-amino "mini-code" inside them.

When Nature tells you something that amazing: pay attention. However, I must confess that my "wonderful confidence numbers" 99.9956% and 99.99999999999988% above may be deceptively optimistic. The problem is that they are based on my "random permutation of the 64 entries of the standard genetic code table" probability model defining my notion of a "random genetic code." That model, although simplest, does not accurately reflect biological reality. It would if each ribosome contained a little green man who kept consulting the 64-entry table inscribed on a little stone slab to decide which amino to glue on next – then it would be reasonable to regard all possible tables on that slab as "equally likely." But in reality, ribosomal translations are performed by tRNAs, not little green men; and I do not know what an appropriate probability model of "random tRNA-mediated genetic codes" should be. But quite likely it exhibits significant differences from my randomly permuted table model.



Also confusing matters is the fact that the genetic code clearly has Darwinianly evolved to try to **maximize safety**, i.e. minimize the chance of a fatal mutation. If a big cause of fatal errors is single-base replacements inside DNA, that means the genetic code wants as many as possible of its codon-triplets for any particular amino acid, to be 1-base-replacement apart. For example, both codes AAU and AAC for asparagine, and both codes AAA and AAG for lysine, lie a single mutation apart, so the genetic code succeeded in making asparagine and lysine maximally safe. On the other hand, the three STOP codons UGA↔UAA↔UAG have only two safe mutations (shown with ↔), *not* the maximum possible number 3, so the genetic code failed to make STOP maximally safe. The total number of "**safe mutations**" (single-letter changes to a codon which still code for the same amino acid) in the standard genetic code is **68**, while the total number of unsafe single-base mutations is 3×64-68=192-68=**124**, meaning 68/192=17/48=**35.4%** of mutations are safe. To explore safety, I generated 2×10^{12} random genetic codes. Every one had a number S of safe mutations obeying 0≤S≤35, with the most common value being S=8. (The observed probability distribution of the number of safe mutations is shown in the figure at right; note the vertical axis is the base-10 *logarithm* of probability.) So the actual genetic code with S=68 is considerably safer than even the safest among 10^{12} random codes: evolutionary optimization works. Nevertheless, it is sub-optimal, as is shown by the following unreal genetic code with 72 safe mutations (72/192=3/8=37.5%, which probably *would* be optimal, and presumably a superintelligent designer – i.e. somebody at least as smart as me – would have done that):

VYLYVHLHVNLLVNLL PPPPAAAAIIIWKFFK RGTRRGTRDGTRDGTR SCMCSQQSSFFSSXXX,

shown in column-raster order with respect to the table at the end of the paper, using single-letter amino abbreviations with "X" for STOP. That kind of evolutionary drive toward safety also highly invalidates my simple "random genetic code" model's uniform probability distribution – the true distribution (e.g. of comparatively "safe codes" only) then ought to be quite *non*uniform. To explore that, I redid the GC:TA variability and mini-code investigations now using, not "random genetic codes," but rather "random 65-safe genetic codes" meaning "codes containing S≥65 safe mutations" (>33.8% safe; this is a much smaller set of genetic codes; and generating random ones, which here means "start with

a random code, then perform random non-worsening element-swaps until the code contains ≥65 safe-mutations," is algorithmically a lot slower). Then instead of 99.9956%, I find **99.52%**; and **>99.999%** of random 65-safe genetic codes have no 2-base→4-amino "mini-code" inside them. If we instead use "random 68-safe genetic codes" then these confidences become **99.23%** and **>99.999%**. Even these diminished confidence numbers remain highly convincing.

RNA hydrolysis and Hypothesized effects of high pressure

Main Hypothesis: High pressure tends to stabilize RNA (or some molecule resembling it, or some exponentially large subset of such molecules), thus permitting L.E.Orgel's "RNA world" picture of the origin of life, to happen deep underground, even though hydrolysis of RNA at time scales much shorter than the doubling time of the first life, would have refuted that picture at 1 atm pressures.

Extension: That also is true for other biochemically-important polymers (and dimers) that are thermodynamically unstable against hydrolysis in water at 1 atm pressure, e.g. polypeptides and polysaccharides (and nucleosides and nucleotides).

I do not need the Hypothesis to hold for *all* RNA molecules, just a substantial (e.g. exponentially large) subset of them. That is because if some RNA are stable and some unstable, then the pre-biotic Earth, over time, would build up large amounts of the stable ones, while losing the unstable ones, which still would be good enough (in view of our information estimates) to start Orgel's "RNA world." Also, if it did not work for RNA, but rather for some variant form (e.g. with some extra methyl groups, or TNA, etc), that could also be ok.

These hypotheses can and should be tested experimentally. I've been urging such tests for over 30 years (apparently futilely), and will sketch ways to perform such experiments in the next section. However, without such experiments, I have evidence, both theoretical and experimental (which I'll now discuss) supporting my hypotheses. That should be screaming to biochemists: *do the experiments*!

1. Correct sign: pressure *encourages* polymerization. "In practice, molecules always tend to polymerize upon compression, [called] pressure-induced polymerization (PIP)" (I quoted this highly-general rule of thumb from Fang Li, Jingqin Xu, et al 2021). As three simple examples:

1. Wieldraaijer et al. 1983 observed the polymerization of ethylene at 2.5 GPa and 330 K without needing catalyst or initiator.
2. Similar remarks apply to polystyrene.
3. Acetylene polymerizes under the action of Ziegler-Natta catalysts to form chains with alternating single and double bonds. At pressures above 3.5 GPa, this polymerization occurs at room temperature without catalyst, producing *trans*-polyacetylene (Aoki, Isuba et al 1988).

Thereotically speaking, it should be obvious that high pressure ought to encourage polymerizations of the above kinds. I.e: dimerizations A+B→AB where it seems obvious the dimer AB has smaller "volume" than A+B, *benefit* at high pressures P from an extra energy "reward" $\Delta E = P\Delta V$ where ΔV is the lost volume. Indeed, *almost every* reaction with **more reactants than products**, should become favored at high enough pressures. Two biochemically-relevant examples:

Instability of (the simple sugar) Glucose. Consider the fact that glucose ($C_6H_{12}O_6$) releases about 115 kJ/mol of energy by splitting into two lactic acid ($C_3H_6O_3$) molecules. This is the basis of anaerobic respiration inside a sprinter's muscles (which extract two usable ADP→ATP energy units as a side effect). A different kind of anaerobic respiration (performed by yeast), also yielding two ATPs, is $C_6H_{12}O_6 \rightarrow 2\ C_2H_5OH + 2\ CO_2$. So obviously, glucose is thermodynamically unstable. The decay time of aqueous sugars can be gauged by food industry recommendations:

- "If stored in a cool, dry place with the lid sealed tightly, an *opened* bottle of corn syrup usually lasts up to 2 years beyond the 'best used by' date."
- "Coca cola syrup should be consumed within 75 days from the date of manufacture."

However, under enough pressure, those reactions prefer to run backward and glucose *becomes* thermodynamically

stable against both kinds of decay. How much pressure? The densities of glucose and lactic acid (as anhydrous crystals) are 1.562 and 1.209 gram/cm^3. So the volume of a mole of glucose (180.156 grams) is 115.34 cm^3, which increases to 149.01 cm^3 if converted to lactic acid. So ΔV=33.68 cm^3 per mole of glucose. So solving for P in $P\Delta V=\Delta E\approx$115 kJ we find that P≈3.41 GPa≈33704 atm is the pressure needed to cause (solid anhydrous crystals of) glucose to become stabler than lactic acid.

We similarly could consider Glycoaldehyde (arguably the "simplest sugar," sweet-tasting, known to occur in outer space, see Zellner et al 2020) HO-CH_2-CH=O. Glycoaldehyde's chemical formula $C_2H_4O_2$ is "one-third of glucose." Its density is 1.065 gram/cm^3, even further below glucose's 1.562, and hence again glucose is energetically favored versus glycoaldehyde at all sufficiently high pressures. In fact, 1 mole of glucose has lower energy than 3 moles of glycoaldehyde even at zero pressure.

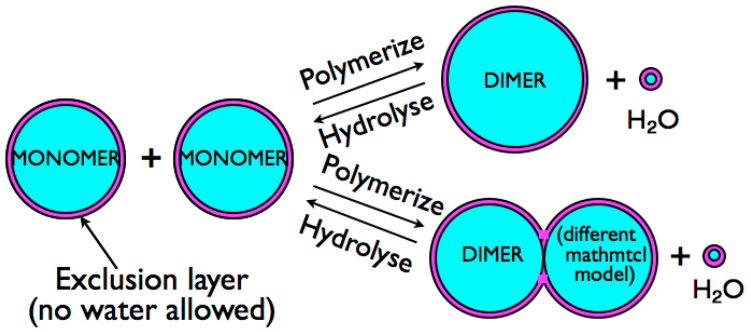
Instability of (the amino acid) Glutamic acid ($C_5H_9NO_4$). Supposedly various bacteria, e.g. Acidaminococcus fermentans (Rogosa 1969) and Steptomyces viridochromogenes (Romano & Nickerson 1958) are capable of fermenting glutamic acid to acquire energy. Supposedly they convert 5 glutamic acid and 6 water molecules to 5 NH_3 + 5 CO_2 + 6 acetic acid (CH_3COOH) + 2 butyric acid (C_3H_7COOH) + H_2. Unfortunately according to my calculations that reaction actually *loses* energy! But I think these bacteria do not really release H_2, but rather react it with something else in order to acquire a net energy gain; and/or perhaps the reaction they perform is not really understood. (For example Again since 5+5+6+2+1=19>11=5+6 this reaction would prefer to run in reverse at all high-enough-pressures. (Both this and yeast's reaction have considerably larger ΔV and hence should have much smaller turnaround pressures than the glucose→2lactic decay.)

However, examples 1-3 above (except perhaps #1, which could yield lipids in the pre-biotic solar system) probably lack *bio*chemical interest since they involve no water and invoke an order of magnitude higher pressures than those I am considering. What is more interesting for biochemistry is a different kind of polymerization and dimerization: namely in the case of dimerization, reactions of the form A+B→D+H_2O, where D is a dimer and A and B are monomers. For these reactions, since the number of reactants (2) now **equals** the number (also 2) of products, it no longer seems obvious that high pressures P will favor dimerization. Why wouldn't they instead encourage the reverse ("hydrolysis") reaction? (Or have no effect?) Read on – I soon shall answer those questions.

The polymerizations – or their time-reverse "hydrolysis" – reactions I am most interested in correspond to hydrolysis of RNA, polypeptides, and polysaccharides into nucleotides, amino acids, and mono-saccharide simple sugars. The dimerizations-via-dehydration I am most interested in are the joining of the RNA nucleo-bases (the pyrimidines cytosine and uracil and purines adenine and guanine) to the 5-carbon sugar ribose, to yield the RNA nucleosides (cytidine, uridine, adenosine, and guanosine) and then their further re-dimerization via an additional dehydration with phosphoric acid to form the RNA nucleotides: cytidine, uridine, adenosine, and guanosine mono-phosphates. (Or: we could first dimerize the sugar and phosphate to sugar-phosphate.) All of these reactions have not been seen in Miller-Urey type experiments and hence were regarded by many, including Miller himself, as "the great obstacle." (Benner 2014 called these instabilities to hydrolysis the "water paradox," with a side order of what he called the "asphalt paradox.")

The most highly studied reaction of this kind (although it is industrial, not biological) is the polymerization of polyesters from monomers of form HO-R-COOH (or from two monomers, one of form HOOC-R-COOH, the other a glycol), with release of water as a waste product. Maurer, Shaw, Smith et al 2000 found these reactions indeed happen rapidly, even without catalysts, under pressures 1-170 atm at temperatures 220-320°C.

Theoretical evidence. I now state three **simplified mathematical models** of polymerization/hydrolysation reactions involving equal numbers of reactants and products, in which I shall prove that all sufficiently high pressures will drive polymerization.



Imagine that two monomers, each balls with diameter D filled with matter at density ρ and each swimming in a water bath, combine into a dimer – a larger ball of the same density – by expelling one H_2O molecule (modeled as a ball with small diameter=2ε<D).

We assume the monomers and dimers are surrounded by thin constant thickness=ε "exclusion zones" in which no water molecule's center can lie.

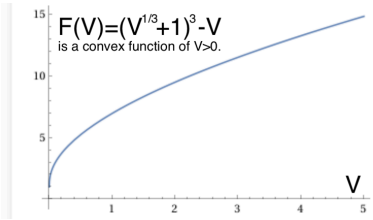
If ε obeys $0<\varepsilon\ll D$, then the summed volume of the original two exclusion zones is $2\pi D^2\varepsilon/3$ (neglecting terms of orders $\varepsilon^2 D$ and ε^3). This exceeds the volume $2^{1/3}\pi D^2\varepsilon/3$ of the final exclusion zones by an additive amount $\Delta V=(2-2^{1/3})\pi D^2\varepsilon/3$. Then because $2>2^{1/3}\approx 1.26$, we see that ΔV is *positive*.

Therefore, at all sufficiently high hydrostatic pressures P, polymerization will be energetically favored because |ΔE|, where $\Delta E=-P\Delta V<0$, will exceed any fixed chemical bond energy penalty (if any).

Note that even for possibly-differing monomers, it *always* is a theorem that the final volume (including exclusion layer) is less than the original volume provided the two monomers each exceed the size of a water molecule. Basically, this is a consequence of the concave-∩ nature of the surface-area function $\text{SurfArea}(V)=V^{2/3}$. To be more formal about it:

Theorem: If $A+B=X+Y$ with $0<A<X<Y<B$, and $\delta>0$, then

$$(A^{1/3}+\delta)^3-A + (B^{1/3}+\delta)^3-B < (X^{1/3}+\delta)^3-X + (Y^{1/3}+\delta)^3-Y.$$



Proof: This follows from Jensen's inequality in view of the fact that the function $F(V)=(V^{1/3}+\delta)^3-V$, which is the volume of the region between two concentric spheres, the inner one having volume V, the outer with radius expanded by the constant additive amount $\varepsilon=(4\pi/3)^{1/3}\delta>0$, is a *concave*-∩ function of $V\geq 0$ (see figure; this may be shown by using the binomial theorem with exponent=3 and then the differential calculus). **Q.E.D.**

Model #2: The monomers are not *balls* of diameter D, but rather *cubes* of sidelength D, and we model the dimer as a 2×1×1 rectangular brick (scaled so the products have the same volume as the reactants). Then the additive change $\Delta V\approx D^2\varepsilon$ in the total exclusion-layer volume due to hydrolysation again must be positive.

To be a little more formal about it: The volume of the region within distance ε from an A×B×C rectangular brick is

$$V(A,B,C) = ABC + 2(AB+BC+AC)\varepsilon + (A+B+C)\pi\varepsilon^2 + 4\pi\varepsilon^3/3.$$

Model #3: The dimer is not one larger ball, but rather *two tangent* monomer-balls. Even in this model, the final two exclusion zones must have smaller total volume than the original two, but now with $\Delta V\approx\pi\varepsilon^2 D$, provided ε is sufficiently small compared to the monomer sizes, i.e. if we as usual neglect terms of order ε^3.

For readers who want exact formulas without any approximations: I deduce from the formula for "sphere-sphere intersection volume" given by Eric Weisstein in his math encyclopedia (his EQ 16) that the volume V of the region within distance ε from the union of two interior-disjoint tangent balls of radii A and B, is

$$V(A,B) = (4\pi/3)([A+\varepsilon]^3+[B+\varepsilon]^3) - 4\pi\varepsilon^2\, AB / (A+B) - 4\pi\varepsilon^3 / 3 ;$$

and if $0<\varepsilon<A\le B$ and $7\varepsilon\le 3AB/(A+B)=3/(A^{-1}+B^{-1})<3A$, then I find that this function obeys the inequality $V(A,B)+V(\varepsilon,0) <V(A,0)+V(B,0)$. Indeed

$$V(A,0) + V(B,0) - V(A,B) - V(\varepsilon,0) = 4\pi AB\varepsilon^2/(A+B) - 28\pi/3\ \varepsilon^3$$

which plainly is positive for all ε obeying $7\varepsilon<3AB/(A+B)<3A$ and hence certainly for all ε with $\varepsilon<3A/14$.

Polypeptide experimental evidence: Otake, Taniguchi, et al 2011 found that the two simplest amino acids glycine and alanine are stabilized by high pressure, and even without catalysts will form polypeptides up to 5 aminos long at temperatures 180-400°C and pressures 1.0-5.5 GPa. (Kawamura & Yukioka 2001 had already detected dimerization of alanine at 225-275°C at pressures up to 100 atm.) Does this stop at 5, or should we expect much longer lengthening if they'd waited longer, such as millions of years? Obviously, the latter – but in case that was not obvious enough, further insight about that is provided by the "sausage catastrophe." That is: suppose you want to pack N equal interior-disjoint balls in 3-space to try to minimize the volume of their convex hull. If N≤55, the best packing appears to be to make all their centers lie on a line, with spacing equal to the ball-diameter ("sausage" configuration). But if N=56, and for all N>65, there is a "collapse" and the best packings become roughly spherical "blobs." My point is that simpleminded mathematical models like my three above, already indicate that energy-rewards that increase with N will come from N-mer-ization for every N≥2 at all high-enough pressures P, even for polymers that geometrically are collinear "sausages." If those sausages collapse once N gets large enough (via "protein folding") then we expect *additional* pressure-P energy-rewards.

Does that only work for glycine and alanine, or should we expect the same pressure effect for the other 18 amino acids? Well, glycine and alanine are the two amino acids with the fewest atoms, namely 10 and 13. The atom-counts for the other 18 range from 14 for serine and cysteine up to 27 for tryptophan. Our simplistic mathematical models of amino acids as, e.g, "balls larger than a water molecule," therefore predict (since the other 18 are "larger balls") that this pressure effect ought to work *better* in all the other 18 cases! And indeed, based on tabulated crystal densities and molecular weights: glycine, alanine, serine, and cysteine are the four aminos with smallest molar volumes, while the three with the greatest molar volumes are Arg, Phe, Trp (all 7 in increasing-volume order). On the other hand, glycine and alanine both are among the 9 "hydrophobic" aminos, presumably enjoying wider "water exclusion zone" widths ε in our simplistic models. So it would be more conservative only to predict this behavior for the 9 hydrophobic aminos Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val. However, Zamyatnin 1972 provides displaced-water molecular-volume estimates for aminos *in aqueous solution*, finding the four with least volume are Gly, Ala, Ser, Cys, while the two with greatest volume now are Lys and Arg (all 6 in increasing-volume order) – suggesting you do not need to be "conservative." Amino acid molecular volumes range from 2 to 8 times the volume of a water molecule. (Of course, experiments would be more desirable than these theoretical arguments.)

Polysaccharide evidence (or at least, "non-evidence for failure"!): If plant starches, such as pea starch, are placed under high hydrostatic pressure (4000 atm) at room temperature, then they do *not* hydrolyse, but their properties change, e.g. they "gelatinize." This is believed to be because the starch no longer crystallizes, but rather becomes amorphous with "hydrogen pseudobonds" to nearby water molecules. My point: this suggests the *sign* of the effect does *not* favor hydrolysis at high pressures.

2. Correct order of magnitude: P≈4000 atm causes |ΔE|=|ΔV|P energy shifts of the *same order* as RNA, DNA, peptide, and polysaccharide hydrolysis energies. Under standard biochemical conditions (1 atm, room temperature)

1. "The hydrolysis of peptide bonds in water releases 8-16 kJ/mol (2-4 kcal/mol) of Gibbs energy" according to wikipedia. (This should be compared with $k_BT\approx 2.5$ kJ/mol whene T=300°K.)
2. The energy release from hydrolysing one RNA or DNA phosphodiester link is about 5.3 kcal = 22.2 kJ per mole of bonds cleaved, according to Dickson, Burns, Richardson 2000.
3. "The addition of glucose to glycogen is an endergonic process that requires energy" (and hence hydrolysing glycogen into glucoses is an exothermic process) according to a book titled Glycogenesis. And according to Tombari et al. 2007, "On hydrolysis, the enthalpy decreases by 14.4 kJ/mol of sucrose at 310°K." That is 42 joules per gram. For instead hydrolysing trehalose, the enthalpy decrease is 4.73±0.41 kJ/mol, according to

Tewari & Goldberg 1991.

The sign-fact that all these polymers are thermodynamically unstable to hydrolysis in 1 atm room temperature water, also is obvious once you know that animals digest starches, proteins, and RNA/DNA by simply putting them in contact with amylases, proteases, and nucleases in saliva and digestive juices, to passively catalyse their decay into monosaccharides, amino acids, and nucleotides (which the animal then absorbs). When, on the other hand, animals and plants *synthesize* these polymers from monomers, they consume energy in the form of ATP to drive the reaction. Namely DNA polymerases input ATP and other nucleotide triphosphates, and actually cut off *two* phosphates (i.e. going from tri- to monophosphate) for extra drive (since cutting off just one would cause too large error-rates; organisms like to be very careful with their DNA). Protein sythesis in ribosomes consumes one ATP→ADP per peptide bond formed. Glycogen synthesis consumes one UTP→UDP per glucose added.

To take that argument further: it also is obvious that, at room temperature and pressure, the sugar-phosphate bond in nucleotides is unstable to hydrolysis because of the existence of "nucleotidease" digestive enzymes which break that bond with no energy input needed. Similarly the sugar-base bonds in nucleosides are unstable to hydrolysis because of the existence of "nucleosideases." The latter reasoning is confirmed by Carnici, Sgarella et al 1980's measurement of "the standard Gibbs free energy of hydrolysis" for ribose-1-phosphate to be -5.4±0.15 kcal/mol = -22.6±0.6 kJ/mol.

A rough way to estimate the absolute value of the size of the effects of pressure on these energy changes E is simply to look up tabulated *densities* and use them to compute volume changes |ΔV|, for some well-studied sample reactions involving only well-characterized chemicals. Then our rough size estimate is |ΔE|≈|ΔV|P. We now perform some calculations of this kind.

Hydrolysing sucrose (disacccharide) to fructose+glucose monosaccharides. Density of glucose($C_6H_{12}O_6$)=1.562 gram/cm^3. Density of fructose($C_6H_{12}O_6$)=1.694 g/cm^3. Those each have molecular weight=180.156. Density of sucrose($C_{12}H_{22}O_{11}$)=1.587 g/cm^3. Molecular weight=342.2965. Water+sucrose molec.wt.=360.3118 since water's molecular weight is 18.0153. The volume of 1g glucose + 1g fructose = 1/1.562+1/1.694 = 1.2305 cm^3. The volume of the corresponding amount of sucrose+water is 2((18.0153/360.3118)/1+(342.2965/360.3118)/1.587) = 1.2972 cm^3. Since the latter is greater, that indicates that pressure *encourages* hydrolysis – "wrong" sign! However, that is not relevant biologically because lifeforms do not store their fructuse, glucose or sucrose as solid crystals, but rather in aqueous solution, and then the volume effects of solvent attraction, repulsion, and expulsion must also be considered, in which case you'd realize that all high-enough pressures instead encourage sucrose to form, despite hydrolysis being thermodynamically favored at 1 atm pressure. Here I will ignore all that, and the wrong sign; I am interested purely in the *size* of |ΔV|, which for 2 total grams of mono-saccharides is |ΔV|=1.2972-1.2305=0.0667 cm^3. Then the hydrostatic pressure P required to cause |ΔV|P to reach 84 joules (the enthalpy release from hydrolysing 2 grams of sucrose according to Tombari et al 2007) is 1.259 GPa ≈ 12430 atm.

Hydrolysing trehalose (disacccharide) to two glucoses. Trehalose ($C_{12}H_{22}O_{11}$) has density=1.58 gram/cm^3 and molecular weight 342.296 grams/mol. ΔV = 342.296/1.58 + 18.015/1 - 2×180.156/1.562 = 3.985 cm^3/mol, again with "wrong sign," which again is irrelevant because lifeforms do not store their glucose and trehalose as anhydrous crystals. The standard enthalpy of hydrolysis of trehalose is ΔE=-4.73±0.41 kJ/mol, according to Tewari & Goldberg 1991. The pressure estimate P≈|ΔE/ΔV| is 11700 atm.

Hydrolysing L-Alanyl-L-glutamine to alanine and glutamine. L-Alanyl-L-glutamine is a commercially sold dipeptide with formula $C_8H_{15}N_3O_4$ and molecular weight 217.225. Its density allegedly is 1.305 g/cm^3. Its two component amino acids alanine ($C_3H_7NO_2$, molecular weight 89.094) and glutamine ($C_5H_{10}N_2O_3$, molecular weight 146.146) have densities 1.424 and 1.47 g/cm^3 respectively. (Again the sign is "wrong," which again is irrelevant because lifeforms do not store their aminos and dipeptides as anhydrous crystals. Incidentally, if these quoted densities pertain to racemic mixtures rather than pure single-chiral enantiomers, that does not matter for our crude purposes of merely computing order of magnitude estimates.) The volume of 1 mole of alanine therefore is 89.094/1.424=62.566 cm^3 and of 1 mole of glutamine is 146.146/1.47=99.419 cm^3. Meanwhile, the volume of 1 mole of Alanyl-glutamine is 217.225/1.305=166.456 cm^3 and the volume of 1 mole of water is 18.015 cm^3. So |ΔV|=184.471-161.985=22.486

cm^3 per mole. In order to cause |ΔV|P to reach peptide-bond hydrolysis enthalpies of 8-16 kJ, we need P=356-712 MPa=3511-7022 atm.

Hydrolysing uridine (nucleoside) to uracil and ribose. Ribose ($C_5H_{10}O_5$) has density=1.1897 gram/cm^3 and molecular weight=150.13. Uracil ($C_4H_4N_2O_2$): density=1.32, molecwt=112.08676. Uridine ($C_9H_{12}N_2O_6$): density=0.99308, molecwt=244.20. Water (H_2O): density=1.00, molecwt=18.015. . We find ΔV = 244.20/0.99308 + 18.015/1 - 112.08676/1.32 - 150.13/1.1897 = 52.81 cm^3/mol. (As usual the sign is "wrong," which again is irrelevant because lifeforms do not store ribose, uracil, and uridine as anhydrous crystals.) This ΔV would at P=7400 atm amount to |PΔV|=39.6 kJ/mol, which is roughly the correct value for this hydrolysis |energy| based on claimed enthalpies of formation for all 4 chemicals.

Hydrolysing D-ribose 5-(dihydrogen phosphate), "5RP," to phosphoric acid and ribose. Phosphoric acid (H_3PO_4): density=1.6845, molecwt=97.994. 5RP ($C_5H_{11}O_8P$): density=1.803, molecwt=230.11. We find ΔV = 150.13/1.1897 + 97.994/1.6845 - 230.11/1.803 - 18.015/1 = 38.72 cm^3/mol. (Finally an example with the "right" sign: Ribose+H_3PO_4 has greater volume than 5RP+water.) This ΔV would at P=1450 atm amount to PΔV=5.69 kJ/mol, agreeing with the value for this hydrolysis energy measured by Tewari, Steckler, et al 1988.

Conclusion: All five of these calculated |pressure| estimates range between 1450 and 12430 atm, i.e. indeed have the *same order of magnitude* as 4000 atm; and (helpfully, if this is not a delusion) the pressure estimates are smaller for the RNA-related cases than they are for proteins and sugars. **But** there are several problems:

i. It is rather jarring that four of our five ΔV calculations (based on anhydrous crystal densities collected from chemical datasets on the internet) gave the *wrong* sign.
ii. These all were merely crude theoretical estimates, not experimental measurements.
iii. When I said "same order of magnitude as **4000 atm**" that was an uncomfortably large number. It would be better if only 1500 atm were needed. (Which it might be.)

Our excuse for (i) is that our calculations about *anhydrous crystals* are irrelevant to biological reality, so their wrong signs hopefully mean nothing, and are useful only as signless "order of magnitude estimates." Even if their signs might have a little to do with biological reality, one expects when tossing 5 coins to get ≥4 heads with chance 6/32=3/16=18.75%, suggesting those "4 wrong signs" were a statistically insignificant reason for worry. Re (ii&iii): Until high pressure **experiments** – or experiments with very careful volume-change measurement (which might be possible using ultra-precise laser interferometric techniques to measure the position of a water surface pre- and post-hydrolysis) – are done, the proposition that in real *aqueous solutions* (which is where all hydrolysis enthalpies were measured) the ΔV signs become *right*, while its absolute value's order of magnitude usually remains about the same, and correspond to values achieved in early life, are merely plausible imprecise hopes.

Importance of *thermodynamic* stability. The pre-life period on Earth was ≥100 Myr – more than long enough for mixing to occur between the atmosphere, oceans, and "deep biosphere" region; and for some regions to concentrate prebiotic chemicals. The longer the timespan, the more *thermodynamic* stability matters. If there is a synthesis pathway (even one with kinetics thousands or even millions of times slower than the rates of enzyme-catalyzed reactions, or the reactions observed by the likes of Miller & Urey) to RNAs, polypeptides, etc, then they are going to accumulate in places where they are stable. (I.e, under our hypothesis, high-pressure places.) Unstable chemicals will not be there unless continually getting abiotically synthesized by processes fueled by energy sources like sunlight, lightning, radiation, and chemical energy frozen into primordial rocks at rates comparable to or exceeding decay rates. Essentially *all* macromolecules inside today's life are thermodynamically *un*stable under Earth-surface conditions; and even if they can be synthesized abiotically, that is slow. *That makes it clear that the first life could not have arisen near Earth's surface. It must have arisen deep, and my Hypothesis must be true.*

"Synthetic paths"? S.N.Platts in 2004 proposed "PAH world," while A.G.Cairns-Smith earlier had proposed clays, both as possible ways synthesis of RNA or something like it (e.g. GNA, PNA, TNA) could have been catalysed. Although both sound mildly plausible, as of year 2025, neither the clay or PAH theory has managed to acquire any experimental support. Clays occur everywhere and consist of layers of alumino-silicate (plus impurities) a few atoms thick with water between, providing enormous total surface area. The layers contain charged atoms, which can electrostatically attract stuff. These are known to be capable of concentrating chemicals dissolved in the water and

catalyzing reactions among them. There are an enormous number of possible impurity-atom patterns on clay surfaces, presumably capable (if you find the right pattern of the right impurities) of catalyzing virtually anything. For this reason I am not worried about the problem of finding a synthetic path. It seems to me, virtually *any* thermodynamically-favored reaction, is going to have synthetic paths *somewhere* in the vastness of worldwide clay sediments over 100 Myr of earth history. In particular, if something catalyzes "add one more monomer" then we will get polymers whenever they are thermodynamically favored; and that something, if a lucky randomly-created clay pattern, almost-certainly would exhibit steric favortism (e.g. preferring right-handed monomers), explaining why life employs chiral molecules usually with only one handedness.

Two kinds of observational proof of my Hypothesis. Although, as far as I know, nobody has done the high pressure experiments I want, there nevertheless are senses in which those experiments already have been done and the results prove my Hypothesis!

A. Observers have claimed that

1. The "deep biosphere," e.g. water leaking into 2 mile deep South African gold mines, contains many active prokaryotes and viruses (Kyle, Eydal, et al 2008; Eydal, Jägevall, et al 2009). Engelhardt, Sahlberg, et al 2011 found "functional viruses" in 420m-long drill cores below 150 to 5100m-deep seafloors, finding virus:cell ratios 3-21 that *increased* with depth, which they hypothesize is due to *greater preservation of virus particles at greater depths*.
2. In their natural environment, especially in the absence of nutrients tracing to photosynthesis, the doubling times of those bacteria are ≳1000 years.

Think about that. Viruses are made of either RNA or DNA, usually inside a protein envelope. They cannot self-repair. They must last for timespans at least comparable to the host prokaryote doubling time before they get to infect their next victim. *These viruses plainly are lasting ≳1000 years*.

I conclude from this (even merely the observation of numerous *dead* viruses would suffice!) that both proteins, and either DNA or RNA (or both) are *stable* for ≳1000 year time spans at observed depths underground. The hypothesis is validated.

B. There also have been many publications finding *living* archaea trapped inside fluid inclusions inside deep buried NaCl crystals, 34kyr, 97kyr, 150kyr, 123Myr, and 250Myr old, as well as nonviable(?) archaea 830Myr old. (See special references subsection on this topic.) Those findings have been disputed by those claiming "it must have been contaminants introduced by shoddy lab procedures." However, these findings have kept coming, over and over again over decades, from disjoint author teams, reproduced by multiple labs, using standard well-accepted decontamination procedures. Normally in science, results reproduced several times by independent authors are accepted.

Graur & Pupko 2001's "shoddy contamination" criticism of Vreeland, Rosenzweig, Powers 2000 was based on the claim VR&P's "250 Myr old" halophile had 16S RNA too similar to a modern genome's (only two bases differed), based on this G&P quote: "Rate of substitution for 16S ribosomal DNA in prokaryotes is remarkably uniform among diverse lineages and ranges between (1 and 5)×10^{-8} substitutions per site per year." If that G&P quote were true, then *every* site would get substituted in 16S RNA after 1 Gyr, rendering the entire use of 16S RNA sequence phylogenetics by a vast community of people to trace out the multi-Gyr-old origins of life, in particular *all* today's lifeforms that Ciccarelli, Doerks, et al 2006 claim lie within a few tree-edges from the tree-root, completely useless. In Ochman, Elwyn, Moran 1999 we find "a rate of 16S rRNA divergence of 1-2% per 50 million years" which is 25-250× slower than the rate-range claimed by Graur & Pupko. Also, Karnachuk et al 2019 found, 2km down a Siberian borehole, the archaean Desulforudis audaxviator which is killed if exposed to oxygen, which they found to have remarkably similar genome to another sample collected by a different team from water 2 miles deep in a S.African gold mine, despite these two populations having been isolated from each other for "hundreds of millions of years." They attributed that remarkable genomic similarity either to some amazing unsuspected seemingly-impossible mechanism for worldwide distribution, or to very small mutation rates in deep very-slow-dividing archaeans. In view of these two counter-facts, G&P's criticism loses some or all of its force.

The trapped halophiles also have been disputed *theoretically* by those claiming "DNA and other important organic biomolecules could not have survived longer than a few centuries." No satisfactory answer has ever been provided to

that theoretical complaint – until now. Why do I say "until now"? Because not one among all those complainers ever had even one sentence considering the effects of *pressure* and *salinity* on those chemical instabilities! All those halite samples came from 500-1550m deep underground, and hence from under 130-400 atm pressure, and (obviously) maximum possible salinity. Only one way has been proposed to allow the two sides in that dispute *not* to be locked in irreconcilable conflict: my Hypothesis, which is therefore confirmed. (In mathematical parlance: "proof by contradiction.") The more certainty those two conflicting sides proclaim, the more confidence we acquire that they jointly experimentally confirmed my Hypothesis.

High pressure techniques & phenomena, and suggested experiments

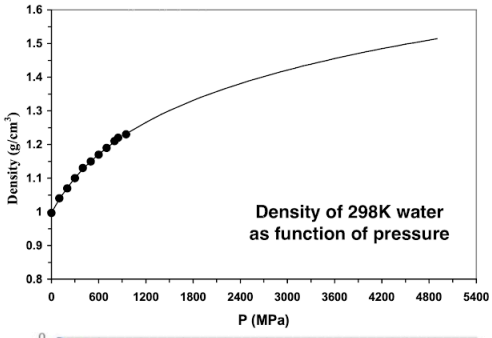
Helpfully for life, **water is a rather remarkable liquid**. First of all, hydrogen and oxygen are among the most common atoms, making H_2O very common. Second, water is very "polar" thanks to its bent molecular structure and high H-O electronegativity difference. This causes ions, and molecules containing charged atoms, to be "hydrophilic" and to tend to be soluble. It also causes "hydrophobicity" effects for, e.g. hydrocarbons. Those two effects are essential for cell-wall formation/maintenance and the self-assembly of "amphiphilic" molecules into vesicles observed by Luisi and Deamer,. It also is crucial for other "self assemblies" such as virus particles and ribosomes, and for automatic "protein folding." Third (relatedly), water has very high heats of vaporization and freezing, causing its liquid state to be comparatively stable. Fourth (relatedly), ice is lighter than liquid water, causing it to rise – as opposed to almost all other liquids, which become denser when they solidify. If water did not enjoy that property, then the oceans would freeze from the bottom up rather than top down, then stay frozen forever, causing Earth to be far more hostile to life, and quite possibly preventing life, or at least today's sophisticated multicellular surface life, altogether. Some people have proposed the idea that somewhere in the universe, there is life *not* based on water as the solvent, but rather on liquid ammonia NH_3. Ammonia also is polar. However, I object:

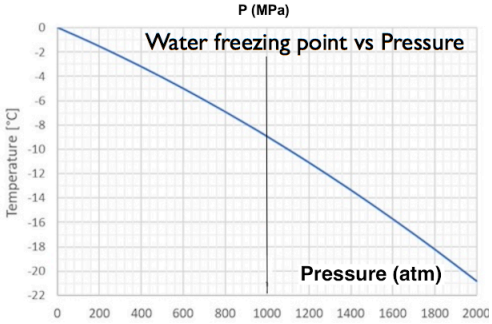
1. Oxygen has "doubly even" atomic nuclei, indeed with O-16 being "doubly magic," causing oxygen-16 to be 10× more common in the universe (and perhaps 2000× more common on Earth) than nitrogen (for essentially the same reason, C-12 is 99× more common than C-13).
2. Solid ammonia is denser than liquid ammonia, so on any planet with oceans made of ammonia, those oceans would freeze from the bottom up, then stay almost fully frozen forever.

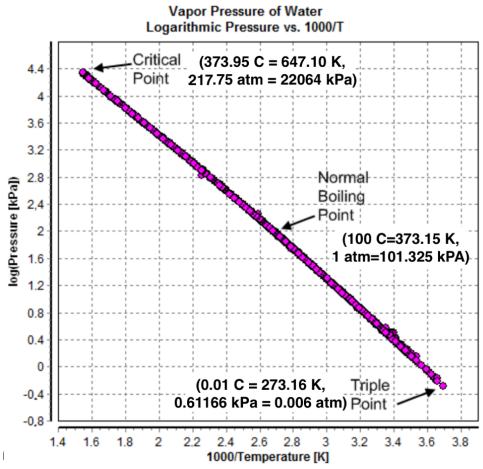
Other possible solvents include liquid CO_2 (which can only exist at pressures≥5.2 atm and temperatures≤216.6°K) liquid N_2, or simple liquid hydrocarbons. However, all these are nonpolar hence sacrifice the very useful ability of polar liquids to generate "hydrophobic/philic" pseudoforces; and again the solid is denser than the liquid so any oceans made of these would freeze from the bottom up.

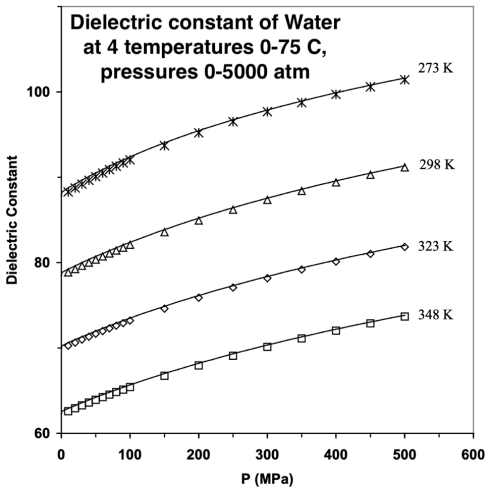
Now consider the data plots and formulas below about the **effects of pressure on water**. We see that within the pressure range 0-4000 atm of interest to us, greater pressure *increases* the polar nature of water (as quantified by its dielectric constant relative to that of vacuum) thus *enhancing* hydrophilicity effects. Pressure also strengthens the excluded-volume effects responsible for hydro*phob*icity. Therefore, genesis-friendly effects, such as lipid-vesicle self-assembly (and the so-called "discotic self-assembly" that S.N.Platts hypothesized as the crucial idea in his "PAH world"), and protein-self-folding and RNA-self-folding reliability, and the effectiveness of any catalysis that is due in part to hydrophobicity/philicity effects (including those hypothesized by A.G.Cairns-Smith as part of his "clay world") all should be enhanced at high pressures. The boiling point of water also increases with pressure from 100°C at 1 atm to 374°C at 218 atm. Also, water's freezing point decreases from 0°C at 1 atm to -21°C at 2000 atm. So high pressures increase the stability of *liquid* water.

Implications for life & genesis. All those effects should enable high-pressure life to survive much hotter and colder temperatures T than 1 atm life. That also should greatly improve the chances for genesis both by increasing the set of allowed temperatures and because, as I'd said before, a rule of thumb is that biochemical reaction rates double for each 10°C temperature increase. [More accurately, the logarithms of reaction rates often should be straight lines if plotted versus 1/T, the same behavior exhibited by the logarithm of water's vapor pressure, which makes sense if we regard water's "vapor pressure" as the $H_2O(\ell)\rightarrow H_2O(g)$ "reaction rate."] Therefore, the hotter temperatures enabled by increased pressures should be very genesis-friendly.









The **vapor pressure** P of water in kPa, as a function of temperature T in Kelvin, is approximated by $\log_{10}(P)\approx 7.70164-2.15176X$, and better by $\log_{10}(P)\approx 7.13375-1.65824X-0.0952593X^2$, where X=1000/T. [I chose this quadratic(X) to get perfect agreement with experiment at the triple, critical, and standard boiling points; my linear(X) yields imperfect agreement at these three points, but with pressure relative error≤18%.] Above the critical temperature there no longer is a clear distinction between "liquid" and "gaseous" water. Water's dielectric constant ε (in units such that vacuum has ε=1) falls with rising temperature T, but rises with pressure P. Floriano & Nascimento 2004 gave the formula (compared with data in their plot) ε(T,P) ≈ ε(T,10MPa) + C·ln((B+P)/(B+10)) which reproduces available data for dielectric constant to 0.1% accuracy for pressures P in MPa between 0.1 and 500 at T=273-348°K = 0-75°C using parameters in the table below:

T(°K)	ε(T,10MPa)	C	B
273	88.28	11.9240	232.5026
298	78.85	14.1113	341.5902
323	70.27	14.3899	386.4825
348	62.59	13.9450	393.2031

For the **density** ρ of water in gram/cm^3, they gave ρ(T,P) ≈ ρ(T,0.1MPa) + C·ln((B+P)/(B+0.1)) with error≤0.1% for pressures below about 10000 atm, where with T=298°K the parameters are:

T(°K)	ρ(T,0.1MPa)	C	B
298	0.997	0.214	476.693.

Equivalently, the 298°K pressure P depends **exponentially** on density ρ in this range: P≈0.314exp([1000ρ-999]/476673)-0.214.

Salinity as pseudo-pressure. Dissolving ionic salts (most importantly in nature, NaCl) in water causes many of the same effects as pressure. At low salt concentrations, freezing point lowering is $\Delta T\approx K_{WaterFreeze}M$ where M is the number of moles of *ions* per liter (for NaCl, *twice* the number of moles of NaCl per liter, but *three* times the number of moles if the dissolved salt instead were $CaCl_2$) and $K_{WaterFreeze}$=–1.86°C per mole. Thus from the fact that seawater freezes at -1.8 to -2.0°C we deduce its salinity is about 0.51±0.025 moles of NaCl per liter; direct measurements show that seawater has about 35 grams salts per kg, which in view of the molecular weight 58.44 for NaCl would be 0.60 moles per kg. Thus we see that (at least as far as the effect on freezing point is concerned) 100 atm of pressure is "approximately equivalent" to a salinity of 0.45 moles NaCl per liter. (Saturated brine contains 6.6-8.1 times as much salt as seawater, boils at 109°C, and freezes at -21°C.) The same formula also governs the increase in saline water's boiling point, except now with $K_{WaterBoil}$=+0.51°C per mole of ions. Thus we see that (at least as far as the effect on boiling point is concerned) increasing the pressure from 1.0 to 1.1 atm (causing water to boil at 102.9°C) is

"approximately equivalent" to adding 2.8 moles NaCl per liter of water.

Also similarly to pressure:

- Saltier water is denser. (1 mole/liter of NaCl increases the density roughly the same as 2000 atm pressure.)
- One might naively imagine that saltier water has greater dielectric constant, thus increasing hydrophilic/phobic effects (which would be the same effect caused by pressure): If we imagine two insulated metal plates, forming a capacitor, immersed in water, then: charging the two plates to positive and negative voltages will cause Na+ and Cl- ions in the water to migrate so (Na+)ions are near the (-)plate, and (Cl-)ions near the (+)plate; thus salt effectively makes the water "more polar," increasing the capacitance. That is certainly true, **but nevertheless** (Hasted, Ritson, Collie 1948; Gadani et al 2012; Gavish & Promislow 2016), at given frequency, the measured dielectric constant of water **decreases** slowly with increase in salinity from 0 to 3.5% (this sign-reversal versus the naive expectation is known as "dielectric decrement"), and also decreases slowly with increase in frequency from 0 to 1.4 GHz for given salinity. Dielectric decrement is attributed to the disruptive effect of the local electric fields generated by the ions: polar water molecules arrange into a shell around each ion, aligning with the local ionic field and diminishing their responsiveness to the external field, thereby reducing the dielectric constant. As far as the effect of salt on zero-Hertz dielectric constant is concerned, 1 molar NaCl is roughly "equivalent" to *negative* 3000 atm of pressure (based on fig.1a from Gavish & Promislow).
- Considering the speeds and sizes of molecules, atoms, and electrons, what biochemistry on molecular or smaller length scales probably cares most about is *not* the dielectric constant relative to vacuum at zero Hertz, but rather at about the frequency of red light (4×10^{14} Hertz), i.e. the **squared index of refraction n^2** of water. And n in fact, does rise with salinity, i.e. there is no sign-reversal anymore, so salinity again acts like pseudo-pressure. For temperatures T in °C ($0\le T\le30$), light of wavelength λ in nm ($400\le\lambda\le700$), and salinity S in grams NaCl per liter ($0\le S\le35$), Christopher Parrish in 2020 empirically found
$$n \approx -(1.97812-0.006883S)\times10^{-6}T^2 + (1.03223+0.001069S)\times10^{-7}\lambda^2 - (8.58125-0.01433S)\times10^{-6} - (1.54834+0.001041S)\times10^{-4}\lambda + 1.38919+0.0001826S.$$
This predicts that at T=25°C and λ=400 nm, the index of refraction of water rises from 1.33014 for pure water to 1.33987 with 1 mole/liter of NaCl. About that same Δn=0.00700 is caused by 100 atm pressure (fig.2 of Weiss, Tazibt, Tidu, Aillerie 2014).

In summary: Pressure increases refractive index and both hydrophobic and hydrophilic effects; and increases water's density and boiling point; but depresses freezing point. Salinity has all those same six effects, *but* with the asterisk that it decreases water's refractive index n at much lower frequencies. The "**crossover frequency**" at which salinity stops decreasing n and starts increasing it probably lies somewhere in the microwave range, perhaps F_{cross}=10-100 GHz. For the effects we've discussed, 1 mole/liter NaCl has the same effect as anywhere from 0.03 to 2000atm extra pressure (depending on which effect we are speaking of); and indeed salinity directly causes "osmotic pressure" $P\approx MRT$ in the presence of a membrane permeable to water but not ions, where M is the concentration of dissolved ions in moles/liter, R≈8.314 joule/(mole·kelvin) is the ideal gas constant, and T the absolute temperature. At T=298°K (25°C) this means water with 1 mole/liter of NaCl would generate 49 atm of osmotic pressure versus pure water. I believe F_{cross} lies below the frequency range important for biochemistry at molecular and smaller scales, *but* for effects at length scales much *larger* than molecular, e.g. Luisi and Deamer's "self assembly of bilayer vesicles" from 10^8 amphiphilic molecules in water, and osmotic pressure across macroscopic membranes, low frequency dielectric constant should matter. Luisi & Deamer's vesicles have about the same diameter as bacteria, i.e. 10^{-6} meter, 100× larger than even the largest proteins. So this would predict that some such vesicle self-assemblies could stop working if enough ions are added to the water, thus reducing the hydrophilic/phobic forces that drive that assembly on micron length scales. This prediction is true: "fatty acid vesicles are extremely sensitive to metal ions (especially divalent ions like Mg^{2+})" according to Joshi, Steller et al 2021 (effect already noted and worried about by Monnard, Apel et al 2002, but neither had considered the likely possibility that high pressure should be able to overcome that problem).

The pressures P≤4000atm I am considering in this paper are technologically easy to achieve using only fairly common steels, without need for superhard materials. They are tiny compared to the world record sustained pressures achieved with "diamond anvils" (over 10^6 atm!) and also small compared even to the pressure range 3000-30000 atm explored by Percy W. Bridgman using "carboloy" tungsten carbide composite chambers he built at Harvard in the 1910s. (I recommend Bridgman's 1949 book.) Nevertheless, 4000 atm is no joke and equipment at those

pressures should be treated similarly to a loaded gun or live grenade.

Rifle barrels can withstand pressures of 1000-5000 atm. The Sporting Arms and Ammunition Manufacturers' Institute (SAAMI) sets pressure limits for different calibers to ensure safe and standardized ammunition. To fire the 9mm Luger cartridge, for example, your gun barrel needs to be designed to handle 2300 atm. For 5.56×45mm NATO rounds, two pressure specs are 3600 and 4200 atm.

Bridgman's first main invention was "Bridgman seals" for the ends of plungers, which prevent fluid leakage because the more you pressurize the fluid inside the chamber, the more determinedly it seals against leaks. A second was his idea of shaping his pressure chambers like a cone-frustum, so that as they were pressurized by a downward-pushing plunger, they would be pressed into a cone-shaped depression in a massive support, and therefore supported *more* against explosion, the more pressure was applied. This trick enabled designs largely *preventing* pressure-dependent distortions. (Bridgman's third main improvement was switching to "carboloy" instead of hard steels to enable reaching higher pressures, but we shall not need that idea.)

Suppose you purchase a Baleigh industrial Hydraulic Shop Press (HSP-10H; stock number BA9-1004779, price $760.00) which weighs 160 lb and has 7-inch stroke and 13.5 inch width. It will exert up to 10 tons of force, manually powered by a pump-lever. If you build an appropriately mounted and supported gun-barrel-like cylinder and plunger (both hardened steel) with a Bridgman seal, then by pushing the plunger into the barrel with the hydraulic press, you should be able to create pressures up to 4000 atm.

Approximately 2000 atm pressure is achieved by freezing water to -21°C in a (rigid and strong enough) sealed vessel.

Newport Scientific Inc. formerly sold compressors capable of up to 2000 atm, e.g. these two models:

```
P/N 46-13416-2    30000 PSI (2000 atm)   SINGLE-END          net weight 254, ship wt 305 lb   20x22x10in
P/N 46-13427-2    30000 PSI (2000 atm)   TWO-STAGE (10-30)   net weight 286, ship wt 327 lb   29x22x19in
```

both powered by a 1 HP, 1000 RPM, 230/460 volt, 3 phase, 60 Hz electric motor (or 50 Hz at additional cost). You need "superpressure" 1/4in outer diam. tubing for working fluid input/output, which they would sell you, along with a line of valves, connectors, etc. all capable of handling those pressures. Newport did not reveal their prices to me, but I think they lay between $2000 and $10000.

Once you have got the ability to control pressure, temperature, time, and composition (e.g. salinity) in your chamber, let me outline a simple **experiment**. Place RNA for some "naked RNA phage virus" in your chamber along with distilled water and whatever other contaminants you want to consider. The point of using naked RNA phage virus is to make it easy to crudely assay how many RNAs remain intact in your fluid, by simply exposing host bacteria growing on agar in a petri dish, to the phage, then seeing how many "holes" appear in the bacterial mat caused by infections. We now have the ability to vary pressure, temperature, time, and contaminating substances to see how that affects RNA decay. We similarly can and should try all the model reactions I tried to explore mathematically before, at high pressures.

S.Spiegelman in 1965, and later J.W.Szostak, did a lot of experiments where they used enzymes to replicate RNA and tried to select for and thereby breed, "improved" RNAs. Miller & Urey, and many followers, simulated early earth chemistry. Followers of Cairns-Smith tried chemistry experiments with clays. Followers of S.N.Platts presumably want to do chemical experiments with aqueous polycyclic aromatic hydrocarbons related to his "PAH world." All those kinds of experiments need to be tried partly or wholy **at high pressures**; and if you are going to use enzymes, get them from barophiles.

(Too Optimistic) Summary of "Deep salty RNA world" (& confidence assessments)

Writing this and my previous (Smith 2025) biochemistry paper both felt like Sherlock Holmes deducing "whodunit" by accumulating many clues.

> "Accumulating evidence suggests that RNA... could have completed the difficult task of organizing itself into a self-contained replicating system."
> – L.E.Orgel, 1998.

"All the major biopolymers are unstable in aqueous solution with respect to their (deactivated) monomers."
– A.G.Cairns-Smith: *Genetic takeover and the mineral origins of life*, Cambridge Univ. Press 1982.
"The backbone of the first genetic material could not have contained ribose or other sugar because of their instability [e.g. to hydrolysis]."
– Stanley L. Miller (in Larralde et al 1995).
"...this may indicate that primitive life evolved at high temperatures in the first place [C.R.Woese: Microbiol.Rev. 51 (1987) 221-271]. If it did and the archaebacteria [today called archaea] are the earliest forms of bacteria, [they may have] evolved at some depth in the rocks... spread laterally at depth, [then] evolved and progressed upwards to survive at lower temperatures nearer the surface."
– Thomas Gold: *The deep hot biosphere* (1992). This Gold speculation is exactly what I claim must have happened; and that early life was of the "RNA world" variety.

My two main ideas were **(I)** my "deep polymers hypothesis," largely regardable as simply *combining* Gold's "deep biosphere" and Orgel's "RNA world" ideas (albeit neither Orgel, nor his revisits by Neveu et al 2013; Zhou, Ding, Szostak 2021; and Ding, Zhou, Mittal, Szostak 2023; nor Gold, nor his revisit by Colman et al 2017, nor the review by Colwell & D'Hondt 2013, nor numerous books on "origin of life" ideas, mentioned this *combined* idea, so strangely enough, it apparently is introduced here for the first time); and **(II)** thinking more clearly about life's "vital information content." These together overcome the two main obstacles preventing understanding genesis, finally beginning to provide a complete picture. Both can be investigated and confirmed/denied experimentally in ways I described.

Four reasons for (I):

1. First, the **anthropic argument:** You exist. Many other lifeforms exist. A large number of "living fossil" clues found by prior workers show the "RNA world" must once have existed. But the RNA world has two big problems preventing it from working: aqueous RNA's instability to hydrolysis on time scales of minutes to months – *far* shorter than (what must have been) the doubling times of early life – and the inability so far of Miller-Urey type experiments to produce RNA nucleotides and polymerize them into RNA, and indeed to produce all other biopolymers: proteins and polysaccharides. Therefore, there must have been a way those two problems were overcome. The only way I (or arguably anybody) ever thought of is my very simple "deep polymers" pressure-hypothesis. Therefore it must be true.
2. Second, the **safety argument:** The earliest life must have been just barely alive, with very long doubling times, at least 30, and probably 20000, years. The early earth was very unsafe place. There was only one place on the early earth safe enough to permit early life to survive for its doubling timescale: deep underground (or sub-seafloor). Also, going beyond "safety," there is the matter of long-term *constancy* of environmental conditions. Genesis of a lifeform capable of *adjusting* to changing conditions is a lot less likely than life that can handle just *one* unchanging environment. Deep environments change the least. Therefore that was where it must have arisen. And voila: This deepness (under the Hypothesis) is precisely what is needed to make RNA world feasible.
3. Third, **experimental verifications.** Deep life has been found, e.g. in water, shown by Ar isotope dating to have been isolated from the surface for Myrs, leaking out of the walls of South African gold mines 2 miles deep, with doubling times ≳1000 years, and plenty of viruses. Therefore the RNA and/or DNA and protein coats of those viruses are stable for 1000-year timespans in the hot high-pressure water there. Also, archaea trapped in buried halite fluid inclusions for 19 kyr to 250 Myr remain alive (are revivable) today, despite critics claiming biomolecule instabilities should prevent survival for more than a few centuries. Therefore the hypothesized stability of deep polymers is experimentally proven.
4. Finally, I have a **large scattered collection of miscellaneous clues** (list below), which, even without the first three reasons, would have sufficed to make the Hypothesis probable. A future program of chemical experiments in pressure chambers – if anybody does it – ought to be able to solidify and greatly clarify everything. (Or refute it!) There has already been some pressure chamber work by some origin-of-life investigators, but the examples I saw were mostly not investigating the centrally important questions. The present paper's theory is a guide telling you the right questions to investigate.

Meanwhile idea (II) is supported by:

5. **Anthropic argument:** You exist. So genesis happened on Earth. Also, our galaxy is not crawling with "machine life" everywhere (none ever seen); and Mars, and everyplace else in our solar system besides Earth, never had

life (as far as we currently know, and I presume all future work will keep confirming all that). Therefore genesis is an *improbable* event – but not *too* improbable – from those facts I deduced rough numerical upper and lower *bounds* on how improbable it was.
6. But naive attempts to compute the "information content of life" suggest that even the simplest known lifeforms are far **too complicated** to have arisen by chance. The naive information content of the simplest lifeforms is about 800 times that permitted by our observation-based bound-window on improbability! Therefore those naive bit-count estimates must be very wrong.
7. My **mathematical re-analysis** of life's "vital information content" shows the naive bit-count estimates indeed are about a factor 200 too large. As a result, the factor-800 gap shrinks to about 4. That factor 4 is postulated to be acceptable because the earliest life was "much less alive" (far more fragile, far slower-doubling, generally horribly weak and utterly uncompetitive) than any life today. If I handed a biologist a sample of that kind of life, he would not be able even to *recognize* it as *being* "alive." The information content of the simplest kind of that less-alive kind of life, could plausibly be a factor 4 less than that of the simplest life meeting today's minimum standards of robustness, doubling speed, and competitiveness.

My whole information-content analysis is basically just a **mathematical theorem**, hence arguably doesn't really need experimental confirmation. Nevertheless experiments should help clarify/refine both (I) and (II) and hence are highly desirable. I actually now find myself in an interesting position here of "sticking my neck out": it seems to me (a) my confidence calculations below build up enough confidence that my origin-of-life picture must be correct, *but* (b) it could be refuted tomorrow by experiments – embarrassingly for me and my confidence calculations! But that is OK: theorists *should* stick their necks out by proposing theories easily refutable by experiment. If such a refutation does indeed occur, it would be interesting to understand how that managed to happen, and will probably teach some important lesson.

Confidence assessments. *Your* numbers might differ somewhat, but I'll explain *my* confidence assessments. I'm **98%** sure "RNA world" happened, because I find the evidence for it (gathered by others that I recounted; and without even using the other reasons I there called A and C) 98% convincing. So my confidence in the implication "anthropic argument⇒(I)" would be 98% *if* I were completely sure my deep pressure Hypothesis were the only possible way RNA world could have happened. Maybe there is some other way, but neither I nor anybody else has ever been able to think of any that doesn't seem ≥1000× less likely to me. My confidence in "safety argument⇒(I)" is **99.99%** since I find it absolutely inconceivable that the earliest life could have survived on the unsafe early earth for order 1 doubling times anyplace besides the most protected stable place: deep inside the solid part of the Earth. My confidence in "deep mine/borehole observations⇒(I)" is **97%**. I arrived at this 97% by:

- Assigning 85% confidence to the argument based on viruses and long inter-division timespans for deep prokaryotes. The reason for only 85% is that the holes yielding the clearest virus findings, are not the same as the ones yielding the clearest conclusions that part of the deep biosphere is isolated and has very-slow-reproducing prokaryotes with doubling times≳1000 years – so I'd like to see more and better such observations.
- Also assigning 85% confidence to to the completely independent experimental verifications from halite fluid inclusions. The reason for only 85% here is that those papers have been disputed. I happen to think there are enough independent such papers (see special references section) that they are true, but out of respect for the disputors I'm giving this only 85%. (Actually there also are further similar papers allegedly finding very old microbes trapped in shale rocks, etc, instead of halite, which I haven't even mentioned but which would boost confidence further.)

Then $1-(1-0.85)^2>0.97$. Finally, my confidence in "miscellaneous clues⇒(I)" is **97%**:

> Those clues include: ① two indications from the genetic code of life's deep origin, ② the belief of phylogeneticists (Slesarev, Mezhevaya et al 2002 and refs. cited within) that a good candidate for today's lifeform most resembling LUCA is the deep hot oxygen-intolerant halophile methanogen Methanopyrus Kandleri which holds the current record 122°C for the highest temperature at which a microbe has been observed to grow and reproduce (Takai et al 2008 – it was discovered in a 2000m deep ocean "black smoker" off California and later in another near Iceland), ③ the estimates that PΔV pressure-caused adjustments to hydrolysis energies seem to have the right order of magnitude if P=4000 atm, and ④ both

proofs in theoretical models, and ⑤ some observational evidence, suggesting they have the right sign. [If each of those 5 items is reckoned good enough to cut the chance of (I) being wrong by factor 2, then we get confidence=1-2^{-5}≈97%.] A critic could counter that most members of that list could equally well be interpreted as supporting origin inside deep-ocean volcanic "hot vents" such as the *Kandleri*'s black smokers, *not* buried subsurface. However: hot vents are far less stable and less safe, and occupy far tinier volume – any and all of which inherently makes them, a priori, extremely relatively unlikely. To keep (4) independent of (2) we should rely on the volume-argument *alone*, but that should be good enough for ≥10^6 relative likeliness factor, so that criticism doesn't affect my 97%.

Since we have all four {1,2,3,4} acting independently (or, even better, cooperatively), not just one, my confidence in the Hypothesis (I) exceeds **99.999999%**. Any one is enough for ≥97% confidence, any two for ≥99.9%, and any three for ≥99.997%. Critics may regard different numbers as more reasonable, or attack my arguments... but I do not think you are going to be able to push the confidence below 99%.

My confidence is "(5&6)⇒(II)" basically equals my confidence that the Bible's theory of genesis is incorrect. The "homogeneous backtrack tree" model that underlies "7⇒(II)" is suspect, therefore my confidence in (II) is below 100%. (If I knew that model accurately described reality, it would be 100% since it then would just be a math theorem.) However, straightforward experiments I've sketched would clarify that arbitrarily much the more data the experimenter gathered. Putting both 5&6 and 7 together (since either suffice), my net confidence in (II) **exceeds 99%.**

Benner 2014 helpfully compiled 5 "paradoxes" he labeled a-e, which I would instead call "devastating problems" for prior origin-of-life ideas. Our (I) and (II) in "deep salty RNA world" seem able (at least for now – unless and until experimental evidence appears to the contrary) to resolve *all 5* of them.

Caveat: It could be that Hypothesis (I), although technically true, only works for pressures too large to have been involved in early-Earth life (e.g. >7000 atm), so "nobody cares." The only arguments I have against that possibility are (1,3), yielding diminished confidence versus (1,2,3,4).

Am I stupid/deluded? Observed fact: The fraction of published wrong conclusions having "confidence≥99.99%" considerably exceeds 0.01%. So obviously there exist stupid/deluded/overconfident authors. Am I one? Well, (1) if some part of my analysis is wrong, please tell me what it is. But (2) we can skip all that. There are conceptually-simple direct experiments, that should be quite feasible, to overcome that. My present analyses allegedly finding "high confidences," especially if you both (a) can't punch logical holes in them and (b) nevertheless are skeptical, suffice to justify trying those experiments. Do them.

Along the way I also pointed out major problems with the Fox, Woese et al "tree of life" deduction papers (and followers) suggesting they all need to be redone and explained why I consider that field "diseased"; described a possible new way to find early-life "fossils"; pointed out my Hypothesis solves a multi-decade dispute about long-term survival of halophile archaea trapped in NaCl crystals, and suggested the "drill baby drill" idea for possibly *finding* some still-surviving RNA life – which in the unlikely event it succeeded would prove the "RNA world" hypothesis, arguably with my "deepness" addendum too, winning you a Nobel prize.

Acknowledgments. I thank Prof. Erik Winfree (CalTech) for some comments about a preliminary draft of this manuscript, e.g. pointing out some more Szostak papers, and suggesting some reorganization, some of which helped me improve the paper (I thought, albeit Winfree might not necessarily agree). I also thank S.N.Platts (Rensselaer Polytechnic Institute) for some helpful comments and pointers in reaction to v1, e.g. re the Ryugu and Bennu asteroids and discotic liquid crystals.

(Pessimistic) Post Mortem

I originally came up with most of this paper in the late 1990s, then failed for the next 30 years to convince biochemists to do experiments confirming/refuting it, then wrote the paper in 2025. While doing so, I found out (which I had not known back in the late 1990s) that two classes of experiments (albeit probably a better word is "observations," since the "experimenter" was Nature, not humans) had *already happened* – and both seemed to support my "deep RNA" Hypothesis! That energized me greatly; and versions 1 & 2 of this paper, summarized in the preceding section, found

large confidences, featuring many "nines," that I had to be on the right track.

Unfortunately for those nines (now, a month later), I stumbled across a *third* class of (genuine!) experiments in the literature, most of whose results go *against* my "deep polymerization" hypothesis. They all concern "PSV." I'll now explain them. As a result, my confidences in those hypotheses are greatly diminished, so I retitled the preceding "summary" section "(Too Optimistic) Summary," and may need to switch my *Queen* music from We are the champions to Another One Bites the Dust. The third class of experiments seem to contradict the first and second to lesser and greater extents. I'm not sure what to make of that. Should my whole concept be thrown in the garbage? Or should it be rescuscitated – and if so how? How are those contradictions explained?? How did *both* these refutatory reslts happen *and* all those nines happen?? I'll propose a possible partial answer, but don't really know. All three classes of experiments were serendipitous and run by people completely unaware of me and uninterested in my origin of life theories. It would be better to obtain much clearer results from future clearer, more direct, purpose-designed experiments. I'm unwilling at present to call my Hypothesis "refuted"; but I will say that

1. my confidence is much lower,
2. the situation has become more interesting and mysterious,
3. and the desirability of clear experiments has increased.

If my Hypotheses are wrong, then they're wrong – but they deserve consideration and it is important to understand whether and why they are right or wrong.

"Effective molar volume" and "Partial specific volume" (PSV). Both those names denote essentially the same concept. A small but notable subset of experimental chemistry papers address it. Suppose you have some sugar. In my discussions earlier in this paper I was estimating the "volume V of a sugar molecule" by V=(mass of 1 sugar molecule)/(density of sugar crystal). As I'd pointed out, that V is only a crude estimate of the volume that biochemists interested in my "deep" hypotheses actually want to know. You really want to know the volume of water *displaced* by dissolving that sugar molecule *in aqueous solution*. This (multipled by Avogadro's number) has been called the "effective molar volume" or "apparent molar volume" of that sugar in water. Also, one can measure the volume of water displaced *per gram*, rather than *per mole*, of that kind of dissolved sugar. That quantity (usually measured in mL/gram) is called that sugar's "PSV."

Unfortunately (since this confuses matters) PSV is *not* really just one number: it **depends** on temperature, concentration, other things (such as metal ions) dissolved in the water, and pressure. Since those dependencies – especially the latter two (unfortunately since those are the ones I care most about) – have as yet been little-explored by experimenters, I mostly am going to make the (perhaps crude) approximation that PSV *is* just one number, and only concern myself with the low-concentration limit. I do this so that I can get some tentative answers.

In the preceding "optimistic summary" section, I had remarked that if my "deep origin" theory was wrong despite all my alleged nines of confidence, that should teach some important lesson. One candidate for that lesson is the unintuitive fact that **PSVs can be *negative***:

- If you mix water and ethanol, their combined volume *decreases*.
- Magnesium sulfate $MgSO_4$, in aqueous solutions of molality<0.07 mol/kg, has a negative PSV, i.e. dissolving it in water actually *shrinks* the volume of that water. (In contrast, NaCl has positive PSV.)
- Another substance (now with some biological importance!) with negative PSV is trisodium phosphate Na_3PO_4, see fig.3 of Millero et al 2010 where, e.g, volume shrinkage of about 0.9 mL was observed when dissolving (1/9) mole of Na_3PO_4 in 1 liter of salty (0.725 molar NaCl) water at 25°C. This PSV appears to be negative for all Na_3PO_4 concentrations≲0.8 moles/liter.

All my simplistic mathematical models of dimerization and polymerization had involved a parameter I called "ε" which I had implicitly assumed *positive*; and then I had proved theorems in those models showing all sufficiently high pressures would energetically favor polymerization and dimerization. When and if "negative volume" effects can happen, all those theorems are called into question.

Sugars. Birch & Catsoulis 1985 (and others) measured the following apparent molar volumes (**AMV**) of some important sugars (at 10 weight% concentration in water) and related molecules; units mL/mole (monosaccharides

yellow, di/polysaccharides green, nonsugars white background):

Substance	AMV in 20°C water	AMV in 40°C water	AMV 25°C with 0.6mol/kg K_3citrate	Comment
Water	18.05	18.12		AMV=μ/(density of water) where μ=18.01528 grams is the mass of 1 mole of water
Xylose	95.39	97.50	97.12	
Glucose	112.52	113.77	115.81	
Fructose	110.90	112.93	113.96	
Galactose	110.75			
Sucrose	211.59	217.48		∴ Sucrose+H_2O→Fructose+Glucose favored at high pressure
Palatinose	220.36	226.51		∴ Palatinose+H_2O→Fructose+Glucose favored at high pressure
Lactose	211.59			∴ Lactose+H_2O→Galactose+Glucose favored at high pressure
Trehalose	211.59	216.59		∴ Trehalose+H_2O→2 Glucose favored at high pressure
Raffinose	307.29	311.99		∴ Raffinose+2H_2O→Galactose+Glucose+Fructose favored at high pressure

The AMVs at 25°C in water with 0.6mol/kg K_3citrate added are taken from table 2 of Kumar & Sheetal 2016; the others were measured in pure water. For all the di- and trisaccharides tabulated, high pressures ought to favor *hydrolysis* into monosaccharides because the summed AMVs of the products is less than the summed AMV of the reactants – "wrong sign" – precisely the *opposite* of what my dimerization Hypothesis and my simplistic models' theorems had predicted. As example calculations:

1. Raffinose in 20°C water: The "**>**" sign in 307.29+2×18.05=343.39**>**334.17=112.52+110.90+110.75 shows that its hydrolysis should be favored at high pressures.
2. Sucrose in 20°C water: The "**>**" sign in 211.59+18.05=229.64**>**223.42=112.52+110.90 shows that its hydrolysis should be favored at high pressures.

I.e, my Hypotheses all seem to be false for all these sugars, albeit with the caveat that this all was at zero salinity, and it is not clear how that is affected by salinity. (Perhaps this is due to local negative-volume effects near hydrophilic parts of these sugars, invalidating my theorems?)

Also, it has been claimed, rather imprecisely (e.g. without stating experimental conditions, and there are different kinds of glycogen, e.g. with more or less "branching") that the PSV of **glycogen** (polymerized glucose) is **0.62-0.68** mL/gram. (Barber, Harris, Anderson 1966 say 0.62 on p.286; Chee & Geddes 1977 mention 0.63 and 0.68 citing two previous papers.) When we consider that the PSV of water is **1** mL/gram, while the PSV of glucose (molecular weight 180.16) is **0.624**=112.52/180.16 mL/gram at 10 wt% concentration at 20°C (although this increases with concentration, temperature, and salinity) this suggests that glycogen hydrolysis will still be energetically favored even at high pressures. I.e. my pressure⇒polymerization hypothesis tentatively is false for glycogen too.

The reason I began with sugars is that the available AMV experiments yield the clearest conclusions for sugars. That's unfortunate because sugars seem the least important kind of biomolecule for my Hypothesis. But actually even for sugars these results remain somewhat **murky** because as the table shows, the AMVs *change* with temperature and salinity (and presumably also pressure, which we might guess acts like pseudo-salinity) apparently by amounts of the same order needed to change the sign of our conclusion!

Nucleobases, nuclosides, nucleotides, and RNA & DNA. Birch & Catsoulis unfortunately did not measure AMV for ribose and deoxyribose. However table 2 of Kumar & Sheetal 2016 measured the AMV of ribose to be 97.74 mL/mole

in 25°C water containing 0.2 mole/kg of tripotassium citrate. (They found that sugar AMVs usually increase with K_3citrate concentration.) Kishore & Ahluwalia 1990 measured the following AMVs for nucleobases and nucleosides at low concentrations in 25°C water containing 180 gram/kg of dissolved urea:

Substance	AMV (ml/mol)	Comment
Ribose	97.74	Kumar & Sheetal 2016 but using K_3citrate not urea in their water
thymine	88.75	
cytosine	74.31	
adenine	89.04	
uracil	72.20	
Water	18.05	
thymidine	168.18	Formation (with release of H_2O) from thymine+ribose favored by pressure
cytidine	115.13	Pressure favors hydrolysis into cytosine+ribose
adenosine	172.44	Pressure favors hydrolysis into adenine+ribose
uridine	153.71	Pressure favors hydrolysis into uracil+ribose
disodium phosphate Na_2HPO_4	14.2	Millero et al 2010; but there is no urea in their measurement
UMP·2Na	174.46	Pressure favors hydrolysis into uridine+Na_2HPO_4

These figures if taken at face value (which might be unwise because of the different conditions in the ribose AMV measurement) indicate that high-enough pressures should favor *hydrolysis* of three of the four nucleosides into bases+ribose, i.e. three "wrong signs" versus only one "right sign." (The "right sign" indicating high pressure should favor dimerization of thymine and ribose into thymidine arises since 168.18+18.05=186.23<186.49=97.74+88.75; the "wrong sign" for uridine arises since 153.71+18.05=171.76>169.94=97.74+72.20.) The differences between the left and right hand sides are small enough that I do not consider those three wrong signs to be "refutations" (nor the "right sign" a confirmation); but they are worrying.

Furthermore, 153.71+14.2=167.9<192.51=174.46+18.05 indicates that the nucleotide UMP·2Na under high pressure would want to hydrolyse into uridine+Na_2HPO_4 (then further into uracil+ribose+Na_2HPO_4), another "wrong sign." The different conditions in the experiments of Millero et al versus Kishore & Ahluwalia may subject this conclusion to doubt, but again, it is worrying.

Necessarily imprecise attempts have been made to measure the PSVs of "average" dsDNA and RNA. For "average RNA" it allegedly is approximately 0.569 mL/gram, while for DNA it is in the range 0.55 to 0.63 mL/gram (e.g. Savelyev 2021 and papers he cites). If these figures are taken at face value then RNA's effective volume in water is about 247.5 mL per mole of bases, while dsDNA's is about 232-259 mL per mole of bases (where note a "base pair" counts as 2 bases). These numbers again suggest the whole "deep RNA pressure⇒polymerization" hypothesis is wrong. (Again perhaps due to local "negative-volume" effects? – e.g. keep in mind that highly charged phosphate ions have been associated with negative PSVs) If so, what does this tell us? E.g. does it mean we should focus on things like PNA and TNA, not RNA?

Polypeptides and amino acids. I have now added Zamyatnin 1984's AMV numbers to the amino acids table at the end of the paper. From that table we see that long proteins weigh about 110.70 AMU per amino. Also, according to Zhao, Brown, Schuck 2011, the average PSV for all human proteins is 0.735 mL/gram, and over several kinds of life (not just humans) their range is 0.735 to 0.744. Therefore proteins have AMV per mole of constituent aminos of (0.735-0.744)×110.70=81.36-82.36 mL/mol. If we add to that the AMV 18.05 mL/mol of 1 water molecule per amino, we get **99.41-100.41**. Meanwhile, the AMVs for amino acids range between 43.2 mL/mol for glycine and 127.3 mL/mol for arginine, with a frequency-weighted average of **86.89** mL/mol. Therefore, high pressure is expected to *encourage* average proteins to hydrolyse into amino acids. In particular, Zamyatnin 1984 on p.159 claims that the 45-amino polypeptide $(pro\text{-}pro\text{-}gly)_{15}$ has PSV=0.74 mL/gram which considering its molecular weight of 84.17 per amino

is 62.29 mL/mol of its constituent aminos, which if we add the AMVs 18.05 of 44 moles of water yields **79.94** mL/mol. This exceeds the average AMV (2×81.0+43.2)/3=**68.40** of 2 prolines and 1 glycine, so high pressures should energetically favor hydrolysis of this particular polypeptide – yet another "wrong sign."

References

Special subsection: references about "prokaryotes trapped in fluid inclusions in halite."

Dan Graur & Tal Pupko: The Permian bacterium that isn't, Molecular Biology & Evolution 18,6 (2001) 1143-1146.

Martin B. Hebsgaard, Matthew J. Phillips, Eske Willerslev: Geologically ancient DNA: Fact or artefact? Trends in Microbiology 13,5 (May 2005) 212-220.

Salla T. Jaakkola, Friedhelm Pfeiffer, Janne J. Ravantti, 7 others: The complete genome of a viable archaeum isolated from 123-million-year-old rock salt, Environ. Microbiol. 18,2 (Feb.2016) 565-579.

T.K.Lowenstein, B.A.Schubert, M.N.Timofeeff: Microbial communities in fluid inclusions and long-term survival in halite, GSA Today 21,1 (2011) 4-9.

Tim K. Lowenstein, Cindy L. Satterfield, Russell H. Vreeland, William D. Rosenzweig, Dennis W. Powers: New evidence for 250 Ma age of halotolerant bacterium from a Permian salt crystal: Comment and reply, Geology 33,1 (Feb.2005) e93-e94; 33,4 (Apr.2005) 265-298. E.Willerslev & M.B.Hebsgaard: (same title, same place)

Melanie R. Mormile & 6 others: Isolation of Halobacterium salinarum retrieved directly from halite brine inclusions, Environmental Microbiology 5 (2003) 1094-1102.

Howard Ochman, Susannah Elwyn, Nancy A. Moran: Calibrating bacterial evolution, Proc.Natl.Acad.Sci. USA 96,22 (Oct.1999) 12638-12643.

Krithivasan Sankaranarayanan, Tim K. Lowenstein, Michael N. Timofeeff, Brian A. Schubert, J. Koji Lum: Characterization of ancient DNA supports long-term survival of Haloarchaea, Astrobiology 14,7 (July 2014) 553-560.

Sara I. Schreder-Gomes, Kathleen C. Benison, Jeremiah A. Bernau: 830-million-year-old microorganisms in primary fluid inclusions in halite, Geology 50,8 (May 2022) 918-922.

Brian A. Schubert, Tim K. Lowenstein, Michael N. Timofeeff, Matthew A. Parker: Halophilic Archaea cultured from ancient halite, Death Valley, California, Environ. Microbiol 12,2 (Feb.2010) 440-454.

Brian A. Schubert, Tim K. Lowenstein, Michael N. Timofeeff: Microscopic Identification of Prokaryotes in Modern and Ancient Halite, Saline Valley and Death Valley, California, Astrobiology 9,5 (June 2009) 467-482.

Brian A. Schubert, Michael N. Timofeeff, Tim K. Lowenstein, Jürgen E.W. Polle: Dunaliella Cells in Fluid Inclusions in Halite: Significance for Long-term Survival of Prokaryotes, Geomicrobiology Journal 27,1 (Jan.2010) 61-75.

Brian A. Schubert; Tim K. Lowenstein; Michael N. Timofeeff; Matthew A. Parker: How do prokaryotes survive in fluid inclusions in halite for 30 kyr? Geology 37,12 (Dec.2009) 1059-1062.

Russell H. Vreeland, W.D.Rosenzweig, D.W.Powers: Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal, Nature 407,6806 (Oct.2000) 897-900. Discussed in Science.org and American Scientist (2000).

All remaining references (skip ahead to F, M, Z):

Ahmet Cemal Alicioglu & Can Alkan: Pairwise sequence alignment with block and character edit operations (arXiv 2024).

Rika E.Anderson, Wm.J.Brazelton, J.A.Baross: The deep virosphere: Assessing the Viral Impact on Microbial Community Dynamics in the Deep Subsurface, Reviews Minerology & Geochemistry 75 (2013) 649-675.

K.Aoki, S.Usuba, M.Yoshida, & 3 others: Raman Study of the Solid-State Polymerization of Acetylene at High Pressure, J. Chem. Phys. 89 (1988) 529-534.

Nicholas T. Arndt & Euan G. Nisbet: Processes on the young Earth and the habitats of early life, Annual Rev. Earth Planetary Sci. 40,1 (May 2012) 521-549.

Albert A. Barber, Warren W. Harris, Norman G. Anderson: Isolation of Native Glycogen by Combined Rate- Zonal and Isopycnic Centrifugation, Nat'l. Cancer Institute Monograph (June 1966) 285-302.

Colin Barras: Buried microbes exist at limit between life and death, New Scientist 214,2866 (May 2012) 12.

Sidney Becker, Hidenori Okamura, Christina Schneider, & 4 others: Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides, Science 366,6461 (Oct.2019) 76-82.

Steven A. Benner: Paradoxes in the Origin of Life, Origins of Life & Evolution of Biospheres 44,4 (2014) 339-343. Benner also stated this *different* list of 4 paradoxes when interviewed by Suzan Mazur in Huffington Post 6 Dec.2013: "We have failed to provide a recipe that gets from the simple molecules that we know were present on early Earth to RNA... we're up against these four paradoxes... First paradox: the tendency of organic matter to devolve and to give **tar**. If you can avoid that, you can start to try to assemble things that are not tarry, but then you encounter the **water problem**, which is related to the fact that every interesting bond that you want to make is unstable, thermodynamically, with respect to water. If you can solve that problem, you have the problem of **entropy**, that any of the building blocks are going to be present in a low concentration; therefore, to assemble a large number of those building blocks, you get a gene-like RNA – 100 nucleotides long – that fights entropy. And the fourth problem is that even if you can solve the entropy problem, you have a paradox that RNA enzymes, which may be catalytically active, are more likely to be active in the sense that **destroys** RNA rather than creates it."

Elwyn R. Berlekamp, John H. Conway, & Richard K. Guy: Winning Ways for Your Mathematical Plays, (2 vols) Academic Press 1982. (2nd ed. 2004, now split into 4 volumes, by A.K.Peters.) The final section includes conlife and explains the universal Turing machine, vanishing-undecidability, and universal assembler "life" constructions.

Max P. Bernstein, Scott A. Sandford, Louis J. Allamandola: Life's far-flung raw materials, Scientific American 281,1 (July 1999) 42-49.

G.G.Birch & S.Catsoulis: Apparent molar volumes of sugars and their significance in sweet taste chemoreception, Chemical Senses 10,3 (July 1985) 325-332,

Jordan T. Bird, E.D.Tague, L.Zinke, 9 others: Uncultured microbial phyla suggest mechanisms for multi-thousand-year subsistence in Baltic Sea sediments, ASM mBio 10,2 (Apr.2019) e02376-18.

G.Borgonie, A.Garcia-Moyano, D.Litthauer, & 6 others: Nematoda from the terrestrial deep subsurface of South Africa, Nature 47,7349 (June 2011) 79-82.

Percy W. Bridgman: Water, in the Liquid and Five Solid Forms, under Pressure, Proc. American Academy of Arts and Sciences 47,13 (Jan.1912) 441-558.

Percy W. Bridgman: The physics of high pressure, 1931; 2nd ed. 1949; reprinted with supplement in 1952 & 1958 by Bell, London; and by Dover 1970. QC131.B7 or QC281.B7. Also, a short introduction to this area is Bridgman's Nobel Prize in Physics 1946 lecture, reprinted pp.53-72 in *Nobel Lectures in Physics 1942-1962* (World Scientific 1998).

Lanlan Cai, Markus G. Weinbauer, Le Xie, Rui Zhang: The smallest in the deepest: the enigmatic role of viruses in the deep biosphere, Natl. Sci. Rev. 10,4 (Jan.2023) nwad009.

L.Cai, B.B.Jørgensen, C.A.Suttle, et al: Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years, ISME (International Society for Microbial Ecology) Journal 13,7 (Mar.2019) 1857-1864.

A.G.Cairns-Smith: Genetic takeover and the mineral origins of life, Cambridge Univ. Press 1982 (QH325.C225); Seven Clues to the Origin of Life: A Scientific Detective Story, Cambridge Univ. Press 1990 (QH324.C24).

Marcella Carnici, Francesco Sgarella, et al: The standard Gibbs free energy change of hydrolysis of ribose-1-phosphate, Archive Biochem. & Biophysics 205,1 (Nov.1980) 191-197.

R.L.Charlebois & W.F.Doolittle: Computing prokaryotic gene ubiquity: Rescuing the core from extinction. Genome Res. 14 (2004) 2469-2477.

N.P.Chee & R.Geddes: The structure of liver glycogen, FEBS letters 73,2 (Feb.1977) 164-166.

Pang C. Chen: Heuristic sampling: a method for predicting the performance of tree searching programs, SIAM J. Computing 21,2 (Apr.1992) 295-315.

Francesca D. Ciccarelli, Tobias Doerks, Christian von Mering, Christopher J. Creevey, Berend Snel, Peer Bork: Toward Automatic Reconstruction of a Highly Resolved Tree of Life, Science 311 (Mar.2006) 1283-1287.

Daniel R. Colman, Saroj Poudel, Blake W. Stamps, Eric S. Boyda, John R. Spear: The deep, hot biosphere: Twenty-five years of retrospection, Proc. Natl. Acad. Sci. USA 114,27 (Jul.2017) 6895-6903.

Frederick S. Colwell & Steven D'Hondt: Nature and Extent of the Deep Biosphere, Reviews in Mineralogy & Geochemistry 75,1 (Feb.2013) 547-574.

C.W.Cook, W.A.Fowler, C.C.Lauritsen, T.Lauritsen: B^{12}, C^{12}, and the red giants, Phys. Rev. 107,2 (July 1957) 508-515.

G.Cornuejols, M.Karamanov, Y.Li: Early Estimates of the size of branch-and-bound trees, INFORMS J. Computing 18 (2006) 86-96.

F.H.C.Crick, J.S.Griffith, L.E.Orgel: Codes without commas, Proc. Nat'l Acad. Sci. USA 43,5 (May 1957) 416-421.

D.W.Deamer & R.M.Pashley: Amphiphilic components of the Murchison carbonaceous chondrite: surface properties and membrane formation, Origins of Life & Evol. Biosph. 19,1 (1989) 21-38.

K.S.Dickson, C.M.Burns, J.P.Richardson: Determination of the Free-Energy Change for Repair of a DNA Phosphodiester Bond, J.Biological Chem. 275,21 (May 2000) 15828-15831 open access.

Dian Ding, Lijun Zhou, Shriyaa Mittal, Jack W. Szostak: Experimental tests of the virtual circular genome model for nonenzymatic RNA replication, J.Amer.Chem.Soc. 145,13 (2023) 7504-7515.

Christopher M. Dobson, G. Barney Ellison, Adrian F. Tuck, Veronica Vaida: Atmospheric aerosols as prebiotic chemical reactors, PNAS 97,22 (Oct.2000) 11864-11868.

Tim Engelhardt, Monika Sahlberg, Heribert Cypionka, Bert Engelen: Induction of prophages from deep-subseafloor bacteria, Environ. Microbiol. Rep. 3,4 (Aug.2011) 459-465.

Evgeny Epelbaum, Hermann Krebs, Dean Lee, Ulf-G.Meissner: Ab Initio Calculation of the Hoyle State, Physical Review Letters 106,19 (2011) #192501. Discussed in Morten Hjorth-Jensen: Viewpoint: The carbon challenge, Physics 4 (2011) 38-39.

Giuseppe Etiope & Barbara Sherwood Lollar: Abiotic methane on Earth, Reviews of Geophysics 51,2 (June 2013) 276-299.

Hallgerd S.C.Eydal, Sara Jägevall, Malte Hermansson, Karsten Pedersen: Bacteriophage lytic to Desulfovibrio aespoeensis isolated from deep groundwater, ISME J. 3,10 (Oct.2009) 1139-1147.

Martin J. Falk, Leon Zhou, Yoshiya J. Matsubara, Kabir Husain, Jack W. Szostak, Arvind Murugan: Suppression of errors in collectively coded information, (arXiv 2025).

H.Fei, D.Yamazaki, M.Sakurai, N.Miyajima, H.Ohfuji, T.Katsura, et al: A nearly water-saturated mantle transition zone inferred from mineral viscosity, Sci Adv. 3 (2017) e1603024.

Da-Fei Feng, Glen Cho, Russell F. Doolittle: Determining divergence times with a protein clock: Update and reevaluation, Proc. Natl. Acad. Sci. USA 94,24 (Nov.1998) 13028-13033.

Martin Ferus, Fabio Pietrucci, Antonino Marco Saitta, et al: Formation of nucleobases in a Miller-Urey reducing atmosphere, Proc. Natl. Acad. Sci. USA 114,17 (Apr.2017) 4306-4311.

Wely Brasil Floriano & Marco A.C. Nascimento: Dielectric constant and density of water as a function of pressure at constant temperature, Brazilian J. Physics 34,1 (Mar.2004) 38-41.

G.E.Fox, K.R.Pechman, C.R.Woese: Comparative Cataloging of 16s Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics, Intl. J. Systematic Bacteriology 27,2 (Jan.1977) 44-57.

G.E.Fox, E.Stackebrandt, R.B.Hespell, J.Gibson, J.Maniloff, T.A.Dyer, R.S.Wolfe, W.E.Balch, R.S.Tanner, L.J.Magrum, L.B.Zablen, R.Blakemore, R.Gupta, L.Bonen, B.J.Lewis, D.A.Stahl, K.R.Luehrsen, K.N.Chen, C.R.Woese: The phylogeny of prokaryotes, Science 209,4455 (Jul. 1980) 457-463. http://pubmed.ncbi.nlm.nih.gov/6771870/

M.Freer & H.O.U.Fynbo: The Hoyle state in 12C, Progress in Particle & Nuclear Physics 78 (2014) 1-23.

Harold N. Gabow: Two Algorithms for Generating Weighted Spanning Trees in Order, SIAM Journal on Computing 6,1 (Mar.1977) 139-150. Another paper on same topic is Kenneth Sörensen & Gerrit K. Janssens: An algorithm to generate all spanning trees of a graph in order of increasing cost, Pesquisa Operacional 25,2 (Aug.2005) 219-225.

Deepak Gadani, V.A.Rana, Sp Bhatnagar, A.N.Prajapati, A.D.Vyas: Effect of salinity on the dielectric properties of water, Indian Journal of Pure & Applied Physics 50,6 (June 2012) 405-410.

Nir Gavish & Keith Promislow: Dependence of the dielectric constant of electrolyte solutions on ionic concentration: A microfield approach, Phys. Rev. E 94,1 (July 2016) #012611.

Rosario Gil & Amparo Latorre: Factors Behind Junk DNA in Bacteria, Genes (Basel) 3,4 (Oct.2012) 634-650.

Eric Girard, Thierry Prange, & 7 others: Adaptation of the base-paired double-helix molecular architecture to extreme pressure, Nucleic Acids Research 35,14 (Jul.2007) 4800-4808.

Geoffrey B. Glasby: Abiogenic Origin of Hydrocarbons: An Historical Overview, Resource Geology 56,1 (2006) 85-98.

Thomas Gold: The deep, hot biosphere, Proc. Natl. Acad. Sci. USA 89,13 (July 1992) 6045-6049, and 235-page book of same title, Copernicus NY 1999, TN870.5.G66. Gold's book was critically reviewed by K.E.Peters in Organic Geochemistry 30 (1999) 473-475 and revisited 25 years later by Colman et al 2017.

Adam P. Goucher: Universal computation and construction in GoL cellular automata, pp.505-517 in book Game of Life Cellular Automata (Ed. Andrew Adamatzky), Springer 2010.

Thomas Hales & 21 others: A formal proof of the Kepler conjecture (ArXiv, Jan.2015).

J.B.Hasted, D.M.Ritson, C.H.Collie: Dielectric Properties of Aqueous Ionic Solutions. Parts I and II, J.Chemical Physics 16,1 (Jan.1948) 1-21.

Rachael Hazael, F.Foglia, L.Kardzhaliyska, 3 others: Laboratory Investigation of High Pressure Survival in Shewanella oneidensis MR-1 into the GigaPascal Pressure Range, Frontiers in Microbiol. 5 (Nov.2014) 612.

Marijn J.H. Heule: Schur Number Five (arXiv 2017).

Marc M. Hirschmann: Water, melting, and the deep earth H_2O cycle, Annual Rev. Earth Planetary Sci. 34 (2006) 629-653.

Sahand Hormoz: Amino acid content of proteins..., Nature Scientific Reports 3 (Oct.2013) #2919.

Fred Hoyle: On Nuclear Reactions Occurring in Very Hot Stars. I. The Synthesis of Elements from Carbon to Nickel, Astrophys. J. Suppl. Ser. 1 (1954) 121-146. The story of how Hoyle predicted his excited state in C-12 is told on pages 205-210 of Simon Mitton: Fred Hoyle: A Life in Science, Cambridge Univ. Press (Illustrated Edition 2011, earlier ed. 2005). QB36.H75M58.

Laura A. Hug, Brett J. Baker, and 15 others: A new view of the tree of life, Nature microbiology 1 (Apr.2016) #16048.

Clyde A. Hutchison III & 22 others: Design and Synthesis of a Minimal Bacterial Genome, Science 351,6280 (Mar.2016)

F.Inagaki & 45 others: Exploring deep microbial life in coal-bearing sediment down to 2.5km below the ocean floor, Science 349,6246 (2015) 420-424.

Bo Barker Jørgensen & Steven D'Hondt: A starving majority deep beneath the seafloor, Science 314,5801 (Nov.2006) 932-934.

Manesh Prakash Joshi, Luke Steller, Martin J. Van Kranendonk, Sudha Rajamani: Influence of Metal Ions on Model Protoamphiphilic Vesicular Systems: Insights from Laboratory and Analogue Studies, Life (Basel) 11,12 (Dec.2021) 1413.

Olga V. Karnachuk, Yulia A. Frank, Anastasia P. Lukina, Vitaly V. Kadnikov, Alexey V. Beletsky, Andrey V. Mardanov, Nikolai V. Ravin: Domestication of previously uncultivated Candidatus Desulforudis audaxviator from a deep aquifer in Siberia sheds light on its physiology and evolution, ISME J. 13,8 (Aug.2019) 1947-1959.

K.Kawamura & M.Yukioka: Kinetics of the racemization of amino acids at 225-275°C using a real-time monitoring method of hydrothermal reactions, Thermochimica Acta 375,1-2 (July 2001) 9-16.

Philip Kilby, John Slaney, Sylvie Thiebaux, Toby Walsh: Estimating Search Tree Size, AAAI 2006, pp.1014-1019.

Adrienne Kish, Patrick L. Griffin, Karyn L. Rogers, Marilyn L. Fogel, Russell J. Hemley, Andrew Steele: High-pressure tolerance in Halobacterium salinarum NRC-1 and other non-piezophilic prokaryotes, Extremophiles 16,2 (Mar. 2012) 355-361.

Nand Kishore & Jagdish C. Ahluwalia: Partial molar heat capacities and volumes of some nucleic acid bases, nucleosides and nucleotides in aqueous urea solutions at 298.15°K, J. of the Chemical Society, Faraday Transactions 86,6 (1990) 905-910.

Donald E. Knuth: Estimating the efficiency of backtrack programs, Maths. of Computation 29,129 (Jan.1975) 121-136.

A.Kolesnikov, V.G.Kutcherov, A.F.Goncharov: Methane-derived hydrocarbons produced under upper-mantle conditions, Nature Geoscience 2,8 (2009) 566-570. Discussed in Peter Aldhous: Black gold causes a stir, Nature 353 (Oct.1991) 593.

Harsh Kumar & Sheetal: Densities and Speeds of Sound of d(+)-Glucose, d(âˆ’)-Fructose, d(+)-Xylose and d(âˆ’)-Ribose in Aqueous Tripotassium Citrate Solutions at Different Temperatures, J. Chem. Engin. Data 61,7 (2016) 2244-2259.

Jennifer E.Kyle, H.S.C.Eydal, F.G.Ferris, K.Pedersen: Viruses in granitic groundwater from 69 to 450m depth of the Äspö hard rock laboratory Sweden, ISME J. 2,5 (May 2008) 571â€“574.

Jurij Lah & San Hadzi: Thermodynamic Origin of the Linear Pressure Dependence of DNA Thermal Stability, J. of Physical Chemistry Letters 15,35 (Aug.2024) 9064-9069.

Rosa Larralde, M.P.Robertson, S.L.Miller: Rates of decomposition of ribose and other sugars, Implications for chemical evolution, Proc.Natl.Acad.Sci. USA 92,18 (Aug.1995) 8158-8160.

Mark A. Lever, Karyn L. Rogers, 6 others: Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations, FEMS Microbiology Reviews 39,5 (Sep.2015) 688-728.

Matthew Levy, Stanley L. Miller, Karen Brinton, Jeffrey L. Bada: Prebiotic Synthesis of Adenine and Amino Acids Under Europa-like Conditions, Icarus 145,2 (June 2000) 609-613.

Fang Li, Jingqin Xu, et al: Pressure-Induced Polymerization: Addition and Condensation Reactions, Molecules 26,24 (Dec.2021) 7581.

Jongdoo Lim, Mauri Kostiainen, Jan Maly, & 4 others; Synthesis of Large Dendrimers with the Dimensions of Small Viruses, J.Amer. Chem. Soc. 135,12 (Mar.2013) 4660-4663.

Barbara Sherwood Lollar, T.D.Westgate, J.A.Ward, G.F.Slater, G.Lacrampe-Couloume: Abiogenic formation of alkanes in the Earth's crust as a minor source for global hydrocarbon reservoirs, Nature 416 (Apr.2002) 522-524.

Bente Aa Lomstein, A.T.Langerhuus, S.D'Hondt, B.B.Jørgensen, A.J.Spivack: Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment, Nature 484,7392 (Mar.2012) 101-114.

C.C.McCune, F.Wm.Cagle Jr, S.S.Kistler: The effect of hydrostatic pressure on the rate of racemization of 6-nitro-2,2'-carboxybiphenyl, J. of Physical Chemistry 64,11 (Nov.1960) 1773-1775.

Frank J. Millero, Fen Huang, Antonio Lo Surdo, Faina Vinokurova: Partial Molal Volumes and Compressibilities of Phosphoric Acid and Sodium Phosphates in 0.725 Molal NaCl at 25°C, The Journal of Physical Chemistry B 114,49 (Dec.2010) 16099-16104.

Pierre-Alain Monnard, Charles L. Apel, Anastassia Kanavarioti, David W. Deamer: Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium, Astrobiology 2,2 (Summer 2002) 139-152.

Charles J. Maurer, Gordon Shaw, Vicky S. Smith, & 4 others: Pressure polymerization of polyester, US patent 6127493A (Jan.2000).

Edmund R.R. Moody, S.Alvarez-Carretero, & 17 others: The nature of the LUCA and its impact on the Early Earth, Nature Ecology & Evolution 8 (July 2024) 1654-1666.

Edmund R.R. Moody, Tara A. Mahendrarajah, & 7 others: An estimate of the deepest branches of the tree of life from ancient vertically evolving genes, eLife (Feb 2022).

S.J.Mojzsis, T.M.Harrison, R.T.Pidgeon: Oxygen-isotope evidence from ancient zircons for liquid water on Earth's surface 4300 Myr ago, Nature 409,6817 (Jan.2001) 178-181.

Marc Neveu, Hyo-Joong Kim, Steven A. Benner: The "strong" RNA world hypothesis: fifty years old, Astrobiology 13,4 (Apr.2013) 391-403.

Tullis C. Onstott: Deep life: the hunt for the hidden biology of Earth, Mars, and beyond, Princeton Univ. Press 2017. QP82.O57. Onstott's book unfortunately often is more of an "adventure" than "science" book, and has a bad habit of citing "references" not in his reference list pp.463-472. A much shorter and more careful paper (albeit older and out of date) which was the embryo of this book was James K. Fredrickson & Tullis C. Onstott: Microbes Deep inside the Earth, Scientific American 275,4 (Oct.1996) 68-73.

L.E.Orgel: Evolution of the genetic apparatus, J.Molecular Biology 38,3 (Dec.1968) 381-393.

L.E.Orgel: RNA catalysis and the origins of life, J.Theoretical Biology 123 (1986) 127-149.

L.E.Orgel: The origin of life – review of facts & speculations, Trends Biochem. Sci. 23,12 (Dec.1998) 491-495.

Tsubasa Otake, Takashi Taniguchi, Yoshihiro Furukawa, Fumio Kawamura, Hiromoto Nakazawa, Takeshi Kakegawa: Stability of amino acids and their oligomerization under high-pressure conditions: implications for prebiotic chemistry,

Astrobiology 11,8 (Oct.2011) 799-813.

Tommy J. Phelps, E.M.Murphy, S.M.Pfiffner, D.C.White: Comparison between geochemical and biological estimates of subsurface microbial activities, Microbial Ecology 28,3 (Jan.1994) 335-349. See also 351-364.

Simon Nicholas Platts: Contributions to chemical questions in origins of life, PhD thesis, biochemistry, Rensselaer Polytechnic Institute, 2006. Advisor: Gerald M. Korenowski. 221pp.

Paul W. Purdom: Tree Size by Partial Backtracking, SIAM J. Computing 7,4 (1978) 481-491.

Lydia Robert, Jean Ollion, & 4 others: Mutation dynamics and fitness effects followed in single cells, Science 359,6381 (Mar.2018) 1283-1286.

Morrison Rogosa: Acidaminococcus gen. n., Acidaminococcus fermentans sp. n., Anaerobic Gram-negative Diplococci Using Amino Acids as the Sole Energy Source for Growth, J.Bacteriology 98,2 (May 1969) 756-766.

Tomas Rokicki, Herbert Kociemba, Morley Davidson, John Dethridge: The diameter of the Rubik's cube group is twenty, SIAM J. Discrete Maths. 27,2 (2013) 1082-1105. Also TR & MD wrote God's Number is 26 in the Quarter Turn Metric (cube20.org, 2017).

Anthony H. Romano & Walter J. Nickerson: Utilization of amino acids as carbon sources by Steptomyces fradiae, J.Bacteriology 75,2 (1958) 161-166.

Alexey Savelyev: Assessment of DNA Partial Specific Volume and Hydration Layer Properties from CHARMM Drude Polarizable and Additive MD Simulations, Physical Chemistry Chemical Physics 23,17 (Apr.2021) 10524-10535.

Dana Shapira & James A. Storer: Edit Distance with Move Operations, J.Discrete Algorithms 5,2 (June 2002) 85-98.

Anurag Sharma, James H. Scott, & 5 others: Microbial activity at gigapascal pressures, Science 295,5559 (Feb.2002) 1514-1516.

Alexei I. Slesarev, Katja V. Mezhevaya, Kira S. Makarova, & 13 others: The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens, PNAS 99,7 (Apr.2002) 4644-4649.

Warren D. Smith: Hypothesized Evolutionary Purpose of Endocannabinoid System, (Aug.2025).

Ken Takai & 9 others: Cell proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation, Proc. Natl. Acad. Sci. USA 105,31 (July 2008) 10949-10954.

T.Tashiro, A.Ishida, M.Hori, M.Igisu, M.Koike, P.Mejean, et al: Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada, Nature 549 (2017) 516-518.

Connor J. Taylor, Alexander Pomberger, & 7 others: A Brief Introduction to Chemical Reaction Optimization, Chemical Reviews 123,6 (2023) 3089-3126.

Yadu B. Tewari, David K. Steckler, Robert N. Goldberg, William L. Gitomer: Thermodynamics of hydrolysis of sugar phosphates, J. Biological Chemistry 263,8 (Mar.1988) 3670-3675.

Yadu B. Tewari, Robert N. Goldberg: Thermodynamics of hydrolysis of disaccharides, Biophysical Chemistry 40,1 (Apr.1991) 59-67

E.Tombari, G.Salvetti, C.Ferrari, G.P.Johari: Kinetics and thermodynamics of sucrose hydrolysis from real-time enthalpy and heat capacity measurements, J.Phys.Chem.B 111,3 (Jan.2007) 496-501.

Fumio B. Ueno, Akira Nagasaka, Kazuo Saito: Pressure effect on the intramolecular racemization of Tris(acetylacetonato)germanium(IV)perchlorate, Inorganic Chemistry 20,10 (Oct.1981) 3504-3506.

Dietrich Vanlint, Rachael Mitchell & 5 others: Rapid acquisition of gigapascal-high-pressure resistance by Escherichia

coli, Mbiology 2,1 (Jan.2011) e00130-10.

Aurele Vuillemin, Sergio Vargas, 6 others: Atribacteria Reproducing over Millions of Years in the Atlantic Abyssal Subseafloor, ASM Journals mBio 11,5 (Oct.2020) e01937-20.

Laurent Weiss, Abdel A. Tazibt, A. Tidu, M. Aillerie: Water density and polarizability deduced from the refractive index determined by interferometric measurements up to 250 MPa, Journal of Chemical Physics 136,12 (Mar.2012) 124201. [Erratum J.Chem.Phys. 141,14 (Oct.2014) 149904.]

Madeline C. Weiss, Filipa L. Sousa, Natalia Mrnjavac, Sinje Neukirchen, Mayo Roettger, Shijulal Nelson-Sathi, William F. Martin: The physiology and habitat of the last universal common ancestor, Nature Microbiology 1,9 (Jul. 2016) 16116. See also reply by J.P.Gogarten & D.Deamer: 16229 and Weiss et al's counter Reply to 'Is LUCA a thermophilic progenote?', Nature Microbiology 1 (Nov.2016) 16230.

William B. Whitman, David C. Coleman, William J. Wiebe: Prokaryotes: The unseen majority, Proc.Natl.Acad.Sci. USA 95,12 (1998) 6578-6583.

H.H.Wieldraaijer, J.A.Schouten, N.J.Trappeniers: Investigation of the Phase Diagrams of Ethane, Ethylene, and Methane at High Pressures, High Temp.-High Press. 15 (1983) 87-92.

Carl Woese: The universal ancestor, Proc. Natl. Acad. Sci. USA 95,12 (Jun.1998) 6854-6859.

Tobias Wöhrle, Iris Wurzbach, 9 others: Discotic Liquid Crystals, Chemical Reviews 116,3 (Oct.2015) 1139-1241. An earlier review by many of the same authors is Sabine Laschat & 10 others: Discotic liquid crystals: from tailor-made synthesis to plastic electronics, Angewandte Chemie Int'l. Edition in English 46,26 (2007) 4832-4887.

John Q. Wu & Robert B. Macgregor Jr: Pressure dependence of the melting temperature of dA.dT polymers, Biochemistry 32,46 (Nov.1993) 12531-12537; Pressure dependence of the helix-coil transition temperature of poly[d(G-C)], Biopolymers 35,4 (Apr.1995) 369-376.

Hubert P. Yockey: Information Theory, Evolution, and the Origin of Life, Cambridge Univ. Press 2005. QH506.Y634.

K.Zahnle, N.Arndt, C.Cockell, A.Halliday, F.Nisbet, F.Selsis, et al: Emergence of a habitable planet, Space Sci. Rev. 129 (2007) 35-78.

Kevin Zahnle, Laura Schaefer, Bruce Fegley: Earth's Earliest Atmospheres, Cold Spring Harbor Perspect Biol. 2,10 (Oct.2010) a004895.

Alexander A. Zamyatnin: Protein volume in solution, Progr. Biophys. Molec. Biol. 24 (1972) 107-123; Amino Acid, Peptide, and Protein Volume in Solution, Annual Review of Biophysics & Bioengineering 13,1 (Feb.1984) 145-165. See also the follow-up papers by Q.Yuan et al: J.Chemical Thermodynamics 38,1 (Jan.2006) 20-33 and Christelle Pommie, Severine Levadoux, Robert Sabatier, Gerard Lefranc, Marie-Paule Lefranc: IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties, J.Molec.Recognit. 17,1 (Jan-Feb.2004) 17-32 and their online amino acid "Aide-memoire" property tables.

Alexander Z. Zelikovsky: An 11/6 approximation algorithm for the network Steiner problem, Algorithmica 9,5 (1993) 463-470.

Alexander Z. Zelikovsky: A faster approximation algorithm for the Steiner tree problem in graphs, Information Processing Lett. 46 (1993) 79-83.

Nicolle E.B. Zellner, Vanessa P. McCaffrey, Jayden H.E. Butler: Cometary Glycolaldehyde as a Source of Pre-RNA Molecules, Astrobiology 20,11 (2020) 1377-1388.

Huaying Zhao, Patrick H. Brown, Peter Schuck: On the distribution of protein refactive index increments, Biophysics Journal 100,9 (May 2011) 2309-2317.

Lijun Zhou, Dian Ding, Jack W. Szostak: The virtual circular genome model for primordial RNA replication, RNA 27,1

(Jan.2021) 1-11.

Answer to puzzle:

```
In the beginning God created the heavens and the earth.
Now the earth was formless and empty, darkness was over the surface of the deep,
and the Spirit of God was hovering over the waters.
And God said "Let there be light."
```

Amino Acids

The 20 standard biological amino acids, their abbreviations, observed frequencies in the proteome, #codons in standard genetic code, #atoms, molecular weight, AMV, and type. Inside proteins the molecular weight of each non-first amino gets reduced by 18.015 (the molecular weight of water) so that long proteins weigh about 110.70 AMU per amino. Here "AMV" stands for "apparent molecular volume" in mL/mol in water as given by Zamyatnin 1984 in his table 3. These 20 supply a wide palette of chemical characteristics. Based on the frequencies shown, and them alone, the entropy of protein sequences would be 4.167 bits/amino, close to $\log_2(18.00)=4.170$.

Amino acid	3-letter abbrev.	1-letter	freq(%)	#codons	#atoms	molec.wt.	AMV	type
Alanine	**Ala**	A	8.8	4	13	89.1	60.4	hydrophobic
Arginine	**Arg**	R	5.8	6	26	174.2	127.3	+ polar
Asparagine	**Asn**	N	3.9	2	17	132.1	78.0	polar amide
Aspartic acid	**Asp**	D	5.5	2	16	133.1	73.8	– Brønsted base
Cysteine	**Cys**	C	1.4	2	14	121.16	73.4	thiol, Brønsted acid
Glutamine	**Gln**	Q	3.9	2	19	146.15	93.9	polar
Glutamic acid	**Glu**	E	6.3	2	20	147.13	85.9	– Brønsted base
Glycine	**Gly**	G	7.0	4	10	75.1	43.2	hydrophobic
Histidine	**His**	H	2.3	2	20	155.16	98.8	weak +, Brønsted acid & base
Isoleucine	**Ile**	I	5.5	3	22	131.2	107.5	hydrophobic
Leucine	**Leu**	L	9.7	6	22	131.2	107.5	hydrophobic
Lysine	**Lys**	K	5.2	2	24	146.2	108.5	+ Brønsted acid, thioether
Methionine	**Met**	M	2.3	1	20	149.2	105.3	hydrophobic
Phenylalanine	**Phe**	F	3.9	2	23	165.2	121.2	hydrophobic, aromatic
Proline	**Pro**	P	5.0	4	17	115.13	81.0	hydrophobic
Serine	**Ser**	S	7.1	6	14	105.1	60.3	polar, hydroxyl
Threonine	**Thr**	T	5.5	4	17	119.1	76.8	polar, hydroxyl
Tryptophan	**Trp**	W	1.3	1	27	204.2	143.9	hydrophobic, aromatic
Tyrosine	**Tyr**	Y	2.9	2	13	181.2	123.1	Brønsted acid, aromatic
Valine	**Val**	V	6.7	4	19	117.15	90.8	hydrophobic
STOP		X		3	0	0	0	
(overall)			∑=100.0	avg=3.2	wtavg=18.28 avg=18.65	wtavg=128.85 avg=136.90	wtavg=86.89 avg=92.17	

Standard Genetic Code Table for mRNA

Note that, in the standard code (plus in 28 of the 30 variant codes given by Wikipedia as of Oct.2025; those two exceptions code CGG→Trp to enjoy even more GC:TA ratio-varying capability) only the 4 aminos Gly, Ala, Asp, Pro (not 8), can be generated by CG-only DNA. (In contrast, TA-only DNA can code Lys, Ile, Asn, Phe, Leu, Tyr, and STOP.) These 4 aminos form a "**mini-code**," highlit yellow, determined by the first two RNA bases in each triplet *alone*. That sub-code suggests that originally RNA world's protein-creating capability involved only the 4 aminos Gly, Ala, Asp, Pro and only the two RNA bases C and G. (Or if A/U were involved, it was only for start/stop, or perhaps "comma," not coding, purposes.) If so, this also suggests *deep hot* origin because precisely the kind of DNA which **"melts" hottest is GC-only**, typically 41°C hotter than TA-only DNA at P=1 atm pressure; and this melting point difference increases with P.

The 9 "essential" amino acids, which humans must eat since they cannot synthesize them, are starred (*). If we re-ordered the bases UACG instead of UCAG, that would bring the mini-code together to form a contiguous sub-block on the lower right, and also bring the *s together to make them contiguous mainly on the left side (except Trp) – a layout some may prefer.

1st Base	2nd Base				3rd Base
	U	C	A	G	
U	Phe* (UUU)	Ser (UCU)	Tyr (UAU)	Cys (UGU)	U
	Phe* (UUC)	Ser (UCC)	Tyr (UAC)	Cys (UGC)	C
	Leu* (UUA)	Ser (UCA)	STOP (UAA)	STOP (UGA)	A
	Leu* (UUG)	Ser (UCG)	STOP (UAG)	Trp* (UGG)	G
C	Leu* (CUU)	Pro (CCU)	His* (CAU)	Arg (CGU)	U
	Leu* (CUC)	Pro (CCC)	His* (CAC)	Arg (CGC)	C
	Leu* (CUA)	Pro (CCA)	Gln (CAA)	Arg (CGA)	A
	Leu* (CUG)	Pro (CCG)	Gln (CAG)	Arg (CGG)	G
A	Ile* (AUU)	Thr* (ACU)	Asn (AAU)	Ser (AGU)	U
	Ile* (AUC)	Thr* (ACC)	Asn (AAC)	Ser (AGC)	C
	Ile* (AUA)	Thr* (ACA)	Lys* (AAA)	Arg (AGA)	A
	Met* (AUG)	Thr* (ACG)	Lys* (AAG)	Arg (AGG)	G
G	Val* (GUU)	Ala (GCU)	Asp (GAU)	Gly (GGU)	U
	Val* (GUC)	Ala (GCC)	Asp (GAC)	Gly (GGC)	C
	Val* (GUA)	Ala (GCA)	Glu (GAA)	Gly (GGA)	A
	Val* (GUG)	Ala (GCG)	Glu (GAG)	Gly (GGG)	G