THE MATHEMATICS OF ORIGIN

Mathematical Basis for Probability Calculations Used in *Origin*

The film's animated sequence depicting an amoeba hauling the entire universe—one atom per round trip, while waiting for a single usable protein to self-assemble by chance— draws on a number of concepts and calculations. Here, we provide the background and mathematical basis for this illustration.

Today, origin of life research continues under the assumptions of materialism. Researchers believe that if they can explain the formation of a building block or a possible energy source, they are making progress toward solving one of the most baffling mysteries of science. The major factor they consistently fail to address is the source of the *information* that is the hallmark of life. It's not enough to get the building blocks of a cell any more than it is to get iron ore for a skyscraper. The building blocks need to be assembled and arranged in a purposeful way. That's the *sequencing problem* for RNA, DNA and proteins.

PROTEINS

Since protein machines do most of the work in living cells (both modern and primordial), their existence merits explanation.

As discussed in ***Origin***, proteins are constructed from precisely sequenced chains of amino acids. Most proteins in the simplest life forms (Achaea) range from 156 to 283 amino acids in length.[a] Some shorter proteins exist (more accurately called "polypeptides"), but most of them have simpler roles in the cell, acting as signaling molecules or cofactors. Some proteins contain many hundreds or thousands of amino acids. We chose a smaller-than-average protein of 150 amino acids to illustrate the difficulty of sequencing *any* protein by chance—including those required in the first living cell.

PROBABILITY

The estimated probability for a 150-amino-acid protein comes from the work of Douglas Axe and Stephen Meyer. Axe published a paper in 2004 that calculated the fraction of useful proteins in random chains of amino acids.[b] A "useful" protein must be able to fold into a stable structure to perform any function. Compared to the huge number of random chains that would not fold, the number of proteins with this ability is miniscule.

After carefully measuring the tolerance to change in particular enzymes, Axe estimated that only one in 10^{74} chains of 150 amino acids would fold and be functional. This implies that you would have to search through 10^{74} chains of that length to find a single useful protein. So we start by looking for one protein (any chain of 150 amino acids) that *could* be useful in a primitive cell by spontaneously folding into a stable shape.

In his book, *Signature in the Cell*,[c] Stephen Meyer recognized two additional constraints for the chance origin of a protein. First, amino acids need to be "one-handed." In nature, amino acids (except the simplest, glycine) come in two forms: left-handed and right-handed. All living things use only the left-handed form. This is what gives proteins the ability to fold. Experiments show that random chains using both hands become useless lumps of molecules. Indeed, living cells cannot tolerate wrong-handed amino acids, and employ quality controls to ensure their amino acids are left-handed. The earliest imaginable life form could have used either left-handed or right-handed building blocks, but would have

to select one or the other--and stick with it. With that constraint, Meyer assumes an additional 1-out-of-2 chance the correct form would be selected at each point in the chain (after the positioning of the first building block). For 150 amino acids, that becomes 0.5^{-149} or one chance in 10^{45}.

The second constraint concerns the type of bond that must form between amino acids. Proteins use peptide bonds in which the H atom on one end joins with the OH atoms on the other end, releasing H_20—a water molecule. (This, incidentally, is why proteins cannot be expected to form spontaneously in water, because peptide bond formation would go against chemistry's law of mass action; such bonds would be far more likely to break than join.) Other bonds between amino acids, however, are possible. They result in clumps of useless "tar" as biochemists call it. Meyer assumed another 1-out-of-2 chance that each bond would be a peptide bond (a generous assumption). That decreases the probability by another 10^{45}.

Putting the probabilities together means adding the exponents. The probability of getting a properly folded chain of one-handed amino acids, joined by peptide bonds, is one chance in $10^{74+45+45}$, or one in 10^{164} (Meyer, p. 212). This means that, on average, you would need to construct 10^{164} chains of amino acids 150 units long to expect to find one that is useful. How long would that take?

THE TIME PROBLEM

The amoeba illustration used in ***Origin*** is adapted from the book *Evolution: Possible or Impossible?* by James F. Coppedge (ch. 6-7).[d] He constructed a fictional scenario to give readers a way to visualize the powerlessness of chance. In his memorable illustration, he set up a race between a world filled with amino acids and an amoeba .

To give protein formation (by chance) the best possible circumstances to succeed, Coppedge imagined the Earth stocked with sets of amino acids, using all the available atoms of nitrogen, carbon, and oxygen on the planet—for a total of 10^{41} possible sets. Although no one knows the exact quantities of these atoms, the 10^{41} sets of 20 different amino acids estimated for our experiment is a very generous total. [e]

We then calculated the self–assembly rate of 150-amino-acid chains at one chain per second. In the assumed 4.6-billion-year age of the earth, we could expect the construction of at least 1.45×10^{58} chains. That is far short of the 10^{164} chains that would have to self-assemble (on average) to expect one useful protein.

So, now we can ask, how long would you have to wait before the single lucky protein forms? The answer can be calculated by letting the experiment run as long as necessary to expect, on average, a successful chain to form.[f] Dividing the number of trials needed by the rate of formation yields a waiting time of 3.15×10^{115} years. That's far, far longer than the assumed age of the entire universe (13.7×10^9 years). In ***ORIGIN***, we set out to visualize how much time would be involved.

THE TRAVELING AMOEBA

First, we construct a bridge across the universe and position an amoeba that will carry a single atom from one end to the other. Using 90 billion light years as the assumed diameter of the observable universe, it would take the amoeba (traveling one foot per year), 5.7×10^{27} years to travel across the universe, drop the atom off, and return for more. Since there are approximately 10^{80} atoms in the universe, [g] the amoeba could transport them all in 5.7×10^{107} years at one atom per round trip.

But that number is seven orders of magnitude smaller than the time required for the protein to form by chance. Dividing one by the other, we learn that in the time we could expect chance to build one protein, the amoeba would have plenty of time--traveling just one foot per year, one atom per round trip--to haul over 56 million universes!

As noted in ***Origin***, even if you could imagine one protein forming by chance, you would not have life. You would have one small protein, just a lifeless arrangement of amino acids. The simplest living cell we know of has over 300 different types of proteins (to say nothing of nucleic acids, fats, sugars, and a membrane to keep them together). And all of these cellular components would have to be available in the same tiny place at the same time, before life could begin. Subsequent proteins would be even less probable, since they would have to match the first.

For anyone who desires to cross check our calculations, below are the mathematical factors and steps involved:

Let T_p represent the average time it would take for a usable protein of length *n* to form.

Let T_u represent the time it would take for the amoeba to haul the whole universe one atom per round trip.

Then the quantity *U* (number of universes hauled) is $T_p \div T_u$. If $T_p > T_u$, then the amoeba can haul more than 1 universe. If $T_p < T_u$, the amoeba can only haul a fraction of the universe.

Let's look in detail at the factors involved in calculating T_p and T_u.

a = number of amino acid types used in proteins (20, only considering canonical types, since those are the only types that remained after the origin of life.

n = number of amino acids in the target protein chain. Set as a constant, 150.

r = rate of chain formation in each set per year. At 1 per second, that's 3.15×10^7 chains per year per set (60 seconds x 60 minutes x 24 hours x 365 days).

s = number of sets forming chains, limited by C, N, O on earth = 10^{41} (see Coppedge, 109, and footnote).

P = inverse probability of usable protein of length *n*. For n = 150 amino acids, this is 10^{164} (Meyer, p. 212). This is the number of trials required, on average, in which to expect a success.

T_p is calculated as: Pr·sPr·s

d = diameter of the universe in inches. 90 billion light-years (a common estimate) is 3.4×10^{28} inches.

u = number of atoms in the universe. We use 10^{80}atoms as stated above.

i = rate amoeba travels in inches per year. We use 1 foot, which is 12 inches per year.

T_u is calculated as: 2d·ui2d·ui

$U = T_p / T_u$ is therefore: Pr·s2d·uiPr·s2d·ui

Simplifying, this becomes P·i2d·u·r·sP·i2d·u·r·s

Substituting quantities, this results in:

U = 10164·122·(3.4·1028)·(1080)·(3.15·107)·(1041)10164·122·(3.4·1028)·(1080)·(3.15·107)·(1041)

= 5.6 x 10^7 = 56,000,000 universes (56 million)

It's important to note that with the probability so vanishingly small, no amount of reasonable adjustment of the inputs will make a difference to the conclusion. For instance, if the waiting time for a protein to self-assemble was reduced drastically so that the amoeba could only haul one universe, or half a universe, chance still would fail miserably.

WHAT OTHERS HAVE SAID

Many scientists have recognized the impotence of chance to explain the first living cell. Among them are:

- The Wistar Institute Symposium "Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution" (1966) concluded that chance was hopelessly inadequate to account for life, and that new physical laws were needed. Stanislaw Ulam said,
    - "[I]t seems to require many thousands, perhaps millions, of successive mutations to produce even the easiest complexity we see in life now. It appears, naively at least, that no matter how large the probability of a single mutation is, should it be even as great as one-half, you would get this probability raised to a millionth power, which is so very close to zero that the chances of such a chain seem to be practically non-existent." (cited in Evolution News & Views, 7/11/06).
    - On the 50th anniversary of the Wistar symposium (4/25/16), Dr Paul Nelson discussed the enduring significance of the gathering on Evolution News & Views
- Ilya Prigogine said in 1972, "The probability that at ordinary temperatures a macroscopic number of molecules is assembled to give rise to the highly ordered structures and to the coordinated functions characterizing living organisms is vanishingly small. The idea of spontaneous genesis of life in its present form is therefore highly improbable, even on the scale of the billions of years during which prebiotic evolution occurred." Prigogine, Nicolis and Babloyantz, "Thermodynamics of evolution," *Physics Today* 25 (November, 1972), p. 23
- Astronomers Fred Hoyle and Chandra Wickramasinghe in their book *Evolution From Space* (Simon and Schuster, 1981) calculated the chance of getting all enzymes needed for a cell at 1 in $10^{40,000}$—"an outrageously small probability that could not be faced even if the whole universe consisted of organic soup" (p. 24). Other quotations:
    - "Biochemical systems are exceedingly complex, so much so that the chance of their being formed through random shufflings of simple organic molecules is exceedingly minute, to a point indeed where it is insensibly different from zero" (p.3).
    - "There is no way in which we can expect to avoid the need for information, no way in which we can simply get by with a bigger and better organic soup, as we ourselves hoped might be possible a year or two ago. The numbers we calculated above are essentially just as unfaceable for a universal soup as for a terrestrial one" (p. 31).

- In his latest book *Undeniable: How Biology Confirms Our Intuition That Life Is Designed* (Harper One, 2016), Douglas Axe proves that "fantastically" improbable things (such as the chance formation of a functional protein) are physically impossible, because there are not enough resources in the universe to allow them to occur.

FOOTNOTES

[a] The average protein length found in 97 species of archaea was 283 amino acids, according to Tiessen et al., "Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species….," *BioMed Central Research Notes*, 2012. The most frequent occurrences were lengths between 150-250 amino acids.

Brocchieri and Karlin found the median protein length for 16 archaea species to be 247 amino acids. The shortest (~156 amino acids median) were involved in transcription; the longest (~282 to 360 amino acids) were involved in cell division and lipid metabolism. Brocchieri and Karlin, "Protein length in eukaryotic and prokaryotic proteomes," *Nucleic Acids Research* 33:10, 2005.

Zhang found average protein lengths for 5 completely sequenced archaea species to be between 237 and 282 amino acids (average 270). J. Zhang, "Protein length distributions for the three domains of life," *Trends in Genetics*, 2000.

[b] Douglas D. Axe, "Estimating the prevalence of protein sequences adopting functional enzyme folds," *Journal of Molecular Biology* 2004 Aug 27;341(5):1295-315.

[c] Stephen C. Meyer, *Signature in the Cell*, Harper One (2009), pp. 210-212.

[d] James F. Coppedge, *Evolution: Possible or Impossible?*Center for Probability Research in Biology, 1995, ch. 6-7.

[e] Since the other elements are more abundant, the limiting factor would be nitrogen. No one knows the amount of nitrogen on earth, but one source estimates the abundance of nitrogen in the crust to be approximately .002% the mass of the Earth's crust.

http://periodictable.com/Properties/A/CrustAbundance.an.html

Calculating .002% of the mass of the Earth's crust (2.4×10^{22}kg) yields 4.8×10^{17} kg of nitrogen. Converting to atomic mass units, that becomes 2.9×10^{44} amu. Dividing by 14, the atomic mass of nitrogen, yields 2×10^{43} atoms of nitrogen on the earth. This value is in the ballpark with Coppedge's estimates by one order of magnitude. If we add in all the nitrogen in the atmosphere, the value rises to 9×10^{43} atoms of nitrogen on earth.

To figure the number of sets available, we note that the average amino acid contains 1.35 atoms of nitrogen. So we divide the number of nitrogen atoms by 1.35 and also by 20, the number of amino acids required per set, and by 150, the length of the chain. This yields 2.2×10^{40} sets, which we can round up

to 10^{41} sets to make it easier for chance to succeed (the more trials, the higher probability of success). This number is in close agreement with Coppedge's original estimate.

[e] The waiting time to expect a success, on average, is the inverse probability divided by the number of trials per year (formation rate). The probability is one in 10^{164}, meaning that 10^{164} chains would need to be generated, on average, to expect a success. The formation rate is the number of sets times the number of chains each set makes per year (10^{41}sets times 3.15 x 10^{7} chains per year per set, which is 3.15 x 10^{48} trials per year). Dividing 10^{164} by 3.15 x 10^{48} yields 3.15 x 10^{115} years, the waiting time to expect one success on average.

An article at Universe Today estimates the number of atoms in the universeat somewhere between 10^{78} and 10^{82}. We use the middle of this range, 10^{80} atoms.