

## *Probability and the First Proteins*

*Even the simplest of these substances [proteins] represent extremely complex compounds, containing many thousands of atoms of carbon, hydrogen, oxygen, and nitrogen arranged in absolutely definite patterns, which are specific for each separate substance. To the student of protein structure the spontaneous formation of such an atomic arrangement in the protein molecule would seem as improbable as would the accidental origin of the text of Virgil's "Aeneid" from scattered letter type.<sup>1</sup>*

—A. I. Oparin

THE QUESTION MAY have occurred to the reader: If evolution is so improbable, why do many scientists and other intelligent people accept it? This naturally is a central issue, and chapter 11 will be devoted to it.

It should be mentioned now, however, that even the most intelligent people may not get around to concentrating their intelligence on some specific issues. It is customary to accept without question many conclusions that are supposed to be "common knowledge" among scientists. No one has time to look into every subject. It is easy for even a scientist to take for granted that certain principles or axioms have actually been proved.

It is currently the conventional attitude among scientists never

---

<sup>1</sup> A. I. Oparin, *The Origin of Life* (New York: Dover Publications, 1953 edition), pp. 132, 133.

to question the notion that Darwin really proved the main idea of evolution.

An avowed communist like Oparin can freely admit, as in the quotation above, that it is futile to expect chance to put together anything so highly ordered. He resorts instead to his philosophy of dialectical materialism.

“Rationally,” Oparin writes, “life can be materialistically comprehended only as a special form of the motion of matter, arising in a regular manner at a certain stage in the development of matter.”<sup>2</sup>

For the noncommunist materialist, this makes a great difficulty. Many evolutionists have followed Oparin’s careful step-by-step plan of evolving life from nonliving chemicals. They see no reason, however, to accept his communist philosophy. This leaves one in an intolerable position, as it turns out. The customary procedure seems to be to hope that nobody will notice the predicament, and to glide cheerfully over the problem.

The difficulty is that there is nothing left but chance to do the whole job of creating life from nonlife. Even to consider the thought of an intelligent Creator is completely out of the question, “an idea which of course has been long exploded.”

As seen in the previous chapter, natural selection would have no way to come to the rescue at that stage, for lack of an accurate duplicating process.

### *Wrong Ideas About Probability*

Probability theory does not deal with absolutes. It never says that a thing is “impossible” unless it is completely outside the realm of subjects which involve uncertainty. Charles-Eugène Guye, who was the eminent Professor of Physics at the University of Geneva, delineated this truth (using the word “evolution” in the general sense of any process of change, rather than as we are defining it):

---

<sup>2</sup> A. I. Oparin, *Genesis and Evolutionary Development of Life* (New York: Academic Press, 1968), p. 6.

This communist belief, as Tresmontant indicated (see page 66), is but a form of pantheism, the idea that “everything is God,” a false faith which involves no responsibility to a higher Power. In the view of communists, matter by its very nature must develop. When life finally evolves as a matter of course, new biological laws come into play as just one stage in this development. This view actually is a religion, and “matter in motion” serves as its god. This, in effect, turns out to be a form of the “vitalism” which Dr. Oparin deplors in the writings of some other scientists (such as Erwin Schrödinger, the physicist who added much to our knowledge through his hypothesis on the wave nature of matter).

The evolution of a phenomenon tends to take place according to the greatest probability, and it follows that physico-chemical laws are viewed as statistical laws, very exact it is true on account of the law of large numbers, but no longer possessing that character of absolute determination which it has been customary to attribute to them.<sup>3</sup>

Knowing this, many scientists easily assume that since a thing is *theoretically* not impossible, then it can be expected sooner or later to occur.<sup>4</sup>

Scientists are human and are not immune to some common misconceptions. One such mistaken notion is that even extremely improbable events become certain to occur if you have a few hundred million years to wait. One Harvard Professor in 1954 wrote a widely quoted article proposing that life began by chance: "It becomes increasingly probable," he said, "as the trials are multiplied. Eventually the event becomes virtually inevitable."<sup>5</sup>

But does it? If you apply probability reasoning, you may well doubt if even that well-known scientist did much careful figuring before making such a statement. Hundreds of other biologists, reading his article, doubtless took for granted that someone had actually checked this. If anyone goes to the trouble to do so, however, it soon becomes evident that there is no basis for such a conclusion, although the writer may have been quite sincere.

Thus "myths" are propagated and become almost universal in acceptance. It is usually no one's fault in particular. It is in one sense like the train wreck immortalized by the ballad about Casey Jones, an Illinois Central railroad engineer. When several freight trains were unable to crowd onto one siding, some of the boxcars of the last train were left protruding out on the main line in the path of Casey's oncoming express. Each brakeman from the freight trains supposed someone else had gone back up the line to signal the passenger train to stop.

In the matter we're considering, many a scientist or other

---

<sup>3</sup> Charles-Eugène Guye, *Physico-Chemical Evolution* (New York: E. P. Dutton and Co., 1925), pp. 3, 4.

<sup>4</sup> Professor Guye, on the other hand, stated that the probability of an event may be so slight as to amount to *impossibility* in practical terms. "To assert," he said, "that a phenomenon is impossible or to declare that its chance of occurring is one in a hundred million is 'practically' to say the same thing" (*ibid.*, p. 104). This is understood in the sense of an improbable event within a time limit such that its practical probability is nil.

<sup>5</sup> George Wald, "The Origin of Life," *Scientific American* (August, 1954), p. 47.

educated person supposes that surely someone has investigated these commonly believed scientific dogmas carefully.

With the information now at hand, you will find it possible to check firsthand for yourself the probability of certain popularly held conclusions.

### *Some Proteins Described*

We now come to the exciting application of the laws of chance to this major question: What is the probability that a protein molecule might have been aligned by chance?

All known life on the earth consists largely of these giant molecules. "The chemical basis of all life," says the *Encyclopaedia Britannica*, "is protein in a watery medium."

To learn the secrets of the composition of such an entity as a protein molecule was no simple undertaking. Biologists spent years in painstaking, patient study. One of the first molecules to be mapped was insulin, which is the smallest molecule qualifying as a protein. It has fifty-one amino acid "links," in two chains—one with twenty-one and the other with thirty amino acids. The two chains are joined together by "sulfur bridges."

A method was discovered whereby an X-ray beam could be projected through a crystallized protein molecule. The individual atoms of the molecule scattered the beam as it passed through. From the pattern of this diffraction or deflection, it was possible to figure out the position of the various atoms.

John C. Kendrew completed mapping the myoglobin molecule in 1959. Then, by 1967, Max Perutz and others at Cambridge succeeded in solving the structure of the hemoglobin molecule, the most important protein in blood. Larger than myoglobin, it consists of four chains fastened together and folded in a complex pattern. In all, hemoglobin has 574 amino acid links. The molecule has 10,000 atoms!<sup>6</sup>

Hemoglobin is just one of the multitudinous kinds of proteins.

---

<sup>6</sup> There are some 280,000,000 hemoglobin molecules per red blood cell (M. F. Perutz, "The Hemoglobin Molecule," in *The Molecular Basis of Life* [San Francisco: W. H. Freeman & Co., 1968], p. 39). A drop of blood may contain 35 million red blood cells. Imagine the task of mapping the 10,000 atoms of just one hemoglobin molecule when there are around 10,000,000,000,000,000 such molecules in a drop of blood! The average adult has around 27 trillion mature red cells (George C. Simpson and William S. Beck, *Life, An Introduction to Biology*, shorter edition [New York: Harcourt, Brace, and World, 1969], p. 211). This means that an adult has 75 hundred million trillion hemoglobin molecules!

The length of the average protein in the smallest known living thing is at least 400 amino acid links, containing more than 7,000 atoms.

*Protein Links Must Be in Correct Order*

Most proteins include all of the usual twenty kinds of amino acids. Each protein has a specific exact sequence of these units.<sup>7</sup> Of the hemoglobin molecule, John C. Kendrew wrote (using British spelling):

Thus, practically everyone reading this book contains identically the same kind of haemoglobin molecules in his or her blood, identical down to the last amino acid and the last atom. Anyone having a different haemoglobin would be seriously ill or dead, because only the very slightest changes can be tolerated by the organism.<sup>8</sup>

Research has continued on whether variation can be tolerated. I checked, therefore, with Dr. Kendrew by telephone (November, 1971) for his understanding at that time on this matter. He indicated that the picture is about the same as in the above quotation, although it is now known that occasionally a variation can exist which does not seem to impair health. Most such changes, he noted, would never be discerned, but they probably do not occur very often.

Just how exact a sequence must be in order to work effectively as a protein is of course a crucial question. Even the simplest theoretical living system requires a high degree of precision and delicate regulation of timing and quantities of its protein molecules. Because of the teamwork that is essential, little variation is consistent with the degree of organization that must be present.

This involves a real problem for those who must explain the origin of life without planning, but valiant attempts continue to be made. Appendix 4 will give consideration to some of the interesting ideas which have recently been advanced to try to provide for molecular evolution.

It will be evident from calculations later in this chapter and in chapter 10 that *even if free substitution were allowable at most of the sites* in a chain of amino acids or DNA, the odds

---

<sup>7</sup> In chapter 12, we will list the actual sequences for insulin, toxin in bee sting, and cytochrome *c*, and present intriguing facts about comparing the sequences of different animals and plants.

<sup>8</sup> John C. Kendrew, *The Thread of Life* (Cambridge, Mass.: Harvard University Press, 1966), pp. 32, 33.

are astronomical against random production of a set of usable proteins for minimal life. There's no way without design.

We may assume from present experimental knowledge that many places in a protein chain might allow substitution. It appears probable, however, that only about one such substitution may be tolerated in an entire chain.<sup>9</sup> Moreover, any substitute may have to be similar to the one it replaces.<sup>10</sup> There can be absolutely no substitution in positions in the chain which comprise what is known as the "active site." Any substitution

---

<sup>9</sup> This is considered as an average, because some substituents work, but less efficiently. This conclusion (one allowable substitution per chain) is based in part on the actual observed variations in human hemoglobin, which involve with few exceptions just one amino acid replacement per individual with abnormal hemoglobin. "So delicately balanced are the equilibria involving the hemoglobin tetramer that single amino acid replacements can exert a very profound effect on the structure and reactivity of the whole protein," wrote Margaret Dayhoff (*Atlas of Protein Sequence and Structure 1972*, National Biomedical Research Foundation, Washington, D.C., p. 78).

Different species of organisms have different sequences of amino acids in their hemoglobin chains, but all species apparently have roughly half of the sites identical. In a 1971 conversation with Dr. Dayhoff, we seemed in agreement that the hemoglobin sequence of a particular kind of organism may work best for that organism because of the complex reactions involved between hemoglobin and other molecular entities in the blood of that species.

In the case of a simpler protein, insulin from some species will function when placed in other species, at least in its gross effectiveness. It likely will be found, however, that there are sophisticated or hidden advantages in a species' own characteristic insulin sequence. (A protein as simple as insulin would be more likely to have some characteristics of "universal fit" than more complex ones like hemoglobin. A pair of pliers fits many situations, whereas a complex wrench may fit one certain need. The difference in sequences of insulin in various species amounts in some cases to only one or two amino acids, anyway.)

In chapter 12, the assumed proof of evolution by sequence comparisons will be examined in the case of the protein, cytochrome c.

It will become clear later in this chapter that even if all the differences in protein sequences of different species came about by mutations through the ages, there would still be no practical possibility of getting a set of proteins by random action at the beginning under the rules of probability. (See pages 113, 165, 166.)

Of single substitutions and other mutations, Margaret Dayhoff wrote: "It is generally believed that many of the abnormal chains are deleterious to the individual, at least in a subtle way" (*Ibid.*, p. 67).

In order for hemoglobin evolution, for example, to have occurred, there would have had to be a long series of such substitutions. Each would have had to be at least viable. This would doubtless require matching changes in other inter-related molecular entities in the organism, with synchronized timing of the mutations. At least in the end result, the entire mechanism would have to be advantageous enough to monopolize the gene pool, as is the observed situation in each species. Such precisely matched, intricately coordinated mutations are not a reasonable possibility, of course, under the laws of probability even when assisted by natural selection.

(Dr. Dayhoff would not necessarily concur in these general conclusions, since it seems clear from her writing that she is committed to the standard evolutionary interpretation as to the presumed *origin* of sequence differences. We did not discuss origins.)

<sup>10</sup> Dayhoff, *Protein Sequence*, p. 98, and Kendrew, personal communication, November, 1971.

anywhere is likely to be harmful if not lethal. All twenty of the amino acids are listed in Figure 3.

The present goal will be to find the probability that amino acids might happen to be in the right order by chance if natural forces were able to line them up. It will be considered a success if the resulting chain is *any usable protein molecule*. Also to be determined are the odds that the amino acids might be aligned correctly for a *group of 239 proteins* needed for the minimal living entity described by Morowitz.<sup>11</sup>

Every effort will be made to avoid getting overtechnical, so that the nonbiologist reader may follow the calculations and know for himself that the outcome is sound. That is important in the attainment of certainty. Even if one finds the figuring itself uninteresting or difficult, it is helpful to his progress toward certainty just to know the evidence is there if he ever

FIGURE 3

The Twenty Amino Acids Used in Proteins  
With Their Abbreviations and Atoms  
(in alphabetical order)

Alanine	Ala	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>	Leucine	Leu	C <sub>6</sub> H <sub>13</sub> NO <sub>2</sub>
Arginine	Arg	C <sub>6</sub> H <sub>14</sub> N <sub>4</sub> O <sub>2</sub>	Lysine	Lys	C <sub>6</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub>
Asparagine	Asn	C <sub>4</sub> H <sub>8</sub> N <sub>2</sub> O <sub>3</sub>	Methionine	Met	C <sub>5</sub> H <sub>11</sub> SNO <sub>2</sub>
Aspartic Acid	Asp	C <sub>4</sub> H <sub>7</sub> NO <sub>4</sub>	Phenylalanine	Phe	C <sub>9</sub> H <sub>9</sub> NO <sub>2</sub>
Cysteine	Cys	C <sub>3</sub> H <sub>7</sub> SNO <sub>2</sub>	Proline	Pro	C <sub>5</sub> H <sub>9</sub> NO <sub>2</sub>
Glutamic Acid	Glu	C <sub>5</sub> H <sub>9</sub> NO <sub>4</sub>	Serine	Ser	C <sub>3</sub> H <sub>7</sub> NO <sub>3</sub>
Glutamine	Gln	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub>	Threonine	Thr	C <sub>4</sub> H <sub>9</sub> NO <sub>3</sub>
Glycine	Gly	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	Tryptophan	Trp	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>
Histidine	His	C <sub>6</sub> H <sub>9</sub> N <sub>3</sub> O <sub>2</sub>	Tyrosine	Tyr	C <sub>9</sub> H <sub>9</sub> NO <sub>3</sub>
Isoleucine	Ile	C <sub>6</sub> H <sub>13</sub> NO <sub>2</sub>	Valine	Val	C <sub>5</sub> H <sub>11</sub> NO <sub>2</sub>
Alternate abbreviations for some amino acids are:					
Asparagine	Asp-NH <sub>2</sub>		Isoleucine	Ileu	
Glutamine	Glu-NH <sub>2</sub>		Tryptophan	Try	

The atoms, listed by atomic symbol are: carbon, hydrogen, sulfur, nitrogen, and oxygen. The subscript numbers indicate the number of atoms of that element. (The atomic symbol for these atoms is merely the initial letter.)

Leucine and Isoleucine have the same numbers of atoms of each kind, but are different in spatial arrangement of the atoms.

<sup>11</sup> Harold J. Morowitz (Yale University Biophysics Dept.), personal communications, October and November, 1971.

needs to dig into it further. The next chapter will be easier to read.

### *The Odds Against a Single Insulin Molecule*

Before figuring for average size proteins, just for practice we may calculate the odds for the random alignment of the amino acid units for insulin, since insulin is usually considered the smallest protein, with 51 amino acids. Even insulin, it turns out, is not as simple as it first appears.

The insulin molecule is composed of two strands that must be linked together in an exact manner by sulfur bridges. To bring this about, the cell first constructs a longer chain of more than 80 amino acids called *proinsulin*. It ranges in length from around 81 to 86 in various animals. We will consider an insulin molecule of length 84 amino acids (of which an example is the pig.) This extended sequence of 84 units causes the chain to fold and cross-bond correctly, and then a particular section of 33 units is cut out by special enzymes, leaving the final 51 amino acids in two chains properly oriented with cross links between them.

Chance will therefore need to align 84 amino acids in correct order to form proinsulin, as the *precursor* for insulin.

Since each of the 84 positions in the chain could be occupied by any one of the 20 kinds, the total possible arrangements is  $20^{84}$ , which, after conversion to base 10, is roughly  $10^{109}$ . The different arrangements are considered equally probable; so the probability of any one molecule being in the correct order for insulin is 1 in  $10^{109}$ . Allowing for one substitution (to be tolerated) makes it a bit easier for chance, and brings the probability down to 1 in  $10^{106}$  approximately.<sup>12</sup>

Going back to Dr. Eden's statement that the total number of protein molecules that ever existed on earth might be  $10^{52}$  as a liberal estimate, we will give chance another big boost by assuming that the  $10^{52}$  are all different and are all the proper length for insulin. We can now figure the probability that any one of those would by chance be in correct order for insulin.

Émile Borel, French mathematician, tells the rule for such cases: "If an event can occur in several different ways which are

---

<sup>12</sup> The formula for probability allowing one substitution is  $\frac{(a-1)n+1}{a^n}$  where  $a$  is the number of kinds available and  $n$  is the number of units per chain. In this case, with 20 kinds and 84 in the chain, the result is roughly 1 chance in  $10^{106}$ .



mutually exclusive, its probability equals the sum of the probabilities corresponding to the different alternatives."<sup>13</sup> Any one of those  $10^{52}$  is a different way that *might* fulfill the event of a chain being in the order for insulin. Therefore, the sum of the different ways is  $10^{52}$ . The probability that any one of all that ever existed on earth would be insulin is therefore  $10^{52}/10^{106}$ . Dividing in order to simplify the fraction, the probability is 1 in  $10^{54}$ .

Therefore, *the odds are a million trillion trillion trillion trillion to one that not a single protein molecule of all that ever existed on earth would by chance be in the correct order for an insulin molecule!* (Adjusting for different kinds of insulin would have little effect on a figure this size.)

#### *Usable Proteins and Nonsense Chains of Amino Acids*

In this test, chance is supposed to come up with *any usable protein*. How many of the total number of possible arrangements could be considered usable proteins? There is obviously no way on earth that anyone could say for sure. One way to get a tentative idea is by comparison with other systems which have a number of parts that work only when in specific orders. The best analogy is our alphabet with its 26 letters. By putting them in various sequences, we can make multitudes of messages.

The letters are mere nonsense, however, unless they are in a *meaningful order*. What percentage of the possible orders of the alphabet letters are meaningful?

To answer that question, an experiment was conducted at the Center for Probability Research in Biology. Thirty thousand letters were drawn at random, and then all meaningful sequences were listed—words, phrases, or messages that would carry a meaning to the average American. The results are given below.

FIGURE 4  
Meaningful Orders in 30,000 Random Letters

Meaningful sequences 7 letters in length	1
Meaningful sequences 6 letters in length	3
Meaningful sequences 5 letters in length	17
Meaningful sequences 4 letters in length	139
Meaningful sequences 3 letters in length	1,113
Meaningful sequences 2 letters in length	4,890

<sup>13</sup> Emile Borel, *Elements of the Theory of Probability* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965), p. 19. (First published in France in 1950.)

In 30,000 letters drawn at random, the longest meaningful sequence contained only 7 letters. The letters ABC appeared in order only once in the entire experiment. AB occurred 41 times.

It is quite clear that chance arrangements are predominantly nonsense arrangements. Even many of the sequences that could be said to be meaningful require a stretch of the imagination. For example, the six-letter sequences were: RUINPA, WEETED, and AGMCAP.

A working hypothesis derived from the experiment is that there is between 1/4 and 1/5 as good a chance of getting a meaningful sequence for each added letter of length, on the average. The probability of getting a 400-letter sequence that would be meaningful would then be between 1 in  $4^{400}$  and 1 in  $5^{400}$ . (If one considers different languages using the same alphabet, not many such languages would be possible without changing the value of the letters, so it would not affect the calculations more than a few orders of magnitude. Perhaps only one language should be considered, since the information value for the various letters is different in different languages, and amino acids only have to fit the one "protein language.")

For lack of any other way to estimate the proportion of meaningful amino acid sequences, let it be assumed that a similar probability exists as in the alphabet. We will, however, use the figure of 1 in  $4^{400}$ , because the amino acids are fewer than the alphabet letters, there being only 20.

On that basis, for an amino acid chain 400 long, the probability of getting a usable protein would be 1 in  $4^{400}$ , which is 1 in  $10^{240}$ .

In using this formula based merely on the alphabet analogy, there is, of course, an uncertainty factor. There are many similarities, however, between the alphabet and the 20 amino acids. Some letters are much more "reactive" than others, like amino acids, and some are used sparingly.

This uncertainty affects only the first stage of our study, namely, the first protein molecule to be produced by chance. There must be at least 239 matching protein molecules for a set, in order to provide the minimum number for the smallest theoretically possible living entity. The second of these and

all the ones following would not involve this alphabet formula, and hence the uncertainty factor would be eliminated.<sup>14</sup>

### *Allowing Concessions to Make Things Easier for Chance*<sup>15</sup>

It has already become apparent that chance is very backward when it comes to producing an ordered result. There clearly will be no hope at all of arranging a protein by chance unless some extreme advantages are allowed.

Two approaches will be used. First, we will allow some exceedingly helpful concessions, for the present, to assist chance in arranging an average length protein, and to obtain one minimum set of such proteins for the theoretical smallest living thing. Afterward, we will check on the probability of much simpler protein chains resulting from such random alignment. In the latter case, the concessions will be reduced a bit, but chance will still be given numerous advantages that would not have actually existed at the time of the presumed evolution of the first living thing.

It can be seen at a glance that most of the following fourteen concessions could not have actually been true, but chance will have enough of a task even when offered all this help. If chance fails under such extreme conditions, it should indicate clearly that perhaps it is unreasonable to rely on it at all in the quest for the way life began. Here are the fourteen assumptions for this purpose. Some of these are extreme, and some are not.

1. Assume that the primitive atmosphere was as evolutionists claim.

2. Suppose that all of the twenty amino acids did form naturally and in the right proportions, by the action of ultraviolet rays, lightning, and heat.

3. Presume that the amino acids were formed in only the left-handed configuration.

4. In calculations that follow, consider that the average protein molecule is 400 units in length,<sup>16</sup> which is shorter than the

---

<sup>14</sup> Perhaps it should be emphasized that the uncertainty factor in the first one does not carry over any uncertainty at all to the others in the set. These others would have to be not just any protein that might work *somewhere*, but an exact sequence to fit here.

<sup>15</sup> Chance is constantly being used herein as if the word implied a conscious entity. Charles Darwin once excused a similar usage of words thus: "Every one knows what is meant and what is implied by such metaphorical expressions; and they are almost necessary for brevity." (*Origin of Species*, Mentor Edition, [New York: New American Library, 1958], p. 88.)

<sup>16</sup> Avram Goldstein, Dora B. Goldstein, and Louise Lowney, "Protein Synthesis

445 average length computed earlier for the smallest theoretical cell from Morowitz' data.

5. Postulate *that all the atoms on earth have been used to form amino acids*. That is, all the carbon, nitrogen, oxygen, hydrogen, and sulfur atoms in air, water, and crust of the earth have been made up into amino acids for this all-out effort to get proteins by random alignment.

6. Consider *that all of these amino acids are grouped in sets*. Each set contains one of each kind available at each position of the forming chain. These groupings may be pictured as being in the form of coacervate droplets described by Oparin, or any other way so that they are together.

7. Let it be granted *that these groupings are permanently protected in some manner from the destructive effect of ultraviolet rays*. It is widely recognized that ultraviolet rays would be lethal to the life being formed unless protected in some way. These rays, particularly those in the wavelength range near 2600 Å, are "highly toxic (absorbed by protein and nucleic acids),"<sup>17</sup> with lethal chemical changes resulting.

---

at 0° Centigrade in *Escherichia coli*," *Journal of Molecular Biology*, Vol. 9 (1964), p. 234.

It should be noted that, although many proteins are shorter than that, a living organism requires a full set of many kinds of varied lengths, and the figure mentioned represents *average* length for a minimal cell as described earlier.

<sup>17</sup> Roger Y. Stanier, Michael Doudoroff, and Edward A. Adelberg, *The Microbial World*, 3rd. ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1970), p. 217.

8. Concede also that the amino acids would automatically unite, even though this would require going against an "energy-gradient," and the complex system which unites them in all known living things would be absent.<sup>18</sup>

9. Suppose that one substitution is allowable in each chain. In this concession, it will not be required that the active site be exempt from substitution, and it will be considered permissible for any amino acid to substitute for any other at any point. (See discussion of substitution earlier in this chapter.) If future discoveries ever widen the viable limits of substitution, the extreme concessions we are allowing, such as in number ten below, would take up the slack. In some cases a protein with substitutions may be partially functional.

10. Assume that the rate of chain formation is fantastically rapid, such that an entire chain requires only one-third of a ten-million-billionth of a second! This is around 150 thousand trillion times the normal speed in living things which itself is quite fleet.<sup>19</sup>

---

<sup>18</sup> If the bonding is considered as having taken place in water, a further problem is that amino acids give up a molecule of water when joined. If they are already surrounded by water, that presents an obstacle. For this reason, Fox and his coworkers tried bonding them outside of water. This, however, would have left hypothetical early cells at the mercy of lethal rays.

<sup>19</sup> The much-researched bacterium, *Escherichia coli*, requires only five seconds per protein chain of 400 length at the congenial temperature of 37° C. (Goldstein, et al., "Protein Synthesis," p. 234.) Slower rates have been reported in vertebrates, e.g. one minute for a chain only 100 amino acids long. (Dayhoff, *Protein Sequence*, p. 54.)

The enormous speed we have chosen in concession 10 is some 1200 times the limit of speed for atomic processes, said to be close to 1/10<sup>16</sup> second. (Harold J. Morowitz, *Energy Flow in Biology* [New York: Academic Press, 1968], p. 12.) From data by Pauling, it can be figured that a hydrogen electron orbits its nucleus around 10<sup>16</sup> times per second, which is ten million billion times per second. (Linus Pauling, *The Chemical Bond* [Ithaca, N.Y.: Cornell University Press, 1967 edition], p. 11.) Concession 10 allows 400 units to be joined in 1/3 of 1/10<sup>16</sup> second.

A normal rate in living things depends upon ideal temperature, the proper solvent, and the best concentration of hydrogen ions, as well as conveniently available all the working parts of the protein assembly line to be described in chapter 9, including all twenty of the amino acids in sufficient supply.

Actually, in the opinion of some prominent scientists, solar luminosity affecting the earth 4 billion years ago would have been 60% of present value, with the consequence that "the global temperatures on the earth were substantially below the freezing point of water at the time of the origin of life." (Philip Handler, ed., *Biology and the Future of Man*, [New York: Oxford University Press, 1970], p. 174.)

Consider that idea in relation to the following: "The total temperature span within which organisms can grow is a narrow one, extending from about -5 to +80° C." (Stanier et al, *The Microbial World*, p. 315.) They can survive at lower temperatures, but not grow. Lowering of temperature from an organism's ideal range has the effect of slowing down reactions. From the quotation just noted, at about 5° below zero C, the reactions would be so slow that there would be no growth. Yet the primitive temperature was supposed to have been "sub-

11. For each set of amino acids, let it be figured *that every unusable chain is immediately dismantled and another one made* at the same rate of around 30 million billion per second, which is a trillion trillion ( $10^{24}$ ) per year in each set.

12. Assume that nothing will interfere, so that chance will have an ideal opportunity,<sup>20</sup> and that if a usable sequence is ever obtained, the action will stop so that it may be preserved. (In the matter of trying for a set of 239 proteins, regardless of the speed of trials, it would of course be necessary for 239 contiguous sets to obtain right sequences at the same time. However, even if there were a way to arrange for a long time of overlapping existence of each sequence that occurred, with staggered timing of different sets, it would not make enough difference to affect the outcome.)

13. Consider further *that if 239 proteins in contiguous sets are ever obtained, they will be able to merge into one group of proteins ready for working together in a living system.*

14. For our present purpose, assume *that the age of the earth is five billion years, and that the age of the universe is fifteen billion years.* From an evolutionary standpoint, these rounded figures are more or less standard at the present time. (As we will see, evidence is growing for a *much* younger age.)

With the extreme concessions listed above it might be thought that chance should easily arrange many protein chains in the course of earth's history. On the contrary, we will find it otherwise.

The reader will recognize that many of these concessions are either definitely impossible or else have a probability that is "vanishingly small," to put it mildly. They are made merely to offer chance such a favorable opportunity that its failure to produce should be conclusive.

#### *Using All the Atoms on Earth for Making Proteins by Chance*

In the list of concessions, it was assumed that all of the appropriate atoms on earth's surface, including air, water, and crust of the earth, were made into amino acids and arranged conveniently in sets to make it easier for chance to come up

---

stantially below the freezing point of water at the time of the origin of life." (The freezing point of water is 0° C.) This is very interesting.

<sup>20</sup> A difficulty with many theories on the origin of life is that hope depends on postulating many unusual conditions. As will be seen, however, even this concession will not bring success in seeking a chance origin of proteins.

with a usable protein. It can be estimated that there would be about  $10^{41}$  such sets available.<sup>21</sup>

With each of these sets making a total of  $10^{24}$  different chains per year as assumed in concession 11, that gives a total of  $10^{24} \times 10^{41}$  chains produced on earth in a year's time, which is  $10^{65}$ . Under concession 14, the total chains made since the earth began would be  $5 \times 10^{74}$ , which we will round off to  $10^{75}$ .

*The Odds Against an Average Protein  
by Chance Since the Earth Began*

We have just seen that chance could have made  $10^{75}$  different protein-length chains at the speed assumed during the entire time the earth has existed. Using the formula from the alphabet, we can now estimate how many of those might be considered usable protein molecules.<sup>22</sup> First, we should allow for one substitution per chain. This would have the effect of changing that  $1/10^{240}$  formula to around  $1/10^{236}$ . The probability, then, for usable protein molecules in this total of  $10^{75}$  produced since the world began is  $10^{75}/10^{236}$ . Simplifying the fraction, we get 1 in  $10^{161}$  as the probability that even one would be usable, on the average.

Therefore, *the odds are  $10^{161}$  to 1 that not one usable protein would have been produced by chance in all the history of the earth*, using all the appropriate atoms on earth at the fantastic rate described. This is a figure containing 161 zeroes.

It might be well to recall that even if one molecule *were* obtained, it would not help at all in arranging the second protein molecule unless there existed an accurate duplication process. Even if there were such a process, there are many other *kinds* of proteins needed before there can be a living organism.

---

<sup>21</sup> There are about  $2.5 \times 10^{44}$  nitrogen atoms and  $9 \times 10^{44}$  carbon atoms thus available. Since there is considerable variation in the data from which this can be figured, it will help chance if we use totals calculated from the most liberal estimates, namely:  $9.35 \times 10^{44}$  nitrogen atoms (C. C. Delwiche, "The Nitrogen Cycle," *Scientific American* [September, 1970], p. 140), and  $4.6 \times 10^{45}$  carbon atoms (Morowitz, *Energy Flow in Biology*, p. 69) available on earth. Since 1.35 nitrogen atoms are required per average amino acid in some bacteria, nitrogen would run short in supply before carbon, of which around five atoms are needed per amino acid.

To provide one of each of the 20 amino acids at each point in the desired 400-length chain, there will be enough nitrogen and other required elements for around  $10^{41}$  sets.

<sup>22</sup> The total of possible arrangements 400 amino acids long is  $20^{400}$ , which is the same as  $10^{520}$ . The percentage of usable ones, 1 in  $10^{240}$ , being smaller, is the figure to be used for the first protein. (This figure comes from the alphabet analogy just described.)

In Morowitz' minimal cell, the 239 protein molecules required include *at least 124 different protein species*.<sup>23</sup>

### *To Obtain a Set of Protein Molecules for Minimal Life*

It has just been calculated that the probability of a single protein molecule being arranged by chance is 1 in  $10^{161}$ , using all atoms on earth and allowing all the time since the world began. What is the probability of getting *an entire set* of proteins for the smallest theoretical living entity?

The second protein molecule will be far more difficult than the first, because it has to be part of a matching set. The protein molecules of a cell are quite specifically adapted to work together as a team. We assumed that the first protein could be *any* usable protein that might be good somewhere. Once that first one is specified, the rest of the set has to match it exactly. It is like assembling a car from a crate of automobile parts. Once the kind of automobile is determined by the first component used, then all the other items must be of that same matching group. Nothing else on earth will fit, in most cases, except the part made for that particular purpose.

After the first protein molecule is obtained by chance (if it ever happens), then the others must be quite specific in the same way as the automobile components.

### *The Second Protein Molecule Is More Difficult*

The probability of getting the first protein molecule was influenced by the formula taken from the alphabet analogy. The second one is more difficult to obtain, we have just seen, because it has to be more exact to match the first one, instead of being just any protein.

The total number of possible orders in a chain of 400 amino acids of 20 kinds is  $20^{400}$ . (The formula is: the number of kinds to the power of the number of units in the chain.) As stated above,  $20^{400}$  is the same as  $10^{520}$ .

Considering the first one as already obtained, we need 238 more. The second one could be any one of those 238. The probability is therefore  $238/10^{520}$ . The third one could be any of the 237 still needed, so its probability would be  $237/10^{520}$ . Calculating all of these, and allowing for one substitution per chain,

---

<sup>23</sup> Morowitz, personal communication, October, 1971.



we arrive at a probability of 1 in  $10^{122470}$ . (See note 24 below.) Even if almost a trillion different sequences might work in each protein, the probability resulting is 1 in  $10^{119614}$ . (See note 25 below.)

This figure represents the second through the 239th protein molecules. Multiplying in the first one, which was at a probability of  $1/10^{236}$ , we arrive at the final figure for the minimum set needed for the simplest theoretical living entity, namely, 1 chance in  $10^{119850}$ .

Earlier, we obtained the figure of  $10^{75}$  which was the total number of chains made since the earth began. In order to allow for overlapping sets of 239 each, we will use that same figure to represent the total protein sets formed. Dividing into the big figure just calculated, we learn that *the odds against one minimum set of proteins happening in the entire history of the earth are  $10^{119775}$  to 1.* (See note 26 below.)

Even if such a set could be obtained, we would not have life. It would simply be a helpless group of nonliving molecules alone in a sterile world, uncaring and uncared for, the end of the line. In chapter 10 we will see that even if unlimited substitution is allowed in 9/10 of all the positions, the odds against one minimum set of proteins are still beyond comprehension.

#### *No Conceivable Chance of Success*

The odds against such a correctly ordered set of proteins for the smallest conceivable living entity are thus hopelessly large beyond understanding. The next chapter will make it easier to realize the gigantic size of even the smaller figures with which we have been dealing.

Chance was given every possible concession in this investigation. Its failure was miserable beyond power to describe. Yet chance is the only natural way that the first proteins could have been arranged. In all of the careful efforts of Oparin and his

<sup>24</sup> For one substitution, we again use the formula  $\frac{(a-1)n+1}{a^n}$ , adapting it to the situation just described.

<sup>25</sup> The calculation of this was  $\frac{7601^{238} \times 238! \times 10^{12} \times 238}{10^{520} \times 238}$  yielding approximately  $1/10^{119614}$ . (The  $238!$  means 238 *factorial* and signifies that we must multiply  $238 \times 237 \times 236 \dots \times 1$ . Actually, we used an over-simplification in this instance, which helps chance even more. Instead of all 239 being different kinds, which would require the use of the factorial, the 239 protein molecules include around 124 different protein species.)

<sup>26</sup> The number  $10^{119850}$  divided by  $10^{75} = 10^{119775}$ .

followers in their attempts to get life from nonlife, *no real way has been found to get away from chance. It is the only possible source of the sequence, in the absence of planning,* in spite of the millions of words and the valiant efforts expended in the attempts that have been made.

There is no real reason at present to believe that any living thing has ever existed that is simpler than the *Mycoplasma hominis* H39, which is the smallest living entity known.<sup>27</sup> Although some viruses may be smaller, they do not qualify as autonomous self-replicating systems. A virus cannot duplicate itself without the help of a host cell whose machinery it must use to manufacture its proteins and nucleic acids. Instead of being an earlier step on the ladder of evolution, viruses are now thought by some scientists to be just the opposite, a deterioration or setback rather than part of the line of progress. Could it be that they are intimately involved with the curse that the Bible describes as having come upon the earth as a result of man's sin?

Both the *Mycoplasma hominis* H39 and the theoretical smallest living thing have proteins averaging at least 400 amino acids of the 20 common varieties.<sup>28</sup> Dr. Morowitz' eminent work on the lower limits of size and complexity for living things was done, in part, for the National Aeronautics and Space Administration. Obviously the reason was to enable space missions to recognize any possible forms of life that might be simpler or different from our own. There seems to be no actual scientific evidence that would indicate a more primitive, simpler method of duplication than the one in use in existing living things.

Neither is there any adequate reason to think that there was ever a lesser number of amino acid types used in proteins. Many vital functions require proteins with all the twenty kinds included. Even viruses use all twenty. The same twenty are part of the theoretical minimal cell.

It is significant that the Yale team working on this subject has revised its figures upward at least twice. As research has progressed, it has become clear that the minimal living thing requires more parts than was at first thought. From 45 different proteins estimated as the lowest minimum in 1967, the number

---

<sup>27</sup> Harold J. Morowitz, "Biological Self-Replicating Systems," *Progress in Theoretical Biology*, ed. Fred M. Snell, Vol. 1 (1967), pp. 52, 54.

<sup>28</sup> Morowitz, personal communications, November, 1970, and November, 1971.

needed has risen to at least 124 different protein species. The average molecular weight per protein has also been revised upward slightly.<sup>29</sup>

If words are adequate, the situation appears utterly hopeless for a random linkup in a usable order. Even the hormones which the modern Lamarckian theory requires are likewise outside the realm of reasonable probability, as we saw in the case of insulin. Some hormones are complex proteins much larger than insulin.

*Another Extreme Concession to Help Chance, but All in Vain*

Just suppose that there are only ten kinds of amino acids, all left-handed. Imagine further that the average protein requires only twelve units per chain,<sup>30</sup> and that one substitution is allowed in any of ten of those twelve positions. Let it be granted that a living cell requires only ten proteins. Assume a speed of polymerization of one chain in three seconds, or 10,000,000 per year per set. Let all the other extreme concessions listed earlier remain in effect (use of all atoms on earth, automatic joining, the initial chain qualifying as "usable" under the formula developed on page 104, etc.).

Even with these preposterous assumptions, the probability of one set being produced during the history of the earth is only 1 in 10<sup>32</sup>. The odds, therefore, are 100 million trillion trillion to 1 that not one set of even this simple kind of "proteins" would have been obtained by chance since the earth was born. So chance fails again.

This section should be remembered in connection with any further research, if, for example, it is ever discovered that more substitution in regular-length proteins is allowable. *The margin by which chance fails is so vast that no conceivable amount of new discovery along this line could change the basic conclusion that complicated working systems do not arise by chance.* In order to obtain and continue to have certainty, one must remember and apply this key principle whenever new challenges are met.

<sup>29</sup> Calculated by data from Morowitz, personal communication, November, 1971.

<sup>30</sup> Actually, a protein could not be that short, so this is an extreme concession. According to A. D. McLachlan of the Laboratory of Molecular Biology at Cambridge in England, "There is probably also a critical length of about 50 amino acids below which it is difficult to form a protein structure which is stable in solution under normal conditions." (A. D. McLachlan, "Repeating Sequences and Gene Duplication in Proteins," *Journal of Molecular Biology*, Vol. 64 [March 4, 1972], p. 431.)

### A More Realistic Calculation

Those extreme concessions we listed on page 105 and following were used to show that it is useless to expect a protein to result from random linkups even under such artificially ideal conditions. Before leaving this subject, it may prove helpful if we now cut some of those concessions down to a more realistic picture (while remaining generous to chance so that there will still be no doubt about its failure).

When refigured thus, these conclusions can be reached:

*The probability of a protein molecule resulting from a chance arrangement of amino acids is 1 in  $10^{287}$ .* A single protein molecule would not be expected to happen by chance more often than once in  $10^{262}$  years on the average, and the probability that one protein might occur by random action during the entire history of the earth is less than 1 in  $10^{252}$ .

For a minimum set of the required 239 protein molecules for the smallest theoretical life, the probability is 1 in  $10^{119879}$ . *It would take  $10^{119841}$  years on the average to get a set of such proteins. That is  $10^{119831}$  times the assumed age of the earth and is a figure with 119,831 zeroes, enough to fill sixty pages of a book this size.*<sup>31</sup>

We are again prompted by the evidence to realize that Some-

<sup>31</sup> We suspect that some readers will not be eager to check the calculations for this section, after all the other figures in this chapter. Just for the individuals who may wish to do so, this brief summary of adjustments in the concessions which lead to the above figures is given: Concession 2, estimate  $1/10^6$  probability that all would occur and in proper proportions. Cancel concession 3 because amino acids formed in a primitive environment would be racemic (see chapter 3 and appendix 1). Figuring at 400 length, 32 glycine, 32 following glycine or first in the chain, 336 others at 3/4 preference for the same hand, the probability of all being left-handed is  $1/10^{51}$ . (This, when multiplied by the  $1/10^{236}$  obtained under the extreme concessions, produces the  $1/10^{287}$  figure above.) Concession 5,  $1/10$ ; concession 6,  $1/10^3$ ; concession 7,  $1/10^4$ ; concession 8, 10, and 11 together,  $1/10^{26}$  as follows:  $1/10^{16}$  to adjust from our enormous speed down to normal top speed in bacteria; then  $1/10^{10}$  to adjust to nonenzymatic speed ("...  $10^{10}$  times as long, this being a rough estimate of the ratio of reaction rates with and without enzymatic catalysis. . . ." —Henry Quastler of Yale, in his *Emergence of Biological Organization*, [New Haven, Conn.: Yale University Press, 1964], p. 6). The total of these adjustments (not counting concession 3 made earlier) is  $1/10^{40}$ . Multiplying by the  $1/10^{287}$  and dividing by rate per year from extreme concessions ( $10^{65}$ ), the result is  $10^{202}$  years. Dividing by  $10^{10}$  as a rounded figure of assumed earth-age, the conclusion  $10^{252}$  times the age of the earth.

Regarding sets, adjust concession 13 by  $1/10^{10}$ , and concession 12 by  $1/10^2$ . Totalling all the adjustments on all the concessions, we have  $1/10^{103}$ . This, when multiplied by the  $1/10^{119776}$  obtained under extreme concessions, gives  $1/10^{119879}$ , the "realistic" probability of one set by chance. Dividing by  $10^{38}$  as the realistic rate per year, we get  $10^{119841}$ , which is  $10^{119831}$  times the age of the earth.

Regarding concessions 1 and 14, although there is insufficient evidence to require them, they were left in effect, since materialists must have them, and, as we discovered, chance failed completely anyway,

