Does Protein Specificity Destroy the Theory of Evolution?

A review of David Foster's 'The Philosophical Scientists' *by* Gert Korthof.

updated 9 Apr 2018 (first published: 12 Sep 1997)

"*I did not set out to destroy Darwinism*" (David Foster)

Darwinism relies exclusively on random mutations, natural selection and time to explain life. Critics attempt to demonstrate that these three factors cannot possibly do what Darwinists claim. David Foster is one of the critics who calculated the probability of the origin of life by chance.

I will only discuss the chapters on Darwin's theory of evolution and on the specificity of proteins and DNA in Foster's book *The Philosophical Scientists* (1)

In the Chapter "Monkeys and typewriters", Foster refutes an argument he attributes to Thomas Huxley: "six monkeys typing randomly for millions of millions of years would type all the books in the British Museum." It was supposedly what Darwinists claimed about the powers of chance in evolution (18). It is great fun that Foster did calculate the probably: "Huxley's six typing monkeys typing for the duration of the universe would type 36 letters of sense in one of the books in the British Museum" (p97). This is caused by the extremely high number of permutations of a single line of text of 50 characters: 8.5 x 10^{49} (based on an alphabet of 26 letters). However Huxley could not have told the story about typewriters in 1860, because typewriters did appear 14 years later (in 1874) on the market! (11). Furthermore in 1860 it was not known that genetic information is a linear arrangement of a small number of symbols, so the whole argument could not have arisen in Darwin's time. So, although Foster was wrong in attributing the argument to Huxley/1860, he correctly pointed out *the computational limits of a random process producing meaningful information.*

By specificity Foster means the information content of proteins and DNA. Proteins contain information because there is a linear arrangement of 20 different amino acids. The possible arrangements can be calculated and are of astronomical magnitude. The sequence of the basic elements of proteins is biologically important: it enables a subclass of proteins, selectively speed up specific biochemical reactions resulting in increased amounts of a specific product. This idea is still central to biochemistry.

The following section has been revised 27 Mar 2018 (22)

The example Foster uses is hemoglobin. It consists of a chain of 564 amino acids. Since there are 20 possible choices for each position, a protein of only 3 amino acids long has 20×20×20 = 8000 possible configurations. On the basis of this kind of reasoning Foster calculates the specificity as 10^{650}. That means there are 10^{650} possible linear arrangements of the amino acids, of which hemoglobin is only one. Trying out all these possibilities to find the right one takes longer than the age of our *universe* (25). Of course the sum of the specificity of all proteins of an organism is much higher. So life cannot evolve by chance, Foster concludes. However, in a *postscript* to the chapter, Foster mentioned the discovery of neutral (non-specific) amino acids. This lowers the amount of significant amino acids to 516 and reduces hemoglobin's actual specificity. This is the only correction Foster permits himself.

The problem with Foster's calculation is that he assumes that (**1**) the origin of life equals the origin of proteins from random amino acids; (**2**) that the total sequence is unique and has no repeated units; (**3**) that each amino acid is specific and irreplaceable; and (**4**) the specificity of a sequence is the relevant biological property. These assumptions are wrong.

1. The first assumption is wrong because life did not originate from a random assembly of proteins and certainly not of big proteins like hemoglobin. No scientist claims that this is how life started. Proteins are not formed by throwing dice. This invalidates the whole argument. The argument is irrelevant for real-life situations. An additional error is that his argument depends on the presumption that all 20 amino acids are randomly utilized in equal frequency. Obviously, this is not the case in real proteins. Relative amino acid counts vary from 1.35% (tryptophan) to 9.68% (leucine) (1/20 = 5% only is a statistical mean value) (24).
2. If we accept Foster's calculation just for the sake of argument, the second assumption is wrong because human hemoglobin is made up of 4 subunits, two alpha subunits of 141 amino acids and two beta subunits of 146 amino acids (2). That reduces the number of unique amino acids to 141+146=287. The fact that hemoglobin consists of sub-units also suggests an evolutionary *mechanism* of origin: duplication and variation of sub-units. Furthermore, single-domain globins do exist (26).
3. Again for the sake of argument, his third error is the number of non-specific amino acids. The hemoglobins of different species can differ as much as 81% (3). Since the whole protein is functional, only 19% of the sequence seems to be significant (mostly the "active site" of the folded protein). What really matters is not the sequence but the 3-dimensional structure of the protein. For example hemoglobin of mouse and men match for 80% on the level of the sequence, but are identical when viewed as 3-dimensional structures (13). This lowers drastically the number of specific amino acids and increases the probability that hemoglobin could evolve by a stepwise selection of random mutations. See further: Did hemoglobin evolve to carry oxygen?
4. The fourth error is that the specificity of a sequence is the ultimate relevant biological property. However, the specificity that really matters is not the one-dimensional linear sequence, but the 3-dimensional shape of the protein, which in turn determines its function in the cell. Proteins with very different sequences can have the same shape. The universe of all different protein shapes is called "shape space" or "shape library". A finite number of different molecules, about a hundred million, can constitute a universal shape library (23). So, there could be trillions of sequences, but what matters is shape.

> In order to arrive at a general conclusion, we must consider proteins in general. Athel Cornish-Bowden wrote the following about that issue:
>
> "Comparisons between the sequences of enzymes fulfilling the same function in different species give at least a *minimum* estimate of how many *related* sequences are capable of doing any given task. In most cases (histone IV being an outstanding exception) we find at least one or two differences between the sequences found even in closely related organisms. So the number of sequences possible for an enzyme that occurs in all species should be at least as large as the number of species, making millions of sequences (even if we admit only species that exist on earth today, but many more if we include extinct species)." (16)

If we follow Foster's calculations for the sake of argument, he should explore the possibility that proteins don't necessarily need those 20 amino acids occurring today in proteins. There is evidence that functional proteins could be constructed from less than the current 20 amino acids. Proteins can be constructed from 8-10 different amino acids. This again drastically lowers the number of specific amino acids and increases the probability that a useful protein could evolve by a stepwise selection of random mutations.

Furthermore, it would not be disadvantageous if the first enzymes were inefficient, slow and would have low-specificity, because the very first forms of life did not have competitors. Also, they would have needed fewer proteins, because they did not need defense mechanisms against poisonous predators.

How many possible oxygen binding proteins are there?

A very important question Foster does not ask is: What are all the possible protein structures that also could bind oxygen? If there were hundreds or thousands of possible protein structures doing the same as hemoglobin, then it would be much easier to produce an oxygen-binding protein by random variations.

According to Stuart Kauffman (4) the right question is: What is the probability of finding *any one* of a possible set of 2000 enzymes for 2000 particular reactions, that are necessary for life? Not just the one set which happens to have been found by evolution.

From recent studies in the creation of artificial proteins, it appears that one does not need such a complex molecule as hemoglobin to transport oxygen: "The ease with which globin-like properties can be reproduced in a completely unrelated and simply engineered maquette indicates that the relatively complex globin fold is for the most part unremarkable, and may be common in nature not because of a uniquely capable design for oxygen binding, but simply because it is good enough." (18).

So far we have looked exclusively at proteins. However, if we turn to DNA, the situation changes dramatically. If there are millions of possible DNA sequences that encode exactly the same protein, the probability that a random DNA sequence would produce the protein would be much larger. The redundancy of the genetic code implies that a typical 300-amino-acid protein can be encoded in about 10^{151} ways! (17).

Did hemoglobin evolve to carry oxygen?

updated: 9 April 2018

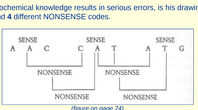
There is another hidden assumption in Foster's reasoning: hemoglobin was designed for oxygen transport and it must be able to do so right from the start. However, the theory that hemoglobin evolved to carry oxygen around the body may need to be reconsidered in light of another way in which molecules related to nitric oxide, NO, released from hemoglobin, help the brain control respiration. Given the results of recent experiments, we may legitimately question whether hemoglobin first evolved to carry oxygen or to ferry NO to key locations in the body. It turns out that hemoglobin's original task was to detoxify nitric oxide, and that its ability to carry oxygen came later. Other work supports this view (14).

Hemoglobin has been found in plants (Leghemoglobin is an oxygen-carrying Phytoglobin) and even in bacteria (!) (15). The green alga *Chlamydomonas reinhardtii* has what is known as a "truncated" haemoglobin (21). Transport of oxygen cannot be the function of hemoglobin in bacteria. This is a general principle in evolutionary biology: the current function need not be the reason it evolved in the first place.

See wikipedia for the Evolution of vertebrate hemoglobin. There is not one unique hemeglobin protein with one unique sequence. For example: there are adult hemoglobin, fetal hemoglobin and several non-pathological hemoglobin variants in humans. In vertebrates alone 8 globins are known to occur: androglobin, cytoglobin, globin E, globin Y, hemoglobin, myoglobin, and neuroglobin (26).
From a biochemical point of view: one of the best-known families of porphyrin complexes is heme, the pigment in red blood cells, a cofactor of the protein hemoglobin. Hemes are also found in a number of other biologically important heme proteins such as myoglobin, cytochromes, catalases, heme peroxidase, and endothelial nitric oxide synthase. Conclusion: A molecule such as hemoglobin is evolutionary and biochemically part of a family of molecules, and should not be considered in isolation.

Is Darwin's theory of evolution wrong?

One example where Foster's lack of biochemical knowledge results in serious errors, is his drawing on page 74 where he displays DNA with 3 SENSE codons AAC, CAT, ATG and **4** different NONSENSE codes.



(figure on page 74)

"The DNA may typically have a string of letters such as the top line, but the problem is if these group into syllables (codons) of three letters, how are the syllables separated?"

On the next page he clearly states that there are:

20 codons for 20 amino acids
40 codons for comma or full-stop effects
4 unusable codons
—
64 codons total

However: *the genetic code is commaless and spaceless. There are no equivalents of commas and spaces in DNA* (7). Those 40 codons 'for comma or full-stop effects' do not exist. There are no more than **3** stop-codons or non-sense codes, which function as start/stop for reading, and there are 61 codons which code for amino acids. A surprising aspect of Foster's presentation of the genetic code is that *that* particular view of the genetic code was published by Crick, Griffith, and Orgel as "Codes without commas" in **1957** and which received an immediate and almost universal acceptance (p. 40). A coincidence? However, the most stunning aspect is, that the hypothesis was refuted by experimental evidence in **1961**! So, 32 years after the discovery, Foster (1993) still believes in this refuted hypothesis (20). This must imply that Foster did not touch a biology or genetics textbook since the sixties! The discovery of the structure of DNA and of the genetic code are the central discoveries in biology of this century and Foster misrepresents them. It shows Foster's attitude to science.

Notwithstanding grave errors, Foster points to the information content of DNA and proteins and tries to calculate it. The information content of proteins is one of the things one needs to know to estimate how long it will take a random trial and error process to generate it. If time is to short, the information could not be generated by a trial and error process. So the combination of a random trial and error process and the amount of time, are potential falsifiers of Neo-Darwinism. One rarely encounters this kind of approach in textbooks on evolution (10). The concept of specificity is not explained. The approach is present in Denton (1986,page 323). However, if Foster's falsification is meant to be a serious one, he has to use a 'nut-and-bolt' biochemical correctness. Foster fails to do that. He admits he is not a molecular biologist (p.60), and this affects the reliability of his conclusions negatively. But also, the calculation of the information content of proteins is much better done by Hubert Yockey (9).
Foster announced that he did not set out to destroy Darwinism. On page 82 he concludes: "Darwin's theory of evolution is wrong", because "Darwin totally underestimated the time and Foster misrepresents them. It shows Foster's attitude to science.
the universe". This proof of the impossibility of evolution reminds me of Zeno's convincing proof of the impossibility of movement (6). However impressive the proof may be, we simply observe movement, so there must be something wrong with Zeno's argument.

> "Thus not only can we apparently prove that life could not have originated on earth, we can prove almost as easily that it could not have originated anywhere. Yet this conclusion must be wrong: we are here to discuss it, so any proof otherwise must be flawed." (18).

I found another 'impossibility' in Ian Stewart (12). A protein is a linear chain of amino acids, which has to be folded to work. To determine the optimal fold is called 'the protein-folding problem'. Mathematicians estimated that the calculation of the optimal fold for cytochrome-c would take about 10^{127} years on a supercomputer. Longer then the age of the universe. Impossible. The funny thing is that living organisms can fold a protein containing a thousands amino acids in about a second. So there must be something wrong with the calculation.

CONCLUSION: I reject Foster's calculation because the calculation is not realistic. It ignores biochemical knowledge and is based on an imaginary mechanism (dice). Darwinism survived another impossibility: that of Lord Kelvin's claim that the age of the earth was too short for evolution to occur, which turned out to be wrong (6).

Is DNA 'programmed from the sun'?

This is the title of a weird chapter in *The Philosophical Scientists*. It is Foster's solution to the problem where the huge amount of information in DNA is coming from: beaming information by gamma radiation from the sun. I sum up only a few of the problems his 'solution' faces.
Is the receiving organism just an empty cell with no DNA at all? In any case a recording mechanism is needed to receive and translate the information into DNA-sequences and to incorporate the sequences into chromosomes. A major problem with this hypothetical mechanism is a vicious circle: if the 'receiver-mechanism' is also coded by DNA (and how could it be otherwise?), how did *that* DNA get into the organism? This mechanism does not exist (anymore!) in today's organisms. So a designer should have constructed the mechanism and destroyed it after the information was received. This is a weird way to provide genetic information to organisms. The Designer could as easily put the information right away into the organism: Let There Be DNA! There are also 'minor' problems: how to focus the message to the right organism? How does a proto-organism know which information is meant for it? If there is a continuous flow of information: when to start recording? When to stop? What about organisms that live in total darkness? How could information be stored in the sun anyway? Why was the information in the sun not used for giving birth again to the millions of extinct species?

Foster's book is unfinished. There is no index and no bibliography. Foster uses postscripts to chapters, in stead of integrating it in the chapter. The book could have been improved if he had discussed the chapters on DNA and hemoglobin with a molecular biologist before committing them to print. *The Philosophical Scientists* is a serious error of Barnes & Noble.

The Philosophical Scientists
by David Foster.
Barnes & Noble Books.
1993
New York
179 pages.
ISBN 0-88029-624-0

Contents: (partial)

- Part One. The 1930 Cambridge club
- Part Two. The symmetrical Paradox.
  - Ch 8. Darwin's Theory of Evolution
  - Ch 10 Monkeys with Typewriters
- Part Three. **Specificity & Molecular Biology.**
  - Ch 12. Specificity in Physics and Chemistry
  - Ch 13. Amino-Acids, Proteins and DNA
  - Ch 14. The Specificity of Haemoglobin
  - Ch 15. Stocktaking - Facts and Mysteries.
- Part Four. Approach to LOGOS.
- Part Five. Where Angels fear to tread.
  - Ch 27. Is DNA Programmed from the Sun?

No index.
No Reference list.

■ **About the author:** David Foster, now (1993) retired from a career as a scientific consultant, received his technical training at King's College London and has the degrees of M.Sc. and Ph.D. This is his seventh book.

Notes:

1. The title of the book refers to A. Eddington, J. Jeans, B. Russell and A.N. Whitehead ('The 1930 Cambridge club').
2. See Michael Behe(1996) 'Darwin's Black Box', page 174 and 206.
3. See Michael Denton(1986) 'Evolution: a theory in crisis', page 284.
4. Stuart A Kauffman(1993) 'The origins of order', page 22 and Stuart A Kauffman(1995) 'At home in the universe', page 44.
5. Zeno was a Pre-Socratic Greek philosopher. He proved that movement is impossible, because one cannot traverse an infinite number of points in a finite time.
6. See Philip Kitcher(1982) 'Abusing Science', page 100,101.
7. See for example: I.M. Lerner & W.J. Libby (1976) 'Heredity, Evolution and Society', page 96. There is a clear disadvantage of a commaless code: nonsense and frameshift mutations disrupt the code. Frameshift mutations are particularly devastating.
8. Brian Hayes "The Invention of the Genetic Code", American Scientist, January-February 1998. A very stimulating historical account of ingenious but wrong hypotheses of the genetic code told by a computer scientist.
9. Hubert Yockey: 'Information theory and molecular biology'. See: review on this site.
10. I found the argument of the improbability of a 100 amino acids long protein arisen by chance, discussed in *The Problems of Biology* by John Maynard Smith, 1986, page 48. JMS rejected the argument.
11. This was pointed out to me by Robert Holloway [email: 24 June 2000]. The 'qwerty'-keyboard was designed by Charles Lotham Sholes in the 1870s.
12. Ian Stewart(1998) *Life's other secret*, p64.
13. Prof. Gert Vriend, CMBI, Nijmegen, The Netherlands. Additionally, amino acids can be grouped in 3 classes: charged (4 amino acids), polar (8 amino acids) and hydrophobic (8 amino acids). source. That means amino acids within groups are highly interchangeable.
14. Stuart Lipton: "Physiology: Nitric oxide and respiration", Nature 413, 118-121 (2001) (news and views).
15. Ross Hardison(1999) The evolution of Hemoglobin, American Scientist, March-April 1999.
16. Athel Cornish-Bowden (2004) *The pursuit of perfection. Aspects of biochemical evolution*, Oxford University Press, page 138 and 136. I am grateful to the author for a copy of the book.
17. J. Robert Coleman et al (2008) 'Virus Attenuation by Genome-Scale Changes in Codon Pair Bias', Science, 27 Jun 2008.
18. Ronald L. Koder et al (2009) 'Design and engineering of an O2 transport protein', *Nature* 458, 305-309 (19 March 2009).
19. The argument against that spontaneous the origin of cosmos is refuted in a splendid way by Denis Diderot (1713 – 1784) in his Diderot's Early Philosophical Works. Philosophic Thoughts, paragraph XXI, page 38. He uses the probability argument (infinite number of trials). Homer's "Iliad" could be created by an infinite number of trials of random letters. It is not about the origin of life, but it could easily be applied to the problem.
20. Because the 1993 edition is a (unmodified) reprint of the 1985 first edition, it is actually 24 years after the discovery.
21. ScienceDaily: When green algae run out of air: Single cell organisms need haemoglobin to survive in an oxygen-free environment June 21, 2013. Green algae has another eleven haemoglobin genes.
22. See guestbook at 19.02.2018 21:42, Ed Townley [does not exist anymore]
23. Stuart Kauffman (2000) 'Investigations', page 13. For example, endorphin and morphine bind equally well to the endorphin brain receptor, altough both have completely different structures.
24. Joji M. Otaki *et al* (2005) Availability of short amino acid sequences in proteins
25. "... takes longer than the age of our universe": If the universe exists infinitely, then there is no problem. See: "Thus Hoyle and Wickramasinghe (1981) concluded that the improbabilities for the non-random assemblage of living proteins and nucleic acids are so huge (1 part in $10^{40,000}$) that maybe an 'Infinite Universe' or super intelligent God would be required to produce a living miracle, which then spread and evolved on a Cosmic scale." source.
26. Globin wikipedia article (May 5 2022).

Further Reading:

- the Quackery and Logic-Chopping of David Foster's *The Philosophical Scientists* (part 1-11). A very thorough and devastating review by Richard Carrier. Especially relevant is part 9: The Odds of Life Evolving by Chance. (*thank-you* for Steve Entringer who notified me that this site had moved. 5 Aug 2006).
- See for an overview of the origin of life literature: Introduction.

This book was a gift from Sid King

First published: 12 Sep 1997

Updated. 4 Feb 2021 Notes: 5 Apr 2018