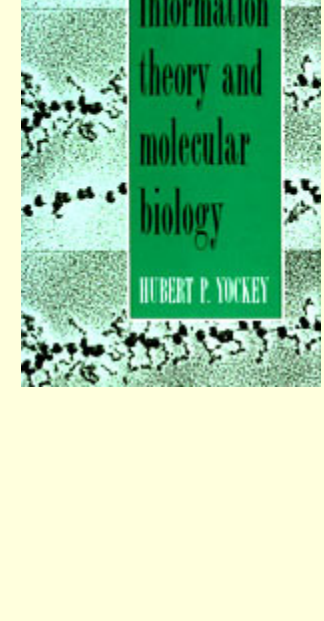


Information is the difference between life and matter

A review of Hubert Yockey's 'Information theory and molecular biology' by Gert Korthof
24 Aug 1998 (updated 18 May 2010)



"Building a theoretical biology based on mathematical foundations. That is what this book is all about."
Hubert Yockey

Hubert Yockey is a physicist who worked under Robert Oppenheimer and worked on the Manhattan Project (production of the first atomic bomb). In the fifties he published about effects of radiation on living systems and started to work on the application of information theory to genetics and evolution. Yockey published 7 articles in the *Journal of Theoretical Biology* from 1974 - 1995 and was organiser of the *Symposium on Information Theory in Biology*.

"Information theory and molecular biology"
by Hubert Yockey
Cambridge University Press, 1992 (16)
ISBN 0 521 35005 0 hardback
408 pages
Contents:
Part I: The basic mathematical ideas
1 Basic ideas in probability theory
2 The role of entropy: a quantitative measure of information, uncertainty and complexity
3 The principle of maximum entropy
4 Coding theory and codes with a Central Dogma
5 The source, transmission and reception of information
Part II: Applications to problems in molecular biology
6 The information content or complexity of protein families
7 Evolution of the genetic code and its modern characteristics
8 The early earth and the primeval soup
9 Did life emerge by chance from primeval soup?
10 Self-organization origin of life scenarios
11 Error theories of aging
12 Information theory and molecular evolution
Epilogue, References(42p), Author index, Subject index

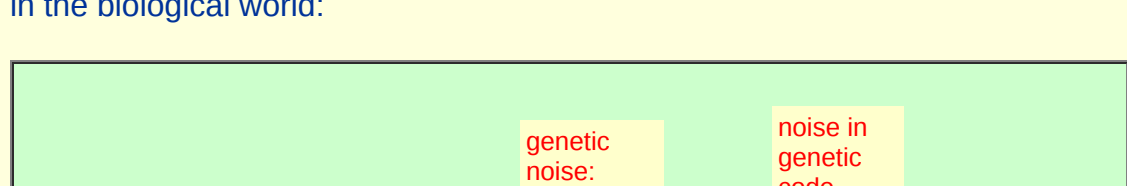
■ The difference between life and matter is information

The information is located in the one-dimensional DNA-sequence of four letters A,T,C,G which is translated with the help of the genetic code to the one-dimensional sequence of 20 different amino acids in proteins [12]. This one-dimensional sequence of the protein determines the 3-D structure of proteins. The 3-D structure of proteins enables specific biochemical reactions to be speeded up. This sustains structures essential to life. In the end, information is the difference between life and matter, between biology and physics. Information is the ultimate explanation of life. Information is the secret of life [12]. This view of life is an oversimplification. Other scientists point out that life consists of 3 subsystems: a chemical motor, a double-layer membrane and an information storing system [13]. It is fortunate that information theory in mathematics and engineering exists. Yockey explains this theory in part I of his book and applies it in part II to problems in biology: protein complexity, the genetic code, the primeval soup theory, the origin of life, theories of aging, and molecular evolution.

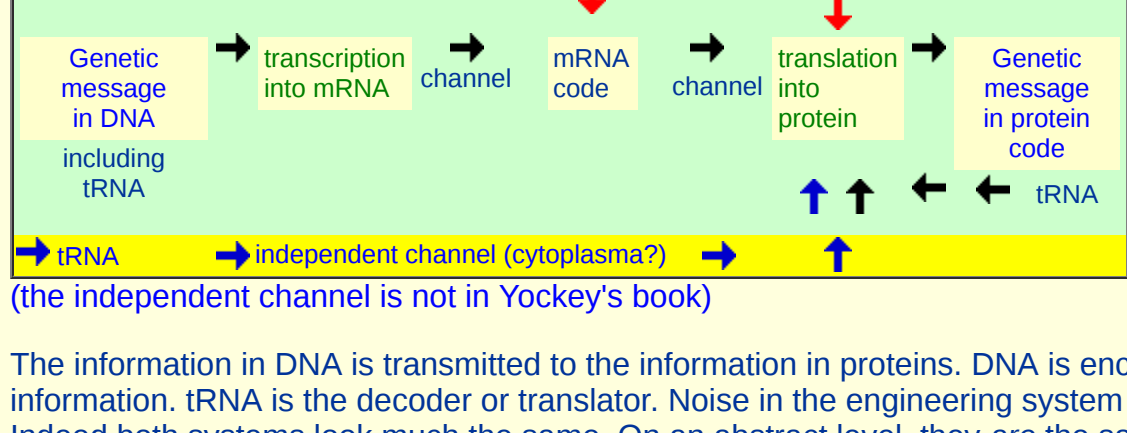
■ DNA as a message

In his book, Yockey uses communication theory to study the DNA-RNA-protein system in living organisms. Yockey uses the theory of communication systems not only as a metaphor, but also as a theory to describe, explain and predict phenomena in molecular biology. Here we have a communication system (telephone or CD player)

in the engineer's world:



in the biological world:



The information in DNA is transmitted to the information in proteins. DNA is encoded information. Proteins are decoded information. tRNA is the decoder or translator. Noise in the engineering system equals mutation in the biological system. Indeed both systems look much the same. On an abstract level, they are the same. However what is not clear from this picture and Yockey's text is that the analogy breaks down at two points:

1. If 'decoding' is defined as information in DNA is translated into nature, so 'encoding' is defined as information in proteins is translated into DNA, then encoding does not exist in nature. So, contrary to engineering systems, there is no **encoding process** in the biological world. DNA is only **decoded**. However, metaphorically speaking, one could say that **evolution is encoding information in DNA** [15]. This does not reflect a concrete biochemical process. In artificial systems, decoding implies encoding. Not so in genetic systems. There is no Master Mind encoding information in DNA. Therefore, the encoding part of the 'information metaphor' is misleading.
2. Contrary to engineering systems, the **decoder device** (the genetic code) is itself transmitted through the same channel as the message! The genetic code is a coupling of tRNA and one amino acid. Each tRNA/amino acid coupling is catalysed by an 'assignment enzyme', because they need to be synthesised repeatedly. Those 'assignment enzymes' themselves are encoded in DNA and so are part of the encoded message. DNA is the only thing that is transmitted (inherited). Proteins are not inherited. So the message and the decode instructions are transmitted via the same DNA channel. *However, both are encoded.*

■ Vicious circle

This is a vicious circle: the **message** and the **decode instructions** are both **encoded**. How to start decoding? Yockey does not mention the problem at all. Both biological systems and engineering systems need to solve it. Consider for example a CD player. The instructions how to translate the information on a CD are not stored on the CD itself, but in the CD player. Alternatively, consider for example the Morse code with its dots and dashes. Could the meaning of the Morse code itself (the meaning of the dots and dashes) be transmitted *together* with the message? Clearly an impossibility: to decode the decode-instructions one needs the decode-instructions before *anything* can be decoded. This is solved for the Morse code by sending the decode instructions on a piece of paper, so by a separate channel. Another analogy is the boot process of a computer. Still another analogy: try to read a Chinese book with a dictionary which is also in Chinese! I can only guess how this problem is solved in the biological world. According to the textbooks, heredity means only DNA is transmitted from one generation to the next. Although the decode instructions are stored in DNA, they are encoded themselves. So the first cell of the embryo needs at least one molecule of each of the 64 translators (transfer-RNA) to start (boot up). Once the embryo has functional translators, it can produce more of them. The only solution seems to be to transmit the tRNAs (the decode instructions) via an extra **independent channel**. But how? They must be present in every cell of every individual; otherwise, no DNA could be translated (or: no message could be decoded). Going back in the life of an individual we end up in the zygote, and going back further: sperm and egg. It seems that the egg is most suitable to 'transmit' those decode instructions from one generation to the next generation. They must be present in the cytoplasm of the egg (outside the nucleus) as ready-to-use translator molecules. Once booted, the translation process produces its own translators and enough translators can be produced and transmitted to all the cells of the growing body. Amazingly, this implies that there must be an *unbroken chain* of transmission of tRNA molecules going back in time to the first organism with a genetic code. I never found this clearly stated in the literature. It also means that this boot problem needs to be solved at the origin of life. The origin of life looks like a boot problem, a kind of vicious circle itself.

The vicious circle is broken

When amino acids spontaneously associate with specific RNA's at the origin of life, we do not need genes containing the associating of RNA and amino acid. In other words: the circle is broken. Exactly this is described by Michael Yarus (2010) *Life from an RNA world. The Ancestor within*. The function of genes is to speed it up. [18 May 2010]

■ The origin of life

A priori, it is not clear what information theory has to do with the origin of life. The origin of life seems to be chemical problem, not of information theory. Why would anybody want to measure the information content of organisms in number of bytes, unless one wants to store that information on a hard disk? The point is that we need to quantify information to say anything useful at all about the information stored in organisms. Therefore, we need a technical definition of information. Only then can we state the problem of evolution as the **increase** of information. Only then can we state the problem of the origin of life as the **origin** of information. Only if we have a quantified concept of information, can we see the **magnitude** of the problem. Then the information content of genomes, genes and proteins can be calculated in an objective and reproducible way. That is what Yockey did (Ch 6). Molecular biologists mean by 'genetic information' DNA that is translated into proteins, but Yockey goes beyond merely counting bases. To keep things 'simple' he started with the information content of one molecule **cytochrome-C** (113 amino acids long). The information content of the cytochrome c family is 233-373 bits. (Please note that the number of base pairs of a gene is not the same as the number of bits of information in a gene). One molecule of iso-l-cytochrome c can be formed spontaneously with a probability of 0.95 in 1.5 x 10⁴⁴ trials. He adds some further conditions that lowers the probability and concludes that even if we believe that the buildings blocks are available, they do not spontaneously make proteins, at least not by chance [1]. It must be clear by now why Yockey is so interested in calculating the information content of proteins: it shows that life cannot arise by chance.

The discussion above demonstrates clearly, however, that the minimum information content of the protobiont must be in the range of hundreds of thousands to several million bits. Scenarios on the origin of life must show how a complexity of that magnitude, which is characteristic of organisms, was generated. (p.244)

That such scenarios do not exist is the basis of William Dembski's book *Intelligent Design*. The subtle difference is that Yockey probably would say that we never can know whether such scenarios are true or not. The reader of Yockey's book soon notices that Yockey attacks paradigms as: the existence of primeval soup (Ch 8), the origin of life from primeval soup (Ch 9) and self-organization (Ch 10). Yockey claims and demonstrates that "there is no evidence for a primeval soup" (ch 8.4) and is very critical of the current theories of the origin of life:

Origin of life:
possible positions:
1. we know: natural
2. we know: supernatural
3. we do not know
4. we cannot know

The belief that life on earth arose spontaneously from non-living matter, is simply a matter of faith in strict reductionism and is based entirely on ideology. (p. 284)

This is identical to what a creationist such as Phillip Johnson [2] believes. The difference is that Yockey's statements are inspired by a thorough analysis of the information content of DNA and proteins and his subsequent realisation that there is too much information to have arisen by chance alone. Furthermore, Yockey does not conclude 'design'. Despite his agnosticism, Yockey claims that we know that life *originated* on Earth. The explanation may be beyond human reasoning powers (agnostic). Scientists should admit ignorance. We must accept the existence of life as an axiom (p.335). Yockey claims that all textbooks written for college undergraduates present the primeval soup paradigm as an established fact. I checked Ridley (1996) [3]: indeed it is present. However, it is present in the textbooks because "The growing consensus is now that both extraterrestrial delivery and in situ Miller-Urey chemistries contributed to the formation of the rich, prebiotic soup of organic materials necessary for life to form". [14].

Closely connected to the origin of life is the origin of the genetic code. Yockey cites a few writers who state that the origin of the genetic code is 'practically inscrutable' and 'baffling'.

"Many papers have been published with titles indicating that their subject is the **origin** of the genetic code, whereas the content actually deals only with its **evolution**." (p. 178). [emphasis is mine].

Pointing to this type of facts, has been a favourite theme of creationists (Michael Behe). There is a clear distinction between problems we *have not solved* and problems we *cannot solve*. It seems contrary to the character of empirical science to state a priori that we cannot solve certain problems (mathematics is different). We simply do not know enough, to conclude that we can *never* know the origin of life.

Is Yockey a creationist?

Yockey's application of information theory to living organisms resulted in a widening of the gap between life and nonlife. Although orthodox science postulates natural mechanisms that generate information, Yockey's view of life blocks a smooth and gradual route from chemicals to the first forms of life. That is his choice. Since he furthermore declares the origin of life inaccessible to science, the question 'Is Yockey a creationist?' seems to be justified. Yockey quotes from the Bible (including the exact location of the quotes) (p284; p336). Since atheists usually do not quote from the Bible, Yockey probably is a Christian [1]. Yockey probably believes that first life was created by God. This is reflected in the subjects Yockey chooses to discuss in his book: the conclusions he draws and the words he uses to describe orthodox scientists. It all shows a very sceptical attitude towards all theories of the origin and evolution of life. Ignorance about life's origin leaves the door open to supernatural intervention, but Yockey nowhere states this! Yockey surely is not a 'creation-scientist', because 'creation-science' is an oxymoron (p. 288) and he is not a young-earth-creationist because he accepts that life is 3.8 billion years old. He further attacks creationists:

"The opinion that evolution was contrary to the second law of thermodynamics was pushed by scientists who did not accept evolution. In the twentieth century this alleged conflict has been a favourite theme of the Biblical creationists and of the creation-science advocates (Wilder-Smith,1981; Gish,1989)". (p. 310).

However, "even a scientist as eminent as Edgington believed there is a conflict between the second law of thermodynamics and evolution." ... "Organisms cannot defy the second law: there has never been any question in sober minds that organisms are not perpetual motion machines." (p. 312). "Thermodynamics has nothing to do with Darwin's theory of evolution." (p.313). "Therefore creationists, who are found of citing evolution as being in violation of the second law of thermodynamics (Wilder-Smith, Gish), are hoist by their own petard: evolution is not based on increasing *order*, it is based on increasing *complexity*." (p. 313). [5], [6], [7].

So, Yockey is not a (young-earth-)creationist, but he is certainly *interested* in the critics of evolution; and creationists themselves (for example Dean Overman) are interested in Yockey. Yockey has a couple of books about creationism in his reference list: "*Scientists confront Creationism: Beyond neo-Darwinism*", "*Science and Creationism*", "*Creation/Evolution*", and Kitcher's "*Abusing Science: The case against creationism*" [8], but he does not discuss them. I did not find a rejection of Darwinism. His discussion of protein evolution (chapter 12) seems to imply that Darwinian evolution is possible. It looks as if once life got started, there are no big obstacles to further evolution. So he is not opposed to evolution understood as common descent and evolution as a fact. Although a clear discussion of the arguments for common descent are missing in his book. He is opposed to scientists who hold beliefs contrary to the facts (primeval soup).

Conclusion: It is difficult to place Yockey in a category. Yockey is not a creationist as far as the argumentation in this book concerns, but his views are compatible with the creationism of Johnson (1993) [2] and Denton (1986) [9], but not with Denton (1998) [10]. 'Compatible': as long as the recommended scientific ignorance does not imply religious agnosticism about the origins question. For, if one cannot know, one cannot know, including knowledge of God. Yockey certainly does not belong to the 'inference-to-design' club of William Dembski and Michael Behe. Dembski and Behe follow Paley, whereas Paley is entirely absent from Yockey's book. That is a huge difference with the Intelligent Design movement and all other creationists. Yockey has a profound critical attitude towards all the origin of life myths and uses words as 'faith' and 'ideology' to characterise those who assume a natural origin of life. I cannot label him a neo-Darwinist [11], because he is clearly not using *Information theory and molecular biology* to contribute to the theory of evolution in the classical sense (population genetics is entirely missing in his book). On the contrary. He primarily uses it to criticise orthodox theories of the origin of life. By the way: the concept of information is not a central concept in neo-Darwinism. A few anti-YEC remarks are not enough to place him in the category 'Anti-Creationism'.

Although he mentions in a short paragraph that the evolution of the Earth is very sensitive to the Earth-Sun distance (if 5% smaller and a greenhouse effect would occur and if 1% bigger a run-away glaciation would have occurred), he does not draw a design conclusion like 'Fine Tuners' Denton (1998) and Overman (1997). Yockey has no alternative theory for evolution or the origin of life, so he is not in the group of alternative theories.

The fact that I discuss creationism in this review, should not distract the reader from the understanding that *Information theory and molecular biology* is a very thorough and solid textbook of information- and coding-theory and its application to molecular biology. It is not an introduction to the theory of evolution. Yockey did not attempt to integrate his theory of information into the neo-Darwinian framework. His book is not a popular science book; most chapters are written on the level of articles for scientific journals, and some did in fact appear in peer-reviewed journals. Several writers have been influenced by Yockey's work (remarkably, William Dembski does not belong to that group).

Finally, one example what information theory can do: we can learn from information theory that the famous 'Central Dogma' is not a first principle of molecular biology at all. The Central Dogma is a property of any code in which the source alphabet is larger than the destination alphabet. "Although many people feel that the Central Dogma belongs only to biology, we must, nevertheless, render unto biology that which is biology's and to mathematics that which belongs only to mathematics". These delightful insights illustrate the originality of Hubert Yockey. ■

■ Notes:

(This book was recommended to me by dr S. King, by the way: someone in favor of Intelligent Design)

1. p257. Creationist/fine tuner **Dean Overman** frequently quotes Yockey, especially the probability of spontaneous generation of a single molecule cytochrome-c protein.
2. *review of Darwin on Trial*, Phillip Johnson.
3. *review of Evolution*, Mark Ridley.
4. In an *email* Yockey says that quotes from the Bible do not prove anything. That is a rather misleading reply considering the fact that Yockey is in the *Board of Advisors of the Christian "Truth Journal"*. [New url: *Message from Professor Hubert Yockey* Board of Advisors of the Christian Leadership Ministries.]
5. Stephen J. Gould makes the same mistake in *Full House* on page 24, where he attacks the psychologist Peck: "But this final fate does not preclude a long and local buildup of **order** in that little corner of totally called earth." (p24). [emphasis mine]. What he should say is: buildup of **complexity**. Remarkably Stuart Kauffman published *The Origins of Order* instead of *The Origins of complexity!*
6. There is a well-documented well-illustrated online article about two types of entropy and the relation with evolution: [The Second Law of Thermodynamics](#) by Brig Klyce.
7. See for specified complexity my *review* of William Dembski.
8. However Denton(1986) is not in his list.
9. Michael Denton: *Evolution: A Theory in Crisis*. ([review](#) on this site).
10. Michael Denton: *Nature's Destiny*. ([review](#) on this site).
11. However in the *email* from Hubert Yockey to me, Yockey states: "That will explain why the recent doubt on the genomes of human and other organisms provide a mathematical proof of Darwinism beyond a reasonable doubt". Furthermore he states that he is an anti-creationist.
12. Independent of Yockey, the new discipline of 'Artificial Life' claims that "life is about function, not form. What separates the living from the dead is not a matter of matter but resides in patterns of information". John L. Casti reviews Steve Grand's *Creation: Life and How to Make it*, *Nature*, **409**, 17-18 (2001). The idea that information is the essence of life dates back to John von Neumann, the inventor of the computer (See: Mark Ward (1999) *Virtual Organisms*, p.66).
13. Life does not and can not consist of information alone. A system with a chemical motor + membrane is alive in a restricted sense. (**Tibor Gánti**). In a sense the chemical motor + membrane subsystem are primary. This is important for the origin of life question.
14. Kevin W. Plaxco & Michael Gross (2006) *Astrobiology: A Brief Introduction*, p.84.
15. This was suggested to me by Jim Sowder [30 Sep 06]
16. Also from Yockey: *Information Theory, Evolution, and The Origin of Life*, 2005. Cambridge University Press, 272 pages. It seems this is not a second edition because the 1992 title contains 408 pages. Usually a second edition contains more pages than the first.
17. John Walker (2014) *Frederick Cooper (1918–2013)*, *Nature*, **505**, 27 (02 January 2014): "In the 1950s, many thought that the amino acids within a protein were arranged randomly, but Sanger proved beyond doubt that they instead form a unique sequence. Although he made light of this conclusion, saying that those who knew about proteins expected this outcome, the knowledge that proteins had a precise sequence suggested that this information must be codified in DNA."

■ Further Reading:

- [email from Hubert Yockey](#)
- Intelligent Design Creationist William Dembski agrees with all of Yockey's conclusions except his agnosticism which he replaces with 'intelligent design'. See [review](#) on this site.
- Lily Kay (2000) devoted *Who wrote the Book of Life? A History of the Genetic Code* to showing why the metaphors of "code", "language", and "information" can be so misleading. R.C. Lewontin agrees, see [his review in Science](#) Feb 16 2001: 1263-1264.
- Hubert P. Yockey (2002): *More light on pioneers of electrochemistry*, *Nature*, **415**, 833 (21 Feb 2002).
- Bada and Lazeano (2002): *Miller revealed new ways to study the origins of life*, *Nature*, **416**, 475 (4 Apr 2002) is a reply to Yockey.
- Hubert P. Yockey (2004) *Information Theory, Evolution and the Origin of Life*, Cambridge University Press (Apr 18, 2005) This is the second edition of the book reviewed on this page.
- Werner L. Loewenstein (1999) *The Touchstone of Life*, Oxford University Press. Reviewed in *Science*, Vol 284, Issue 5422, 1935, 18 June 1999. "Information flow, not energy per se, is the prime mover of life". "The origin of life is approached from the standpoint of theoretical physics".
- [A review of Information Theory and Molecular Biology](#) by Brian D. Harper, 26 Feb 1996
- Jeffrey Shallit (2009) *Test Your Knowledge of Information Theory*, posted: January 11, 2009.
- Peter R. Wills (2016) *DNA as information*, Opinion piece, Royal Society.