

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

Information theory, evolution, and the origin of life

Information Theory, Evolution, and the Origin of Life presents a timely introduction to the use of information theory and coding theory in molecular biology. The genetical information system, because it is linear and digital, resembles the algorithmic language of computers. George Gamow pointed out that the application of Shannon's information theory breaks genetics and molecular biology out of the descriptive mode into the quantitative mode, and Dr. Yockey develops this theme, discussing how information theory and coding theory can be applied to molecular biology. He discusses how these tools for measuring the information in the sequences of the genome and the proteome are essential for our complete understanding of the nature and origin of life. The author writes for the computer competent reader who is interested in evolution and the origins of life.

Hubert P. Yockey is a former director of the Pulsed Radiation Facility at the US Army's Aberdeen Proving Ground, Maryland. He is the author of *Information Theory and Molecular Biology* (1992).

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

Information theory, evolution, and the origin of life

HUBERT P. YOCKEY



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
0521802938 - Information Theory, Evolution, and the Origin of Life
Hubert P. Yockey
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA
www.cambridge.org
Information on this title: www.cambridge.org/9780521802932

© Hubert P. Yockey 2005

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2005

Printed in the United States of America

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Yockey, Hubert P.

Information theory, evolution, and the origin of life / Hubert P. Yockey.
p. cm.

Includes bibliographical references (p.).

ISBN 0-521-80293-8 (hardback : alk. paper)

1. Molecular biology. 2. Information theory in biology. 3. Evolution (Biology)
4. Life – Origin. I. Title.

QH506.Y634 2004

572.8 – dc22

2004054518

ISBN-13 978-0-521-80293-2 hardback

ISBN-10 0-521-80293-8 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this book and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

It must be considered that there is nothing more difficult to carry out nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things. For the reformer has enemies in all those who profit by the old order, and only lukewarm defenders in all those who would profit by the new order, this lukewarmness arising partly for fear of their adversaries, who have the laws in their favor; and partly from the incredulity of men, who do not truly believe in anything new until they have had actual experience of it.

Niccolò Machiavelli (1469–1519), *The Prince*, Chapter 6.

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

Contents

Preface	<i>page</i> ix
1 The genetic information system	1
2 James Watson, Francis Crick, George Gamow, and the genetic code	8
3 The Central Dogma of molecular biology	20
4 The measure of the information content in the genetic message	27
5 Communication of information from the genome to the proteome	33
6 The information content or complexity of protein families	57
7 Evolution of the genetic code and its modern characteristics	93
8 Haeckel's <i>Urschleim</i> and the role of the Central Dogma in the origin of life	114
9 Philosophical approaches to the origin of life	149
10 The error catastrophe and the hypercycles of Eigen and Schuster	158
11 Randomness, complexity, the unknowable, and the impossible	164
12 Does evolution need an intelligent designer?	176
13 Epilogue	182

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

<i>Contents</i>	viii
Mathematical appendix	191
Glossary	213
References	219
Index	251

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)

Preface

This book introduces the general reader and the specialist to the new order of things in evolution, the origin of life on Earth, and the question of life on Mars and Europa and elsewhere in the universe. Although there are many fields of biology that are essentially descriptive, with the application of information theory, theoretical biology can now take its place with theoretical physics without apology. Thus biology has become a quantitative and computational science as George Gamow (1904–68) suggested. By employing information theory, comparisons between the genetics of organisms can now be made quantitatively with the same accuracy that is typical of astronomy, physics, and chemistry.

Spacecraft send messages to Earth as they pass the outer planets – Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto – in spite of the small amount of energy available. Enormous amounts of data and information flow about on the Internet. Huge sums of money are transferred every day. Errors in these communications cannot be tolerated. Claude Shannon (1916–2001) showed that this is accomplished because communication is *segregated, linear, and digital* so that sufficient redundancy can be introduced in communication codes to overcome errors. Furthermore, he showed that these signals, which contain messages, can be measured in *bits* and *bytes*, terms that are familiar to computer users.

Watson and Crick discovered that there is a genetic message, recorded in the digital sequence of nucleotides in DNA, that controls the formation of protein and of course all biological processes. The message in the genetic information system is *segregated, linear, and digital* and can be measured in *bits* and *bytes*. Computer users will notice the isomorphism between the

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)*Preface*

x

program in computer memories and the genetic message recorded in DNA (Chaitin, 1979).

The genetic information system is essentially a digital data recording and processing system. The fundamental axiom in genetics and molecular biology, which justifies the application of Shannon's information and coding theory, is the *sequence hypothesis* and the *digital* rather than the *analog* or *blending* character (Jenkin, 1867) of inheritance as Darwin (1809–82) and his contemporaries believed (Fisher, 1930).

Watson and Crick's solution of the structure of DNA and its application in biology would not have been so important if it had not been for their famously coy remark:

It has not escaped our notice that the specific pairing we have postulated immediately suggests a copying mechanism for the genetic material. (*Nature* 1953)

A fundamental question in genetics is, how does the cell divide into two cells both containing the same genetics? Here, at one stroke, was the solution nicely framed by reductionism!

I show in this book that only because the genetic message is *segregated*, *linear*, and *digital* can it be transmitted from the origin of life to all present organisms and will be transmitted to all future life. This establishes Darwin's theory of evolution as firmly as any in science. The same genetic code, the same DNA, the same amino acids, and the genetic message unite all organisms, independent of morphology.

The genetic message recorded in the DNA of every living organism is unique to that individual. The relationship and evolution among animals and plants can now be determined by comparing DNA sequences rather than relying on morphology. Genetic information is being applied to genomic medical practice and genetic counseling for the benefit of patients. Sickle-cell anemia is a blood disorder that is an important example of the role of DNA in the placement of amino acids in the sequences of amino acids that form hemoglobin. It is so named because the red blood cells that are normally round are shaped like a sickle. Hemoglobin is composed of four chains of amino acids. Fundamental to this disease, at site 6 in the β chain, glutamic acid is replaced by valine. The identification of this genetic disorder in the structure of hemoglobin with the symptoms of sickle anemia was made by Linus Pauling (1901–94) and is one of his more important discoveries (Pauling, 1949).

Cambridge University Press

0521802938 - Information Theory, Evolution, and the Origin of Life

Hubert P. Yockey

Frontmatter

[More information](#)*Preface*

xi

DNA now plays a role in forensics identification that is far more important than fingerprints. Forensics has reached new levels of certainty. A number of guilty people have been convicted, and others, falsely accused by conventional methods, have been vindicated.

This is a monograph and not an encyclopedia so I have not considered it necessary to call attention to papers I believe did not make an important contribution or those which are incorrect. I have included in the references only those I felt contributed to the point I was making. Some readers may think that I have neglected an important paper here and there. I acknowledge that this may be the case, but there are times when one must hew to the line and let the chips fall where they may.

This monograph follows my interest in the subject, which was first attracted by the work of Dr. Henry Quastler (1908–63). With his collaboration I organized the *Symposium on Information Theory in Biology* at Gatlinburg, Tennessee, in October 1956. I am indebted to the late Professor Thomas Hughes Jukes (1906–99), whose strong recommendations resulted in my original papers being published. Many of Professor Jukes' important contributions to molecular biology have shaped the ideas presented in this book, particularly those concerning the evolution of the genetic code. I am grateful to Dr. Gregory J. Chaitin, whose original and seminal work in algorithmic information theory is reflected throughout the book. I appreciate the efforts of Dr. David Abel and Mr. John Tomlinson, who read the manuscript and made important corrections and comments. My daughter, Cynthia Ann Yockey, edited this manuscript from proposal to final draft and contributed much to improve the clarity and organization of the material. My editor at Cambridge University Press, Dr. Katrina Halliday, organized the review of the manuscript and arranged for the publication. I appreciate her patience during the writing. Without the contribution of these people I would not have been able to write this book. The reference material is up to date as of February 20, 2004.

Hubert P. Yockey
Bel Air, Maryland, USA