Progress in Molecular and Subcellular Biology 6

6 Progress in Molecular and Subcellular Biology

With Contributions by
F. C. Cabello · S. N. Cohen · B. M. Cox
J. F. P. Dixon · L. Ebringer · A. Goldstein
F. E. Hahn · S. L. Hajduk · H. Kleinkauf
H. Koischwitz · R. L. O'Brien · J. W. Parker
K. N. Timmis

Edited by: F. E. Hahn, H. Kersten
W. Kersten, W. Szybalski

Advisors: T. T. Puck, G. F. Springer
K. Wallenfels

Managing Editor: F. E. Hahn

Springer-Verlag Berlin Heidelberg New York 1978

Professor Fred E. Hahn, Ph. D.
Division of Biochemistry
Walter Reed Army Institute of Research
Washington, D. C. 20012, USA

With 65 Figures

ISBN-13: 978-3-642-66858-6 e-ISBN-13: 978-3-642-66856-2
DOI: 10.1007/978-3-642-66856-2

Library of Congress Catalog Card Number 75-79748.

2131/3130-543210

Contents

List of Contributors

Cabello, Felipe, C., Department of Microbiology, New York Medical College, Valhalla, New York 10595, USA

Cohen, Stanley N., Department of Medicine/S-155, Stanford University School of Medicine, Stanford, California 94305, USA

Cox, Brian M., Addiction Research Foundation, 701 Welch Road, Palo Alto, California 94304, USA

Dixon, James F.P., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Ebringer, Libor, Department of Microbiology, Comenius University, Páričkova 3, 80100 Bratislava, Czechoslovakia

Goldstein, Avram, Addiction Research Foundation, 701 Welch Road, Palo Alto, California 94304, USA

Hahn, Fred E., Division of Biochemistry, Walter Reed Army Institute of Research, Washington, D.C. 20012, USA

Hajduk, Stephen L., Department of Zoology, University of Georgia, Athens, Georgia 30602, USA

Kleinkauf, Horst, Max-Volmer-Institut für Physikalische Chemie und Molekularbiologie, Technische Universität Berlin, Franklinstraße 29, 1000 Berlin 10

Koischwitz, Hans, Max-Volmer-Institut für Physikalische Chemie und Molekularbiologie, Abtl. Biochemie, Technische Universität Berlin, Franklinstraße 29, 1000 Berlin 10

O'Brien, Richard L., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Parker, John W., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Timmis, Kenneth N., Max-Planck-Institut für Molekulare Genetik, Ihnestraße 63-73, 1000 Berlin 33

Genes Within Genes

Fred E. Hahn

In the Editorial to the first volume of this Progress series (Hahn, 1969), a proposition of Stent (1968) was discussed that molecular biology was essentially of the past and that all that remained for the field in the academic phase that it had attained was "the need to iron out the details." Subsequently, an apocryphal *bon mot* was widely circulated that a molecular biologist was a former *Escherichia coli* geneticist in search of a new field of study.

Kornberg (1976) observed: "Ten years ago, fashionable biochemists were examining .... the molecular details of gene expression in replication, transcription and translation. There was emphasis on using bacteria and their phages, the simplest and most rewarding systems for basic biomedical and genetic studies. Then interest shifted to bigger and more complex things: ..... animal cells were in, and bacteria were out."

Kornberg then described one of his "most useful" research decisions, *viz.*, to abandon work on *E.coli* DNA replication and, instead, turn to the biochemistry of much simpler bacteriophage DNA. Toward the end of his writing, he singled out a major concern: "research is diverted by fashion and by funding pressures to a premature attack on overly complex problems."

Recent advances in the biochemical genetics of phage ϕX174, however, indicate that molecular biology may have entered into another fundamental pioneering phase through the study of a small and simple object and that some generally held basic assumptions about the nature of the genetic code may need to be revised.

Gamow (1954) suggested the first model of a genetic code with overlapping triplet codons and also presaged the possibility of codon ambiguity, meaning that several different triplets might code for one and the same amino acid. These ideas were focused upon by Crick in 1955 in a paper entitled On Degenerate Templates and the Adaptor Hypothesis. Unfortunately, Crick's manuscript was not published but only circulated among a small number of friends; only the adaptor hypothesis was eventually published as a discussion remark (Crick, 1957).

The designation of the code as being "degenerate" is perhaps an unfortunate choice of expression not only from the cryptographic viewpoint but because of its evolutionary connotation. Nevertheless, of the 4^3=64 possible triplets, three are stop signals (UAA, UAG, and UGA), and the codons AUG (for methionine), or less frequently GUG (for valine), are parts of a more complicated initiation signal in the translation of the code into protein. This leaves 61 different codons to specify 20 amino acids. Different codons which specify the same amino acid are called synonyms (leucine and arginine, for example, each are specified by 6 different synonyms). The biological significance or purpose of the extensive use of synonyms in the code has gone unrecognized although it explains the existence of organisms whose DNA composition ranges from 30 to 70 per cent [G+C].

It now turns out that the abundance of synonyms renders it possible to write structural genes which are entirely contained in the nucleotide sequence of larger structural genes. Two such genes are read in different reading frames or "phases" during phenotypic expression (Barrell et al., 1976; Sanger et al., 1977).

The genome of the small *E.coli* virus ϕX174 consists of one single strand of DNA of a length of 5375 nucleotides; the complete nucleotide sequence has been determined by Sanger and his associates (1977). A genome of this size has a maximal coding capacity for proteins of an aggregate molecular weight of approximately 200,000 daltons. However the nine gene products of ϕX174 DNA have a combined molecular weight of 250,000. How is this excess explained?

The nine ϕX174 genes carry the designations A, B, C, D, E, J, F, G and H. The 260 nucleotides of the B-gene are totally contained within the 1546 nucleotides of the A-gene (Sanger et al., 1977) and the 273 nucleotides of the E-gene are totally contained within the 456 nucleotides of the D-gene (Barrell et al., 1976). This is accomplished by the placement of synonymous triplets whose sequence can be read in two reading frames or "phases", each messenger making sense and giving rise to translation into a biologically functional protein.

Up to that time, it was generally assumed that individual genes in genomes are contiguous and separated by termination and initiation "punctuation" and that a shift in the reading frame (as in frameshift mutations) would, from its locus on down, cause a nonsensical transcription and, hence, translation into non-functional protein. The translation of a message was thought to be non-overlapping, the correct reading frame was set at a defined starting point and the message then sequentially read off, groups of three letters at a time. For the gene E which has been shown (Barrell et al., 1976) to lie completely within the nucleotide sequence of gene D it is established that it is read in a different "phase" or frame which is displaced one nucleotide to the right, i.e., in the direction of reading. The possibility of complete structural gene overlap is under consideration for additional bacterial viruses (Lewin, 1976).

These recent results of fundamental import bear out the view of Kornberg (1976) that the study of "tiny bacterial viruses proved to be uniquely useful beyond my expectations." For it must be noted that these results were obtained with a phage for which there existed a complete genetic analysis; a complete sequence analysis of its genome, and complete amino acid sequence analyses for several of its gene products.

The central question remains, of course, if this type of compact genetic organization is merely a peculiarity of certain bacterial viruses or is of wider biological distribution and significance. It will be difficult and, currently, is impossible to obtain conclusive experimental evidence for or against such a genetic organization in bacteria whose genomes and number of gene products are orders of magnitude greater than those of the small viruses.

Two arguments can be advanced which favor the idea of a more general occurrence of genetic overlap. In evolutionary terms, prokaryotes (and probably their parasites) are much older than eukaryotes. It would be paradoxical if the obvious advantages of high genetic information density should have been lost during evolution in favor of a strictly sequential arrangement of individual genes.

More important, the high content of synonyms in the codon catalogue may serve the *purpose* of organizing overlapping structural genes. No other

persuasive purpose of synonymy has been discovered or proposed. Since the code is univeral, it is difficult to evade the speculation that the potential and advantage of the overlapping type of cryptography may be equally universal.

References

Barrell, B.G., Air, G.M., Hutchinson, C.A. III: Overlapping genes in bacteriophage ϕX174. Nature 264, 34 (1976)
Crick, F.H.C.: Discussion. Biochem. Soc. Symp. 14, 25. Cambridge Univ. Press, 1957
Gamow, G.: Possible relation between deoxyribonucleic acid and protein structures. Nature 173, 318 (1954)
Hahn, F.E.: On molecular biology. In: Progress in Molecular and Subcellular Biology, Vol. 1, p. 1. Berlin-Heidelberg-New York: Springer 1969
Kornberg, A.: Research, the lifeline of medicine. N. Engl. J. Med. 294, 1212 (1976)
Lewin, B.: DNA sequences coding for more than one protein. Nature 264, 11 (1976)
Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A. III, Slocombe, P.M., Smith, M.: Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265, 687 (1977)
Stent, G.S.: That was the molecular biology that was. Science 160, 390 (1968)