INTERNATIONAL

Review of Cytology

EDITED BY

G. H. BOURNE

J. F. DANIELLI

ASSISTANT EDITOR

K. W. JEON

VOLUME 93

Genome Evolution in Prokaryotes and Eukaryotes

EDITED BY

D. C. REANNEY

PIERRE CHAMBON

INTERNATIONAL

Review of Cytology

EDITED BY

G. H. BOURNE
St. George's University School of Medicine
St. George's, Grenada
West Indies

J. F. DANIELLI
(Deceased April 22, 1984)

ASSISTANT EDITOR

K. W. JEON
Department of Zoology
University of Tennessee
Knoxville, Tennessee

VOLUME 93

Genome Evolution in Prokaryotes and Eukaryotes

EDITED BY

D. C. REANNEY
Gene Link Australia Limited
South Melbourne, Victoria
Australia

PIERRE CHAMBON
Laboratoire de Génétique Moléculaire
des Eucaryotes du C.N.R.S.
Institut de Chimie Biologique
Faculté de Medecine
Strasbourg, France

ACADEMIC PRESS, INC. 1985
(Harcourt Brace Jovanovich, Publishers)
Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 52-5203
ISBN 0-12-364493-3

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

GIUSEPPE ATTARDI (93), *Division of Biology, California Institute of Technology, Pasadena, California 91125*

C. C. F. BLAKE (149), *Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3PS, England*

THOMAS R. CECH (3), *Department of Chemistry, University of Colorado, Boulder, Colorado 80309*

D. J. FINNEGAN (281), *Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland*

J. D. ROCHAIX (57), *Departments of Molecular Biology and Plant Biology, University of Geneva, 1211 Geneva 4, Switzerland*

JOHN H. ROGERS (187), *MRC Laboratory of Molecular Biology, Cambridge 2B2 2QH, England*

J. A. SHAPIRO (25), *Department of Microbiology, The University of Chicago, Chicago, Illinois 60637*

EDWARD B. ZIFF (327), *Kaplan Cancer Center and Department of Biochemistry, New York University Medical Center, New York, New York 10016*

Preface

Future historians of science may say that the past decade has produced some of the most startling discoveries in the history of modern biology—discoveries that challenged both conventional wisdom and the intuitive understanding of general genetics built up by the study of bacterial models such as *Escherichia coli*.

Perhaps the most unexpected of these discoveries was the finding that most genes in higher cells were "split" in the sense that coding modules (exons) were interrupted by ungrammatical sequences (introns) which, for the most part, appeared functionless. Equally unsettling was the accumulation of evidence for the view that the mobile genetic elements ("transposons") familiar to bacteriologists had counterparts in the chromosomes of plant and animal cells.

These new data have engendered a ferment among molecular biologists. The traditional view of phylogeny, which simplistically puts prokaryotes at the bottom and eukaryotes at the top, has been replaced by a growing conviction that the streamlined genomes of prokaryotes are a derived feature and the "split" structure is probably the original format for genes. The philosophical framework of the debate has been colored by the suggestion that transposons are a kind of molecular parasite: hence the epithet "selfish DNA" now commonly applied to any genetic unit with the capacity for autonomous multiplication within the genome.

The editors believe the time is now ripe for a "wide lens" view of the way this new information has restructured our thinking on evolution. There has been no lack to date of specialist reviews on isolated topics such as "transposition" or "mechanisms of RNA splicing" but we are unaware of any comprehensive interpretation of genome evolution in the light of the mass of new data now available on issues which go so directly to the crux of the problem. This volume then is an attempt to fill a major perceived need.

The volume has been divided into three sections. The book opens with T. Cech's provocative chapter on the implications of self-splicing RNA for the origin of genetic systems. This chapter occupies a section by itself at the beginning of this volume because an increasing number of biologists now appear to accept Crick's view that the self-replicating RNA "machines" were the original forms of life. This chapter also introduces the basic principles of "splicing," a central theme of the entire volume, in what is mechanistically their simplest format.

Section II turns to the modern repository of genetic information, DNA. The first chapter by J. Shapiro reviews mechanisms of DNA reorganization in bacteria in a way that makes the underlying unities and principles especially clear. The next two chapters deal with subcellular organelles—chloroplasts (J. Rochaix) and mitochondria (G. Attardi). While it is still reasonable to discuss the evolu-

tion of chloroplasts and mitochondria in terms of the endosymbiont theory (which holds that each organelle was originally a free-living "prokaryote") the present contributions illustrate the multiple way these small DNAs reflect the specialized requirements of their intracellular environments. Chloroplast and mitochondrial DNAs have been included in Section II because they combine features of the genetic organization of both prokaryotes and eukaryotes and hence form a bridge between the postulated "primitive" RNA-based splicing system and the more "typical" splicing processes which occur in the nuclei of eukaryotes. Genes in both subcellular organelles contain introns, i.e., both kinds of organelle can have an RNA splicing system.

Section III, dealing with the implications of RNA splicing for eukaryote evolution, opens with an interesting evaluation by Blake of the role of exons in protein evolution. This is followed by a cogent explanation of the mechanisms of RNA splicing that occur in the processing of transfer RNAs, mitochondrial and chloroplast RNAs, and nuclear RNAs by J. Rogers. The same author then goes on to review "retroposons," nuclear elements that appear to have been generated by reverse transcription of RNA sequences into DNA form. This section emphasizes the potential or actual mobility of many retroposons, thus setting the stage for the chapter by D. Finnegan which addresses itself specially to the issue of transposable genetic elements in the chromosomes of higher cells. The volume closes with a concise account by E. Ziff of the ways in which selection has exploited the versatility of the RNA splicing mechanism to accommodate the compacted expression of viral genes.

While this volume itself has set an ambitious goal the editors believe the standard of the contributions and the standing of the authors in their respective specialties go far toward achieving the desired purpose. The volume is not a collation of vaguely related data bundled between one set of covers; rather each author has been asked to organize his material around the evolutionary context set by the volume's theme. As an up-to-date treatment of genome evolution the volume should be invaluable not only to geneticists but to all biologists who wish to understand the molecular basis of evolution.

The volume does not, however, make any claim to produce "final" answers to the problems it addresses. Molecular biology is itself in a state of rapid evolution and new concepts must constantly adapt to the changing environment produced by novel data. If, over the next few years, this volume stimulates biologists to ask new questions and seek new solutions it will have fulfilled its function.

D. C. REANNEY
P. CHAMBON

Contents

Animal Mitochondrial DNA: An Extreme Example of Genetic Economy

GIUSEPPE ATTARDI

Section III. Split Genes and the Genetic Evolution of Eukaryotes

Exons and the Evolution of Proteins

C. C. F. BLAKE

The Origin and Evolution of Retroposons

JOHN H. ROGERS

Transposable Elements in Eukaryotes

D. J. FINNEGAN

Splicing in Adenovirus and Other Animal Viruses

EDWARD B. ZIFF

Section I

RNA and Early Genome Evolution

INTERNATIONAL REVIEW OF CYTOLOGY, VOL. 93

Self-Splicing RNA: Implications for Evolution

THOMAS R. CECH

Department of Chemistry, University of Colorado, Boulder, Colorado

I. Introduction: Ribosomal RNA Genes in *Tetrahymena*

Tetrahymena, a genus of ciliated protozoa, has some unique advantages for studies of rRNA genes and their expression. The rRNA genes (rDNA) reside in the nucleoli of the macronucleus as linear, extrachromosomal molecules (Gall, 1974; Enberg *et al.*, 1974; reviewed by Blackburn, 1982). Unlike chromosomal rDNA in many higher eukaryotes, where the rDNA is heterogeneous due to variations in the nontranscribed spacer, the *Tetrahymena* rDNA molecules are essentially homogeneous in size and nucleotide sequence. The rDNA is present in $\simeq$10,000 copies per macronucleus. It comprises 2% of the total cellular DNA, so it is readily isolated either as purified DNA or as a minichromosome.

Each rDNA molecule contains about 20 kb (kilobase pairs). The molecules

ISBN 0-12-364493-3

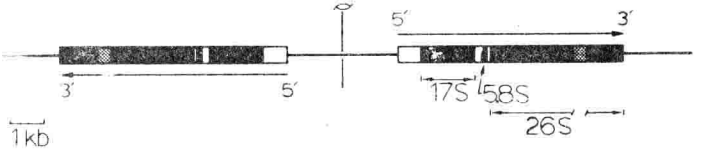


FIG. 1. Transcription map of the extrachromosomal rDNA of *Tetrahymena thermophila*. The wide bars represent the sequences coding for the primary transcript. Sequences present in mature rRNA are shown in black. Sequences removed from the precursor during processing are shown in white, except for the IVS which is stippled.

have a palindromic (inverted repeat) sequence symmetry (Karrer and Gall, 1976; Engberg *et al.*, 1976). As shown in Fig. 1, each half contains sequences coding for the 17 S, 5.8 S, and 26 S rRNAs.

Tetrahymena, like the other ciliated protozoa, contains two types of nuclei within the same cell. The transcriptionally active macronucleus is polyploid and divides amitotically during vegetative growth. The germinal micronucleus is diploid and divides mitotically. During conjugation, the old macronucleus is destroyed and a new one differentiates from a division product of the zygotic nucleus. The 10,000 extrachromosomal rDNA molecules in the macronucleus are formed by amplification of a nonpalindromic copy integrated in a micronuclear chromosome (Yao and Gall, 1977; Pan *et al.*, 1982).

Wild and Gall (1979) found that in some strains of *T. pigmentosa*, the 26 S rRNA coding region of the rDNA was split by a 0.4 kb intervening sequence (IVS, or intron). In any one strain, the rDNA was homogeneous: either all copies were interrupted, or all were continuous. Such homogeneity is a consequence of the amplification process. If the micronuclear copy that is amplified contains an IVS, all macronuclear copies will be IVS$^+$. *Tetrahymena thermophila*, *T. hyperangularis*, and *T. cosmopolitanis* all have a 0.4 kb IVS in their rDNA, while several other species that have been examined have uninterrupted rDNA (Din and Engberg, 1979).

The discovery of strains in which all copies of the rDNA were split by an IVS immediately led to the proposal that RNA splicing was taking place (Wild and Gall, 1979). This proposal was soon substantiated by the identification of an IVS-containing pre-rRNA *in vivo* (Cech and Rio, 1979; Din *et al.*, 1979) and the observation that excision of the IVS from pre-rRNA was a posttranscriptional event in isolated nuclei (Zaug and Cech, 1980; Carin *et al.*, 1980).

Splicing of pre-rRNA is not common. The only other nuclear pre-rRNA known to be transcribed and spliced is that of the slime mold *Physarum polycephalum* (Gubler *et al.*, 1980), which like *Tetrahymena* has extrachromosomal rDNA. While *Drosophila* and other dipteran flies have interrupted copies of rDNA, their genomes also contain uninterrupted copies; the interrupted units contribute little or nothing to rRNA production, and there is no real evidence that

rRNA splicing takes place (reviewed by Beckingham, 1982). There are more examples of split rRNA genes in organellar genomes: chloroplast rDNA in *Chlamydomonas* (Rochaix and Malnoë, 1978) and mitochondrial rDNA in *Saccharomyces, Aspergillus, Neurospora,* and *Kluyveromyces* (Bos *et al.*, 1978; Netzker *et al.*, 1982; Burke and RajBhandary, 1982; Michel *et al.*, 1982). RNA splicing is clearly implicated in these cases, because the organelles contain only interrupted copies of rDNA. While more examples of split rRNA genes, nuclear and organellar, whill undoubtedly be found, it is probably significant that in all higher eukaryotes that have been studied (e.g., *Xenopus*, mouse, human) the nuclear and mitochondrial rRNA genes are not interrupted.

II. *Tetrahymena* Pre-rRNA Splicing Is Self-Catalyzed

By "self-catalyzed" I mean that the molecule lowers the activation energy for a chemical reaction in which that same molecule is a reactant. Lowering the activation energy results in a very great acceleration in the rate of the reaction. Those interested in the development of the idea of self-splicing RNA are directed to the papers by Cech *et al.* (1981) and Kruger *et al.* (1982). In this section, I will avoid such a historical account and discuss the most compelling evidence that we currently have for self-catalyzed splicing.

Unspliced pre-rRNA can be synthesized by incubation of isolated nuclei under conditions where the endogenous RNA polymerase I on the rDNA is active. Inhibition of splicing during transcription is accomplished by the inclusion of polyamines and Ca^{2+} in the transcription mixture and by keeping the concentration of monovalent cations low. Presumably these di- and polycations bind to the RNA and stabilize some splicing-incompetent conformation(s). The unspliced RNA is then purified by SDS-phenol extraction and sedimentation in a formamide-sucrose gradient (Cech *et al.*, 1981). When the resulting RNA is incubated in a solution containing only a monovalent cation (e.g., 200 m*M* NH_4^+ or Na^+), a divalent cation (5–10 m*M* Mg^{2+} or Mn^{2+}) and a guanosine compound ($\geq$ 1 μM guanosine or GTP), accurate RNA splicing occurs. In addition, the excised IVS RNA is converted to a covalent circular form. These reactions are described in detail in Section III.

The splicing and cyclization activities copurify with the RNA through gel electrophoresis in 8 *M* urea. In addition, the cyclization activity has been shown to copurify with the RNA through Cs_2SO_4 density gradient centrifugation. The activities are resistant to SDS-phenol extraction, boiling, treatment with 2% SDS, treatment with high concentrations of two nonspecific proteases, pronase and proteinase K, and combinations of the above treatments, such as boiling in SDS and then incubation with proteinase K and 37°C (Grabowski *et al.*, 1983). Such properties are clearly inconsistent with the splicing activity being a protein.

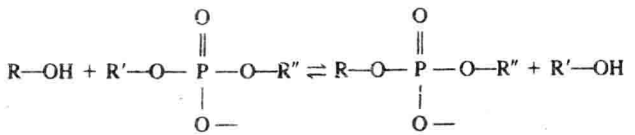
Final confirmation that the activity resides in the RNA was obtained using recombinant DNA. A portion of the *Tetrahymena* rRNA gene, containing 261 bp of the 5′ exon (sequences preceding the IVS), the entire 413 bp IVS, and 943 bp of the 3′ exon was cloned in an *E. coli* plasmid downstream from a promoter for *E. coli* RNA polymerase. RNA was transcribed from this template *in vitro* with purified *E. coli* RNA polymerase, thereby avoiding exposure of the RNA to any eukaryotic protein. When this RNA was deproteinized, it was found to undergo the same splicing reaction as characterized with the unspliced RNA from *Tetrahymena* nuclei (Kruger *et al.*, 1982).

III. Reactions Catalyzed by the RNA

A. Intramolecular Recombination

Any model for the mechanism of *Tetrahymena* pre-rRNA splicing must explain the following features of the reaction: (1) a guanosine nucleotide becomes attached to the 5′ end of the excised IVS RNA via a normal 3′,5′-phosphodiester bond; (2) cleavage occurs such as to produce 5′-phosphate and 3′-hydroxyl termini; (3) RNA ligation occurs without hydrolysis of ATP or any other energy cofactor; (4) covalent circularization of the excised IVS RNA requires only Mg^{2+} and a temperature $\geq 37°C$. A mechanism that is consistent with these facts is shown in Fig. 2.

The mechanism involves phosphoryl transfer reactions:



Because this is a transesterification reaction, it is expected to be highly reversible and should not require an energy cofactor. Splicing could take place by two such transesterification reactions. In the first, the nucleophile R—OH is guanosine. In the second, R—OH is the 5′-exon. IVS RNA circularization would require a third transfer, with the 3′-terminal guanosine residue serving as the R—OH.

An important prediction of the model was that RNA circularization must take place by an attack of the 3′-hydroxyl on some internal phosphodiester linkage, resulting in a circle that is smaller than the linear. The finding that the first 15 nt of the linear IVS RNA are released upon cyclization and that the circle has the 3′-terminal G joined to the A at position 16 (Zaug *et al.*, 1983) therefore provided strong support for the model.

Strand breakage, strand switching and reunion are the essence of nucleic acid

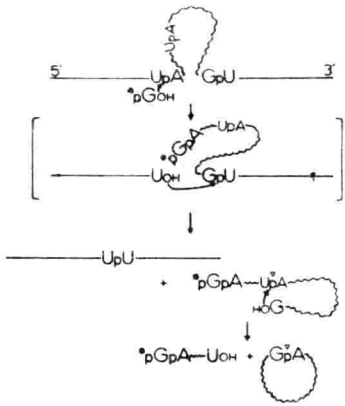


FIG. 2. Transesterification mechanism for *Tetrahymena* pre-rRNA splicing and IVS cyclization. Solid lines, exons; wavy line, IVS; asterisk, labeled guanosine cofactor, in this example 5′-GMP; ∇, phosphate that closes the circle. Square brackets denote a postulated intermediate that has not been isolated. (From Zaug *et al.*, 1983; reprinted by permission from *Nature* (London), **301**, 578–583, copyright (c) 1983 Macmillan Journals Limited.)

recombination. The reactions diagrammed in Fig. 2 involve exactly these steps and can therefore be considered to be intramolecular recombination reactions.

B. Intermolecular Recombination

The linear IVS RNA is capable of forming linear and circular dimers, trimers, and higher oligomers *in vitro* (Zaug and Cech, 1985). The reactions require prior denaturation of the RNA, a treatment which presumably facilitates the formation of intermolecular base-paired complexes. The ligation mechanism is an intermolecular analog of cyclization. That is, the 3′-terminal G of one molecule undergoes transesterification with the A at position 16 of another molecule, producing a molecule that is 15 nt smaller than a true dimer and releasing one 15-mer. Like the cyclization reaction, oligomerization requires only Mg^{2+}. Each oligomer still contains a 3′-terminal G—OH and a 15-mer at its 5′-end, and is therefore capable of circularization. None of these species has been detected in nuclear RNA from *Tetrahymena* grown at either 30 or 39°C (Brehm and Cech, 1983), so they are probably not important in normal RNA metabolism. It remains possible that such oligomerization could have occurred under prebiotic conditions.

C. Circle Autoreopening

When circular IVS RNA is incubated at 42°C at pH 7.5, it is converted to a linear form with a half-time of ~7 hours. At first we thought that the RNA was

undergoing alkaline hydrolysis, but further investigation revealed that only one of 400 phosphodiester bonds in the RNA was being hydrolyzed, the bond that closed the circle (G_{414}–A_{16}). Furthermore, unlike alkaline hydrolysis, the RNA termini produced in circle reopening were 5′-phosphate and 3′-hydroxyl (Zaug *et al.*, 1984). The reaction is similar to a reversal of RNA cyclization, except the phosphoryl transfer occurs to water instead of to the 15-mer. The occurrence of this reaction indicates that the structure of the RNA is making one of the 400 phosphates in the molecule chemically special such as to make it preferentially reactive.

D. Is Self-Splicing Chemically Reasonable?

We must consider both thermodynamics, which determines whether the reaction is possible, and the activation energy barrier, which determines whether the reaction will occur at a reasonable rate.

As far as thermodynamics is concerned, the transesterification reactions are expected to be isoenergetic. Each transesterification produces no net change in the number of phosphodiester bonds. There is therefore no need for ATP or GTP as an energy source. The splicing reaction can be driven to completion by a high molar ratio of the guanosine cofactor (e.g., GTP) to pre-rRNA. *In vivo*, circularization or degradation of the IVS RNA would also help splicing go to completion by removing an end-product. The circularization reaction might be entropically driven, since it involves cleavage of one molecule into two.

RNA self-splicing is as thermodynamically reasonable as other transesterification reactions catalyzed by DNA topoisomerases, resolvases, and the λ *int* protein. These reactions are also ATP independent. (DNA gyrase is normally considered to be an ATP-dependent topoisomerase, but it requires ATP for the introduction of superhelical turns, not for its DNA nicking-closing activity). The enzymes involved in these DNA reactions of course do not alter the free energy of the reaction, but serve to lower the activation energy barrier.

In the case of self-splicing RNA, it is clearly the structure of the RNA molecule that lowers the activation energy for the transesterification reactions. The reactions do not occur in the presence of high concentrations of denaturing agents such as formamide or urea. IVS RNA circularization is completely inhibited by the presence of 30 μg/ml ethidium bromide, which intercalates in double helical regions (N. K. Tanner, unpublished data).

How does the structure of the RNA promote the reaction? First, it provides a binding site for the guanosine cofactor (Bass and Cech, 1984). The binding site has not been localized within the tertiary structure of the molecule, but presumably it places the guanosine with its 3′ hydroxyl group poised to attack the phosphate at the 5′ splice junction of the pre-rRNA. Many enzymes work by approximation, i.e., by binding substrates in proximity to each other and in a