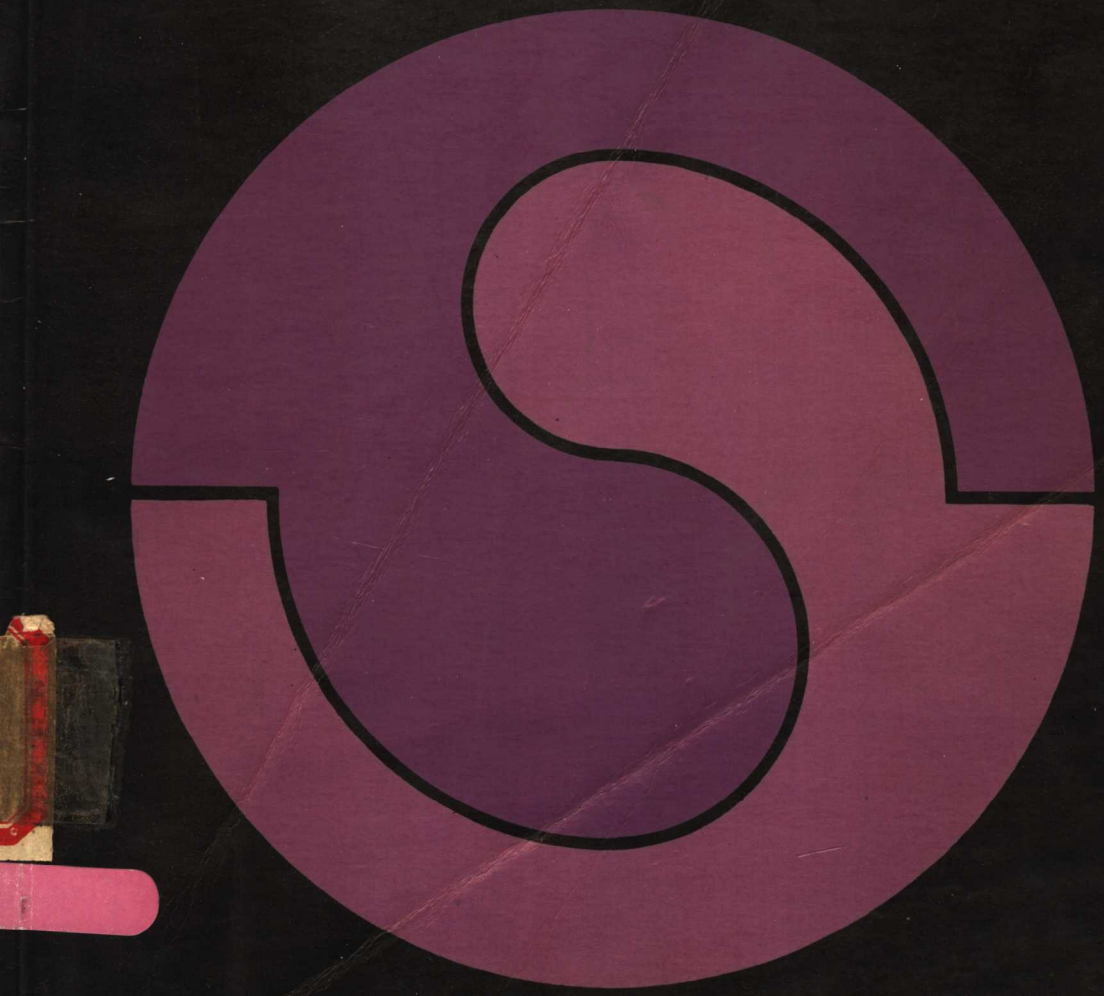
Outline Studies in Biology

RNA Biosynthesis

R. H. Burdon



List of titles

Already published

Cell Differentiation	J.M. Ashworth
Biochemical Genetics	R.A. Woods
Functions of Biological Membranes	M. Davies
Cellular Development	D. Garrod
Brain Biochemistry	H.S. Bachelard
Immunochemistry	M.W. Steward
The Selectivity of Drugs	A. Albert
Biomechanics	R. McN. Alexander
Molecular Virology	T.H. Pennington, D.A. Ritchie
Hormone Action	A. Malkinson
Cellular Recognition	M.F. Greaves
Cytogenetics of Man and other Animals	A. McDermott
RNA biosynthesis	R.H. Burdon
Protein Biosynthesis	A.E. Smith

In preparation

The Cell Cycle	S. Shall
Biological Energy Transduction	C. Jones
Control of Enzyme Activity	P. Cohen
Metabolic Regulation	R. Denton, C.I. Pogson
Polysaccharides	D.A. Rees
Microbial Metabolism	H. Dalton
Bacterial Taxonomy	D. Jones
Molecular Evolution	W. Fitch
A Biochemical Approach to Nutrition	R.A. Freedland
Metal Ions in Biology	P.M. Harrison, R. Hoare
Nitrogen Metabolism in Plants and Microorganisms	A.P. Sims
Cellular Immunology	D. Katz
Muscle	R.M. Simmons
Xenobiotics	D.V. Parke
Plant Cytogenetics	D.M. Moore
Human Genetics	J.H. Edwards
Population Genetics	L.M. Cook
Membrane Biogenesis	J. Haslam
Biochemical Systematics	J.H. Harborne
Biochemical Pharmacology	B.A. Callingham
Insect Biochemistry	H.H. Rees

OUTLINE STUDIES IN BIOLOGY

Editor's Foreword

The student of biological science in his final years as an undergraduate and his first years as a graduate is expected to gain some familiarity with current research at the frontiers of his discipline. New research work is published in a perplexing diversity of publications and is inevitably concerned with the minutiae of the subject. The sheer number of research journals and papers also causes confusion and difficulties as assimilation. Review articles usually presuppose a background knowledge of the field and are inevitably rather restricted in scope. There is thus a need for short but authoritative introductions to those areas of modern biological research which are either not dealt with in standard introductory textbooks or are not dealt with in sufficient detail to enable the student to go on from them to read scholarly reviews with profit. This series of books is designed to satify this need. The authors have been asked to produce a brief outline of their subject assuming that their readers will have read and remembered much of a standard introductory textbook of biology. This outline then sets out to provide by building on this basis, the conceptual framework within which modern research work is progressing and aims to give the reader an indication of the problems, both conceptual and practical, which must be overcome if progress is to be maintained. We hope that students will go on to read the more detailed reviews and articles to which reference is made with a greater insight and understanding of how they fit into the overall scheme of modern research effort and may thus be helped to choose where to make their own contribution to this effort. These books are guidebooks, not textbooks. Modern research pays scant regard for the academic divisions into which biological teaching and introductory textboocks must, to a certain extent, be divided. We have thus concentrated in this series on providing guides to those areas which fall between, or which involve, several different academic disciplines. It is here that the gap between the textbook and the research paper is widest and where the need for guidance is greatest. In so doing we hope to have extended or supplemented but not supplanted main texts, and to have given students assistance in seeing how modern biological research is progressing, while at the same time providing a foundation for self help in the achievement of successful examination results.

J.M. Ashworth, Professor of Biology, University of Essex.

RNA Biosynthesis

R. H. Burdon
Department of Biochemistry,
University of Glasgow

LONDON
CHAPMAN AND HALL

A Halsted Press Book
JOHN WILEY & SONS, INC., NEW YORK

First published in 1976
by Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE

Printed in Great Britain by
William Clowes & Sons Ltd.,
London, Colchester and Beccles

ISBN 0 412 14050 0

Distributed in the U.S.A.
by Halsted Press, a Division
of John Wiley & Sons, Inc. New York

Library of Congress Cataloging in Publication Data
Burdon, Roy Hunter
RNA biosynthesis

(Outline studies in biology series)
"A Halsted Press book."
Includes bibliographies and index.
1. Ribonucleic acid synthesis. I. Title.
[DNLM: 1. RNA—Biosynthesis. QU58 B951r]
QP623.B87 574.8'732 75-33182
0–470–12547–0

Contents

1 Introduction

The genetic information for the development and functioning of organisms is encoded in the linear sequences of deoxyribonucleotides that make up its DNA, or deoxyribonucleic acid. It is the *transcription* of such sequence information from this polymer that is the first step in the process of gene expression.

The sequences of deoxyribonucleotides that comprise DNA, whilst physically contiguous, can be envisaged as being operationally divided into discrete regions, or genes. These regions comprise the deoxyribonucleotide sequences required to specify individual macromolecular cell products. The products of DNA *transcription* are also long polymers but made up of ribonucleotides rather than deoxyribonucleotides and are termed ribonucleic acids (RNA). During the *transcription* process the DNA deoxyribonucleotide sequences serve as 'templates' and are reproduced, but in terms of the ribonucleotide sequences of RNA molecules. Certain DNA deoxyribonucleotide sequences, or genes, when transcribed give rise to transfer RNA (tRNA) and others ribosomal RNA (rRNA). The ribonucleotide sequences of these are such as to cause them to fold extensively and, in the case of rRNA to associate with specific proteins to yield the essential cell components called ribosomes.

Transcription of yet other, but special, sequences results in the biosynthesis of messenger RNAs (mRNAs) which contain ribonucleotide sequences which will ultimately be *translated* into new types of sequences, namely, the amino acid sequences of protein molecules (e.g. enzymes) required for cellular function. This switch to a different variety of molecular sequence is complex, but basically each sequence of three ribonucleotides specifies the insertion of one particular amino acid into the polypeptide chain under construction. Whilst mRNA might be considered the means whereby genetic information is actually transmitted from the genome (the DNA) and placed in the appropriate cytoplasmic sites for *translation* into protein, it is important to realise that the actual *translation* machinery (dealt with in detail elsewhere in this series by A.E. Smith in *Protein Biosynthesis*) depends not only on mRNAs for its function but also the presence of other transcription products.

rRNA is necessary as a structural component of the ribosomes upon which translation actually takes place and tRNA is required in amino acid activation, as an adaptor in mRNA-directed amino acid specification and in binding the growing protein chains to the ribosomes. (see Fig. 1.1).

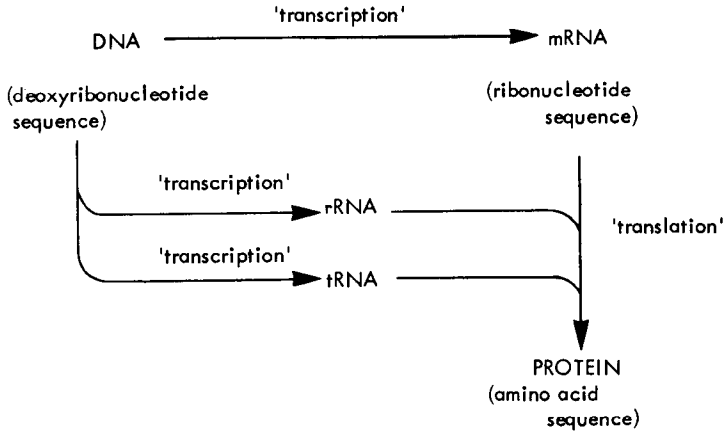


Fig. 1.1 The products of transcription and their cellular role.

2 The products of transcription

2.1 General properties of RNA molecules

An examination of the total RNA isolated from both eukaryotes and prokaryotes shows it to be a collection of polymers of various, but defined, lengths. Chemically they are in fact quite similar to DNA from which they are transcribed. All are long unbranched molecules basically containing four types of nucleotide linked together by 3′–5′ phosphodiester bonds. Unlike DNA however, the nucleotide sugar involved is ribose rather than deoxyribose, and the base uracil usually replaces thymine.

For most RNA molecules the amount of the base adenine (Ade) does not usually equal the amount of uracil (Ura), and the levels of guanine (Gua) and cytosine (Cyt) bases also usually differ from one another. This is quite unlike most DNAs where there is an equivalence of adenine with thymine and guanine with cytosine. This observation was the first indication that most RNA molecules do not possess a regular hydrogen bonded structure like that encountered in DNA, which is composed of complementary polydeoxyribonucleotide sequences arranged in the form of a double helix around a common axis. Such a helical molecule of DNA is maintained in this configuration by the aid of specific hydrogen bonding between the complementary base moieties of the deoxyribonucleotides comprising the two strands, the base adenine (Ade) forming what is known as a complementary base pair with thymine (Thy), and guanine (Gua) forming a complementary base pair with cytosine (Cyt).

With regard to the structures of RNA molecules, or polyribonucleotides, (see Fig. 2.1) it has become usual to use the convenient 'shorthand' system illustrated in Fig. 2.2. A vertical line denotes the carbon chain of the sugar with a base attached at C–1′. A diagonal line from the middle of a vertical line indicates the phosphate link at C–3′ while one at the bottom of a vertical line denotes the phosphate link at C–5′. An even simpler system is also used. A phosphate group is denoted by p. When placed at the right of the nucleoside symbol the phosphate is esterified at the C–3′ of the ribose moiety, whereas when it is placed to the left of the nucleoside symbol, the phosphate is esterified at the C–5′ of the ribose moiety. Thus pUpU is a dinucleotide with one phosphate esterified at the C–5′ of a uridine residue and a phosphodiester bond between C–3′ of that residue and C–5′ of the adjacent residue. Sometimes the letter p is replaced by a hyphen, e.g. pU–U.

Whilst DNA has a firmly established helical structure, the nature of the secondary and tertiary structure of RNA molecules is less well defined. In solutions of low ionic strength, RNA molecules behave like typical highly swollen polyelectrolyte chains, but an increase in ionic strength causes the chains to contract upon themselves so as to display relatively low intrinsic viscosities and high sedimentation rates. This suggests the existence of at least some helical

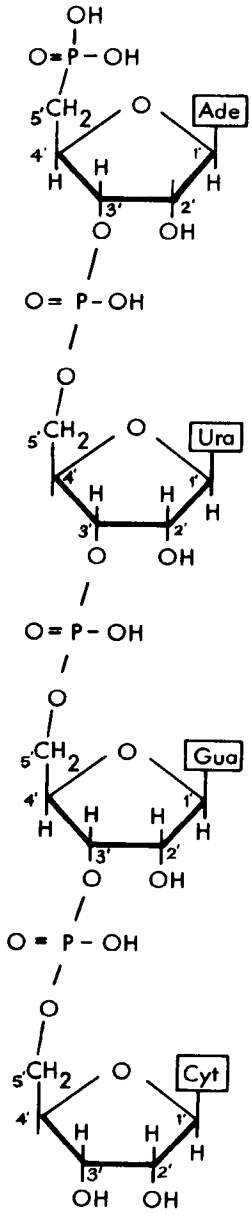
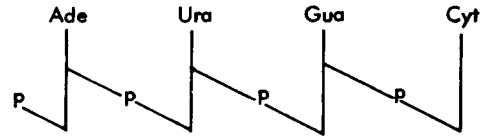


Fig. 2.1 Structure of a hypothetical polyribonucleotide.



or pApUpGpC

or PA-U-G-C

Fig. 2.2 Shorthand notations for hypothetical polyribonucleotide shown in Fig. 2.1. Nucleoside abbreviations; A, adenosine; G, guanosine; U, uridine; C, cytidine.

regions with complementary base pairing in certain regions of RNA nucleotide chains. Enzymological techniques now permit the formation in the test-tube of RNA molecules in which the only base is adenine. These are referred to as poly (A) (polyadenylic acid). RNAs whose only base is uracil can also be synthesised, namely poly (U) (polyuridylic acid). When equimolar amounts of these somewhat artificial RNAs are mixed in solution they form a complex known as poly (A) · poly (U) see Fig. 2.3 in which the adenine bases of one strand are linked by hydrogen bonds to the complementary uracil bases of the other strand [1]. Indeed the X-ray diffraction pattern of this complex indicates a double helical structure as in DNA with 10 bases per turn of the helix, the pitch of which is 3.4 nm.

In fact this RNA helix behaves like a DNA helix in many ways. For instance it shows the phenomenon of 'molecular melting' or 'helix-coil' transition. When heated in 0.15M NaCl at neutral pH the absorbance at 260 nm rises sharply at a temperature of around 60°, the so-called melting temperature, or T_m. At the same temperature the specific optical rotation at 589 nm decreases rapidly. These effects are due to the separation of the two

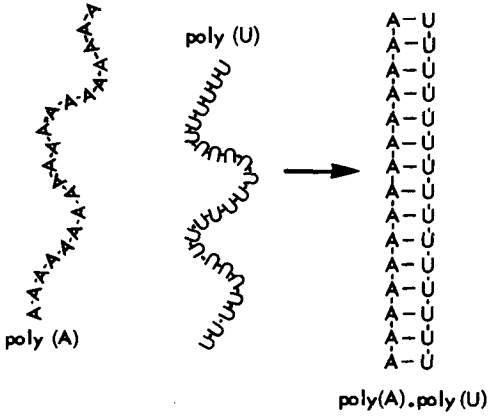


Fig. 2.3 The association of single strands of poly (A) and poly (U) to form the helical poly (A) · poly (U) complex in which two strands are linked by hydrogen bonding between A and U.

complementary strands of the helix on heating. Cooling reverses these effects and the helix is reformed [1].

When solutions of naturally occurring RNAs are heated quite similar, but less pronounced changes occur, which suggest that the RNA chains are folded back upon themselves in a number of places to form short helical regions [2]. Our understanding of these structural considerations is becoming clearer now that several distinct species of cellular RNA have been isolated and extensively characterised.

2.2 The intracellular locations of RNA

As already mentioned there is a variety of cellular RNA species and these often have quite discrete intracellular locations. The basic cytological features of typical prokaryotic and eukaryotic cells are summarised in Figs. 2.4 and 2.5. (A detailed summary of modern cytological work can be obtained in [3]).

One difference between prokaryotic and eukaryotic cells is the complex systems of membranes and membrane bound organelles present in the eukaryotic cell. Another basic difference is that in prokaryotic cells there is no nucleus. The chromosome comprises a single very long DNA molecule (often cyclic) not clearly associated with any other cell component other than the membrane at one region. In eukaryotic cells the DNA is first enclosed within the nucleus and furthermore it is there associated with considerable amounts of basic (histone) and other proteins to form the chromosomes which condense and become clearly visible at the time of mitosis. However during interphase (as in Fig. 2.4) there is much less structural detail to be seen either by light or electron microscope. Nevertheless two different states have been recognised namely *euchromatin,* which consists of loosely coiled fibres of DNA and protein, and *heterochromatin,* which appears to comprise tightly coiled fibres [4].

While the composition of cellular components can be studied *in situ* with the use of various sophisticated histochemical and cytochemical techniques, considerable information on cellular processes such as RNA biosynthesis has been obtained by study of cell components *after* disruption of the cells followed by fractionation involving differential centrifugation. There is no standard method however that is applicable to all cells. Nevertheless the methodology that has been developed for rat liver cells [5] has proven remarkably versatile and adaptable to other tissues and organisms [6].

Basically the tissues (or cells) are firstly disrupted mechanically in suitable media (usually containing sucrose to reduce aggregation problems). This can be done with a tissue grinder, glass homogeniser, a Waring blender or other high speed mixer with the aid of abrasives or glass beads. The disrupted cell preparation can then be centrifuged at around 200 g to remove nuclei and general cell debris

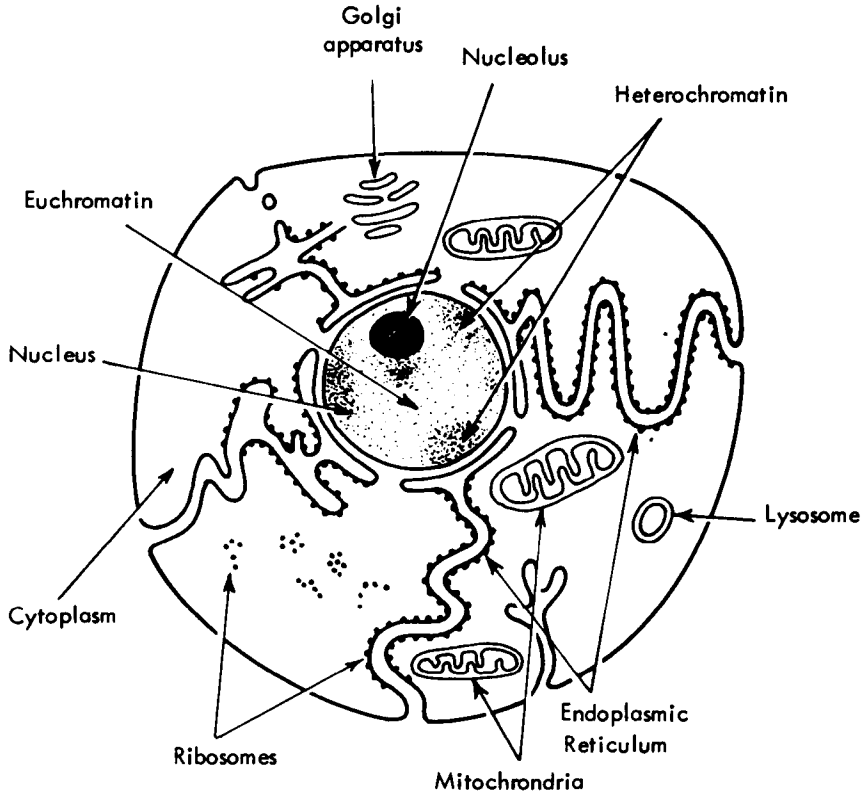


Fig. 2.4 A highly schematic representation of a eukaryotic (e.g. mammalian) cell.

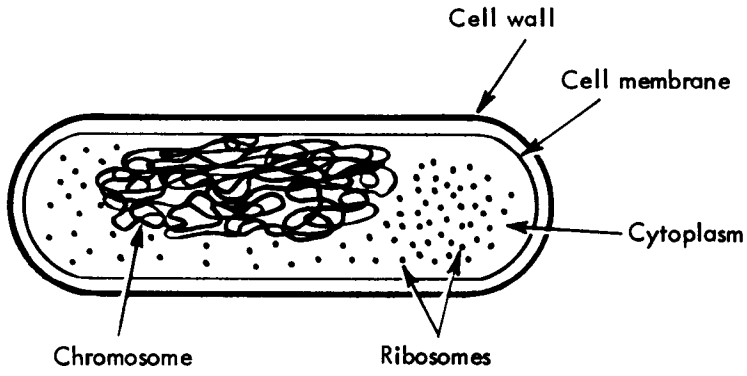
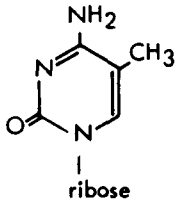


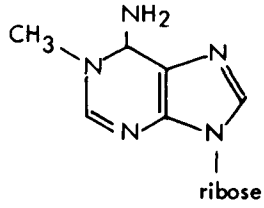
Fig. 2.5 A highly schematic representation of a prokaryotic (e.g. bacterial) cell.

including unbroken cells. After removal of the nuclear fraction the remaining material is centrifuged at 8500 g for 10 min or so to bring down mitochondria, and at 140 000 g for 90 min to sediment microsomes and any 'free' (or 'unbound') ribosomes. The clear supernatant fluid is said to correspond to the cell sap (or cytosol).

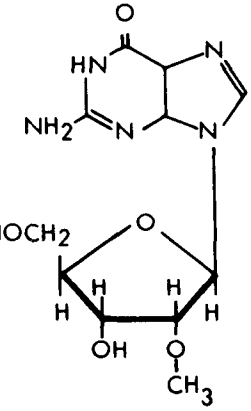
Microsomes, (which correspond to fragments of endoplasmic reticulum) when treated with the detergent, sodium deoxycholate, are further disrupted into a non-sedimentable portion derived from the membranous component which contains most of the protein and phospholipid, and a particulate portion sedimentable at 14 000 g which corresponds to the



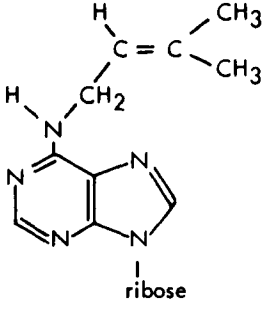
5-methylcytidine



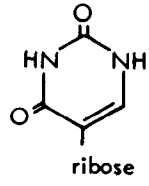
N¹-methyladenosine



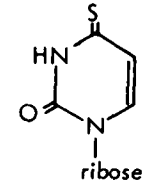
2'-o-methylguanosine



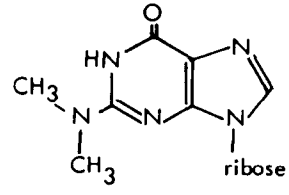
N^6(Δ²-isopentenyl)adenosine



pseudouridine (Ψ)



4-thiouridine



N^2-dimethylguanosine

Fig. 2.6 The structure of a few modified nucleosides.

'bound' ribosomes (i.e. the ribosomes originally bound as part of the endoplasmic reticulum as distinct from those 'free' in the cytoplasm).

As stated already the basic procedure of differential ultracentrifugation has many variations and for further details of the technique the reader is referred elsewhere [6].

When the above mentioned subcellular fractions from both prokaryotic and eukaryotic cells are examined for the occurrence of nucleic acids, the bulk of cellular RNA is to be found in ribosomes. However some also occurs in the cell sap. In eukaryotic cells the nuclei also contain a small amount of RNA, as do mitochondria as well as the chloroplasts of photosynthetic organisms.

2.3 Transfer RNA

Historically the first type of cellular RNA to be characterised was transfer RNA (tRNA) (see Chapter 1).

This class of molecules is predominantly located in the soluble portion of the cytoplasm (cell sap, or cytosol) and accounts for about 10 to 15 per cent of the total cellular RNA in prokaryotes and eukaryotes. In the case of rapidly dividing bacterial cells there are about 4×10^5 tRNA molecules of perhaps fifty or so different varieties. The precise number of varieties is not yet known but there is at least one for every different type of amino acid. In a mammalian cell the total number of tRNA molecules per cell can be as high as 10^8.

tRNAs can be readily extracted from the cell sap of most cells with buffered aqueous phenol and are found to be relatively small, sedimenting in the 4S region on zonal ultracentrifugation. The precise chain length of the different varieties appears to vary over the rather narrow range of 76-85 nucleotides. However, despite the general remarks about structure made earlier in Section 2.1, tRNAs are slightly exceptional. Although they contain the four common nucleosides they contain also a variety of unusual or 'minor' ribonucleosides (see Fig. 2.6). Whilst the presence of these 'minor' nucleosides may at first sight be both puzzling and confusing, it will be seen later in this book that they arise as simple structural modifications to the primary nucleotide sequence of newly made RNA molecules. Despite this chemical knowledge, the functional significance of these molecular modifications is not yet appreciated although a modified nucleoside next to the 'anti-codon' sequences does appear to be essential for recognition purposes. It was clear from physical studies that tRNAs had a certain percentage of double helical structure or secondary structure [2]. For instance, the 'melting' curves, whilst gradual, were nevertheless reversible, resembling the melting of other simple double helical molecules (e.g. poly (A) · poly (U)). However, when the sedimentation properties, viscosities and UV absorbance of total tRNA were measured at various temperatures, a large change in conformation was noted between 20°–40°, but with only a small loss in secondary structure. Thus it seemed that tRNA had a structure of higher order that is more compact and more stable than just a loose combination of helical segments. In other words, there were good grounds for supposing a tertiary structure for tRNA [2].

Since all tRNAs have quite similar properties and are all about 80 nucleotides long a combination of fractionation methods is required to purify single species. An early approach was to use counter current distribution and this permitted the isolation of tRNA from yeast that was specific for alanine (yeast $tRNA^{Ala}$) whose complete sequence was the first to be elucidated by Holley and his colleagues [7]. This separation technique is still widely used but often in combination with other column chromatographic methods.

While the original sequence determination

required fairly large amounts of pure tRNA, only 0.5 mg of highly labelled [^{32}P]–tRNA is required for the extremely rapid sequencing method developed by Sanger and his colleagues – the so-called 'fingerprinting' technique [8]; The methods employed in primary sequence determinations consist essentially of the controlled degradation of the RNA with enzymes and separation of the products by chromatography [7], or by two-dimensional electrophoresis in the case of the Sanger technique [8].

The determination in 1965 of the complete sequence of yeast tRNA^{Ala} has been followed by the sequence of a further forty tRNAs [9]. Just about all these sequences can be fitted to the same hydrogen bonded secondary structure of loops and short helical regions as shown Fig. 2.7. The two parts of the structure which have a known function are (a) the 3′–terminal adenosine residue to which the amino acid is esterified, and (b), the three adjacent bases carrying the *anticodon* which in all tRNAs occupy the same position in the clover leaf.

Whilst there is a lot of evidence to support the clover-leaf arrangement in two dimensions, only recently has there been any unambiguous data regarding tertiary structure. The results of a systematic crystallisation study of yeast tRNA^{Phe} were crystals suitable for X-ray analysis at 3Å resolution using the method of isomorphous replacement [10]. A schematic diagram is shown in Fig. 2.8 of the basic three dimensional arrangement found.

2.4 The RNAs of the ribosomes

Whilst a great deal is known about tRNA, its structure and function, the bulk of cellular RNA (about 80 per cent) is contained in the minute cytoplasmic particles known as ribosomes. These have a diameter of around 20 nm, contain protein as well as RNA, and are found in all types of living cell both 'free' and 'bound' to membrane components (e.g. endoplasmic reticulum).

It is customary, although perhaps somewhat prosaic, to characterise the ribosomes by their sedimentation coefficients expressed in Svedberg units. In mammalian cells there are around 5×10^6 ribosomes each sedimenting around 80S; however in bacteria the basic ribosome is only 70S and there are only $15–18 \times 10^3$ per cell.

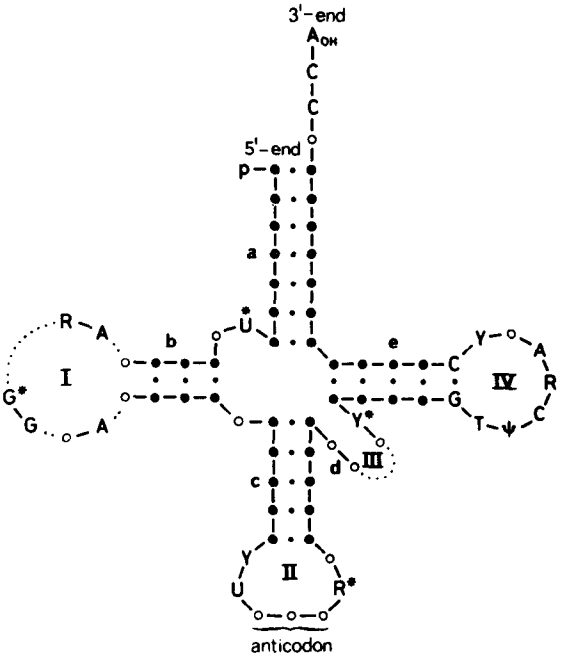


Fig. 2.7 A generalised version of the clover leaf model for tRNA showing constant features. ●, any base; R, purine; Y, pyrimidine; T, ribothymidine; ψ, pseudouridine; R*, modified adenine.

Ribosomes from all sources have very similar structures [11]. They are roughly ellipsoidal complexes of RNA and proteins (ribonucleo-

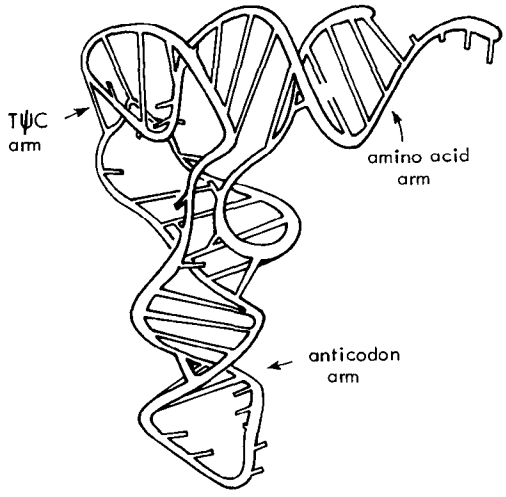


Fig. 2.8 A schematic model of yeast phenylalanine transfer RNA (the ribose phosphate backbone is drawn as a continous cylinder with bars to indicate hydrogen bonded base pairs).

proteins) of molecular weight ranging from 2.5×10^6 in bacteria to roughly 4×10^6 in mammals. Basically they comprise two subunits of which the larger is 2 to 2.5 times the size of the smaller. The two subunits associate together and function as an integrated unit in protein synthesis. They dissociate between rounds of protein synthesis *in vivo* and can be made to dissociate *in vitro* by exposure to low Mg^{2+} concentrations, high concentrations of mono-valent cations, or EDTA. The relative S values for the ribosome and its subunits (large and small) are: in bacteria 70S, (50S and 30S) and in higher organisms 80S, (60S and 40S) (See Fig. 2.9).

Somewhat more than half the mass of each ribosomal subunit consists of RNA, the remainder being made up of protein. The small subunit comprises a single RNA molecule (0.55×10^6 daltons – or 16S–in bacteria and 0.75×10^6 daltons – or 18S – in mammalian ribosomes) together with several proteins (20 in *E.coli*). The large subunit contains one large RNA molecule (1.1×10^6 daltons – or 23S – in bacteria, and 1.75×10^6 daltons – or 28S – in mammalian ribosomes) as well as one small RNA molecule called the 5S RNA (120 nucleotides long) and a further group of proteins (36 in *E.coli*). Fractionation of the protein from animal cell ribosomes also indicates a high degree of complexity [11]. Bacteria, actinomycetes, blue green algae and higher plant chloroplasts however all have ribosomal RNAs of molecular weight 1.1×10^6, whereas the corresponding values for higher plants, ferns, algae, fungi and some protozoa are 1.3×10^6 and 0.7×10^6 [14]. The 0.7×10^6 component (18S) is common to all animals, but the large (28S) component has evolved with each major step of animal evolution from 1.4×10^6 in sea urchins to 1.75×10^6 in mammals [14]. Whilst the ribosomal RNA from most organisms possess broadly similar base compositions with guanine plus cytosine contents of between 50–60 per cent there are some exceptions to this e.g. Drosophila (40 per cent), Tetrahymena (43 per cent) [11].

As was the case for tRNA a small number of specific nucleotides of ribosomal RNAs are modified following transcription. *E.coli* 16S and 23S RNA possess 22 and 27 methyl groups respectively. Most of these groups are on various base moieties (see Fig. 2.10), only a few being on the ribose [12]. Whilst the large RNAs (28S and 18S) of the mammalian ribosome are also methylated (71 and 46 methyl groups respectively), unlike the bacterial situation, most (95 per cent) are on ribose moieties and the rest on bases [11].

The question of whether or not the polynucleotide chains of ribosomal RNA (rRNA) are covalently continuous has been extensively examined. The sedimentation properties of