PROGRESS IN

Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Volume 28

PROGRESS IN

Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Volume 28

1983

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London
Paris San Diego San Francisco São Paulo Tokyo Toronto

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

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

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63-15847

ISBN 0-12-540028-4

PRINTED IN THE UNITED STATES OF AMERICA

83 84 85 86 9 8 7 6 5 4 3 2 1

Contributors

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

ABRAHAM K. ABRAHAM (81), *Department of Biochemistry, University of Bergen, Arstadveien 19, 5000 Bergen, Norway*

RICHARD BRIMACOMBE (1), *Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany*

STEVEN G. ELLIOTT (143), *Department of Biological Chemistry, California College of Medicine, University of California, Irvine, California 92717*

CALVIN S. MCLAUGHLIN (143), *Department of Biological Chemistry, California College of Medicine, University of California, Irvine, California 92717*

PETER MALY (1), *Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany*

SUSUMU NISHIMURA (49), *Biology Division, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104, Japan*

EDDA F. ROBERTS (211), *Department of Chemistry, Wichita State University, Wichita, Kansas 67208*

JONI K. SHAW (177), *Department of Chemistry, Wichita State University, Wichita, Kansas 67208*

RAM P. SINGHAL[1] (75, 177, 211), *Department of Chemistry, Wichita State University, Wichita, Kansas 67208*

ALAP-RAMAN SUBRAMANIAN (101), *Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany*

VIKRAM N. VAKHARIA (211), *Department of Chemistry, Wichita State University, Wichita, Kansas 67208*

CHRISTIAN ZWIEB[2] (1), *Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany*

[1] Present address: Laboratory of Cellular and Molecular Biology, Gerontology Research Center, National Institute on Aging, Baltimore, Maryland 21224.

[2] Present address: Department of Physiological Chemistry, Brown University, Providence, Rhode Island.

Abbreviations and Symbols

All contributors to this Series are asked to use the terminology (abbreviations and symbols) recommended by the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) and approved by IUPAC and IUB, and the Editor endeavors to assure conformity. These Recommendations have been published in many journals (*1*, *2*) and compendia (*3*) in four languages and are available in reprint form from the Office of Biochemical Nomenclature (OBN), as stated in each publication, and are therefore considered to be generally known. Those used in nucleic acid work, originally set out in section 5 of the first Recommendations (*1*) and subsequently revised and expanded (*2*, *3*), are given in condensed form (I–V) below for the convenience of the reader. Authors may use them without definition, when necessary.

I. Bases, Nucleosides, Mononucleotides

1. *Bases* (in tables, figures, equations, or chromatograms) are symbolized by Ade, Gua, Hyp, Xan, Cyt, Thy, Oro, Ura; Pur = any purine, Pyr = any pyrimidine, Base = any base. The prefixes S–, H_2, F–, Br, Me, etc., may be used for modifications of these.

2. *Ribonucleosides* (in tables, figures, equations, or chromatograms) are symbolized, in the same order, by Ado, Guo, Ino, Xao, Cyd, Thd, Ord, Urd (Ψrd), Puo, Pyd, Nuc. Modifications may be expressed as indicated in (1) above. Sugar residues may be specified by the prefixes r (optional), d (=deoxyribo), a, x, l, etc., to these, or by two three-letter symbols, as in Ara-Cyt (for aCyd) or dRib-Ade (for dAdo).

3. *Mono-, di-, and triphosphates of nucleosides* (5′) are designated by NMP, NDP, NTP. The N (for "nucleoside") may be replaced by any one of the nucleoside symbols given in II-1 below. 2′-, 3′-, and 5′- are used as prefixes when necessary. The prefix d signifies "deoxy." [Alternatively, nucleotides may be expressed by attaching P to the symbols in (2) above. Thus: P-Ado = AMP; Ado-P = 3′-AMP] cNMP = cyclic 3′:5′-NMP; Bt_2cAMP = dibutyryl cAMP, etc.

II. Oligonucleotides and Polynucleotides

1. Ribonucleoside Residues

(a) Common: A, G, I, X, C, T, O, U, Ψ, R, Y, N (in the order of I-2 above).

(b) Base-modified: sI or M for thioinosine = 6-mercaptopurine ribonucleoside; sU or S for thiouridine; brU or B for 5-bromouridine; hU or D for 5,6-dihydrouridine; i for isopentenyl; f for formyl. Other modifications are similarly indicated by appropriate *lower-case* prefixes (in contrast to I-1 above) (*2*, *3*).

(c) Sugar-modified: prefixes are d, a, x, or l as in I-2 above; alternatively, by *italics* or boldface type (with definition) unless the entire chain is specified by an appropriate prefix. The 2′-*O*-methyl group is indicated by *suffix* m (e.g., -Am- for 2′-*O*-methyladenosine, but -mA- for 6-methyladenosine).

(d) Locants and multipliers, when necessary, are indicated by superscripts and subscripts, respectively, e.g., -m_2^6A- = 6-dimethyladenosine; -s^4U- or -^{4}S- = 4-thiouridine; -ac^{4}Cm- = 2′-*O*-methyl-4-acetylcytidine.

(e) When space is limited, as in two-dimensional arrays or in aligning homologous sequences, the prefixes may be placed *over the capital letter*, the suffixes *over the phosphodiester symbol*.

2. Phosphoric Residues [left side = 5′, right side = 3′ (or 2′)]

(a) Terminal: p; e.g., pppN . . . is a polynucleotide with a 5′-triphosphate at one end; Ap is adenosine 3′-phosphate; C > p is cytidine 2′:3′-cyclic phosphate (*1*, *2*, *3*); p < A is adenosine 3′:5′-cyclic phosphate.

(b) Internal: hyphen (for known sequence), comma (for unknown sequence); unknown sequences are enclosed in parentheses. E.g., pA-G-A-C(C_2,A,U)A-U-G-C > p is a sequence with a (5′) phosphate at one end, a 2′:3′-cyclic phosphate at the other, and a tetranucleotide of unknown sequence in the middle. **(Only codon triplets should be written without some punctuation separating the residues.)**

3. Polarity, or Direction of Chain

The symbol for the phosphodiester group (whether hyphen or comma or parentheses, as in 2b) represents a 3′-5′ link (i.e., a 5′ . . . 3′ chain) unless otherwise indicated by appropriate numbers. "Reverse polarity" (a chain proceeding from a 3′ terminus at left to a 5′ terminus at right) may be shown by numerals or by right-to-left arrows. Polarity in any direction, as in a two-dimensional array, may be shown by appropriate rotation of the (capital) letters so that 5′ is at left, 3′ at right when the letter is viewed right-side-up.

4. Synthetic Polymers

The complete name or the appropriate group of symbols (see II-1 above) of the repeating unit, **enclosed in parentheses if complex or a symbol**, is either (a) preceded by "poly," or (b) followed by a subscript "n" or appropriate number. **No space follows "poly"** (2, 5).

The conventions of II-2b are used to specify known or unknown (random) sequence, e.g.,

polyadenylate = poly(A) or A_n, a simple homopolymer;

poly(3 adenylate, 2 cytidylate) = poly(A_3C_2) or $(A_3,C_2)_n$, an *irregular* copolymer of A and C in 3 : 2 proportions;

poly(deoxyadenylate-deoxythymidylate) = poly[d(A-T)] or poly(dA-dT) or $(dA\text{-}dT)_n$ or $d(A\text{-}T)_n$, an *alternating* copolymer of dA and dT;

poly(adenylate,guanylate,cytidylate,uridylate) = poly(A,G,C,U) or $(A,G,C,U)_n$, a random assortment of A, G, C, and U residues, proportions unspecified.

The prefix copoly or oligo may replace poly, if desired. The subscript "n" may be replaced by numerals indicating actual size, e.g., $A_n \cdot dT_{12\text{-}18}$.

III. Association of Polynucleotide Chains

1. *Associated* (e.g., H-bonded) chains, or bases within chains, are indicated by a *center dot* (not a hyphen or a plus sign) separating the *complete* names or symbols, e.g.:

poly(A) · poly(U) or $A_n \cdot U_m$

poly(A) · 2 poly(U) or $A_n \cdot 2U_m$

poly(dA-dC) · poly(dG-dT) or $(dA\text{-}dC)_n \cdot (dG\text{-}dT)_m$.

2. *Nonassociated* chains are separated by the plus sign, e.g.:

$2[\text{poly(A)} \cdot \text{poly(U)}] \rightarrow \text{poly(A)} \cdot 2\,\text{poly(U)} + \text{poly(A)}$

or $2[A_n \cdot U_m] \rightarrow A_n \cdot 2U_m + A_n$.

3. Unspecified or unknown association is expressed by a comma (again meaning "unknown") between the completely specified chains.

Note: In all cases, each chain is completely specified in one or the other of the two systems described in II-4 above.

IV. Natural Nucleic Acids

RNA	ribonucleic acid or ribonucleate
DNA	deoxyribonucleic acid or deoxyribonucleate
mRNA; rRNA; nRNA	messenger RNA; ribosomal RNA; nuclear RNA
hnRNA	heterogeneous nuclear RNA
D-RNA; cRNA	"DNA-like" RNA; complementary RNA

mtDNA	mitochondrial DNA
tRNA	transfer (or acceptor or amino-acid-accepting) RNA; replaces sRNA, which is not to be used for any purpose
aminoacyl-tRNA	"charged" tRNA (i.e., tRNA's carrying aminoacyl residues); may be abbreviated to AA-tRNA
alanine tRNA or $tRNA^{Ala}$, etc.	tRNA normally capable of accepting alanine, to form alanyl-tRNA, etc.
alanyl-tRNA or alanyl-$tRNA^{Ala}$	The same, with alanyl residue covalently attached. [*Note:* fMet = formylmethionyl; hence $tRNA^{fMet}$, identical with $tRNA_f^{Met}$]

Isoacceptors are indicated by appropriate subscripts, i.e., $tRNA_1^{Ala}$, $tRNA_2^{Ala}$, etc.

V. Miscellaneous Abbreviations

P_i, PP_i	inorganic orthophosphate, pyrophosphate
RNase, DNase	ribonuclease, deoxyribonuclease
t_m (not T_m)	melting temperature (°C)

Others listed in Table II of Reference 1 may also be used without definition. No others, with or without definition, are used unless, in the opinion of the editor, they increase the ease of reading.

Enzymes

In naming enzymes, the 1978 recommendations of the IUB Commission on Biochemical Nomenclature (*4*) are followed as far as possible. At first mention, each enzyme is described *either* by its systematic name *or* by the equation for the reaction catalyzed *or* by the recommended trivial name, followed by its EC number in parentheses. Thereafter, a trivial name may be used. Enzyme names are not to be abbreviated except when the substrate has an approved abbreviation (e.g., ATPase, but not LDH, is acceptable).

References*

1. *JBC* **241**, 527 (1966); *Bchem* **5**, 1445 (1966); *BJ* **101**, 1 (1966); *ABB* **115**, 1 (1966), 129, 1 (1969); and elsewhere.†
2. *EJB* **15**, 203 (1970); *JBC* **245**, 5171 (1970); *JMB* **55**, 299 (1971); and elsewhere.†
3. "Handbook of Biochemistry" (G. Fasman, ed.), 3rd ed. Chemical Rubber Co., Cleveland, Ohio, 1970, 1975, Nucleic Acids, Vols. I and II, pp. 3–59.
4. "Enzyme Nomenclature" [Recommendations (1978) of the Nomenclature Committee of the IUB]. Academic Press, New York, 1979.
5. "Nomenclature of Synthetic Polypeptides," *JBC* **247**, 323 (1972); *Biopolymers* **11**, 321 (1972); and elsewhere.†

Abbreviations of Journal Titles

Journals	*Abbreviations used*
Annu. Rev. Biochem.	ARB
Arch. Biochem. Biophys.	ABB
Biochem. Biophys. Res. Commun.	BBRC

*Contractions for names of journals follow.

†Reprints of all CBN Recommendations are available from the Office of Biochemical Nomenclature (W. E. Cohn, Director), Biology Division, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tennessee 37830, USA.

Biochemistry	Bchem
Biochem. J.	BJ
Biochim. Biophys. Acta	BBA
Cold Spring Harbor Symp. Quant. Biol.	CSHSQB
Eur. J. Biochem.	EJB
Fed. Proc.	FP
Hoppe-Seyler's Z. physiol. Chem.	ZpChem
J. Amer. Chem. Soc.	JACS
J. Bacteriol.	J. Bact.
J. Biol. Chem.	JBC
J. Chem. Soc.	JCS
J. Mol. Biol.	JMB
Nature, New Biology	Nature NB
Nucleic Acid Research	NARes
Proc. Nat. Acad. Sci. U.S.	PNAS
Proc. Soc. Exp. Biol. Med.	PSEBM
Progr. Nucl. Acid Res. Mol. Biol.	This Series

Some Articles Planned for Future Volumes

Hypermodified Nucleosides of tRNA
A. ADAMIAK

RNA Processing in a Unicellular Microorganism
D. APIRION

Nearest Neighbor Effects in the Structure and Function of Nucleic Acids
P. N. BORER

The Elongation Factor EF-Tu and Its Two Encoding Genes
L. BOSCH

Small Nuclear RNAs and RNA Processing
H. BUSCH

Participation of Aminoacyl-tRNA Synthetases and tRNAs in Regulatory Processes
G. NASS

Neuropolyproteins from Brain
D. RICHTER

Viral Inhibition of Host Protein Synthesis
A. SHATKIN

VOLUME 29: GENETIC MECHANISMS AND CARCINOGENESIS

Contents

The Fidelity of Translation

Abraham K. Abraham

Structure and Functions of Ribosomal Protein S1

Alap-Raman Subramanian

The Yeast Cell Cycle: Coordination of Growth and Division Rates

Steven G. Elliott and Calvin S. McLaughlin

Prokaryotic and Eukaryotic 5 S RNAs: Primary Sequences and Proposed Secondary Structures

Ram P. Singhal and Joni K. Shaw

Structure of Transfer RNAs: Listing of 150 Additional Sequences

Ram P. Singhal, Edda F. Roberts, and Vikram N. Vakharia

The Structure of Ribosomal RNA and Its Organization Relative to Ribosomal Protein

RICHARD BRIMACOMBE,
PETER MALY, AND
CHRISTIAN ZWIEB*

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany

The structure of the *Escherichia coli* ribosome was last reviewed by our laboratory in this series in 1976 (*1*). At that time, the emphasis was on the structure and function of the individual ribosomal proteins, and, although a considerable amount of information was already available concerning the ribosomal RNA, progress in this area was hampered by the lack of complete base sequence data for the ribosomal RNA molecules. Since then, as a direct result of the explosion in DNA sequencing technology (*2, 3*), the situation has changed dramatically. Complete nucleotide sequences have now been established, corresponding not only to the ribosomal RNA molecules from *E. coli*, but also to those from a number of different organisms, covering a large portion of the evolutionary spectrum.

* Present address: Department of Physiological Chemistry, Brown University, Providence, Rhode Island.

Progress in Nucleic Acid Research and Molecular Biology, Vol. 28

ISBN 0-12-540028-4

We begin by reviewing briefly the current status of these sequences. Next, we describe how the sequence information has been used to derive convincing secondary structure models for the RNA from both subunits of the *E. coli* ribosome, and we compare the various models that have been proposed. We show how extrapolation of these data to ribosomal RNA molecules of widely differing size classes leads to the clear conclusion that the secondary structures, as well as significant regions in the primary sequences, have been conserved to a large extent throughout evolution. Section IV deals with the three-dimensional organization of the ribosomal RNA and its arrangement with respect to the ribosomal proteins, concentrating once again on the *E. coli* ribosome. In particular, we include a review of the application of cross-linking techniques (both RNA to protein and intra-RNA) to this problem. In general, rather than presenting an exhaustive survey of the literature, we have selected topics or examples to illustrate those problems or points of interest that we consider to be most relevant to the central objective in this field of research, namely, the elucidation of the three-dimensional structure of the ribosomal RNA *in situ* in the ribosome.

I. Ribosomal RNA Sequences

The organization and transcription of ribosomal RNA genes is a complex and fascinating subject, itself worthy of review. However, we confine ourselves here to a discussion of the mature ribosomal RNA (rRNA) species, as they occur in the completed ribosomal particles.

The "standard" bacterial ribosome, as typified by that of *E. coli*, contains in its small subunit a single RNA molecule (16 S), which is about 1540 nucleotides in length (*4*, *5*). The large subunit contains a 23 S RNA (ca. 2900 nucleotides) (*6*) and a 5 S RNA (120 nucleotides) (*7*). In other organisms, the size of these rRNA molecules varies considerably. The smallest so far reported are those from trypanosome mitochondria, which are only 640 and 1230 nucleotides in length, respectively, from the small and large subunits (*8*). Next come the mitochondrial ribosomes from mammals, with RNA molecules of 12 S and 16 S (ca. 950 and 1550 nucleotides) (e.g., *9*, *10*), and these very small ribosomes contain no 5 S RNA. Other types of mitochondria (e.g., *11*), and also chloroplasts (e.g., *12*, *13*), have rRNA molecules corresponding in size to those of the bacterial ribosomes. The chloroplasts and mitochondria from higher plants both contain 5 S rRNA, and in addition the large subunit of higher plant chloroplast ribosomes contains a 4.5 S RNA species (*14*).

The largest ribosomes are those from the cytoplasm of eukaryotes, with RNA molecules of 18 S (ca. 1800 nucleotides) (e.g., *15, 16*) and 26–28 S (up to 4000 nucleotides) (e.g., *17*) in the small and large subunits, respectively. The large subunit contains a 5.8 S RNA molecule (e.g., *18*) as well as 5 S RNA, and a 2 S RNA species is also observed in the large ribosomal subunit from *Drosophila* (*19*). Further, the large subunit rRNA genes for many eukaryotic ribosomes contain introns (e.g., *20, 21*), and in *Drosophila* the final "28 S" transcription product appears in two distinct halves (reviewed in 22). As shown in Section II,B the small rRNA molecules (4.5 S, 5.8 S, and 2 S) all have clear counterparts within the 23 S rRNA from *E. coli*, and they can therefore better be regarded as products of posttranscriptional processing of the rRNA rather than as "extra" rRNA species.

Many complete or partial sequences are now known for all these classes of rRNA, mostly obtained by determination of the corresponding rDNA sequences. A compilation of known 5 S and 5.8 S sequences has been made (*23, 24*),[1] and this list continues to grow at a rate of about one sequence every two or three weeks. [One of the more interesting new sequences here is the "5 S" rRNA from *Halococcus morrhuae*, which contains a 108-base insertion (25).] The sequence of 2 S rRNA from *Drosophila melanogaster* is known (26), and also the sequences of 4.5 S rRNA from wheat (27), tobacco (28), and maize (29).

More important for the purpose of this article are the sequences of the major rRNA molecules, and those sequences currently complete or nearly complete are listed in Table I (*4–6, 9–13, 15–17, 30–40*). It can be seen from Table I that almost every common size class of large rRNA mentioned above is represented by sequences from two or more species. In addition, many partial sequences of the major rRNA molecules have been determined. In the case of the small subunit, sequences are available from a total of about 20 species for the 50–200 nucleotides at the 3′ terminus, and compilations of these have been published (*41, 42*). In the case of the large-subunit rRNA, some partial sequences are available for the regions flanking introns in the rDNA of *Chlamydomonas reinhardii* chloroplast (*43*) and yeast mitochondrial ribosomes (*20, 44*), as well as of *Drosophila virilis* (*45*), *D. melanogaster* (*46, 47*), *Tetrahymena pigmentosa* (*21*), and *Physarum polycephalum* (*48, 49*) cytoplasmic ribosomes. The positions in *E. coli* 23 S rRNA that correspond to the locations of these inserted sequences have been collated (*36*). Similar short sequences are known from the 5′ ends of the large-subunit rRNA of *Aspergillus nidulans* mitochon-

[1] See also Singhal and Shaw in this volume. [Ed.]

TABLE I
SEQUENCES OF RIBOSOMAL DNA (OR rRNA) MOLECULES[a]

Organism	Small subunit rRNA	Large subunit rRNA
Human mitochondrion	12 S (*9*)	16 S (*9*)
Mouse mitochondrion	12 S (*10*)	16 S (*10*)
Rat mitochondrion	—	16 S (*30*)
Saccharomyces cerevisiae mitochondrion	15 S (*11*)	—
Aspergillus nidulans mitochondrion	15 S (*31*)	—
Paramecium primaurelia mitochondrion	—	20 S (*32*)
Escherichia coli	16 S (*4, 5*)	23 S (*6, 33*)
Proteus vulgaris	16 S (*34*)	—
Bacillus brevis	(16 S) (*35*)	—
Bacillus stearothermophilus	—	23 S (*36*)
Zea mays chloroplast	16 S (*12*)	23 S (*13*)
Euglena gracilis chloroplast	16 S (*37*)	—
Saccharomyces cerevisiae cytoplasm	18 S (*15*)	26 S (*17*)
Saccharomyces carlsbergensis cytoplasm	—	26 S (*38*)
Xenopus laevis cytoplasm	18 S (*16*)	(28 S) (*39, 40*)

[a] Known sequences are listed, showing the size (S value in Svedberg units) of the corresponding rRNA molecule; parentheses denote that the sequence concerned is not yet fully complete; a dash indicates that the sequence is not available. All sequences were determined from ribosomal DNA, with the exception of *E. coli* 16 S rRNA (*5*) and 23 S rRNA (*33*), and *P. vulgaris* 16 S rRNA (*34*); these determinations were made directly from RNA. Numbers in parentheses indicate reference numbers.

dria (*50*) and *Euglena gracilis* chloroplasts (*51, 52*) and from the 3′ ends of *Aeromonas punctata* and *Proteus vulgaris* (*53*). Last, a large amount of sequence data is contained in oligonucleotide catalogs made earlier from the rRNA of many species [summarized by Fox *et al.* (*54*)].

It has long been known that the sequence of 5 S rRNA is highly conserved (see, e.g., *55*), and a corresponding pattern of conservation is seen among the major rRNA molecules. For example, *P. vulgaris* 16 S rRNA shows 93% homology to the 16 S rRNA from *E. coli* (*34*), whereas the homology between the 23 S rRNA molecules from *B. stearothermophilus* and *E. coli* is about 75% (*36*). *Escherichia coli* rRNA and Z. *mays* chloroplast rRNA also show about 75% sequence homology, in both subunits (*12, 13*). *Xenopus* 18 S and yeast 18 S rRNA are again about 75% homologous to one another (*16*). The two *Saccharomyces* 26 S rRNA sequences [*S. cerevisiae* (*17*) and *S. carlsbergensis* (*18*)] are virtually identical, differing only in about 20 positions, and this order of difference is also observed between different

ribosomal cistrons of the same organism; *E. coli* 16 S rRNA, for instance, contains 16 single-base cistron heterogeneities (*56*).

When sequences from the different size classes of rRNA (Table I) are compared, it becomes apparent that a number of regions have been conserved in both the large- and small-subunit rRNA molecules of all species, despite the large differences in lengths of the polynucleotide chains. This type of homology becomes even more striking when the secondary structures of the various molecules are considered; it is discussed in more detail in Section II,B.

Since most of the sequence data (Table I) was determined from rDNA, information concerning modified nucleotides in the corresponding rRNA molecules is incomplete in many cases. *Escherichia coli* 16 S rRNA contains 9 methylated bases (*56*), and the 23 S rRNA contains 10 methylated bases and 3 pseudouridine residues (*33*). In *P. vulgaris* 16 S rRNA, 6 methylated bases are in homologous positions to their *E. coli* counterparts (*34*), and the two N^6-dimethyladenine residues near the 3′ terminus of the 16 S rRNA seem to be a universal feature of all small-subunit RNA molecules (reviewed in *42*). Some methylated bases have been localized in rodent mitochondrial rRNA, and these also appear to be highly conserved (*57*). The pattern of nucleotide modification in eukaryotic cytoplasmic ribosomes is rather more complex. *Xenopus* 18 S rRNA contains 40 methylated bases, of which the majority are 2′-*O*-methyl groups (*16*). There is one hypermodified base as well as several base-methylated residues, and in addition there are 40 pseudouridine residues (*58*), which for the most part have not as yet been localized precisely. The corresponding yeast 18 S rRNA molecule has the same number of base methylations, but fewer 2′-*O*-methyl groups (*59*). These have not yet been localized, whereas in yeast 26 S rRNA, 30 out of the 43 methyl groups have been placed (*38*). There are also some data on the corresponding sites of methylation in *Xenopus* 28 S rRNA (*60*).

To conclude this section, it is clear that an enormous amount of sequence information has already been collected, and this provides the raw material for the next stage in the elucidation of the three-dimensional organization of ribosomal RNA, namely the construction of models for the secondary structure.

II. Secondary Structures

There are, broadly speaking, three different approaches by which a nucleotide sequence can be folded into a double-helical secondary structure: the theoretical, the experimental, and the comparative.

The theoretical approach relies on the use of computer programs to select the thermodynamically most favorable structure for the sequence, using thermodynamic parameters derived from the melting properties of model oligonucleotides or RNA fragments (e.g., *61–64*). A large number of computer algorithms generated for this purpose have been reported (e.g., *65–69*), and the newer versions (e.g., *68*) have been considerably improved, in that they can select structures for an entire long sequence rather than for just short sections. However, the approach suffers from the obvious disadvantage that the computer program can be only as good as the thermodynamic data put into it, and these data are by no means comprehensive. The effects of imperfections such as "bulges" and "loops" in the helices cannot be computed very accurately, and—more important—the thermodynamic effects of tertiary structure or interactions with protein can hardly be assessed at all. As a result, this approach can lead to erroneous structure predictions, and it has, for example, been shown (*70*) that the computer may predict quite different structures for two phylogenetically closely related sequences. Nevertheless, when combined with experimental or comparative data, the computer approach is a powerful tool for screening potential secondary structures.

The experimental approach to secondary structure determination has made use of a number of different methods; these include chemical modification (e.g., *71–76*), analysis of enzyme cutting points (e.g., *56, 77*), oligonucleotide binding (e.g., *78, 79*), isolation of base-paired RNA fragments (*80, 81*), and intra-RNA cross-linking (e.g., *82, 83*). In the case of chemical modification, the RNA is treated with base-specific reagents to test the accessibility (or single-strandedness) of individual residues. The reagents that have been used include kethoxal (*71*) and glyoxal (*72*) (G-specific); monoperphthalic acid (*73*), *m*-chloroperoxybenzoic acid (*72*), and diethyl pyrocarbonate (*74*) (A-specific); and methoxyamine (*75*) and bisulfite (*72*) (C-specific). Dimethyl sulfate (*74*) (C- and G-specific) and soluble carbodiimides (*76*) (U- and G-specific) have also been applied. The analysis of enzyme cutting points is, in principle, a very similar method, in which the RNA is subjected to a mild digestion by a single-strand-specific nuclease (such as nuclease S_1, or ribonucleases A and T_1), and the resulting fragments are analyzed to pinpoint those residues at which the polynucleotide chain has been cut (*56*). The converse approach, using the double-strand-specific nuclease from cobra venom, has also been applied with success (*77*). Oligonucleotide binding is a probe for single-stranded regions, in which the putatively exposed sequences can be tested for their ability to bind a short complementary oligonucleotide