PROGRESS IN NUCLEIC ACID RESEARCH AND MOLECULAR BIOLOGY

VOLUME 33

PROGRESS IN

Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN

Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

KIVIE MOLDAVE

University of California
Santa Cruz, California

Volume 33

1986

ACADEMIC PRESS, INC.

Harcourt Brace Jovanovich, Publishers

Orlando San Diego New York Austin
Boston London Sydney Tokyo Toronto

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: 63-15847

ISBN 0–12–540033–0

PRINTED IN THE UNITED STATES OF AMERICA

86 87 88 89 9 8 7 6 5 4 3 2 1

PROGRESS IN

Nucleic Acid Research and Molecular Biology

Volume 33

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^6_2A- = 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[poly(A) · poly(U)] → poly(A) · 2 poly(U) + 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 1984 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 (1984) of the Nomenclature Committee of the IUB]. Academic Press, New York, 1984.
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
Annu. Rev. Genet.	ARGen
Arch. Biochem. Biophys.	ABB
Biochem. Biophys. Res. Commun.	BBRC
Biochemistry	Bchem
Biochem. J.	BJ
Biochim. Biophys. Acta	BBA
Cold Spring Harbor	CSH

Cold Spring Harbor Lab.	CSHLab
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
J. Nat. Cancer Inst.	JNCI
Mol. Cell. Biol.	MCBiol
Mol. Cell. Biochem.	MCBchem
Mol. Gen. Genet.	MGG
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

DNA Polymerase
F. J. BOLLUM

The Structural and Functional Basis of Collagen Gene Diversity
P. BORNSTEIN

UV-Induced Crosslinks in Nucleoprotein Structure Investigations
E. I. BUDOWSKY

Hormonally Regulated Eukaryotic Genes
R. W. HANSON

Translational Control in Eukaryotic Protein Synthesis
J. W. B. HERSHEY

Messenger RNA Capping Enzymes from Eukaryotic Cells
Y. KAZIRO

Foreign Gene Expression in Plant Cells
P. F. LURQUIN

Translocation of mRNA
W. E. G. MÜLLER

Structure and Organization of the Genome of *Mycoplasma capricolum*
S. OSAWA

Intermediates in Homologous Recombination Promoted by *recA* Protein
CHARLES RADDING

Oligonucleotide-Directed Site-Specific Mutagenesis
U. L. RAJBHANDARY

Early Signals and Molecular Steps in the Mitogenic Response
E. ROZENGURT

Chloroplast DNA Genes: Structure and Transcription
K. K. TEWARI

Damage to Mammalian DNA by Ionizing Radiation
J. F. WARD

Contents

DNA Polymerase-α: Enzymology, Function, Fidelity, and Mutagenesis

Lawrence A. Loeb, Philip K. Liu, and Michael Fry

Replication of Superhelical DNAs *in Vitro*

Kenneth J. Marians, Jonathan S. Minden, and Camilo Parada

Aspects of the Growth and Regulation of the Filamentous Phages

Wilder Fulford, Marjorie Russel, and Peter Model

Roles of Double-Strand Breaks in Generalized Genetic Recombination

Franklin W. Stahl

Regulation of Protein Synthesis by Phosphorylation of Ribosomal Protein S6 and Aminoacyl-tRNA Synthetases

J. A. Traugh and A. M. Pendergast

The Primary DNA Sequence Determines *in Vitro* Methylation by Mammalian DNA Methyltransferases

Arthur H. Bolden, Cheryl A. Ward, Carlo M. Nalin, and Arthur Weissbach

The Interferon Genes

Charles Weissmann and Hans Weber

Expression of Plasmid-Coded Mutant Ribosomal RNA in *E. coli:* Choice of Plasmid Vectors and Gene Expression Systems

ROLF STEEN,*
DAVID K. JEMIOLO,*
RICHARD H. SKINNER,*
JOHN J. DUNN,† AND
ALBERT E. DAHLBERG*

* *Section of Biochemistry*
Division of Biology and Medicine
Brown University
Providence, Rhode Island 02912
† *Biology Department*
Brookhaven National Laboratory
Upton, New York 11973

Rapid advances in the field of molecular genetics have now made it possible to construct mutations of almost any type in a cloned gene. Recently these powerful methods have begun to be applied to cloned *E. coli* ribosomal DNA (rDNA) to explore the structure and function of ribosomal RNAs (rRNA). The *rrn*B operon of *E. coli* was initially cloned into the multicopy plasmid pBR322 by Noller *et al.* (*1*). The first mutants using this plasmid, pKK3535, were deletions in 16-S and 23-S rRNA (*2–4*). The plasmid was linearized by digestion with a restriction enzyme, followed by limited treatment with the exonuclease *Bal* 31. Subsequently, point mutations have been produced by bisulfite (*5, 6*) and ethyl methanesulfonate mutagenesis (*7, 8*) and by synthetic oligonucleotide-directed mutagenesis (*9*).

From this early work, it was apparent that many mutations caused drastic reductions in cell growth rates (*2*). In some cases, mutations in certain regions of the operon could not be cloned, presumably because the gene product was lethal to the cells. To solve the problem of studying lethal mutations, we developed or adopted a number of additional vector systems. The description of these vectors comprises the first half of this chapter and includes a plasmid with low-copy-number

Progress in Nucleic Acid Research
and Molecular Biology, Vol. 33

and plasmids in which the wild-type promoters for the *rrn*B[1] operon, P1 and P2, are replaced by inducible promoters, either the lambda promoter, P_L, or a T7 late-promoter. These systems have allowed us to clone otherwise lethal mutations and to study their expression.

A second major problem encountered in the study of ribosomal RNA mutants has to do with analysis of the mutant rRNA set in the background of host-coded rRNA. The expression of the rRNA operon is an involved and complex process. The operon is transcribed as a long primary transcript containing 16-S, 23-S, and 5-S rRNA. This primary transcript must be cleaved, methylated, and bound to 52 different ribosomal proteins for gene expression to be complete. Because processing involves these many complex steps, which have not been reproduced *in vitro*, it is necessary to express the cloned genes *in vivo*. This then raises the problem of distinguishing the cloned gene transcript from wild-type rRNA transcribed from the seven rRNA operons on the host chromosome. We address this problem in the second half of this chapter as we describe two procedures that accomplish the specific labeling of cloned rRNA genes.

Different plasmid vectors and expression systems must be used to answer different questions about rRNA. In what follows, we describe the advantages and disadvantages of each system as it applies to the

[1] Glossary:

rrnB: A transcriptional unit (operon) coding for 16-S, 23-S, and 5-S ribosomal RNA (in that order). Transcription of the *rrn*B operon is initiated at two promoters (P1 and P2) located in front of the gene for 16-S and terminated at two terminators (T1 and T2) located after the gene for 5-S rRNA.

NR1: A 90 kb plasmid described by Taylor and Cohen (*13*). There are two copies of the plasmid per chromosome in *E. coli*. The plasmid confers resistance to chloramphenicol and streptomycin.

P1, P2, P_L: P1 and P2 are the two natural promoters initiating transcription of the *rrn*B operon. P_L is a repressible promoter from bacteriophage lambda that replaces P1 and P2 in front of the *rrn*B operon in plasmid pNO2680 (*16*).

***Lac*UV5:** The lactose operon in *E. coli* coding for β-galactosidase, permease, and acetylase. This operon is transcribed from an inducible promoter that can be activated by lactose or iPrSGal (IPTG).

BL21/DE3: A lambda lysogen of *E. coli* ($r_B^- m_B^- rif^S$) in which the prophage carries a copy of the gene for bacteriophage T7 RNA polymerase under the control of the *Lac*UV5 promoter.

pEMBL9$^+$: A 4 kb high-copy number plasmid carrying the origin of replication both for plasmid replication and for bacteriophage f1. The plasmid confers resistance to ampicillin.

***Col*E1:** A group of plasmids all derived from the *E. coli* plasmid *Col*E1.

CSR603: An *E. coli* strain (*rec, uvr, phr*) unable to repair UV-damaged DNA.

HB101: An *E. coli* strain (*rec, str, pro*).

p23S: The precursor of 23-S rRNA.

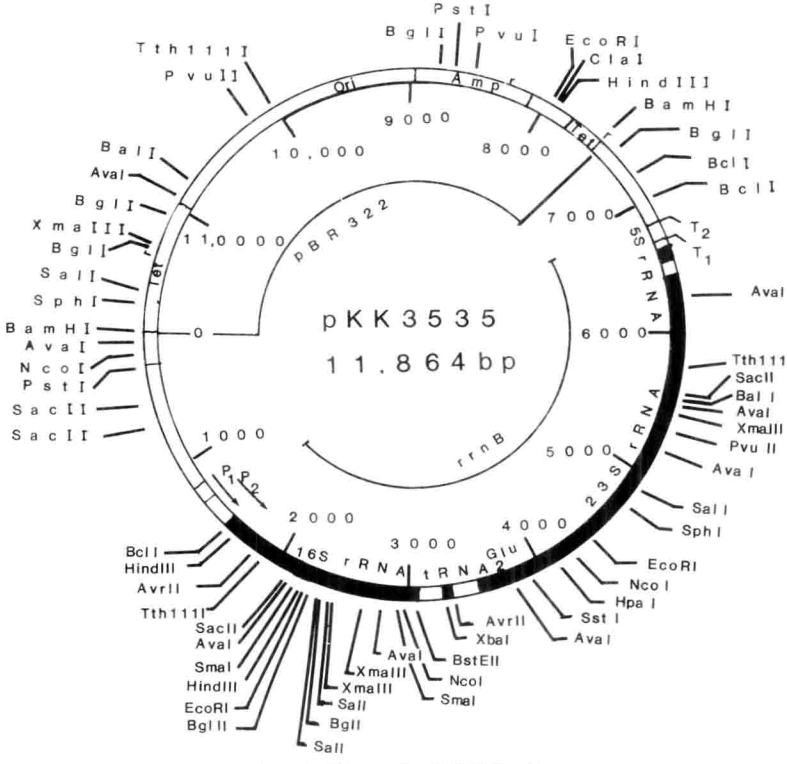

FIG. 1. Plasmid pKK3535 (*1*).

analysis of the structure and function of these large and complex macromolecules, the rRNAs.

I. Expression of Mutant Ribosomal RNA from Wild-Type Promoters, P1 and P2

A. Mutant rRNA Expressed from a High-Copy-Number Plasmid, pKK3535

The initial phase of our work with ribosomal RNA mutagenesis employed the plasmid pKK3535 (Fig. 1). This plasmid was constructed by Noller *et al.* (*1*) as a derivative of the high-copy-number plasmid pBR322, with the *rrn*B operon of *E. coli* inserted at a unique *Bam*HI restriction site (see Fig. 1). The *rrn*B operon contains two tandem promoters, P1 and P2. The plasmid-borne operons are under

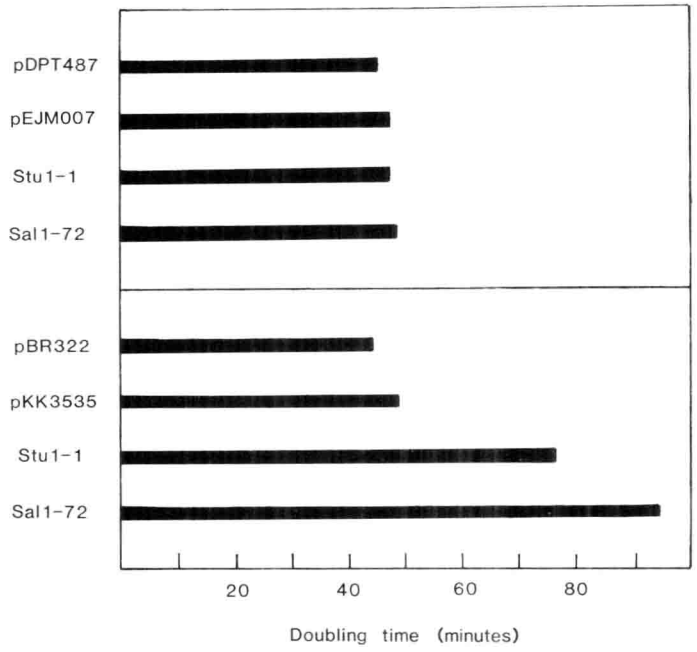

FIG. 2. Doubling times of *E. coli* HB101 strains containing different plasmids. Cell growth rates were measured as described in 2.

stringent control (*10*). Thus the plasmid-borne genes are probably regulated by the same control mechanisms as the host rDNA operons. However, because of the high-copy-number, products from the plasmid-borne genes account for a large fraction of the total cellular pool of rRNA. Plasmid-coded wild-type transcripts may represent close to 50% of the total rRNA, although the level of plasmid-coded mutant transcript may be much reduced in some cases due to instability of the products of the mutant genes (*11*).

We have constructed a number of deletion mutations in the plasmid-borne *rrn*B operon that greatly affect growth rate (2). We have also identified deletions that have no effect on cell growth (5). Examples of the former are deletion mutations within 16-S rRNA of 371 bases (*Stu*I-1, between two *Stu*I sites) and of 53 bases (*Sal*I-72, between two *Sal*I sites). Each of these causes a significant increase in cell doubling time (see Fig. 2). While we are unsure of the exact reason(s) for this effect, several possible mechanisms can be considered. The most interesting possibility involves rRNAs with small deletions, such as the 53-base deletion in *Sal*I-72 that may result in an rRNA product that is impaired in function. It may be assembled into a

ribosomal subunit defective in one or more of the processes involved in translation: initiation, elongation, or termination. Large deletion mutant transcripts may affect cell growth by depleting the cell's resources to produce totally nonfunctional ribosomes. Additionally, they may lack binding sites for ribosomal proteins that regulate their own expression (autogenous regulation, see *12*) thus altering the regulation of expression of ribosomal proteins for host-coded rRNA as well. Translation factors and enzymes involved in rRNA processing (methylases and RNases) may also be tied up with the nonfunctional ribosomal particles. The intracellular turmoil produced by these mutants must be considerable!

In addition to deletion mutants, we have constructed point mutations in plasmid pKK3535 by several techniques, including bisulfite-induced (*5, 6*) and oligonucleotide-directed mutagenesis (*9*). In some cases the mutants had little or no effect on cell growth, while in other cases we failed to recover mutant plasmids, presumably because the products were lethal. For example, in a study of bisulfite-induced mutations near the 3′ end of 16-S rRNA, mutations were commonly found in variable regions of the rRNA sequence (e.g., the stem structure 1409–1491, see Fig. 3), but were rarely isolated in highly conserved regions around 1400 and 1500 (*6*) We were also unsuccessful in producing an oligonucleotide-directed mutation in the mRNA-binding (Shine–Dalgarno) region of 16-S rRNA in plasmid pKK3535. A transition involving C to U at position C1538, within the mRNA binding region, was produced in a cloned fragment in phage M13, but could not be cloned back into pKK3535. Thus plasmid pKK3535 has certain limits to its usefulness in the study of rRNA mutants. While the inability to isolate mutants in certain regions of the rRNA is informative, it provides only indirect evidence of the functional importance of these regions.

In some, if not all cases, the lethal or slow-growth phenotype of the rRNA mutation might be dependent on gene dosage. To determine this, we have cloned mutants into a low-copy-number plasmid that contains the *E. coli* promoters, P1 and P2, and is thus controlled by the cell, much like pKK3535. Using both this system and pKK3535, we can test for the dependence of slow growth (or no growth) phenotype on gene dosage.

B. Mutant rRNA Expressed from a Low-Copy-Number Plasmid, pEJM007

The plasmid pDPT487 is a derivative of the low-copy-number plasmid NR1 (*13*). It carries two genes, chloramphenicol acetyltransferase (EC 2.3.1.28) and streptomycin 3″-adenylyltransferase