PROGRESS IN NUCLEIC ACID RESEARCH AND MOLECULAR BIOLOGY

VOLUME 41

PROGRESS IN
Nucleic Acid Research and Molecular Biology

edited by

WALDO E. COHN
Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

KIVIE MOLDAVE
Department of Biology
University of California
Santa Cruz, California

Volume 41

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers

San Diego New York Boston
London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Academic Press, Inc.
San Diego, California 92101

United Kingdom Edition published by
ACADEMIC PRESS LIMITED
24-28 Oval Road, London NW1 7DX

Library of Congress Catalog Card Number: 63-15847

ISBN 0-12-540041-1 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA
91 92 93 94 9 8 7 6 5 4 3 2 1

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 Editors endeavor to assure conformity. These Recommendations have been published in many journals (*1, 2*) and compendia (*3*) and are available in reprint form from the Office of Biochemical Nomenclature (OBN); they 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 in the frontmatter of Volumes 9–33 of this series. A recent expansion of the one-letter system (*5*) follows.

SINGLE-LETTER CODE RECOMMENDATIONS[a] (*5*)

Symbol	Meaning	Origin of symbol
G	G	Guanosine
A	A	Adenosine
T(U)	T(U)	(ribo)Thymidine (Uridine)
C	C	Cytidine
R	G or A	puRine
Y	T(U) or C	pYrimidine
M	A or C	aMino
K	G or T(U)	Keto
S	G or C	Strong interaction (3 H-bonds)
W[b]	A or T(U)	Weak interaction (2 H-bonds)
H	A or C or T(U)	not G; H follows G in the alphabet
B	G or T(U) or C	not A; B follows A
V	G or C or A	not T (not U); V follows U
D[c]	G or A or T(U)	not C; D follows C
N	G or A or T(U) or C	aNy nucleoside (i.e., unspecified)
Q	Q	Queuosine (nucleoside of queuine)

[a]Modified from *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4 (1986).
[b]W has been used for wyosine, the nucleoside of "base Y" (wye).
[c]D has been used for dihydrouridine (hU or H_2 Urd).

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.† General.
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. Nucleic acids.
4. "Enzyme Nomenclature" [Recommendations (1984) of the Nomenclature Committee of the IUB]. Academic Press, New York, 1984.
5. *EJB* **150**, 1 (1985). Nucleic Acids (One-letter system).†

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. Natl. Acad. Sci. U.S.A.	PNAS
Proc. Soc. Exp. Biol. Med.	PSEBM
Progr. Nucl. Acid. Res. Mol. Biol.	This Series

†Reprints available from the Office of Biochemical Nomenclature (W. E. Cohn, Director).

Some Articles Planned for Future Volumes

Phosphotransfer Reactions of Plant Virus Satellite RNAs
GEORGE BRUENING

Positive and Negative Regulation of Gene Expression by Steroid Agonists and Antagonists
ANDREW C. B. CATO, H. PONTA AND P. HERRLICH

Regulation of Gene Expression in Trypanosomes
CHRISTINE CLAYTON

Oligonucleotides as Antisense Inhibitors of Gene Expression
JACK S. COHEN AND M. GHOSH

The DNA Binding Domain of the Zn(II)-containing Transcription Factors
JOSEPH E. COLMAN AND T. PAN

Specific Hormonal and Neoplastic Transcriptional Control of the Alpha 2u Globulin Gene Family
PHILIP FEIGELSON

Cellular Transcriptional Factors Involved in the Regulation of HIV Gene Expression
RICHARD GAYNOR AND C. MUCHARDT

Correlation between tRNA Structure and Efficient Aminoacylation
RICHARD GIEGE, C. FLORENTZ AND J. PUGLISI

snRNA Genes: Transcription by RNA Polymerase II and RNA Polymerase III
NOURIA HERNANDEZ AND S. LOBO

Regulation of mRNA Stability in Yeast
ALLAN JACOBSON

Recombination Enzymes from *E. coli* and *S. cerevisiae*
RICHARD KOLODNER

Cell Delivery and Mechanisms of Action of Antisense Oligonucleotides
BERNARD LEBLEU, J. P. LEONETTI AND G. DEGLSO

Signal-transducing G Proteins: Basic and Clinical Implications
MICHAEL A. LEVINE

Synthesis of Ribosomes
LASSE LINDAHL AND J. M. ZENGEL

Enzymes of DNA Repair
STUART LINN

RNA Replication of Plant Viruses Comprising an RNA Genome
ANNE-LISE HAENNI, R. GARGOURI-BOUZID AND C. DAVID

Nitrogen Regulation in Bacteria and Yeast
BORIS MAGASANIK

Alkylation Damage Repair Genes: Molecular Cloning and Regulation of Expression
SANKAR MITRA

An Analysis of Intron Splicing in Monocot Plants
RALPH SINIBALDI AND I. METTLER

trp Repressor, A Ligand-activated Regulatory Protein
RONALD L. SOMMERVILLE

Immunochemical Analyses of Nucleic Acids
DAVID STOLLAR

The Structure and Expressions of the Insulin-like Growth-factor Gene
LYDIA VILLA-KOMAROFF AND K. M. ROSEN

Contents

Structural Elements in RNA 131

Michael Chastain and Ignacio Tinoco, Jr.

Nuclear RNA-binding Proteins 179

Jack D. Keene and Charles C. Query

Amplification of DNA Sequences in Mammalian Cells 203

Joyce L. Hamlin, Tzeng-Horng Leu, James P. Vaughn, Chi Ma and Pieter A. Dijkwel

Molecular Structure and Transcriptional Regulation of the Salivary Gland Proline-Rich Protein Multigene Families

DON M. CARLSON,[1]
JIE ZHOU[2]
AND PAUL S. WRIGHT[3]

Department of Biochemistry and Biophysics
University of California–Davis
Davis, California 95616

The proline-rich proteins (PRPs) in mammalian salivary glands are encoded by tissue-specific multigene families whose members have diverged with respect to structure and regulation of expression. A common evolutionary origin of the PRP genes is evident from the extensive conservation of 5′-untranslated regions, coding sequences, and intron/exon organizations. The 42-nucleotide repeat unit CCA CCA CCA CCA GGA GGC CCA CAG CCG AGA CCC CCT CAA GGC has been proposed (*1*) as the ancestral unit, multiples of three bases probably being recruited into, or deleted from, this ancestral sequence during gene duplication. Gene conversion possibly was the mechanism of homogenization of the divergence of the internal repeats.

Two nonallelic mouse PRP genes (*MP2* and *M14*) have essentially identical sequences, with two major differences (*2*). *MP2* has 13 tandemly arranged 42-nucleotide repeats, whereas *M14* has 17 repeats. *M14* has an insertion by transposition of a two-kilobase member of the long, interspersed elements of repeated mouse DNA (LINE family) into intron I. The 5′-untranslated se-

[1] To whom correspondence may be addressed.

[2] Present address: Neurological Sciences Institute, Good Samaritan Hospital and Medical Center, Portland, Oregon 97209.

[3] Present address: Merrell Dow Pharmaceuticals, Inc., Cincinnati, Ohio 45215.

Progress in Nucleic Acid Research
and Molecular Biology, Vol. 41

quences and regions encoding the signal peptides of all PRP mRNAs, regardless of source, are nearly identical.

In another multigene family from rat submandibular glands that encodes contiguous repeat proteins (CRPs) or glutamic acid/glutamine-rich proteins (Glx-rich proteins), the 5′-untranslated sequences and the regions encoding the signal peptides of the mRNAs are 91% identical (nucleotides) and 92% identical (amino acids) to the PRP mRNAs (*3*, *4*). Two mRNA size-classes, each containing multiple PRP mRNAs, are transcripts from PRP gene families of mice (*5*), hamsters (*6*), rats (*7*), and humans (*8*). The CRP or Glx-rich multigene family also encodes two size-classes of mRNAs, and this multigene family has the same intron/exon organization as the mouse and rat PRP genes. Cell-free translations show some unusual differences in PRPs encoded by mRNAs from parotid glands of four mouse strains (BALB/cJ, DBA/2J, CD-1, and C57BL/6J) after isoproterenol treatment (*5*). Reasons for the variations of translation products in these mouse strains after induction of the PRP gene families are unknown.

Repeated administration of the β-agonist isoproterenol causes hypertrophy and hyperplasia of rat and mouse parotid and submandibular glands (*9*, *10*). The morphological changes are accompanied by a dramatic increase, or induction, in the synthesis of PRPs. Typically, these proteins contain 25–45% proline, 18–22% glycine, and 18–22% glutamine and glutamic acid. Aromatic and sulfur-containing amino acids are either very low in amount or absent. Generally, PRPs can be divided into acidic and basic groups, and both groups may be glycosylated and phosphorylated. PRPs may compose more than 70% of the protein in salivary gland extracts after treatment with isoproterenol.

All proteins derived from the nucleotide sequences of PRP cDNAs and PRP genes are characterized by four general regions: a signal peptide region, a transition region, a repeat region, and a carboxyl-terminal region (*11*). The apparent tissue-specific synthesis and the appearance of PRPs in saliva in such large quantities, either constitutive (as in humans) or induced by isoproterenol, suggest biological functions in the oral cavity and the gastrointestinal tract. Several functions, such as calcium binding, inhibition of hydroxylapatite formation, and formation of the dental-acquired pellicle, have been attributed to the human salivary PRPs (*12*). PRPs have an unusually high affinity for such multihydroxylated phenols as tannins; feeding tannins to rats and mice mimics the effects of isoproterenol on the parotid glands (*13*). The induction of PRP synthesis by dietary tannins clearly results in a protective response against the detrimental effects of the tannins (*13*).

Unlike mice and rats, hamsters do not respond to tannins in the diet by the induction of PRPs. Pronounced detrimental effects are observed in weanling hamsters specifically. When these animals are maintained on a 2%

tannin diet for 6 months, they fail to grow (*6*). Tannins are unusually toxic to weanling hamsters; an increase of tannin in the diet to 4% causes death to most animals within 3 days. The association of tannins with pathological problems, including carcinogenesis and hepatotoxicity, and the influences on growth and toxicity in hamsters, have led to the proposal that PRPs may act as a first line of defense against these multihydroxylated phenols (*13*).

This review focuses on the biochemistry and molecular biology of the salivary PRPs; it is not intended to be an overall or complete review of PRPs. To those who have contributed to the PRP literature and whose work is not mentioned, we apologize. Previous reviews are used for many references and studies.

I. Background[4]

Salivary glands of various animals synthesize, or can be induced to synthesize, a group of proteins unusually high in proline, the so-called proline-rich proteins (PRPs) (*12, 14–20*). These proteins collectively constitute the largest group of proteins in human salivary secretions, making up more than 70% of the secreted proteins (*12*). PRPs may be divided into acidic and basic groups, and members of each group may be phosphorylated or glycosylated, or both. These unusual proteins are constitutive in human saliva, but families of similar proteins are dramatically increased or induced in parotid and submandibular glands of rats, mice, and hamsters by isoproterenol treatment (*6, 18, 19, 21*). Profound morphological effects on rat parotid glands by isoproterenol treatment were first observed in 1961 (*9, 10*). Repeated pharmacological doses cause dramatic glandular hypertrophy (Fig. 1). The increase in DNA synthesis with isoproterenol treatment (*25, 26*) probably results mainly from polyploidy; by 4–5 days, more polyploid than diploid nuclei are seen (Fig. 2) (see *27* for a review on the regulation of salivary gland size and the effects of isoproterenol).

The dramatic accumulation of PRPs in the parotid glands of rats treated with isoproterenol was first described in 1974 (*16, 18, 28*). After 7–10 days of treatment (5 mg of isoproterenol per day), PRPs composed about 70% of the total soluble proteins in parotid gland extracts. Initially, an acidic PRP (pI =

[4] Reviews describing mainly the human PRP families are available (*12, 22, 23*). These unusual proteins were first observed in human saliva by Mandel, Thompson, and Ellison (*24*) and were first purified and characterized by Bennick and Connell (*14*) and by Oppenheim, Hay, and Franzblau (*15*). The genetics of this human multigene family were described in a review by Bennick (*23*). Other than for comparisons of the human cDNAs and multigene families, this review focuses primarily on the tissue-specific inducible multigene PRP families of mouse, rat, and hamster.

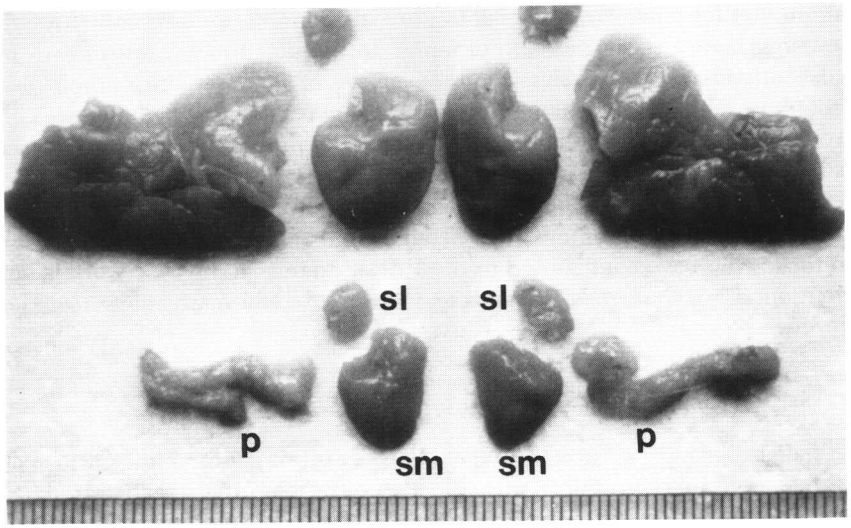

FIG. 1. Hypertrophic effects of isoproterenol treatment on rat salivary glands. Rats (150–200 g of body weight) were injected intraperitoneally with 5 mg of isoproterenol daily for 7 days. The parotid glands (p), submandibular glands (sm), and sublingual glands (sl) were removed from control (bottom) and isoproterenol-treated animals (top). No changes were noted for the sublingual glands, which secrete principally mucous glycoproteins. Parotid glands, which are serous secretors, showed a dramatic increase in weight of about 6- to 10-fold. Submandibular glands are of a mixed cell population and showed an intermediate response to isoproterenol.

4.5) was identified (Ipr-1A2), and this protein was phosphorylated and glycosylated (*16*, *18*, *19*).

Subsequently, six basic PRPs unusually high in proline (40–44%), glutamine plus glutamate (22–25%), and glycine (18–20%), containing varying amounts of lysine plus arginine (7–9%), were isolated and characterized (*18*, *19*). Aromatic and sulfur-containing amino acids were either absent or present in very low amounts. Therefore, PRPs have little or no absorbance at 280 nm. Neither hydroxylysine nor hydroxyproline is present and the treatment of these PRPs with purified prolyl hydroxylase failed to convert proline into hydroxyproline. The molecular weights of the basic proteins, from sedimentation equilibrium, ranged from 15,000 to 18,000, and that of PRP Ipr-1A2 was 25,000. A high MW_{app} (71,000) was observed following chromatography on Sephadex G-100, but the unusually high axial ratio (>25) of these proteins undoubtedly caused this value to be substantially overestimated. *S* values ranged from 1.1 to 1.4. Circular dichroism spectra showed no α-helical or polyproline conformations.

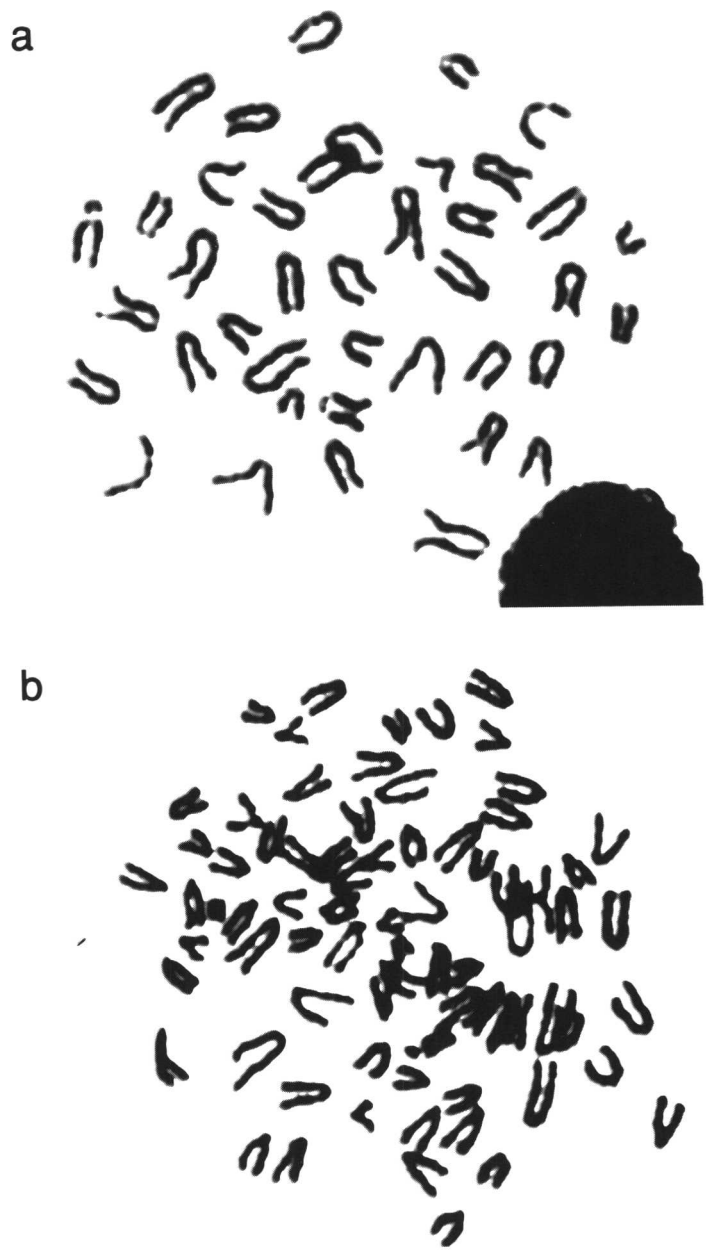

FIG. 2. Karyotypes of (a) a mouse bone marrow cell and (b) a mouse parotid gland cell. The chromosomal display of the mouse bone marrow cell showed the normal $2n$ (= 40) chromosomes after 2 days of isoproterenol treatment. The mouse parotid gland cells (>50% of the cells) showed $4n$ chromosomes after 2 days of isoproterenol treatment. (Courtesy of Christopher Bidwell.)

II. PRP mRNAs and Cell-free Translation Analysis

Studies by cell-free translation analysis using the reticulocyte lysate system and labeling with [^{3}H]proline or]^{35}S]methionine showed dramatic and definitive changes in the patterns of protein synthesis in parotid glands of isoproterenol-treated rats, and PRP mRNAs were highly elevated in the treated animals (*29*). There was very little synthesis of PRPs from poly(A)$^{+}$ RNAs from glands of control rats: poly(A)$^{+}$ RNAs from the glands of treated animals synthesized mainly PRPs; translation patterns with [^{3}H]proline and [^{35}S]methionine gave identical labeling patterns; and PRPs from cell-free translations were all precipitated by antibodies to PRPs. [^{35}S]Methionine was incorporated only into the initiation site, as determined by sequence analysis and by the fact that PRPs synthesized by tissue slices of parotid glands of isoproterenol-treated rats in the presence of [^{35}S]methionine contained no ^{35}S label. Because most PRPs are acid-soluble, a property first used in the purification procedures of rat submandibular gland PRPs (*30*), it is imperative that cell-free translation products be precipitated with a solution containing both trichloroacetic and phosphotungstic acids (*29*).

The induction of PRP mRNAs in the parotid and submandibular glands of both rats and mice by isoproterenol treatment has been demonstrated by Northern and dot-blot hybridizations (*21*). PRP mRNAs either are very low or are not detectable in the glands of untreated rats and mice. After 4–5 days of isoproterenol treatment, mRNAs encoding these unusual proteins compose over 50% of the total glandular mRNAs (*5*). For example, plasmid pRP25 does not hybridize with RNAs from control rats (Fig. 3A), but does hybridize with PRP mRNAs of two size-classes, ranging from 600 to 1100 bases, from isoproterenol-treated animals. These size ranges of mRNAs are consistent with all rat RNA preparations tested. The multiplicity of PRPs encoded by the PRP mRNAs from treated rats is evident from Fig. 3B, since about 12 PRPs were identified by cell-free translation analysis and immunoprecipitation.

The PRP cDNA insert of pUMP40 (*11*), prepared from mRNAs from BALB/cJ mice, has been tentatively identified as the transcript of the mouse PRP gene *MP2* (*1*). However, the nucleotide sequences of *MP2* and the PRP insert of pUMP40 showed only 98% homology (*1*). *MP2* was cloned from a genomic library prepared from chromosomal DNA from the CD-1 mouse strain. In an attempt to reconcile the heterologous regions and base differences between the CD-1 mouse gene *MP2* and the BALB/cJ mouse mRNA, we isolated mRNAs from four mouse strains.

Northern blots of total RNA from the parotid glands of mouse strains CD-1 and BALB/cJ and from strains DBA/2J and C57BL/6J, from both control and isoproterenol-treated mice, were probed with ^{32}P-labeled exon

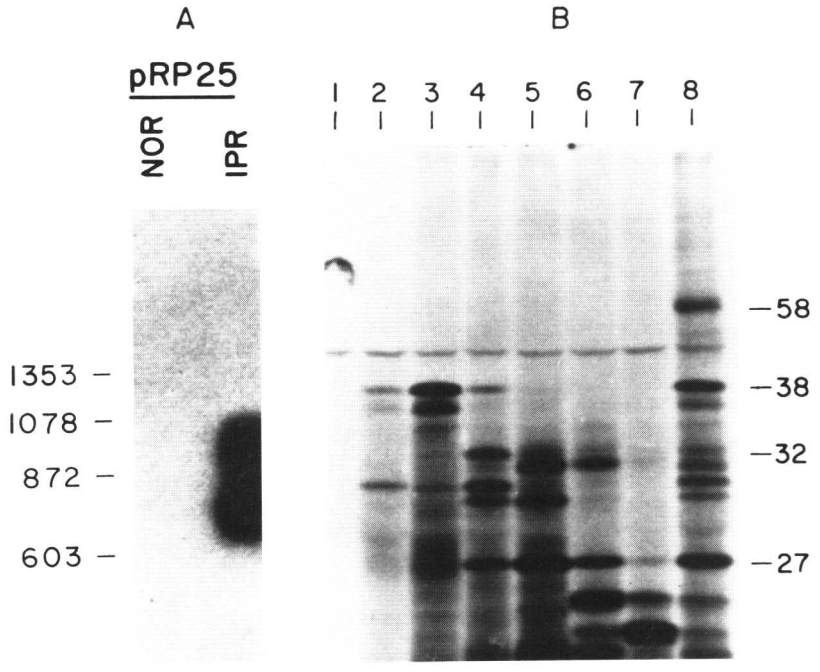

FIG. 3. Northern blot of parotid gland RNA from normal and isoproterenol-treated rats and cell-free translations of "sized" PRP mRNAs. (A) Parotid gland RNAs (10 μg) from normal and isoproterenol-treated rats were electrophoresed on a 1.5% agarose gel containing 5 m*M* methyl mercury hydroxide and transferred to nitrocellulose. The blot was probed with ^{32}P-labeled pRP25 (*11*). (B) RNA was isolated from a methyl mercury denaturing low-melting-point agarose gel and translated *in vitro* with [^{35}S]methionine. The translation products were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Lanes 1 and 8 show ^{35}S label incorporated in the absence of RNA and with total RNA from the parotid glands of isoproterenol-treated rats, respectively. Lanes 2–7 are the translation products obtained from RNA indicated in (A). Molecular-weight standards ($\times 10^{-3}$) are indicated at the right, and nucleotide standards are indicated at the left. [Reprinted with permission from the *Journal of Biological Chemistry* (5).]

IIb (see Fig. 10) of PRP gene *MP2* (5). Two major classes of PRP mRNAs were detected in the treated animals. RNA species of about 1050 and 1300 bases for BALB/cJ and DBA/2J mice and about 1100 and 1200 bases for CD-1 and C57BL/6J mice were observed. Cell-free translations of total RNA from these four mouse strains showed interesting and unusual differences in the PRPs synthesized (Fig. 4). Similar labeling patterns were observed with both [^{3}H]proline and [^{35}S]methionine. The amounts or levels of incorporation varied considerably between controls and treated animals, and α-