BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **228,** 141–147 (1996)
ARTICLE NO. 1629

The Primary Structures of Rat Ribosomal Proteins S3a (The V-Fos Transformation Effector) and of S3b

Yuen-Ling Chan, Joe Olvera, Veronica Paz, and Ira G. Wool[1]

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

Received September 15, 1996

The amino acid sequence of the rat 40S ribosomal subunit protein S3a was deduced from the sequence of nucleotides in two recombinant cDNAs and confirmed by the determination of the NH_2-terminal sequence by Edman degradation. Ribosomal protein S3a has 263 amino acids (the NH_2-terminal methionine is removed after translation of the mRNA) and the molecular weight is 29,794. The protein designated S3b has the same amino acid sequence as S3a except that it lacks the carboxyl-terminal 12 residues. We are unable to determine whether there are separate genes for S3a and S3b, or whether there is a single gene and alternate splicing of the precursor to yield separate mRNAs for S3a and S3b, or whether there is a single gene and a single mRNA whose translation yields S3a which is converted by proteolysis, either physiological or fortuitous, to S3b. The mRNA for S3a is about 1000 nucleotides in length. Hybridization of cDNA to digests of nuclear DNA suggests that there are 8–13 copies of the S3a gene. Rat ribosomal protein S3a is identical to the product of the rat *Fte-1* gene which encodes the V-Fos transformation effector; S3a is also related to the plant protein cyc07, which is encoded by a cell cycle S-phase specific gene.

The sequences of nucleotides and of amino acids in the constituent nucleic acids and proteins is required for the determination of the structure of ribosomes (1, 2). The purpose for the accumulation of this data is its use in arriving at a solution of the structure of the organelle; the structure is in turn needed for a coherent molecular account of the function of ribosomes in protein synthesis. The information on the sequences of amino acids may also help in understanding the evolution of ribosomes, in unraveling the function of the proteins, in defining the rules that govern the interaction of the proteins and the rRNAs, and in uncovering the amino acid sequences that direct the proteins to the nucleolus for assembly on nascent rRNA. As a part of this endeavor we report here the covalent structures of rat ribosomal proteins S3a and S3b; with these amino sequences the primary structure of the 79 proteins in mammalian ribosomes is completed.

S3a forms part of the binding site for the initiation factors eIF-2 (3, 4) and eIF-3 (5); in addition, S3a has been crosslinked to initiator tRNA (6), to synthetic mRNA (7, 8), and to poly(A)$^+$mRNA in polysomes (9). Thus, S3a is a component of the 40S ribosomal subunit domain involved in the initiation of protein synthesis.

S3a is also related to proteins involved in the regulation of growth and of the cell cycle.

METHODS

S3b was isolated from a group fraction (10) of rat liver 40S ribosomal subunit proteins by HPLC on a size exclusion column (Bio-Rad, Bio-Sil TSK-125) with a mobile phase of 0.1 M ammonium acetate, pH 4.1. The sequence of the NH_2-terminal 48 amino acids was determined by Edman degradation in an Applied Biosystems, Model 470A, automated gas phase sequencer.

[1] Correspondence to: Dr. Ira G. Wool, Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637. Fax: (312) 702-0439.

0006-291X/96 $18.00

The recombinant DNA procedures including the methods employed to determine the sequence of nucleotides in nucleic acids have been described or cited (11-13).

It was recognized from the identity of the amino acid sequences at the NH_2-termini of rat S3b and of human S3a (14), the latter kindly provided by Dr. A. Metspalu, that the proteins were related. A pair of deoxyoligonucleotide primers were designed on this assumption. The forward primer, a mixture of 24 oligodeoxynucleotides each 31 bases in length, encoded the first 8 residues, including the NH_2-terminal methionine, of rat S3b and had a *Hin*dIII linker; the reverse primer, a mixture of 36 oligodeoxynucleotides 27 bases in length, was complementary to the DNA encoding the 6 carboxyl-terminal amino acids and the termination codon in human S3a and had a *Bam*HI linker. The primers were used in the polymerase chain reaction with cDNA prepared from rat liver total poly(A)$^+$ mRNA as template. The amplified DNA was made radioactive by the random primer method (15) and was used to screen cDNA libraries for rat S3a and S3b.

The sequence of the carboxyl-terminal amino acids in S3b was determined with carboxypeptidase P. One nmol of S3b was dissolved in 100 μl of 0.1 M acetic acid and 0.02 nmol of carboxypeptidase P. The digestion reaction was at room temperature for 17 h with aliquots taken at 0, 6, and 17 h; the reaction was stopped by lyophilization of the reaction mixture. The identity of the amino acids released by carboxypeptidase P digestion was determined by derivitization with phenylisothiocyanate and analysis by HPLC on a PTH-C18 reverse phase column at 34 °C.

RESULTS AND DISCUSSION

The Sequences of Nucleotides in Recombinant cDNAs Encoding Rat Ribosomal Protein S3a

A cDNA library that had been constructed from regenerating rat liver poly(A)$^+$mRNA (11, 12) was screened with a probe designed to hybridize to clones encoding ribosomal protein S3a. The DNA from the plasmids of the clones that gave a positive hybridization signal was isolated and digested with restriction endonucleases; the cDNA inserts were either approximately 300 or 900 nucleotides in length. The sequences of nucleotides were determined in both strands of the cDNA inserts in two of the plasmids, pS3a-1 and pS3a-8. The insert in pS3a-8 has a 5′ noncoding sequence and an open reading frame but lacks a termination codon and a 3′ noncoding sequence; the insert in pS3a-1 on the other hand lacks a 5′ noncoding sequence but has an open reading frame, a 3′ noncoding sequence, and a long poly(A) tail (Fig. 1). The overlapping sequences in pS3a-1 and pS3a-8 are identical and hence they are likely to be derived from the same gene. Together the two cDNAs encode the entire protein and for convenience we refer to them as pS3a-1,8.

The cDNA insert in pS3a-1,8 has a 5′ noncoding sequence of 24 nucleotides, a single open reading frame of 795, a 3′ noncoding region of 43, and a long poly(A) tail (Fig. 1). The open reading frame starts at an ATG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 793; it encodes 264 amino acids. In the other two reading frames the sequence is interrupted by termination codons. The initiation codon occurs in the context ACC*ATG*G which is the same as the consensus sequence (16). The hexamer, AATAAA, that directs post-transcriptional cleavage-polyadenylation of the 3′ end of the precursor of the mRNA (17) is at positions 819-824, 14 nucleotides upstream of the start of the poly(A) stretch.

There are six pyrimidines, CCTTTT, at the 5′ terminus of the cDNA. Pyrimidine sequences are found at the immediate 5′ end of most, if not all, eukaryotic ribosomal protein mRNAs; the presence of the polypyrimidine stretch can be taken to confirm that the mRNA includes the region where transcription started (1). This polypyrimidine motif may be a promoter that binds a *trans*-acting factor that accounts for the regulation (and perhaps the coordination) of the translation of ribosomal protein mRNAs (8).

The Relationship of S3a to S3b

The protein encoded in the open reading frame of pS3a-1,8 has 99% amino acid identity with the sequence of human ribosomal protein S3a and exactly reproduces the sequence in a rat S3a fragment, residues 24 through 49, determined by Edman degradation. This would appear to establish that pS3a-1,8 encodes S3a. However, the sequence of the NH_2-terminal 48

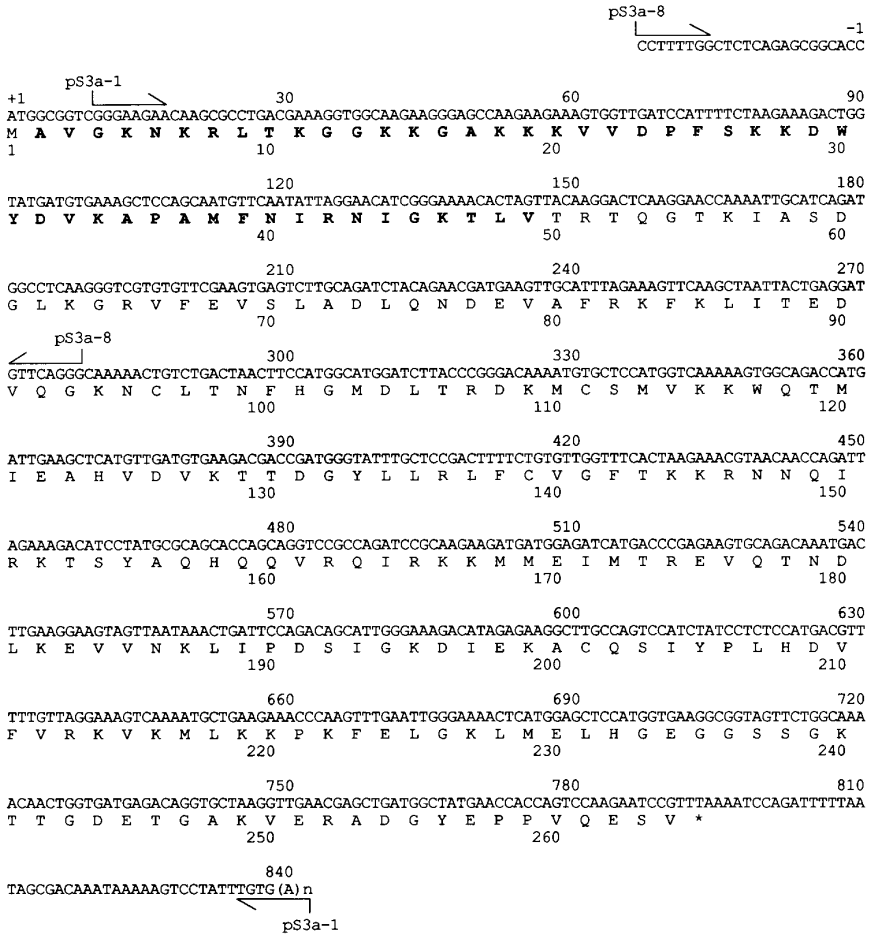

FIG. 1. The sequences of nucleotides in the cDNA inserts in pS3a-1 and pS3a-8 and the amino acid sequence encoded in the open reading frame. The positions of the nucleotides in the cDNA are above the residues; the positions of the amino acids in rat ribosomal protein S3a are below the residues. The initial and terminal nucleotides in pS3a-1 and pS3a-8 are designated by the vertical portion of the bent arrows. The amino acids in S3a determined from the protein by Edman degradation are in boldface.

amino acids of S3b determined from the protein is also the same as that encoded in pS3a-1,8 which raises the question of the relationship of S3a and S3b.

The first question is whether there are separate genes for S3a and S3b. This is not easy to answer, since there are 8-13 copies of the S3a gene (see later) all but one of which are presumed to be retroposon pseudogenes; and because a probe specific for S3b cannot be designed. Experiments were based on the assumption that if there are separate S3a and S3b genes they are likely to have different introns. A series of forward and reverse degenerate primers, in the aggregate spanning the entire coding sequence of the S3a cDNA, were prepared; the pairs were used in PCR reactions with genomic DNA as template and the products were cloned and the sequence of nucleotides determined. We found no evidence for introns differing in size nor in nucleotide sequence. Although, these negative experiments are not conclusive, they provide no evidence for separate S3a and S3b genes.

The position of the S3b spot on two-dimensional gels, i.e. to the southeast of S3a, suggests that S3b is smaller and more basic than S3a. Indeed, the simplest explanation of the relationship

of S3a and S3b is that there is a single gene and a single transcript that encode S3a and that after translation of the S3a mRNA a portion of the product is converted by proteolysis to S3b. This is the way S5 and S5a are related as was shown by translation of the S5 mRNA in a rabbit reticulocyte lysate and observing synthesis of both S5 and S5a (18). The possibility that proteolytic processing of S3a to from S3b is a physiological process, perhaps designed to generate two products form a single gene, is strengthened by the observation that S3b is always present in two-dimensional gels of rat liver ribosomal proteins in amounts, judged from the intensity of staining, roughly equal to that of S3a and to the intensities of the other ribosomal proteins as well. However, S3b is not found in preparations of ribosomal proteins from rat adrenal glands (19), nor from rabbit reticulocytes (20), nor from rabbit liver (20).

S5a lacks the five carboxyl-terminal residues of S5; cleavage is of a glutamylthreonyl peptide bond (18). S3a has a glutamylthreonyl dipeptide 19 residues from its carboxyl terminus (positions 245 and 246). The NH_2 terminus of S3a also has an aspartylproline dipeptide (positions 23 and 24) which is a common site of proteolysis in weak acid (21, 22), but this is unlikely to be where S3a is cleaved since the first 48 residues of S3a and S3b are identical. To assist in resolving the relationship, separate cDNAs were prepared: the first had the complete S3a open reading frame (designated S3a); the second the S3a coding sequence lacking the nucleotides for the NH_2-terminal 23 amino acids (S3a-DP); and the third the S3a open reading frame lacking the nucleotides encoding the carboxyl-terminal 19 amino acids (S3a-ET). The cDNAs were generated in the polymerase chain reaction using pS3a-1 as template and the amplified cDNAs were subcloned in pGEM2 (23). The cDNAs were transcribed *in vitro* with T7 RNA polymerase and the RNA transcripts were translated in a nuclease-treated rabbit reticulocyte lysate containing [^{35}S]-methionine (23). The radioactive products of the translation reactions were analyzed by electrophoresis in two-dimensional polyacrylamide-urea gels (Fig. 2).

The transcription of the S3a cDNA (encoding the entire open reading frame) and translation of the RNA transcript yielded a single major protein that comigrated with authentic S3a (Fig. 2A); there was also a minor protein but it was not S3b. Thus, there is no conversion of S3a to S3b in a rabbit reticulocyte lysate; but rabbit reticulocyte ribosomes do not have S3b (20).

Transcription and translation of S3a-DP cDNA and translation of the RNA gave a single product of the predicted size and charge (Fig. 2B). Transcription of S3a-ET cDNA and translation of the RNA gave a major product that coincided with S3b (Fig. 2C). When the translation reactions were in the presence of the serine protease inhibitor phenylmethylsulfonyl fluoride the amount of minor product was reduced (results not shown).

There is another possibility: that S3b is generated by alternate splicing of a single S3a pre-mRNA and that the processing is both species (rat) and tissue (liver) specific. Inasmuch as we observe only a single band on Northern hybridization with S3a cDNA (see later) this would require that both mRNAs be approximately the same size which, of course, is possible. Primers specific for an appropriate region of the S3a cDNA were prepared so that PCR could be done with a population of cDNAs prepared from poly(A)$^+$mRNA; but products differing in size were not found. Thus, there was no evidence for alternate splicing.

It is also possible that S3b is generated from post-transcriptionally edited S3a mRNA. A conversion of a cytosine to a uridine at position 754 near the 3′ end of the S3a mRNA would create a termination codon and might result in the formation of S3b due to an early termination of translation of the edited S3a mRNA; the protein would lack the carboxyl-terminal 13 residues. We have tested this possibility by scanning the region of the S3a mRNA population from nucleotides 720 to 780 for a cytosine to uridine conversion using primer extension and total poly(A)$^+$mRNA as templates and have found none.

In the absence of evidence to support either separate S3a and S3b genes, or of alternate splicing of mRNA, or of editing of the S3a mRNA, or of specific post-translational processing

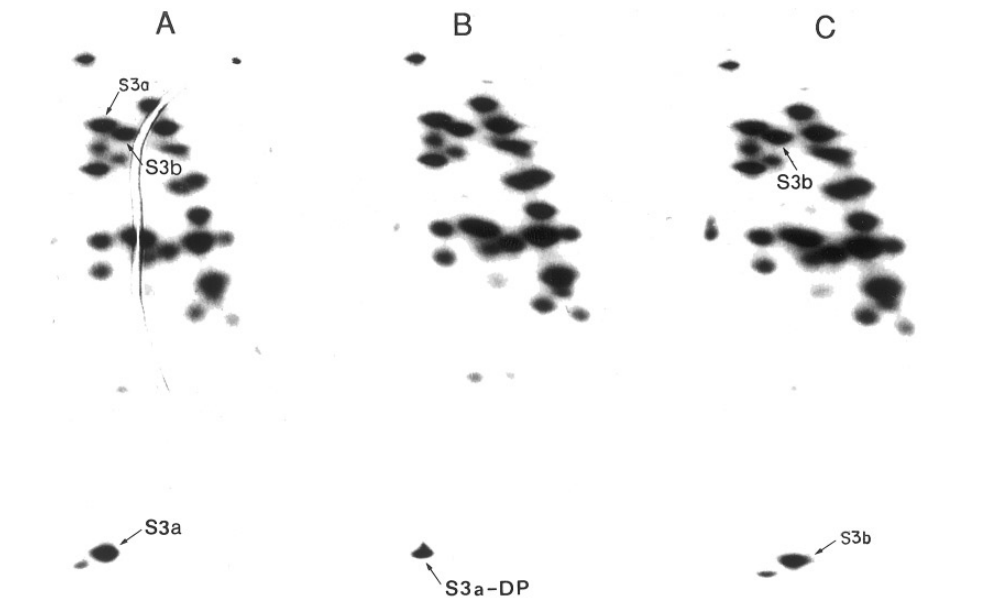

FIG. 2. Electrophoresis in polyacrylamide gels of the products of translation of the RNAs from the transcription of S3a cDNAs. A rabbit reticulocyte lysate (50 μl) was incubated with [^{35}S]methionine and the RNA transcript (1 μg) from the transcription of a cDNA insert in pGEM2 that encodes either: (A) the complete S3a open reading frame; (B) the S3a coding sequence lacking the nucleotides for the NH_2-terminal 23 amino acids; or (C) the S3a open reading frame lacking the nucleotides encoding the carboxyl-terminal 19 amino acids. A sample (15 μl) of the lysate containing the products of the translation reaction was extracted with 67% acetic acid and the protein was precipitated with 90% acetone; this sample was supplemented with 80 μg of 40S ribosomal subunit protein and analyzed by electrophoresis in two dimensions in polyacrylamide gels containing urea. The gels at the top were stained with Coomassie Brilliant Blue; at the bottom are autoradiographs of the stained gels that were enhanced by fluorography.

of S3a, we favor by exclusion the interpretation that S3b is the adventitious product of the proteolysis of S3a. The lingering reservation derives from the observation, recounted above, of the apparent stoichiometry of S3b with S3a and with other 40S ribosomal subunit proteins.

The Primary Structure of Rat Ribosomal Protein S3a

The identity of the ribosomal protein encoded in pS3a-1,8 was established by the correspondence of the amino acid sequence encoded in the cDNA with residues 24 through 49 in S3a determined by Edman degradation; by the congruence of the amino acid sequence encoded in pS3a-1,8 and that of human S3a (14); and by the observation that the transcription of the cDNA and translation of the RNA transcript give a product that migrates in two-dimensional gels to the same position as authentic S3a (Fig. 2A). The molecular weight of rat ribosomal protein S3a, calculated from the sequence of amino acids deduced from pS3a-1,8, is 29,925; however, the NH_2-terminal methionine encoded in the S3a mRNA is removed after translation. Thus, the mature processed S3a has 263 residues and the molecular weight is 29,794.

The carboxyl terminal amino acids of S3b was determined by carboxypeptidase P digestion. The enzyme released amino acids V, K, E, R, and A; the amino acid sequence at positions 248-252 in S3a is AKVER. This result suggests that the carboxyl-terminal amino acid of S3b corresponds to residue 252 of S3a and that S3b has the same primary structure as S3a except that it lacks 12 amino acids at the carboxyl terminus. We note that a transcript from a S3a-ET cDNA which encodes S3a lacking the carboxyl-terminal 19 residues when translated in a

reticulocyte lysate yields a protein that migrates on electrophoresis to the same position as authentic S3b (Fig. 2).

Protein S3a has an excess of basic residues (14 arginyl, 38 lysyl, and 5 histidyl) over acidic ones (17 aspartyl and 15 glutamyl). The total number of charged residues, 84 of 263 (32%), is high. The basic amino acids, as is typical of ribosomal proteins (2), tend to be clustered; there are, for example, seven sites where there are either three consecutive basic residues or three in a sequence of four. There are 90 hydrophobic amino acids (34%). S3a has 10 methionines, an unusually large number for a ribosomal protein.

The Number of Copies of the S3a Gene

The cDNA insert in pS3a-1 was made radioactive and used to probe separate digests of rat liver DNA made with the restriction endonucleases *Bam*HI, or *Eco*RI, or *Hind*III (12). The number of hybridization bands suggests that there are 8-13 copies of the S3a gene (data not shown). Many other mammalian ribosomal protein genes have been found to be present in multiple copies (cf. (2) for references and discussion). However, in no instance has it been shown that more than one of the genes is functional; the presumption is that the other copies are retroposon pseudogenes.

The Size of the mRNA Encoding Rat Ribosomal Protein S3a

To determine the size of the mRNA coding for S3a, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pS3a-1 cDNA. One band of about 1,000 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat S3a with Ribosomal Proteins from Other Species

The sequence of amino acids in rat ribosomal protein S3a was compared, using the computer programs RELATE and ALIGN (24), to the sequences of amino acids in more than 1,400 ribosomal proteins contained in a library that we have compiled; in addition, the program TFASTA (25) was used to search the EMBL DNA data base for nonribosomal proteins related to S3a. The ribosomal protein most closely related to rat S3a is human S3a (14); there are 262 identities in 264 possible matches (99%). Rat S3a is also related to other eukaryotic ribosomal protein; however, no related archaebacterial or eubacterial ribosomal protein has been found.

The Identification of the v-Fos Transformation Effector Protein and the Plant cyc07 Gene Product

The viral p55 *v-fos* oncogene and its cellular homolog, *c-fos,* encode nuclear phosphoproteins that can form heterodimeric complexes with the product of the *c-jun* protooncogene (26). Heterodimers of Jun with c- or v-Fos are regulators of transcription and constitutive overexpression of either the *v-fos* or the *c-fos* gene causes cell transformation (27). Revertants of *v-fos*-transformed rat cells were isolated by disruption of a *v-fos* transformation effector gene, *Fte*-1, and a cDNA for this effector protein was isolated and the sequence of nucleotides determined (27). The protein encoded in the cDNA, i.e. the rat v-Fos transformation effector is identical to that of rat S3a; moreover, human S3a is identical to human Fte-1 (28). It follows that ribosomal protein S3a modulates Fos-induced alterations in gene expression and thereby affects the development of the neoplastic phenotype (27). Whether S3a does this as a component of, or apart from, the ribosome has not been determined.

A protein from the plant, *Catharanthus roseus,* the periwinkle, that is encoded by a gene designated *cyc07* (29) is related to rat S3a. In an alignment of the amino acid sequences of *C. roseus* cyc07 (30) and rat S3a there are 163 identities in 234 possible matches (70%). A

cyc07 cDNA was isolated by differential screening of a library prepared from synchronized plant cells in S-phase (29). Maximal and transient expression of the *cyc*07 gene was at the G_1-S phase boundary. Cyc07 may have a role in cell proliferation especially in entry into, or progression of, the S-phase of the cell cycle (29); this behavior is similar to mammalian genes that play a role in the regulation of the cell cycle and in DNA replication (31-33).

The participation in the regulation of transcription and of the cell cycle by S3a or its homologs are examples of extraribosomal functions of a ribosomal protein of which there are many others (34-35).

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant GM 21769. We are grateful to Anton Glück for assistance with the computer analyses. The sequence has been deposited in the EMBL/GenBank/DDBJ Nucleotide Sequence Database; the Accession No. is X75161.

REFERENCES

1. Wool, I. G. (1979) *Ann. Rev. Biochem.* **48,** 719–754.
2. Wool, I. G., Chan, Y. L., and Glück, A. (1995) *Biochem. Cell Biol.* **73,** 933–947.
3. Bommer, U. A., Stahl, J., Henske, A., Lutsch, G., and Bielka, H. (1988) *FEBS Lett.* **233,** 114–118.
4. Westermann, P., Heumann, W., Bommer, U. A., Bielka, H., Nygård, O., and Hultin, T. (1979) *FEBS Lett.* **97,** 101–104.
5. Westermann, P., and Nygård, O. (1983) *Biochim. Biophys. Acta* **741,** 103–108.
6. Westermann, P., Nygård, O., and Bielka, H. (1981) *Nucleic Acids Res.* **9,** 2387–2396.
7. Westermann, P., and Nygård, O. (1984) *Nucleic Acids Res.* **12,** 8887–8897.
8. Stahl, J., and Kobets, N. D. (1981) *FEBS Lett.* **123,** 269–272.
9. Takahashi, Y., and Ogata, K. (1981) *J. Biochem.* **90,** 1549–1552.
10. Collatz, E., Lin, A., Stöffler, G., Tsurugi, K., and Wool, I. G. (1976) *J. Biol. Chem.* **251,** 1808–1816.
11. Chan, Y. L., Lin, A., McNally, J., and Wool, I. G. (1987) *J. Biol. Chem.* **262,** 12879–12886.
12. Chan, Y. L., and Wool, I. G. (1988) *J. Biol. Chem.* **263,** 2891–2896.
13. Glück, A., Chan, Y. L., Lin, A., and Wool, I. G. (1989) *Eur. J. Biochem.* **182,** 105–109.
14. Metspalu, A., Rebane, A., Hoth, S., Pooga, M., Stahl, J., and Kruppa, J. (1992) *Gene* **119,** 313–316.
15. Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
16. Kozak, M. (1986) *Cell* **44,** 283–292.
17. Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263,** 211–214.
18. Kuwano, Y., Olvera, J., and Wool, I. G. (1992) *J. Biol. Chem.* **267,** 25304–25308.
19. Madjar, J. J., Arpin, M., Buisson, M., and Reboud, J. P. (1979) *Mol. Gen. Genet.* **171,** 121–134.
20. Madjar, M., and Traut, R. R. (1980) *Mol. Gen. Genet.* **179,** 89–101.
21. Landon, M. (1977) *Methods Enzymol.* **47,** 145–149.
22. Suzuki, K., Olvera, J., and Wool, I. G. (1990) *Biochem. Biophys. Res. Comm.* **172,** 35–41.
23. Chan, Y. L., Devi, K. R. G., Olvera, J., and Wool, I. G. (1990) *Arch. Biochem. Biophys.* **283,** 546–550.
24. Dayhoff, M. O. (1978) *in* Atlas of Protein Sequence and Structure (Dayhoff, M. O., Ed.), Vol. 5, Suppl. 3, pp. 1–8, National Biomedical Research Foundation, Washington, DC.
25. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85,** 2444–2448.
26. Ransone, L. J., and Verma, I. M. (1990) *Annu. Rev. Cell. Biol.* **6,** 539–?.
27. Kho, C. J., and Zarbl, H. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89,** 2200–2204.
28. EMBL accession number M84711.
29. Kodama, H., Ito, M., Hattori, T., Nakamura, K., and Komanine, A. (1991) *Plant Physiol.* **95,** 406–411.
30. Ito, M., Kodama, H., and Komamine, A. (1991) *Plant J.* **1,** 141–148.
31. Ayusawa, D., Shimizu, K., Koyama, H., Kaneda, S., Takeishi, K., and Seno, T. (1986) *J. Mol. Biol.* **190,** 559–567.
32. Lee, M. G., Norbury, C. J., Spurr, N. K., Nurse, P. (1988) *Nature* **333,** 676–679.
33. Rickles, R., Marashi, F., Sierra, F., Clark, S., Wells, Stein, J., Stein, G. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 749, 753.
34. Wool, I. G. (1993) *in* The Translational Apparatus (Nierhaus, K. H., Franceschi, F., Subramanian, A. R., Erdmann, V. A., and Wittmann-Liebold, B., Eds.), pp. 727–737.
35. Wool, I. G. (1995) *in* Group I Introns and Ribosomal RNA (Schroeder, R., and Green, R., Eds.), Lande, Austin, in press.