The Primary Structure of Rat Ribosomal Protein L5

A COMPARISON OF THE SEQUENCE OF AMINO ACIDS IN THE PROTEINS THAT INTERACT WITH 5 S rRNA

Yuen-Ling Chan, Alan Lin, James McNally, and Ira G. Wool
From the Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

(Received for publication, March 31, 1987)

The covalent structure of rat ribosomal protein L5, which associates with 5 S rRNA in the organelle, was deduced from the sequence of nucleotides in a recombinant cDNA (pL5-6-4) and confirmed from the sequences of amino acids in portions of the protein. Ribosomal protein L5, encoded by pL5-6-4, contains 296 amino acids and has a molecular weight of 34.298. However, a second recombinant cDNA, pL5-8-5, encodes a protein with an additional methionyl residue at position 236 and may be the product of a second active L5 gene. Rat L5 is homologous to yeast 5 S 13L and to Halobacterium cutirubrum HL13, proteins that also bind to 5 S rRNA. No significant structural similarity, however, was found between rat L5 and other 5 S rRNA-binding proteins; nor with a second H. cutirubrum protein HL19, nor the Escherichia coli ribosomal proteins, L5, L18, or L25, nor the Xenopus laevis transcription factor IIIA. H. cutirubrum HL19, however, has structural identity with E. coli L5 and seems to be related to yeast 5 LS and, hence, may be an evolutionary link between the prokaryotic and eukaryotic 5 S rRNA-binding proteins. A group of ribosomal proteins not known to be associated with 5 S rRNA are also related to L5. They include: rat L39, Euglena gracilis chloroplast S7, Saccharomyces cerevisiae L31 and L46, Homo sapiens L32 and, perhaps, several others as well. There is an especially close interrelationship between rat L5, rat L39, yeast L46, human L32, and mouse L32. These results, and others, suggest that ribosomal proteins form an extended family and that L5 may contain in its structure traces of this affinity.

It is a truism that what is required to solve the structure of the ribosome is information on the chemistry of the constituents and knowledge of how the proteins and nucleic acids interact. Ribosomal 5 S RNA and the proteins that associate with it provide a convenient and tractable means for studying the latter problem since the nucleic acid is small and only a few proteins bind to it. Moreover, ribonucleoprotein complexes containing 5 S RNA appear to be discrete subparticles of the organelle (1–8). The complexes can be reconstituted independently of other ribosomal components and they can be removed from, and added back to, the ribosome as a unit. The hope is that the rules that govern the interactions in this simpler case will be easier to determine and will be generally applicable.

What is essential for this analysis is a means of determining with some precision the region of 5 S rRNA that is in contact with the proteins that bind specifically to the nucleic acid. For this purpose, use has been made of the cytotoxic ribonuclease α-sarcin (9). The toxin has proven nearly ideal for footprinting RNA, something that had been herebefore both tedious and difficult. Thus, the binding site on 5 S rRNA for the Escherichia coli ribosomal proteins L5, L18, and L25 (10), for the rat ribosomal protein L5 (11), and for the Xenopus laevis transcription factor IIIA (12) have been defined precisely. A second necessity for this analysis is the structure of the proteins that associate with 5 S rRNA and, ultimately, the identity and the conformation of the amino acids in the portion of the protein in contact with the binding site on the nucleic acid. We report here, as a part of the undertaking, the primary structure of rat ribosomal protein L5 which associates with 5 S rRNA in the organelle. The sequence of amino acids in L5 was inferred from the sequence of nucleotides in a recombinant cDNA and confirmed by sequencing portions of the protein. A comparison has been made of the sequence of amino acids in rat L5 and in the other proteins that bind to 5 S rRNA. The purpose of this search was to determine if there are common conserved structural features in the several proteins that might lead to the identification of the nucleic acid binding site.

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein L5—A cDNA library of 30,000 independent transformants was constructed from poly(A)+ mRNA prepared from regenerating rat liver. The library was screened for clones that hybridized to an oligonucleotide probe that was synthesized to be complementary to the sequence of nucleotides predicted to be present in the portion of the mRNA that encoded 9 consecutive amino acids near the NH2-terminus of rat ribosomal protein L5 (Fig. 1). A random selection of colonies from the original cDNA library were screened with the L5 probe; 15 gave a positive hybridization signal. The DNA from the plasmids of the 15 transformants was isolated, digested with restriction enzymes, and analyzed by gel electrophoresis. These clones had inserts that...
ranged in length from 0.35 to 1.1 kilobases and Southern blot hybridization with the oligonucleotide probe confirmed that all of the inserts contained cDNA for L5. The anticipated length of the L5 coding sequence, calculated from the molecular weight of the protein, is 885 nucleotides. For this reason, the 2 clones with the largest inserts (1.1 and 1.0 kilobases, designated pL5-6-4 and pL5-8-5, respectively) were selected and the nucleotide sequence determined. Nucleotide sequences from both strands of the DNA, and overlapping sequences for each restriction site, were obtained (Fig. 5).

The cDNA insert in pL5-6-4 contains 1102 nucleotides and includes the homopolymer linkers, the 5' and 3' noncoding sequences, and the nucleotide sequence determined. Nucleotide sequences from both strands of the DNA, and overlapping sequences for each restriction site, were obtained (Fig. 5).

The sequence of nucleotides in the cDNA insert in pL5-6-4 differs from pL5-8-5 in two respects. First, the former (pL5-6-4) lacks the codons for the two carboxyl-terminal amino acids, the 3' non-transcribed region, and the poly(A) tail. Second, and more importantly, pL5-8-5 encodes an additional methionine at position 236. Thus, the sequence in pL5-8-5 is -Pro-Asp-Met-Glu-Glu-Met--; whereas, the sequence in pL5-6-4 is -Pro-Asp-Met-Met-Glu-Glu-Met-. It is, of course, conceivable that the additional methionine codon was inserted in pL5-8-5 as the result of a copying error (i.e. the enzyme stuttered) during the synthesis of either the first or second strand of the cDNA prior to construction of the library. This is conceivable but not likely since the enzymes, reverse transcriptase and E. coli polymerase I, are not known to have trouble copying ATG sequences (polymerase I sometimes makes errors with G or C repeats and reverse transcriptase with A or T repeats); moreover, the error would need to be an insertion of exactly three nucleotides. A second possibility is that there is a single active gene for ribosomal protein L5 but that there are two species of mRNA, perhaps, as a result of a splicing error during processing of the precursor. This possibility would be more plausible if the methionines in question were at an exon-intron junction. Finally, one can imagine that there are two active L5 genes and that they differ only in having 1 or 2 methionyl residues in the relevant region. Unfortunately, we cannot decide between these possibilities from the data on hand since the number of methionyls in L5 is not sufficiently precise. However, there is indirect evidence for 2 methionines and it comes from an alignment of the NH2-terminal regions of rat L5 and the homologous yeast protein YL3 (7). Although, the yeast
protein contains neither 1 nor 2 methionines at this position, 2 residues rather than 1 are needed to properly align the surrounding region (cf. Fig. 7 and the discussion later). This does not mitigate against the possibility that there are two different mRNAs.

Preliminary experiments suggest that the difference between pL5-8-5 and pL5-6-4 is not the reflection of a cloning artifact, but rather that there are two species of mRNA for L5. We synthesized two oligodeoxynucleotides (43-mers) complementary to the region in question. The first encodes 1 methionine as in pL5-6-4 and is referred to as L5(M1); the second has 2 methionine codons as in pL5-8-5 and is referred to as L5(M2). L5(M1) has 22 nucleotides on the 5' side of the methionine codon and 18 on the 3' side; L5(M2) has 22 nucleotides on the 5' side of the 2 methionines and 15 on the 3' side. The two probes were hybridized at 40 °C to poly(A)+ mRNA fixed to nylon filters and then washed at 50, 68, 77, and 85 °C; 50, 68, and 77 °C are the melting temperatures (Tm) for hybrids having 11, 5, or 1 mismatches, whereas, at 85 °C all hybrids will melt. The hybrids formed with L5(M1) and L5(M2) were stable at 50, 68, and 77 °C; of course, neither was stable at 85 °C. At the conclusion of the experiment the probes were rehybridized to the filters containing mRNA at 40 °C and washed at 50 °C. Since rehybridization of the probes was as great as in the first cycle we conclude that little or no mRNA was lost from the filters even at 85 °C. Inasmuch as both probes bound to the poly(A)+ mRNA at 77 °C where even one mismatch should have caused a hybrid to melt it is likely that there are two populations of mRNAs for L5. It now needs to be determined whether there are two functional genes for L5 whose only difference is whether they encode 1 or 2 methionines near position 236, as seems probable to us, or whether a splicing mishap occurs near this position during processing of the primary transcript.

The Primary Structure of Rat Ribosomal Protein L5—The reading frame in pL5-6-4 flanked by initiation and termination codons encodes a protein of 296 amino acids (Fig. 6). This protein was identified as rat ribosomal protein L5 in the following manner. The recombinant cDNA clone (pL5-6-4) was selected using an oligonucleotide probe that was complementary to the codons for a sequence of 9 amino acids near the NH2 terminus of L5 (Fig. 1). Moreover, the amino acid sequence derived from the sequence of nucleotides in the cDNA corresponded exactly to the NH2-terminal 18 residues of L5 determined directly from the protein (Figs. 1 and 6), to the carboxy-terminal 3 residues determined with carboxypeptidase (Fig. 4), and to a pentapeptide (-Arg-Phe-Pro-Gly-Tyr-) derived from L5 (Fig. 6; positions 179-183 in the amino acid sequence and 616-630 in the nucleotide sequence). The amino acid composition inferred from the cDNA is, in general, similar to that previously (29) derived from an hydrolysate of purified L5 (Table 1). There are, nonetheless, discrepancies (for example, in the number of glycyl, histidyl, isoleucyl, leucyl, methionyl, and tyrosyl residues) that we are at a loss to account for, other than to attribute them to experimental error in the original amino acid analysis.

The molecular weight of rat ribosomal protein L5, calculated from the sequence of amino acids deduced from the nucleotide sequence of pL5-6-4, is 34,298, close to that of 32,500 estimated from the migration of the purified protein on sodium dodecyl sulfate gels (29). It is to be noted that since pL-5-8-5 encodes an additional methionine the number of residues in L5 could be 297 and the molecular weight 34,447. Moreover, the NH2-terminal residue in L5 is glycyl (Fig. 1); thus, the initial methionine is removed and the number of amino acids in the processed protein is 295 or 296 and the molecular weight is 34,149 or 34,298, or both depending on whether there are 1 or 2 methionyls near position 236 and whether there are one or two distinct varieties of L5 in the cell.

Protein L5 has an excess of basic (26 arginyl, 33 lysyl, and 5 histidyl) over acidic (15 aspartyl and 20 glutamyl) residues; a total of 64 or 22% of the former and 35 or 12% of the latter. The protein has a strikingly large number of tyrosyl and of methionyl residues, 21 and 10 or 11, respectively. It needs to be remarked that basic amino acids and tyrosine are frequently involved in binding to nucleic acids (quoted in Ref. 46). There are 4 cysteinyl residues (at positions 61, 76, 100, and 144) but whether they are linked in disulfide bridges is not known. The basic and acidic residues are not uniformly or randomly distributed in L5. There are 3 groups of 4 consecutive basic amino acids (at positions 21-24, 195-198, and 261-264) and two regions where they are concentrated. The first is of 20 basic residues in a stretch of 53 near the NH2 terminus (positions 5-58) and the second, near the carboxyl terminus (positions 193-283), is of 27 residues in a sequence of 89. The clustering of basic amino acids in eukaryotic ribosomal proteins has been remarked on before (47, 48) but the significance of the observation, if, indeed, there is any, is not known. There is, perhaps, one region where acidic residues are frequent; at position 184-237, 13 of 53 of the amino acids are acidic.

The Interaction of Ribosomal Proteins with 5 S rRNA—A purpose in determining the sequence of amino acids in L5 was to facilitate an analysis of the nature of the interaction of the ribosomal protein with 5 S rRNA. There is a great deal of data concerning the structure of the regions of 5 S rRNA that interact with proteins (10-12, 49-51). The comparison ranges from the proteins that associate with the nucleic acid in E. coli (10, 50), to those in Helobacterium cutirubrum (52), in Saccharomyces cerevisiae (53), and in rat (11). What has been gleaned from these studies is the rather strong impression of the conservation of the structure of the attachment site for ribosomal proteins on 5 S rRNA (cf. Fig. 8). The similarity is all the more remarkable inasmuch as the number of proteins with an affinity for 5 S rRNA varies. For example, one can affect the release of discrete ribonucleoprotein complexes from the large subunit of ribosomes that contain 5 S rRNA and one to three ribosomal proteins. In the ribonucleoprotein particle from E. coli ribosomes (6) there are three proteins (L5, L18, and L25), in the H. cutirubrum complex (7) there are two proteins (HI13 and HL19), and in the assembly from yeast (YL3) (8) or rat (L5) (1, 5) there is only one. It is curious that the number of ribosomal proteins in primary contact with 5 S rRNA has decreased from three in prokaryotic, to two in archaeobacterial, to one in eukaryotic ribosomes, although, there have been only modest changes in the structure of 5 S rRNA and in the attachment sites for the proteins (cf. Fig. 8). One explanation is that a small number of amino acid domains (each of moderate size) are in contact with the nucleic acid and that these domains can occur in one, two, or three proteins.

Comparison of the Sequence of Amino Acids in the Proteins That Bind to 5 S rRNA—We sought to determine if there had been conservation of the nucleic acid-binding domain in proteins that associate with 5 S rRNA. The first comparison was of rat ribosomal protein L5 and Xenopus transcription factor IIIA. Although it is not a ribosomal protein, the latter forms a complex with 5 S rRNA (54-56). The binding of factor IIIA to an intragenic control region is necessary for the initiation of transcription of 5 S rRNA genes by RNA polymerase III (57-60). During oogenesis in Xenopus factor IIIA also binds
to the transcript, i.e. to 5 S rRNA, to form a ribonucleoprotein particle (54–56) that is part of an inventory that is used later for ribosome biogenesis (61, 62). Thus, transcription factor IIIA binds to 5 S rDNA and to 5 S rRNA. The sequence of amino acids in factor IIIA was deduced by Ginsberg et al. (31) from the sequence of nucleotides in a recombinant cDNA. However, it was Miller et al. (63) who uncovered an extraordinary feature of the protein. Transcription factor IIIA has nine repeats of a 30-residue unit. The most prominent motif in the consensus sequence abstracted from the repeats is two invariant pairs of cysteines and histidines residues. The suggestion (63) is that the 30-residue repeat constitutes a “finger” in which pairs of cysteines and histidines tetrahedrally coordinate one zinc molecule and an extended loop of amino acids forms a DNA-binding region in contact with one-half turn of the helix. This proposal is, of course, most directly pertinent to the binding of factor IIIA to the 5 S gene and the relevance, if any, that it has for the association with 5 S rRNA is not known, although, it has been proposed that factor IIIA binds to a similar helical structure in both 5 S rDNA and rRNA (12). More to the point, the binding site for factor IIIA on 5 S rRNA closely approximates that for L5 (Fig. 8) and, hence, one might anticipate similarities in their structure. Despite this expectation no identity between factor IIIA and rat ribosomal protein L5 was found (Table II). Protein L5 has no repeat sequences, it does not have an unusually large number of cysteine or histidyl residues (Table I), and there is no indication of the presence of a sequence of amino acids related to the consensus derived from the factor IIIA repeat. The latter is not surprising since there is no evidence that L5 coordinates zinc and, hence, it has no need for the pairs of cysteines and histidines that define the factor IIIA repeats. Finally, there are no similarities in the structure of factor IIIA and any of the other ribosomal proteins that bind to 5 S rRNA (Table II).

There is not available a complete sequence of the amino acids in any other eukaryotic 5 S rRNA-binding protein. Nazar et al. (8, 32) have determined 119 of the approximately 318 residues in yeast YL3; 30 at the NH2 terminus (which we refer to as YL3-CN1), and 89 in a cyanogen bromide fragment of 102 residues (YL3-CN2) derived from the carboxyl end of the protein. Even with the limited sequence information available for the yeast ribosomal protein it is apparent that YL3 and L5 are homologous. The term is used here in the strict sense to connote derivation from a common ancestral gene. The score for the comparison of the NH2 termini of YL3-CN1 and L5 is 5.0 standard deviation units (Table II), where 3 S.D. is generally taken to indicate homology (64). For YL3-CN2 the score is 4.8 (Table II). The relatedness is apparent in an alignment of the NH2 termini of rat L5 and yeast YL3 (Fig. 7); there are 18 identical or similar2 amino acids in a sequence of 27 consecutive residues (67% identity). For the YL3-CN2 fragment there are 14 identities in 29 comparisons (48% identity) (Fig. 7). Note that 2 methionyl residues are required at positions 235 and 236 to properly align the 2 amino acid segments.

Smith et al. (7) have determined the sequence of the amino acids at the NH2 termini of the two proteins from H. cuti- rubrum ribosomes, HL13 (30 residues) and HL19 (28 residues), that bind to 5 S rRNA. In the comparison of rat L5 with the archebacterial ribosomal proteins one gets a mixed signal. The extent to which the NH2-terminal portion of HL13 is related to the same region of rat L5 is impressive; the score is 6.2 (Table II). However, there is no indication that the NH2-terminal residues of HL19 are related to rat L5; the score is −0.9 (Table II). These results are, of course, provisional since only a portion of the sequence of proteins HL13 and HL19 is available. Nonetheless, the similarity of the structure at the NH2 termini of HL13 and rat L5 is striking when the alignment is examined (Fig. 7); there are 15 identities2 in 27 comparisons (56% identity). It can be seen that HL13 is related to YL3 as well (Fig. 7); there are 12 identities in 21 comparisons (57% identity).

The comparison of the structure of rat ribosomal protein L5 with the E. coli 5 S rRNA-binding proteins L5, L18, and L25 (43) using the computer program RELATE failed to reveal a relationship (Table II). Nazar et al. (8) had suggested, from inspection of the amino acid sequences, that the NH2 termini of YL3 and of HL13 were related to E. coli L18 and that fragment CN2 of YL3 was similar to E. coli L5. We do not find evidence of either of these identities when the comparison is made using the program RELATE (Table II). We do support, with a more rigorous analysis, another suggestion (7) namely of a relationship of H. cuti rubrum HL19 to E. coli L5. The score for the comparison is 6.6 (Table II) and in the alignment (Fig. 7) there are 11 identities in 20 comparisons (55% identity). Moreover, there is a suggestive relationship of H. cuti rubrum HL19 to yeast YL3-CN1; the score is 2.5. Once again one must be cautious because only a portion of the structure of HL19 has been done. Nevertheless, it does appear that E. coli L5 and HL19 are, at the least in part, related. It is striking that one of the 5 S rRNA-binding proteins in H. cuti rubrum ribosomes, HL13 is related to the eukaryotic proteins, yeast YL3 and rat L5, whereas the second one, HL19, is related to the E. coli protein L5 and perhaps to yeast YL3 as well. This adds support to the surmise that archebacterial ribosomes are evolutionary transition particles between eukaryotes and prokaryotes and have properties of both (65).

We undertook these comparisons in the hope that they might reveal conserved domains in the proteins that could be considered candidate 5 S rRNA-binding sequences. Although, we do find conserved domains we are not yet able to associate them with binding to 5 S rRNA because the amino acid sequences of several of the relevant proteins are not complete, because the number of proteins for which there is sequence data is not sufficiently large, and because there are inconsist-

---

2 We include both strict identities and conservative substitutions, isoleucine/leucine/valine, aspartic acid/glutamic acid, serine/threonine, and lysine/arginine.
encies. For example, there is no relation that we can uncover in the sequence of amino acids in rat L5 and in Xenopus transcription factor IIIA despite the observation that they bind to the same site on 5 S rRNA. There are several possible ways to resolve this paradox. There may be subtle similarities in the primary structure of the two proteins, transcription factor IIIA and L5, that have escaped our notice, or different sequences of amino acids might be able to bind to the same structure in 5 S rRNA, or what is perhaps most likely there are similarities in the secondary or tertiary structures of the two proteins that accounts for their association with the same site on the nucleic acid.

It is not easy to account for the lack of a relationship between rat L5 and the E. coli proteins L5, L18, and L25. It may be that the differences in the secondary structure of prokaryotic and eukaryotic 5 S rRNA, no matter how trivial they seem, is sufficient to require a different sequence of amino acids in the nucleic acid-binding site on eukaryotic and prokaryotic ribosomal proteins, or, once again, the similarities may be in the higher order structure of the proteins.

There is encouragement in the finding that a number of the ribosomal proteins that bind to 5 S rRNA are related, rat L5 to yeast YLS and to H. cutirubrum HL13, and H. cutirubrum HL19 to E. coli L5. When the amino acid sequence of the yeast and H. cutirubrum proteins is completed there may be obvious conserved domains that will allow one to postulate the chemistry of the nucleic acid-binding site on the proteins.

Comparison of the Sequence of Amino Acids in Rat L5 with that in Ribosomal Proteins from Other Species—The sequence of amino acids in the rat ribosomal protein L5 was compared, using the computer program RELATE (30), to the sequence of amino acids in 280 other ribosomal proteins contained in a library that we have compiled, and that includes the complete set of 52 from E. coli (43). High scores were obtained for several proteins (Table III). A score, computed as the distance in standard deviations between the actual comparison and the mean of 100 comparisons of randomized sequences, of at least 3.0 is ordinarily required to assign significance to a relationship (64). The ribosomal proteins related to rat L5 include the following: rat L39, with a score of 4.3 standard deviations; Euglena gracilis chloroplast S7, 3.8; S. cerevisiae L46, 3.5; S. cerevisiae L31, 3.3; and Homo sapiens L32, 3.0. A remarkable aspect of the results of the comparison is the indication that pairs of ribosomal proteins from the same organism are related; i.e. rat L5 and L39, and that a single rat protein (L5) may be related to two proteins in a second organism, i.e. to yeast L46 and to yeast L31. There is, furthermore, a set of ribosomal proteins that gives scores between 2.9 and 2.6 that we interpret as indicating the possibility of a relationship to rat L5 (Table III). We attempted to align the several proteins (Table III) with rat L5 using the program of ALIGN (30) but we were not successful.

There is a group of eukaryotic ribosomal proteins, rat L5 and L39, yeast L46, mouse L22, and human L32, that appear to be closely related (Table IV). Apart from rat L5 none of these proteins are known to be associated with 5 S rRNA. Mouse and human L32 have identical amino acid sequences as had been noted before (36) just as it was known that yeast L46 and rat L39 are homologous (34). What is apparent now is that all of these ribosomal proteins are related to rat L5. There is, for example, one segment of 15 amino acids beginning at position 13 in rat L5 that is very similar to sequences that begin at position 74 in human and mouse L32, at position 8 and again at 33 in rat L39, and at position 34 in yeast L46 (Fig. 9). A second similar fragment, this time of 19 amino acids, is at position 254 in rat L5, at position 24 in human and mouse L32, at position 32 in rat L39, and at position 33 in yeast L46 (Fig. 9). Note that a fragment of rat L39 (starting near position 32) can be aligned with two parts of rat L5 and that two sequences in L39 (beginning at positions 8 and 33) can be aligned with a single segment in L5 (Fig. 9). These results, and others (66, 67), suggest that ribosomal proteins form an extended family and that L5 may contain in its structure traces of the affinity.

Acknowledgments—We are grateful to Veronica Paz for valuable technical assistance and to Ariene Timoschek for help in the preparation of the manuscript. We are indebted to M. Brown for providing us, prior to publication, a copy of a manuscript describing the preparation of mRNA, and to H. Okayama and M. Kawachi for advice and helpful discussions.

REFERENCES

1. Blobel, G. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 1881-1885
2. Lebleu, B., Marbaix, G., Huez, G., Tempner, J., Burny, A., and Chantrenne, H. (1971) Eur. J. Biochem. 19, 284-289
3. Horne, J. R., and Erdmann, V. A. (1972) Mol. Gen. Genet. 119, 337-344
4. Yu, R. S. T., and Wittmann, H. G. (1973) Biochim. Biophys. Acta 324, 375-385
5. Terao, K., Takahashi, Y., and Ogata, K. (1975) Biochim. Biophys. Acta 402, 230-237
6. Chen-Schmeisser, U., and Garrett, R. A. (1977) FEBS Lett. 74, 287-291
7. Smith, N., Matheson, A. T., Yaguchi, M., Willicck, G. E., and Nazar, R. N. (1975) Eur. J. Biochem. 59, 501-508
8. Nazar, R. N., Yaguchi, M., Willick, G. E., Rollin, C. F., and Roy, C. (1979) Eur. J. Biochem. 102, 573-582
9. Endo, Y., Huber, P. W., and Wool, I. G. (1983) J. Biol. Chem. 258, 2662-2667
10. Huber, P. W., and Wool, I. G. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 322-326
11. Huber, P. W., and Wool, I. G. (1986) J. Biol. Chem. 261, 3002-3005
12. Huber, P. W., and Wool, I. G. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1585-1597
13. Lathe, R. (1985) J. Mol. Biol. 183, 1-12
14. Beaucaule, S. L., and Caruthers, M. H. (1981) Tetrahedron Lett. 22, 1859-1862
15. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 599-600
16. Higgins, G. M., and Anderson, R. M. (1981) Arch. Pathol. 12, 186-202
17. Okayama, H., Kawachi, M., Brownstein, M., Lee, F., Yokota, T., and Arai, K. (1987) Methods Enzymol., in press
18. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
19. Laemmli, U. K. (1970) Nature 227, 680-685
20. Lin, A., Collatz, E., and Wool, I. G. (1976) Mol. Gen. Genet. 144, 1-9
21. Michelson, A. M., and Orkin, S. H. (1982) J. Biol. Chem. 257, 14773-14792
22. Deng, G., and Wu, R. (1981) Nucleic Acids Res. 9, 4175-4188
23. Vieira, J., and Messing, J. (1982) Gene (Amst.) 19, 259-266
24. Gronstein, M., and Hoggness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965
25. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, pp. 90-91, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
27. Bartlett, J. A., Gaillard, R. K., and Joklik, W. K. (1986) Biotechniques 4, 208-211
28. Queen, C. L., and Korn, L. J. (1990) Methods Enzymol. 155, 655-699
29. Tsurugi, K., Collatz, E., Wool, I. G., and Lin, A. (1976) J. Biol. Chem. 251, 7940-7946
30. Dayhoff, M. O. (1978) in Atlas of Protein Sequence and Structure, Vol. 5, Supp. 3, pp. 607-619, National Biomedical Research Foundation, Washington, D. C.
31. Ginsberg, A. M., King, B. O., and Roeder, R. G. (1984) Cell 39, 479-489
Supplementary Material to
THE PRIMARY STRUCTURE OF RAT RIBONUCLEAR PROTEIN L5

An Analysis of the Interaction of L5 and Other Proteins with 32P RNA

Yuen-Ling Chan, Alan Lin, James McNally, and Ya Q. Wool

EXPERIMENTAL PROCEDURES

Materials - Oligodeoxyribonucleotides were synthesized in the Research Cell l by polynucleotide kinase reaction using lambda DNA with T4 DNA polymerase, and terminal deoxynucleotidyl transferase from Helobacter-Munitis. Restriction endonucleases were obtained from New England Biolabs and Bethesda Research Laboratories and were used to digest DNA essentially as recommended by the manufacturers. 32P-5'-end labeled dATP was produced by D.S. (The University of Chicago) and was purified by H. L. (University of Chicago). Restriction endonucleases were obtained from New England Biolabs and Bethesda Research Laboratories and were used to digest DNA essentially as recommended by the manufacturers. 32P-5'-end labeled dATP was produced by D.S. (The University of Chicago) and was purified by H. L. (University of Chicago).

RESULTS (of Oligonucleotide Probes) - The purpose was to use the sequence of some nucleotides in a polynucleotide probe to deduce the complementary coding sequence, to synthesize the corresponding oligonucleotide probe (32), a single-stranded DNA that will hybridize to the coding sequence in the protein, and then to use the probe to screen a recombinant cDNA library. To synthesize a probe as efficient as possible, we aligned oligodeoxyribonucleotide primers to the following guidelines: the probe must be at least 15-20 nucleotides in length to allow specific hybridization; a sequence of nucleotides in the long probe was selected that gave the least amount of hybridization to the preloaded cDNA in the probe; and the probe was synthesized using the technique of complementary oligonucleotide probes (32). The strategy that was used to design the probe for the cDNA-expressing rat ribosomal protein L5 is given in Fig. 1, with the sequence of some nucleotides in the long probe and the corresponding cDNA sequence. The probe, a mixture of 15 different oligodeoxyribonucleotides, 26 nucleotides in length, was synthesized as a solid support by the method of polymerase catalyzed using an Applied Biosystems, Model 380A, DNA synthesizer (32) and the oligonucleotides were purified by polyacrylamide gel electrophoresis. The probe mixture was synthesized in the 7-Mi test and in 21-Mi (21-mer) and 27-Mi (27-mer) polyribonucleotide mixtures. The radioactivity of probe was determined using a long column of Sephadex G-50 in 10 mM Tris-HCl, pH 8.0, and 0.5 M NaCl.

Preparation of Poly(rC):Poly(dA) and Oligo(dT)-Transcribed Fragments of the Nucleolar Nucleoprotein - Sedimentation rate measurements were carried out at 39.5-44.5° in water 18% and 18% KCl in the range of 11.7-12.2, respectively. The translation of the product was carried out with a nucleic acid-treated gel column supplemented with [3H]tyrosine. To determine whether the poly(rC):poly(dA) scattered ribonucleoproteins the products of the translation reaction were analyzed by electrophoresis on one-dimensional sodium dodecyl sulfate/10% SDS (37) and in two-dimensional sodium dodecyl sulfate/polyacrylamide gels. The gels were stained with Coomassie brilliant blue and the autoradiograms were prepared with intensification and exposed for an appropriate amount of time at -80°C. In this way, for all rat ribosomal proteins, were identified as being encoded by the preparation of poly(rC):poly(dA).

Conclusion of a Rat Ribosome Library - The synthesis of the first strand of the cDNA was in 35 of 55 of the total of 365 32P RNA ligated with rat ribosomal protein 40S (35), and in 12 of 22 of the rat ribosomes (35) were used by the recombinant cDNA. Ribosomal RNA was then ligated with the restriction endonuclease (35) to the recombinant cDNA. Ribosomal RNA was then ligated with the restriction endonuclease (35) to the recombinant cDNA. Ribosomal RNA was then ligated with the restriction endonuclease (35) to the recombinant cDNA. The cDNA was then ligated with the restriction endonuclease (35) to the recombinant cDNA. The cDNA was then ligated with the restriction endonuclease (35) to the recombinant cDNA. The cDNA design was constructed by 32P (35) poly(rC):poly(dA) and contained 32-000 probes.

The cloning of the cDNA library by Oligo(dT)-transcribed Fragments: An aliquot from the cDNA library was spotted as described above to give 3-5 MILLION colonies grown on 15 mm agar plates with ampicillin (100 mg/ml). The colonies were then added to 10 ml of culture medium 32P (35) and were inoculated in the 21-Mi test and in 27-Mi polyribonucleotide mixture. The probe was then ligated to 80% of the 17-Mi test mixture and in 21-Mi (21-mer) and 27-Mi (27-mer) polyribonucleotide mixtures. The radioactivity of probe was determined using a long column of Sephadex G-50 in 10 mM Tris-HCl, pH 8.0, and 0.5 M NaCl.
Structure of Rat Ribosomal Protein L5

by the method of Gurdon and Hoggness (23) for those inserts that encode rat ribosomal proteins (in this instance L5) with a radioactive end-labeled oligodeoxynucleotide probe synthesized for the occasion (see above). The filters were exposed at 42 °C for 2 h to a solution of [3H]labeled 20% formamide, 5X SSC, 40 mM sodium phosphate pH 6.5, 2X Denhardt's solution (DL: Denhardt's solution 0.2% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% bovine serum albumin, fraction V, 0.1% sodium deoxydextran sulfate, and 3.0 g/ml of succinyl salmon sperm DNA to saturate nonspecific sites. Specific hybridization was for 20 h at 42 °C, with 30 ng (t) x 10⁶ cpm of the radioactive oligodeoxynucleotide probe in 1 ml of the same solution. The filters were washed three-times at 42 °C for 30 min each with a large volume of 5X SSC that contained 0.1% sodium pyrophosphate. (Dow Corning Kry-Fill) with intensely scoring filters were exposed to the filters overnight at -70 °C. The lanes that gave a positive hybridization signal were scored again in the same way but at a lower cell density. The colonies that scored positive a second time were grown in L broth that contained 30 g/ml of ampicillin. The plasmid DNA was isolated (24) and digested with restriction endonucleases PstI, or EcoRI and HindIII, to release the cDNA insert. The cDNA inserts were separated by electrophoresis on 12% agarose gels and identified by Southern blot-hybridization using the radioactive end-labeled oligodeoxynucleotide probe and the same conditions as for the isotopic hybridization.

The determination of the nucleotide sequence in DNA. - The isolated cDNA insert was digested with restriction endonucleases to generate fragments that were cloned in M13 mp18. The sequence of nucleotides in the fragments was determined by the dideoxynucleotide chain-termination procedure (26) using avian myeloblastosis virus reverse transcriptase and [α-32P]dNTP. The conditions for denaturation and for annealing the double-stranded M13 RF I DNA or supercoiled plasmid DNA for the sequencing reactions were as described by Rutert et al. (27). Uncertainties concerning the identity of nucleotides were resolved by repeating the determination of the ambiguous sequence using the method of Maxam and Gilbert (28). To confirm the nucleotide sequence on the P side of position 96, an oligodeoxynucleotide primer corresponding to positions 91-115 (19-19) was synthesized and used in the dideoxyxynucleotide sequencing procedure. The nucleotide sequence was analyzed and converted to an amino acid sequence with the computer program of Queenan and Korn (29).

Determination of the sequence of amino acids in Protein L5. - Rat ribosomal protein L5 was isolated by high-performance liquid chromatography on a gel filtration column (Bio-Rad, Bio-Gel TSK-125, 7.5 x 7.5 cm LID), an acetic acid elution with 5.15 M ammonium acetate pH 5.7 at a flow rate of 0.8 ml/min (Fig. 2). The identity and the homogeneity of L5 was established by electrophoresis on one-dimensional sodium dodecyl sulfate (10% and two-dimensional area SDS) polyacrylamide gels (Fig. 3). The amino acid composition of L5 was determined (29) from an analysis of an acid hydrolysate.

The sequence of amino acids at the NH2 terminus of L5 was obtained with an automated gas-phase sequencer (Applied Biosystems, Model 473A). The identity of the phenylthiohydantoin derivatives of the sequenced peptides was established by high performance liquid chromatography with an instrument designed for the purpose (Applied Biosystems, Model 255A). The sequence of 18 of the 19 amino acids at the NH2 terminus of rat ribosomal protein L5 was determined in this way (Fig. 3). In addition, the three amino acids at the NH2 terminus of L5 were obtained by digestion with carboxypeptidase A (Fig. 4).

The computer program RELATE E30 was used to assess possible evolutionary relationships between L5 and other ribosomal proteins. The scoring matrix was Dayhoff's MSA 10. The ribosomal protein amino acid sequence was divided into nonoverlapping 10-amino acid segments (in 22 fragments) that were subsequently compared by the computer program RELATE E30 with other ribosomal proteins. The amino acid codon composition of rat ribosomal protein L5 was determined (29) from an analysis of an acid hydrolysate of purified rat L5 (31) isolated from the sequence of nucleotides in a recombinant plasmid (28). The amino acid composition of rat ribosomal protein L5 was determined (29) from an analysis of an acid hydrolysate of purified rat L5 (31) isolated from the sequence of nucleotides in a recombinant plasmid (28).

Table 1

| Amino Acids | Number | Percentage |
|-------------|--------|------------|
| Alanine     | 23     | 28         |
| Arginine    | 28     | 30         |
| Aspartic Acid | 26    | 10%         |
| Cysteine    | 26     | 30         |
| Glutamic Acid | 32    | 40         |
| Glutamine   | 19     | 23         |
| Glycine     | 19     | 23         |
| Histidine   | 11     | 14         |
| Isoleucine  | 19     | 23         |
| Leucine     | 22     | 28         |
| Lysine      | 32     | 40         |
| Methionine  | 4      | 5          |
| Phenylalanine | 12   | 15         |
| Proline     | 12     | 15         |
| Serine      | 9      | 11         |
| Threonine   | 9      | 11         |
| Tryptophan  | 1      | 1          |
| Tyrine      | 11     | 14         |
| Valine      | 20     | 25         |
| Residues    | 294    | 36         |

**Fig. 2.** Preparation of rat ribosomal protein L5. Rat ribosomal protein L5 was prepared from group fraction CAI x 150 by high performance liquid chromatography on a Bio-Rad Bio-Gel TSK-125 column (35 ml of 7.5 x 7.5 cm LID); elution was with 5.15 M ammonium acetate pH 5.7, at a flow rate of 0.8 ml/min. Protein L5 eluted after 7.1 min and the fraction designated by a bracket was collected.

**Fig. 3.** Electrophoresis of rat ribosomal protein L5 in polyacrylamide gels. The protein was electrophoresed on a single gel with a cathode at the top (left) and anode at the bottom (right) of the gel. The gel was stained with Coomassie blue and the bands were cut from the gel and digested with pronase. The amino acid sequence of the gel was determined by the method of Edman and Von der Weise. The sequence of amino acids at the NH2 terminus of rat ribosomal protein L5 was determined in an automated gas-phase sequencer. From the possible coding sequences of 9 consecutive amino acid residues (positions 11-17) an oligodeoxynucleotide probe (23-26 mer) was designed that was then used to score the coding frequencies for human ribosomal proteins and, hence, minimized random hybridization. The actual relevant coding sequence in the DNA is given at the bottom of the position of P and P' nucleotides designated by subscript numbers.
Figure 4. The time course of the release of amino acids from rat ribosomal protein L5 with carboxypeptidase. The digestion is 1.25 mmol of rat ribosomal protein L5 with carboxypeptidase A at an enzyme to substrate ratio of 1:50.

Figure 5. A restriction endonuclease map of the cDNA insert cloned into pBR322 and a cartoon of the strategy used to determine the sequence of nucleotides. The upper portion depicts the restriction endonuclease sites in the cDNA insert in the plasmid pBR322 that were used to generate oligonucleotide fragments suitable for the determination of the sequence of nucleotides. The numbers designate the first nucleotide in the restriction enzyme recognition sequence counting from the 5' end of the chain. Each arrow designates the direction and the extent of a nucleotide sequence that was determined. Since the sequence determination was always in the 5' to 3' direction the arrow also specifies the strand. The procedure used for sequencing is also given: single-strand M17 DNA, ---; double-stranded M13 RF1 DNA or supercoiled DNA using the dideoxy procedure, ----; Maxam and Gilbert method, ------.

Figure 6. The kinetics of the release of amino acids from rat ribosomal protein L5 with carboxypeptidase. The digestion is 1.25 mmol of rat ribosomal protein L5 with carboxypeptidase A at an enzyme to substrate ratio of 1:50.

Figure 7. The alignment of related sequences of amino acids in the ribosomal protein L5. The comparison is of: the sequence of amino acids near the N-terminal of rat L5 (amino acids 1-5); the sequence of amino acids near the N-terminal of mouse L5 (amino acids 1-5); the sequence of amino acids near the N-terminal of human L5 (amino acids 1-5); and the sequence of amino acids near the N-terminal of calf L5 (amino acids 1-5). The sequence of amino acids near the N-terminal of rat L5 is in bold letters. The sequence of amino acids near the N-terminal of mouse L5 is in italics. The sequence of amino acids near the N-terminal of human L5 is in bold italic. The sequence of amino acids near the N-terminal of calf L5 is in bold italic.

Figure 8. The alignment of related sequences of amino acids in the ribosomal protein L5. The comparison is of: the sequence of amino acids near the N-terminal of rat L5 (amino acids 1-5); the sequence of amino acids near the N-terminal of mouse L5 (amino acids 1-5); the sequence of amino acids near the N-terminal of human L5 (amino acids 1-5); and the sequence of amino acids near the N-terminal of calf L5 (amino acids 1-5). The sequence of amino acids near the N-terminal of rat L5 is in bold letters. The sequence of amino acids near the N-terminal of mouse L5 is in italics. The sequence of amino acids near the N-terminal of human L5 is in bold italic. The sequence of amino acids near the N-terminal of calf L5 is in bold italic.