The Carboxyl Extension of a Ubiquitin-like Protein Is Rat Ribosomal Protein S30*

(Received for publication, February 25, 1993, and in revised form, April 22, 1993)

Joe Olvera and Ira G. Woolf
From the Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

The amino acid sequence of the rat 40 S ribosomal subunit protein S30 was deduced from the sequence of nucleotides in a recombinant cDNA and confirmed by the determination of the 18 residues at the NH2 terminus of the protein. Unlike the majority of ribosomal proteins, which are unprocessed primary products of the translation of their mRNAs, S30 is formed by cleavage from a larger hybrid protein. The NH2-terminal polypeptide has 38% identity with ubiquitin and contains the characteristic carboxyl-terminal Gly-Gly dipeptide of this family of proteins. S30 has 59 amino acids and the molecular weight is 6,643; the ubiquitin-like sequence has 74 residues and the molecular weight is 7,634. The hybrid protein is encoded in each of the 8-10 members of the family of rat S30 genes; there is, however, only a single species of mRNA which contains the sequences for both proteins. The coding sequence of the hybrid protein occurs in the reverse polarity in the genome of the Finkel-Biskis-Reilly murine sarcoma virus.

Solution of the structure, and understanding of the function, of eukaryotic ribosomes requires knowledge of the chemistry of the molecular components. For this reason we have undertaken to determine the sequence of nucleotides and of amino acids in the constituent nucleic acids and proteins in mammalian (rat) ribosomes (1). As part of this endeavor, we have determined the primary structure of rat ribosomal protein S30 from the sequence of nucleotides in a recombinant cDNA.

The primary structures of more than 60 rat ribosomal proteins have been determined either directly from the protein or deduced from the sequence of nucleotides in recombinant cDNAs (1). Two of these ribosomal proteins are carboxyl extensions of ubiquitin (2-4); all of the others are the primary products of the translation of a simple open reading frame. Rat ribosomal protein S30 is also an exception to the rule; it is formed by cleavage from a polypeptide that diverges from, or deduced from the sequence of nucleotides in a recombinant cDNA, designed with ethanol, resuspended in buffer, digested with the appropriate restriction enzymes (HindIII and BamHI), and separated by electrophoresis in 1% NuSieve GTG agarose (FMC BioProducts). The amplified DNAs were extracted from the agarose gels and ligated to pGEM2 that had been treated with HindIII and BamHI. DH5α cells were transformed with the individual plasmids and three clones, with inserts whose nucleotide sequences depended on the identity of the primers used in the amplification reaction (cf. Table I), were selected (Fig. 2). The first, pGEM2-UBI-S30, has the entire pRS30-

EXPERIMENTAL PROCEDURES

The Isolation and the Determination of the NH2-terminal Amino Acid Sequence of Rat Ribosomal Protein S30—Rat ribosomal protein S30 had not been isolated nor characterized in part because it is lost during conventional two-dimensional electrophoresis in polyacrylamide-urea gels (5). S30 can be identified, however, after electrophoresis in two dimensions in basic-sodium dodecyl sulfate gels (Ref. 6 and Fig. 1A), and this procedure was used to isolate the protein. The proteins from rat liver 40 S ribosomal subunits (3-5 mg in 5% acetic acid) were fractionated first on Sephadex G-100. Fractions containing proteins ranging in molecular weight from 6,500 to 14,000 were collected, and about 750 mg of the mixture was resolved by electrophoresis (Fig. 1B). S30 was identified by staining the gels with Coomassie Blue; the spots containing S30 were cut out of the gels, and the protein was extracted from the gel plugs with 70% formic acid (7). The extract was concentrated in vacuo, resuspended in 200 ml of 0.1% trifluoroacetic acid, and resolved from nonprotein contaminants by high performance liquid chromatography on a size exclusion column (Bio-Rad TSK 125, 300 × 7.5 mm) with 0.1% trifluoroacetic acid as the mobile phase. Fractions were collected, dried, resuspended in 50 µl of 5% acetic acid, and 5 µl from each fraction were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 1C). Fractions containing S30 were pooled and the NH2-terminal amino acid sequence was determined by automated Edman degradation with an Applied Biosystems 470A Protein Sequencer.

The Preparation of Recombinant cDNAs Encoding Rat Ribosomal Protein S30—The recombinant DNA procedures including the design of probes to identify specific cDNAs have been described or cited (8, 9). Probe 1 was a mixture of 342 different oligodeoxynucleotides, each 23 bases in length, predicted on the sequence MKVHGSLA in S30 (residues 1-7; we assumed, since we were not aware at the time that S30 is the carboxyl extension of a fusion protein, that the cDNA encoded a methionine before the lysine found at the NH2 terminus of S30 (this later proved not to be the case)). Probe 2 was a mixture of 288 different oligodeoxynucleotides, each 26 bases in length, based on the sequence AGKVRGQTP (residues 9-17). The oligodeoxynucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer. The sequences of nucleotides in the cDNAs were determined by the dideoxynucleotide chain termination method using modified T7 DNA polymerase (10).

The Identification of the Protein Encoded in the Open Reading Frame in pRS30-12 by Transcription of the cDNA and Translation of the mRNA—The cDNA insert in pRS30-12 (3 ng) was amplified in the polymerase chain reaction in 50 µl of the following: 50 mM KCl, 10 mM Tris-HCl, pH 9.3, 1.5 mM MgCl2, 0.01% gelatin, 200 µM each of the four dNTPs, 1 µM each of two oligodeoxynucleotide primers (Table I), and 1.5 units of Taq polymerase. The reaction mixture was overlaid with 50 µl of mineral oil, and the amplification was carried out for 30 cycles in a Perkin-Elmer Cetus DNA Thermal Cycler programmed for denaturation at 94 °C for 2 min, annealing at 55 °C for 1 min, and extension at 72 °C for 2 min. The amplified DNA was extracted with phenol/chloroform/isoamyl alcohol (24:24:1), precipitated with ethanol, resuspended in buffer, digested with the appropriate restriction enzymes (HindIII and BamHI), and separated by electrophoresis in 1% NuSieve GTG agarose (FMC BioProducts). The amplified cDNAs were extracted from the agarose gels and ligated to pGEM2 that had been treated with HindIII and BamHI. DH5α cells were transformed with the individual plasmids and three clones, with inserts whose nucleotide sequences depended on the identity of the primers used in the amplification reaction (cf. Table I), were selected (Fig. 2). The first, pGEM2-UBI-S30, has the entire pRS30-

* This work was supported by a National Institutes of Health Grant GM-21769. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X62671.

† To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, The University of Chicago, 920 E. 58th St., Chicago, IL 60637. Tel.: 312-702-1341; Fax: 312-702-0439.
Purified and a BamHI sequence (GGATCC) for cloning in pGEM2 and polymerase chain reaction to give templates that encode 1) the entire silver nitrate. primers for each amplification had a HindIII sequence (AAGCTT) sequences are underlined. The primers for the amplification of the trophoresis in 1% agarose gels to assess the length of the transcripts encodes only ribosomal protein S30. The plasmid DNA from the clones was isolated by the alkaline lysis method (11).

12 open reading frame and encodes a ubiquitin-like polypeptide (UbIL)2 and ribosomal protein S30; the second, pGEM2-UbIL, encodes only ribosomal protein S30. The plasmid DNA from the clones was isolated by the alkaline lysis method (11).

Each of the three plasmids was linearized with BamHI and transcribed for 60 min at 37 °C in a reaction mixture (50 µl) containing the following: 80 units of RNAsin (Promega), 40 mM Tris-HCl, pH 7.9, 6 mM MgCl2, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 1 mM each of the four dNTPs; 2 µg of linearized plasmid DNA, and 120 units T7 RNA polymerase. Aliquots were analyzed by electrophoresis in 1% agarose gels or one-dimensional polyacrylamide-sodium dodecyl sulfate gels. The gels were treated with EN3HANCE (Du Pont-New England BioLabs) and analyzed by electrophoresis in two-dimensional polyacrylamide-gels. The gels were treated with ENHANCE (Du Pont-New England Nuclear), dried, and exposed to x-ray film (Kodak X-omat) at −70 °C with intensifying screens.

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein S30—A random selection of 48,000 colonies from two cDNA libraries of 30,000 and of 20,000 independent transformants was screened with two oligodeoxynucleotide probes for rat ribosomal protein S30. Nine colonies gave positive hybridization signals. The DNA from the plasmids of the nine transformants was isolated, digested with restriction endonucleases, and a clone, pRS30-12, was selected. The cDNA insert in pRS30-12 is 505 bases long, has a 5′-noncoding sequence of 58 nucleotides, a single open reading frame of 402, and a 3′-noncoding sequence of 45 as well as a long poly(A) stretch (Fig. 2). In the other two reading frames, the sequence is interrupted by many termination codons. The open reading frame begins with an ATG codon at a position that we designate +1 and ends with a termination codon, TAA, at position 400; it encodes 133 amino acids. The initiation codon is in a context, AACATG, that approximates the optimum, ACCAUGG (12). The 3′-noncoding sequence has, at positions 427–432, the hexamer AATAAAA that directs posttranscriptional cleavage-polyadenylation of the 3′ end of the precursor mRNA (13). The 5′-nontranslated region of pRS30-12 begins with eight consecutive pyrimidines (positions −51 to −58). Pyrimidine sequences are found at the 5′ end of eukaryotic ribosomal protein mRNAs (1), and it is likely that they play a role in the regulation of translation (14, 15). What is notable is that the polypyrimidine stretch also occurs in the mRNA for a hybrid protein (see later) which suggests that the polypeptide mRNA shares a common regulation of translation with the other ribosomal protein mRNAs.

The Identification of the Polypeptide Encoded in pRS30-12—The initial identification of a protein specified by the reading frame in pRS30-12 was from the correspondence to the sequence of amino acids at the NH2 terminus of rat ribosomal protein S30. The sequence, KVHGSLAR-AGKVRGQTPK, determined by Edman degradation with an automated gas-phase sequencer, however, corresponds to residues 75–92 deduced from the nucleotide sequence (Fig. 2). The provisional conclusion from the comparison of the sequences of amino acids encoded in pRS30-12 and that derived from the isolated S30 is that the cDNA codes for a hybrid protein; the NH2-terminal polypeptide is ubiquitin-like (from the similarity of the amino acid sequence to that of ubiquitin; cf. Fig. 3 and Ref. 16) and the carboxyl extension is ribosomal protein S30. This assumption was verified later.

The Primary Structure of Rat Ribosomal Protein S30—The molecular weight of rat ribosomal protein S30, calculated from the sequence of amino acids deduced from pRS30-12 and based on the assumption that its NH2 terminus is at position 75 in this sequence (Fig. 2), is 6,643; the protein contains 59 amino acids. S30 does not have an NH2-terminal methionine

1 The abbreviations used are: UbIL, ubiquitin-like sequence in a fusion protein with rat ribosomal protein S30; FBR-MuSV, Finkiel-Biskis-Reily Murine Sarcoma Virus; UCRP, ubiquitin-like cross reactive protein; AcMNPV, Autographa californica nuclear polyhedrosis virus; v-ub, ubiquitin-like protein encoded in the genome of AcMNPV.
The residues. Amino acid residues 1-74 residues; the positions tides in the cDNA insert in plasmid are of a ubiquitin-like protein; residues in the cDNA insert are given above the pRS30-12 and the amino acid sequence is in philic. Finally, there is a Gly-Gly dipeptide at the carboxyl terminus of the ubiquitin-like sequence as in all of the ubiquitin-like and protein has 38% identity with ubiquitin (16), although, it has because it is the product of posttranslational processing of a mammalian ubiquitin (16); mammalian ribosomal protein genes have been found to be genome only in a fusion gene and the number of bands number of radioactive bands with each of the three probes was the same (Fig. 4), indicating that the ubiquitin-like polypeptide and ribosomal protein S30 are encoded in the rat genome only in a fusion gene and the number of bands suggests that there are 8-10 copies of the gene. Many other mammalian ribosomal protein genes have been found to be present in multiple copies (cf. Ref. 1, for references and discussion). However, in no instance has it been shown that more than one of the genes is transcribed. The presumption is that the other copies are nonfunctional retroposon pseudogenes. It is noteworthy that the UbiL-S30 fusion protein because it is the product of posttranslational processing of a fusion protein (see later). There are, in S30, 21 basic residues (7 arginines, 13 lysines, and 1 histidine), and only 1 that is acidic (Table II). The basic residues tend to be clustered as is the case in all of the ubiquitin-like sequences (Table II). In contrast to S30, the ubiquitin-like polypeptide only); and the cDNA insert in pGEM2-UbiL (which encodes the ubiquitin-like polypeptide to which it is conjugated). The amino acid composition of rat ribosomal protein S30 and of the ubiquitin-like polypeptide to which it is conjugated

<table>
<thead>
<tr>
<th>Amino acids</th>
<th>S30</th>
<th>Ubiquitin-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Arginine</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Asparagine</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Cysteine</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Glutamine</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Glycine</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Histidine</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Leucine</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Lysine</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Methionine</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Proline</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Serine</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Threonine</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Valine</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Residues</td>
<td>59</td>
<td>74</td>
</tr>
</tbody>
</table>

*The amino acid composition (in numbers of residues) was determined from the sequence of nucleotides in the recombinant cDNA pRS30-12.

UbiL-S30 (which encodes the entire ubiquitin-like-S30 fusion protein); the cDNA insert in pGEM2-UbiL (which encodes the ubiquitin-like polypeptide only); and the cDNA insert in pGEM2-S30 (which encodes S30 alone). The pattern and the number of radioactive bands with each of the three probes was the same (Fig. 4), indicating that the ubiquitin-like polypeptide and ribosomal protein S30 are encoded in the rat genome only in a fusion gene and the number of bands suggests that there are 8–10 copies of the gene. Many other mammalian ribosomal protein genes have been found to be present in multiple copies (cf. Ref. 1, for references and discussion). However, in no instance has it been shown that more than one of the genes is transcribed. The presumption is that the other copies are nonfunctional retroposon pseudogenes. It is noteworthy that the UbiL-S30 fusion protein...
gene has this characteristic of mammalian ribosomal protein genes.

The Size of the mRNA Encoding the Hybrid Protein—To
determine the size of the mRNA coding for the hybrid protein,
total poly(A)+ mRNA from rat liver was separated by electrophoresis (9) and screened for hybridization bands using the
same three radioactive cDNAs employed for the genomic blots (i.e., a probe specific for the entire ubiquitin-like sequence, or for the ubiquitin-like protein alone, or for S30 alone). With each probe only a single band of about 1250 nucleotides was detected (Fig. 5). That the genomic blots with the ubiquitin-like and S30 specific probes were identical, and that only a single species of mRNA was detected with the specific probes, suggests that the ubiquitin-like protein and S30 are transcribed from a single functional gene that encodes both. Moreover, the analysis would indicate that this particular ubiquitin-like sequence is not likely to be in the mRNA for any other protein.

The Characterization of the Products of the Transcription
and of the Translation of pRS30–12 cDNAs—Three separate
cDNAs derived from pRS30–12 were transcribed with T7 RNA polymerase, the transcripts were translated in reticulo-

cyte lysates, and the products were analyzed by two-dimen-
sional electrophoresis. Translation of the transcript from the
hybrid protein cDNA (pGEM2-UbiL-S30) yields S30 and a
larger, less basic protein that does not migrate with any 40 S ribosomal subunit protein (Fig. 6a and b) nor for that matter with any 60 S subunit protein; from its apparent molecular weight we presume the latter to be the primary translation product, i.e., the ubiquitin-like-S30 polyprotein. Since S30 is a product of the translation of the fusion protein mRNA it follows that a proteolytic enzyme capable of processing the polyprotein is present in the reticulocyte lysate. The second product of the processing reaction, the ubiquitin-like protein, could not be identified, we presume because it is too acidic and, hence, it is lost into the buffer during electrophoresis in Kaltenschmidt-Wittmann type two-dimensional urea-gels (but see later). It is unlikely that the unprocessed ubiquitin-like S30 fusion protein is incorporated into ribosomes since it migrates to a position on two-dimensional gels (Fig. 6a) that is not occupied by a ribosomal protein.

Translation of the transcript from the S30 cDNA (pGEM2-
S30) in a reticulocyte lysate yields a product that has the same coordinates as authentic S30 (Fig. 6, c and d).

Analysis of the product of the translation of the transcript from the ubiquitin-like protein cDNA (pGEM2-UbiL) was not possible by ordinary two-dimensional gel electrophoresis because of its acidic character (see above) and demonstration of processing in reticulocyte lysates was hindered because globin displaced and obscured the products of the reaction. For these reasons translation of the separate transcripts from pRS30–12 was done in a wheat germ extract, and analysis of the products was by electrophoresis in one dimension in sodium dodecyl sulfate gels (Fig. 6e). The three cDNAs gave separate products: the transcript from pGEM2-UbiL-S30 yielded the fusion protein (Fig. 6e, lane 2); that from pGEM2-
UbiL only the ubiquitin-like protein (Fig. 6e, lane 3); and that from pGEM2-S30 only ribosomal protein S30 (Fig. 6e, lane 4). The wheat germ extract does not have a protease that can process the polyprotein (cf. Fig. 6e, lane 2). However, if an equal volume (5 µl) of reticulocyte lysate was added to a wheat germ extract that had been incubated with the transcript from pGEM2-UbiL-S30, and incubation was continued for 30 min at 37 °C, the fusion protein was processed. The reaction

![Fig. 4. Hybridization of pRS30–12 cDNAs to rat genomic DNA. Rat liver DNA (20 µg) was digested with restriction endonucleases BamHI (lanes 1), EcoRI (lanes 2), or HindIII (lanes 3), and the digests were resolved by electrophoresis in 0.8% agarose gels and transferred to GeneScreen Plus nylon filters. Uniformly labeled radioactive cDNAs were hybridized to the immobilized genomic DNA. In (A), the probe was UbiL-S30 (the cDNA for the entire ubiquitin-like-S30 fusion protein); in (B), the probe was UbiL (the cDNA for the ubiquitin-like protein alone); and in (C), the probe was S30 (the cDNA for ribosomal protein S30 alone). The positions to which DNA standards of the size designated (in kilobase pairs) migrate is shown to the left.

![Fig. 5. Hybridization of pRS30–12 cDNAs to rat liver poly(A)+ mRNA. The rat liver poly(A)+ mRNA was treated with 50% dimethyl sulfoxide and 5.6% glyoxal in 10 mM sodium phosphate, pH 7.0, at 50 °C for 60 min. The glyoxylated RNA was separated by electrophoresis in 1.2% agarose gels in 10 mM sodium phosphate, pH 7.0, and transferred by capillary pressure to GeneScreen Plus nylon filters. The glyoxal was removed from the immobilized RNA by mild alkaline treatment. Uniformly labeled radioactive cDNAs (as in the legend to Fig. 4) were hybridized to the immobilized poly(A)+ mRNA: In (a), the probe was UbiL-S30 (the cDNA for the entire ubiquitin-like S30 fusion protein); in (b), the probe was UbiL (the cDNA for the ubiquitin-like protein alone); and in (c), the probe was S30 (the cDNA for ribosomal protein S30 alone). The hybridizations were to separate filters containing samples of the same preparation of poly(A)+ mRNA. The size of the mRNA was estimated by comparison to the mobility in the same gels of DNA restriction fragments of the size (in kilobases) shown on the left.](image-url)
mixture now had S30 (somewhat displaced by nonradioactive globin) and the ubiquitin-like protein, as well as unprocessed fusion protein (Fig. 6e, lane 5).

Comparison of the Sequence of Amino Acids in Rat S30 with Other Ribosomal Proteins—The sequences of amino acids in the ubiquitin-like S30 hybrid protein and in S30 alone were compared, using the computer program RELATE (17), to the sequences of amino acids in more than 1,200 other ribosomal proteins contained in a library that we have compiled. The comparison did not yield a ribosomal protein that shared significant identity with either the ubiquitin-like polypeptide or with S30.

The ubiquitin-like NH2-terminal portion of the fusion protein has a tetrapeptide repeat, TLEV, at positions 12-15 and 64-67. There are no amino acid repeats in the ribosomal protein S30 sequence.

The Ubiquitin-like S30 Hybrid Protein Is Encoded in the v-fox Gene of a Retrovirus—A search of EMBL/GenBank data base yielded an unexpected finding. The nucleotide sequence encoding the ubiquitin-like-S30 hybrid protein occurs in the reverse polarity between the fos and env genes of the Pinkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) (18) (Fig. 7).

FBR-MuSV is a retrovirus that was isolated from a radiation-induced mouse osteosarcoma (19) that encodes a gag-fos fusion protein that transforms cells in vitro and causes neoplasia in vivo (18, 20). The virus acquired the nucleotide sequence encoding a cellular protein, c-fos, and integrated it between the gag and env genes leading to inactivation of the viral genes (the virus can only reproduce with the assistance of a helper virus) and the formation of a gag-fos fusion protein that is oncogenic (18, 20). The mechanism by which cellular sequences are acquired by transforming retroviruses is not known; nor is the function of the product of the cellular fos gene known. It is thought that certain normal cellular gene products can induce neoplasia 1) when the levels of the gene product are abnormally high, or 2) when a mutant form of the gene is expressed, or 3) when expression of the gene is inappropriate, i.e. expression is either in the wrong cell type or during the wrong stage in development or both. Normal and mutant forms of the c-fos protein cause transformation as does expression of c-fos in an inappropriate cell type (21).

The FBR-MuSV has acquired a second cellular gene, v-fox, that encodes the ubiquitin-like-S30 fusion protein (Fig. 7). The nucleotide sequence was integrated in the reverse transcriptional orientation between fos and env (18). After we had determined the nucleotide sequence of pRS30-12 and identified the ubiquitin-like and S30 proteins, and after we had deposited the information in EMBL/GenBank, the identity of v-fox was confirmed (22) from our submission. There is no
evidence that the v-fox product is necessary for transformation, indeed, there is no indication that it serves any function in the life cycle of the virus.

In an alignment of the nucleotide sequences (Fig. 8), there is 94% identity between v-fox (positions 2703–3161; cf. Ref. 18) and the rat ubiquitin-like-S30 cDNA (positions −11 to 447). There is a five nucleotide direct repeat (ACTGTG) that flanks the ends of v-fox in the FBR-MuSV genome (underlined in Fig. 8), and this may be the site of insertion by homologous recombination. There are 21 nucleotide differences in the open reading frames of v-fox and pRS30–12; of these 13 are silent. Two of the other eight differences may be the result of a transposition of a TC dinucleotide (positions 2913 and 2914) in the v-fox sequence, i.e. it probably should be CT, as in the pRS30–12 sequence (positions 236 and 237), since this would make the codon TCT (Ser) not TTC (Phe). We know the residue at this site in rat S30 is serine since its identity was confirmed from the sequence of amino acids determined directly from the protein.

Ubiquitin and the Ubiquitin-like Proteins—Ubiquitin is a protein of 76 amino acids that has been identified in all eukaryotic cells that have been examined and whose amino acid sequence is conserved; there are only three nonidentities between the ubiquitins of humans and of yeast (cf. Ref. 24 for references and discussion). Genes encoding ubiquitin take two forms (23, 24); polyubiquitin genes having a variable number of coding units arranged in tandem, in an uninterrupted, tail-to-head fashion (the polyubiquitin product is presumably processed to monoubiquitins at Gly–Met junctions (25)). The second class of genes encode a single ubiquitin fused to an extension protein (24). Two of the extension proteins have been characterized; the first is the 40 S ribosomal subunit protein S27a (2, 3) and the second is a large ribosomal subunit protein (4) whose identity, at least in mammalian ribosomes, has not yet been established. The expression of polyubiquitin genes is induced by heat shock and other stresses whereas the fusion genes provide basal levels of ubiquitin (24).

Ubiquitin has no intrinsic enzymatic activity rather it seems to function by marking other proteins. The best characterized of these functions is in protein degradation (26, 27). Ubiquitinated proteins may have either single ubiquitin moieties linked to 1 or more lysine residues in the acceptor protein or several ubiquitins can be attached sequentially to a protein at a single site to form branched ubiquitin-ubiquitin trees in which the carboxyl-terminal glycine of one ubiquitin is joined to lysine 48 of an adjacent ubiquitin (28). The attachment of such polyubiquitin chains to the substrate is critical for nonlysosomal degradation of the protein in a reaction that requires ATP and is catalyzed by specific proteolytic enzymes. Free monoubiquitin is a product of this reaction, and the protein can be reutilized (29). Besides its role in protein turnover ubiquitin may have a function in DNA replication (25), in progression through the cell cycle (30), and in a variety of stress responses (31, 32). In addition, ubiquitin may be conjugated to an acceptor protein, perhaps to modify its function without destabilizing it (a posttranslational modification of the protein that is comparable in some ways to phosphorylation). Acceptors of ubiquitin include histones H2A and H2B (33), actin (34), the lymphocyte homing receptor (35), the platelet-derived growth factor receptor (36), the growth hormone receptor (37), and the intracellular neurofibrillary tangles in neurodegenerative diseases such as Alzheimer’s (38–40).

In contrast to ubiquitin, whose sequence of amino acids is conserved, there is a family of ubiquitin-like proteins that deviate from each other, but have sufficient identity with ubiquitin to establish their relationship. The first of these is a protein that is induced by interferon and that has been given the trivial designation UCRP for ubiquitin cross-reactive protein (41, 42); its function is not known. UCRP has 145 amino acids (41, 42); both the NH2-terminal half has 21 identities in 72 possible matches (29%) whereas the carboxyl-terminal half has 31 identities in 83 possible matches (37%). It would appear that UCRP derives from a duplication of a ubiquitin, or a ubiquitin-like, gene.

A gene on the human X chromosome, located 40 kilobases downstream from the glucose-6-phosphate dehydrogenase gene, encodes a 157-amino-acid protein designated GdX (43). The NH2-terminal 74 residues of GdX share 43% identity with ubiquitin whereas the carboxyl-terminal portion is not related to ubiquitin. This arrangement resembles the organization of the ubiquitin-like-S30 fusion protein and led us to ask whether the carboxyl terminus of GdX is also a ribosomal protein. We were unable, however, to find any ribosomal protein in our library of 1,200 that has an amino acid sequence with significant identity with the carboxyl-half of GdX. Thus, it is unlikely that the carboxyl extension in GdX is a ribosomal
protein; moreover, GdX does not have the Gly-Gly dipetide that is the site of proteolytic processing of the ubiquitin and ubiquitin-like S30 fusion proteins. Finally, the structure of the GdX gene is distinct from that of the ribosomal proteins (44). It is, of course, possible that the carboxyl extension of GdX is a ribosomal protein but that no amino acid sequence for a related ribosomal protein has been determined as yet, and that processing is by a protease that recognizes a different amino acid sequence.

The baculovirus Autographica californica nuclear polyhedrosis virus (AcMNPV) encodes a protein, v-ubi, that has 77 amino acids and shares 76% identity with ubiquitin; 58 identities in 76 possible matches (45). A large number of viruses have ubiquitinated proteins (46); in tobacco mosaic virus one coat protein subunit/virion is ubiquitinated (47). The function of the ubiquitinated viral proteins is not known but their widespread occurrence suggests that they either play a role in the life cycle or that they are a common host response to the stress of infection. The circumstances in the baculovirus AcMNPV are different; v-ubi is not derived from the host but is encoded in the viral genome and it is not ubiquitin but a related protein. The function of v-ubi is unknown. It is expressed in the late phase of baculovirus infection, at the same time as many of the structural proteins, and this has led to the suggestion (45) that v-ubi serves as a molecular chaperone for incorporation of proteins into virus particles.

The several ubiquitin-like proteins associate with each other as would be expected since each is related to ubiquitin (Fig. 9). A pairwise alignment of the amino acid sequences of UbIL (the ubiquitin-like portion of the fusion protein with S30) with the others yields the following results: for UCRP there are 26 identities in 74 possible matches (35%); for v-ubi there are 25 identities in 74 possible matches (34%); and for GdX there are 19 identities in 74 possible matches (26%). The alignment scores for each of the pairwise comparisons are significant. The similarities are even more apparent when the inspection of a simultaneous alignment of the four proteins (UCRP, v-ubi, GdX, and UbIL) and conservative amino acid changes (arginine/lysine, glutamic acid/aspartic acid, threonine/serine, and leucine/isoleucine/valine) are tolerated (Fig. 9). At 16 of the 76 positions in the four proteins there is a similar residue in each of them (it is identical in 5 of these) and there are 24 positions where three of the residues are similar (at most of these 40 positions the residue in ubiquitin is similar or identical also).

The carboxyl-terminal tails of two ubiquitins are ribosomal proteins; a small and a large ribosomal subunit protein (2-4). The suggestion is that the ubiquitin moiety serves as chaperones in ribosome biogenesis (2). The evidence adduced for this proposal is not entirely convincing and the suggestion is not satisfying conceptually. It offers no rationale for the necessity for a covalently attached chaperone for the assembly into subunits of two ribosomal proteins when the 70-80 others appear to manage without assistance. Moreover, the proposal implies that cleavage of the fusion protein occurs on the ribosome after assembly. The site in the cell where the cleavage of the ubiquitin fusion proteins occurs is not known but we have shown here that the ubiquitin-like S30 polypeptide can be cleaved in a reticulocyte lysate in the absence of ribosome assembly. Moreover, no protease activity has been found among the ribosomal proteins although the proteolytic site might only associate transiently with the particle while it is being assembled in the nucleolus. The ubiquitin-like S30 protein clearly is analogous to the two ubiquitin-ribosomal proteins described earlier. Although, we have reservations concerning the proposal that the organization into a polypeptide favors, or is essential for, the assembly of the ribosomal protein into particles we cannot offer a reasonable alternative.

Determination of the sequence of amino acids in rat S30 is a contribution to a set of data which it is hoped will eventually encompass the structure of all the proteins in the ribosomes of this mammalian species. The primary purpose for the accumulation of this data is to use it to arrive at a solution of the structure of the organelle. However, the information 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.

Acknowledgments—We are grateful to Yuan-Ling Chan for advice and for valuable discussions, to William Suzuki for assistance with the computer comparisons, to Veronica Paz for technical help, and to Arlene Timsioscil for the preparation of the manuscript.

REFERENCES

17974

**Rat Ribosomal Protein S30**


