Independent Genes Coding for Three Acidic Proteins of the Large Ribosomal Subunit from Saccharomyces cerevisiae*

(Received for publication, November 16, 1987)

Miguel Remacha, M. Teresa Sáenz-Robles, M. Dolores Vilella, and Juan P. G. Ballesta
From the Centro de Biología Molecular, Consejo Superior de Investigaciones Científicas and Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

The yeast ribosome contains three acidic proteins, L44, L44', and L45, closely related from a structural point of view, that seem to play a functional role similar to that of proteins L7 and L12 in the bacterial ribosome.

By screening a cDNA bank in λgt11 with specific polyclonal and mononclonal antibodies, recombinant phages expressing each one of the acidic proteins have been cloned. A unique copy of each gene is detected using the phage cDNA inserts as probes in nitrocellulose blots of yeast DNA digested with different restriction enzymes. The inserts were subcloned in the plasmid pUC19, and their physical maps and nucleotide sequences were determined. By using the cDNA inserts as probes in genomic DNA banks, DNA fragments carrying the acidic protein genes have been cloned, characterized, and sequenced. The results conclusively show that the three yeast acidic proteins are coded by independent genes and are not the result of a post-translational modification of the product of a unique gene, as in bacteria. Like most ribosomal protein genes, the gene for protein L44' has an intron and two upstream stimulatory boxes (UAS

Primer has an intron and two upstream stimulatory boxes (UAS_primer) fitting closely to the consensus sequence. The genes coding for proteins L44 and L45 lack introns and seem also exceptional in other characteristics of their sequences.

Proteins L44 and L45 have amino acid sequences with about 80% similarity. Protein L44' is only 63% similar to the other two polypeptides. The three proteins have highly conserved carboxyl terminal comprising the last 30 amino acids, and the first 10 amino acids of L44 and L45 are identical.

The results cast doubts about the possibility of a similar role for the different acidic ribosomal proteins.

---

The presence of a set of very acidic proteins is a general feature of ribosomes from all organisms (Juan-Vidales et al., 1981a, 1981b; Ballesta, 1982). These proteins seem to be involved in the interaction of supernatant factors with the ribosome (Kisha et al., 1971; Ballesta et al., 1972; Hamel et al., 1973); although under particular conditions, some factor-dependent ribosomal activities can be performed in their absence (Ballesta and Vázquez, 1972; Hamel et al., 1972; Glick, 1977; Kotelyansky et al., 1977). In Escherichia coli, the two acidic proteins correspond to a unique polypeptide with the amino terminus either free (protein L12) or acetylated (protein L7). They are present as dimers in a complex with ribosomal protein L10, forming a characteristic protuberance of the large ribosomal subunit (see Möller and Maassen (1986) and Traut et al. (1986) for recent reviews).

In higher organisms, two phosphorylated acidic proteins are usually found associated to ribosomes (Juan-Vidales et al., 1981a, 1981b; Ballesta, 1982). In Artemia salina, acidic proteins eL12 and eL12' have been found encoded by independent genes (Maassen et al., 1985). In mammals, the proteins have been called P1 and P2, and their sequences are known in the case of rat and human livers either by protein (Lin et al., 1982) or cDNA (Rich and Steitz, 1987) sequencing studies. The genes for two acidic ribosomal proteins have also been reported in Drosophila melanogaster by independent groups which have called them A1 (Qian et al., 1987) and rp21c (Bigboldus, 1987). One acidic protein, called A1, has been described in Schizosaccharomyces pombe (Beltrame and Bianchi, 1987).

In Saccharomyces cerevisiae, the presence of three acidic proteins, L44, L44', and L45, has been reported (Sánchez-Madrid et al., 1979; Juan-Vidales et al., 1984). The three polypeptides can be monophosphorylated (Sánchez-Madrid et al., 1981a), and this modification affects drastically their affinity for the ribosome and their capacity to reassociate protein-deficient particles (Juan-Vidales et al., 1984). The three proteins have strong structural similarities, such as analogous molecular weights, close amino acid compositions, and common immunological determinants (Sánchez-Madrid et al., 1979; Juan-Vidales et al., 1981b, 1984). On the other hand, tryptic analysis shows differences among the three proteins that may be due to either regions of different amino acid sequence or post-translational modification of a common precursor. In the first case, yeast would resemble higher eukaryotic organisms and in the second case, bacterial cells. Supporting the second alternative is the fact that, as in bacterial proteins L7 and L12, the amino-terminal group of L45 is free, whereas it is blocked in L44 (Juan-Vidales et al., 1981b).

The amino acid sequence of a yeast acidic ribosomal protein, YPA1, has been reported (Itoh, 1980); but it is not known which of the three proteins reported here it corresponds to. This information then does not help to clarify the structural relationship among the three acidic polypeptides.

The availability of both polyclonal and monoclonal antibodies specific against the acidic ribosomal proteins of yeast

---

* This work was supported by institutional grants to the Centro de Biología Molecular from Fondo de Investigaciones Sanitarias and personal grants (to one of us) from Consejo Superior de Investigaciones Científicas. 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) J03780 and J03781.

---

1 The rat liver P1 sequence was provided by Y. L. Chan and I. G. Wool (personal communication).
encouraged us to attempt to clone their genes to study the
structural relationship of the three polypeptides. The results
showing the existence of independent genes for each acidic
protein and the differences among them are presented in this
report.

MATERIALS AND METHODS AND RESULTS

Physical Characterization of cDNA Inserts Coding for Acidic
Proteins—By screening a cDNA library prepared with poly(A)+
mRNA from yeast in λgt11 with specific polyclonal
probes, especially for inserts 242 and 302, strongly
indicated in Fig. 1. Polyadenine stretches are present in the
sequences of inserts 242 and 302, but not in that of insert 171.

Gene Copy Number for Acidic Proteins—The cDNA inserts
were used as probes in Southern blots of total yeast DNA
digested with HindIII (lanes 1) and HindII (lanes 2). Labeled cDNA
inserts 171, 242, and 302 were used as probes.

Cloning and Sequencing of Genes from Genomic DNA
Libraries—The labeled cDNA inserts were also used as probes
for screening genomic yeast DNA libraries in order to clone
the genes encoding the three acidic ribosomal proteins. Only
 clones hybridizing with insert 171 (L44) were obtained from an EcoRI DNA bank in λgt11. A genomic DNA insert of 2.3
 kb was subcloned into plasmid pUC18.

In the screening of a Sau3a DNA library in plasmid YEp13,
positive clones were obtained only in the case of insert 242.
A 5.9-kb DNA fragment was detected in the positive clones,
from which a 3.7-kb EcoRI piece containing the insert 242
hybridizing part was subcloned into plasmid pUC18.

To assure the cloning of the L44' gene, a minibank was
constructed by inserting yeast DNA fragments hybridizing
with insert 302 into the HindIII restriction site of plasmid
pUC18. The fragments were detected by Southern blots in the
1.7–2.2-kb region of HindIII-digested DNA-agarose gels (Fig. 2).
As expected, several positive clones containing inserts of 1.9 kb were isolated by screening the constructed library
with the insert 302.

The restriction enzyme maps of the genomic DNA inserts
subcloned into pUC18 were determined (Fig. 3).

Part of the inserts, including the cDNA hybridizing fragment
and its 3'- and 5'-flanking regions, was sequenced by the
dideoxy chain termination method using complementary
synthetic oligonucleotides as primers (Figs. 4–6). 891, 953,
and 1226 nucleotides were sequenced from genes correspond-
 ing to proteins L44, L44, and L44', respectively. The nucleo-
tide sequences found in the genomic DNA of the three genes
coincide with the sequenced cDNA inserts except at nucleo-
tides 23, 39, and 431 of the L45 gene and nucleotide 487 of
the L44' gene. These positions correspond to cytidines in the
cDNA probably due to errors made by the RNA reverse
transcriptase, but the correction in the genomic DNA does
not alter the amino acid sequence since the positions corre-
spond to third-base positions in the reading frame. The most
important difference in the cDNA sequence is the presence of
an intron containing 300 nucleotides starting at position
114 of the L44' gene. The intron has all the characteristic
sequences found in the intervening regions of yeast ribosomal
protein genes (Planta et al., 1986).

The sequenced 3'-flanking region covers about 160 nucleo-
tides including the polyadenylation site, whereas the 5'-
flanking region ranges from 415 to 460 nucleotides and con-

3 The abbreviations used are: kb, kilobase; IPTG, isopropyl-β-D-
thiogalactoside; mAβ, monoclonal antibody; SDS, sodium dodecyl
sulfate; BSA, bovine serum albumine; PIPES, piperoxone-N',N'-bis(2-
ethanesulfonic acid); CETAB, N-cetyl-N,N,N-trimethylammonium
bromide; HPLC, high pressure liquid chromatography; ELISA, en-
zyme-linked immunosorbent assay.

![Fig. 2. Southern blot of S. cerevisiae Y166 DNA digested with HindIII (lanes 1) and HindII (lanes 2). Labeled cDNA inserts 171, 242, and 302 were used as probes.](image-url)
Genes for Acidic Ribosomal Proteins in Yeast

Fig. 4. Sequence of DNA fragment containing gene coding for protein L45 and its flanking regions. The nucleotides have been numbered starting at the first adenine in the translation initiation codon, which is double underlined, as is the termination codon at position 331. The putative UAS boxes, starting at nucleotide -365 and -241, respectively, have also been marked, with their orientations indicated by horizontal arrows. The polyadenylation signal in the 3'-noncoding region is also underlined. Vertical arrows indicate the transcription start sites (positions -79 to -80) and the polyadenylation site (position 481) in the 5'- and 3'-flanking regions, respectively.

Fig. 5. Nucleotide sequence of DNA fragment coding for protein L44' and its flanking regions. Numbering and symbols are as described for Fig. 4; and, in addition, the splice junctions (positions 114 and 414) and the TACTAAC box (positions 349-358) in the intron have been marked by broken lines. UAS boxes start at positions -279 and -257; transcription starts around nucleotide -113, and the polyadenylation site is at position 718.

Table I

| Gene | Direct Inverted Sequences | Inverted Direct Sequences |
|------|---------------------------|---------------------------|
| L44  | hR 48, ACTGAA, AT -40     | -264, CAACTAT, -160       |
|      | ATTTAA, TTTATG, -111      |                           |
|      | TACTAC 307, GCTATAT, -368 |
|      | ATTGAC, CATTTG, -426      |
| L44' | AAATGT, ATATAG -353       | -131, TACCACCTA, 233      |
|      | TACCACT, ATATAC -352      |                           |
|      | GCACGTC, GGAGG -103       |
|      | AATATAT, TAGTAT -123      |
|      | TTGGTTA, TTATTT, 268      |

Amino Acid Sequence of Three Acidic Proteins—As deduced from the nucleotide sequences of their genes, proteins L44, L44', and L45 contain 106, 106, and 110 amino acid residues and have molecular weights of 12,747, 12,805, and 12,993, respectively. The amino acid sequences of the three polypeptides have been compared using a sequence comparison program from the Wisconsin University Biotechnology Center software package (Fig. 8). Proteins L44 and L45 are about 80% similar to each other, but only 63% similar to L44'. The main bands were found in the three cases (Fig. 7), which have been marked with vertical arrows in the respective sequences (Figs. 4-6). 5'-Leader regions of approximately 113, 80, and 13 nucleotides were found in mRNA corresponding to proteins L44', L44, and L45, respectively.
The higher expression of the acidic proteins as compared with probably explains the difficulties encountered in detecting the somal proteins, generally encoded in more than one gene copy by the anti-acidic protein antibodies. This rapid degradation fact distinguishes the acidic proteins from most other ribosomal proteins. L45 is detected by Southern blot analysis of DNA. This fusion protein in lysates from liquid media-grown cells.

The three acidic proteins of the S. cerevisiae ribosome, proteins L44, L44', and L45, are expressed separately from an in the absence of the inducer isopropyl-β-D-thiogalactoside. It is interesting to note that, in spite of the fact that these fused proteins are under the control of the lacZ gene promoter (Young and Davis, 1983), expression of the polypeptides takes place in the absence of the inducer isopropyl-β-D-thiogalactoside. The functionality of the promoter is, however, clear in the yeast genome (Fried et al., 1985), and at the same time interesting fact since different biochemical results suggest that the cytoplasmic pool seems to be lower for protein L44' than for proteins L44 and L45, which do not have introns in their genes. However, other ribosomal proteins coded by intron-free genes (proteins L3, L16, S24, and S33) do not seem to accumulate in the cell, indicating that additional mechanisms must control the expression of ribosomal proteins. Moreover, splicing control does not work in all intron-containing genes (Warner et al., 1985).

Ribosomal protein genes are also characterized by the usual presence of two conserved sequences (UASex boxes) which are separated by no more than 40 nucleotides, at higher than 20 nucleotides upstream from the translation initiation site, and which act as transcription activators (Leer et al., 1985; Rothenberg and Woolford, 1986). Using a sequence-fitting computer program, two boxes that fit the UASex consensus sequence (Woudt et al., 1987) can be found in the three genes (Figs. 4-6). The quality of the fit is different for all three cases (high for L44', reasonable for L44, and poor for L45), meaning that, probably in the last case, the sequences do not function as activators. In the L44' gene, the sequences are present in the coding strand; whereas in the L44 and L45 genes, the boxes are in different strands in opposite orientation and, in the case of L45, separated by more than 100 nucleotides.

It seems therefore that, also from the point of view of the UASex boxes, protein L44' is closer to the typical ribosomal protein genes than the other two proteins. This is a very interesting fact since different biochemical results suggest that protein L44' seems to function like a typical ribosomal component, contrary to proteins L44 and L45. Thus, in addition to the different cytoplasmic pools of the three proteins discussed above, we have found that protein L44' requires a higher salt concentration to be removed from the ribosomal particle.

It is interesting to note that the ribosomal protein genes

---

**Fig. 8. Comparison of amino acid sequences of S. cerevisiae acidic ribosomal proteins as deduced from their nucleotide sequences.** Comparison was carried out using the GAP program for sequence comparison in the GCG software from the Wisconsin University Biotechnology Center.

codon usage in the three genes show the typical yeast bias corresponding to a highly expressed protein (Sharp et al., 1986).

### DISCUSSION

The three acidic proteins of the *S. cerevisiae* ribosome, proteins L44, L44', and L45, are expressed separately from recombinant λ phages containing cDNA inserts. These results exclude the possibility, suggested by the strong physicochemical similarity of the three polypeptides (Juan-Vidales et al., 1986), that they were the result of post-translational modification of a common precursor, as in bacteria. This situation is consistent with data from higher eukaryotic cells, where genes coding for each ribosomal acidic polypeptide are detected (Maasen et al., 1985).

The three proteins are expressed from recombinant λ phages as fusion polypeptides joined to β-galactosidase. It is interesting to note that, in spite of the fact that these fused proteins are under the control of an important result is the presence of an intervening region only in the gene encoding protein L44'. The intron has splice junctions that correspond to the consensus sequences found in eukaryotic spliced genes (Mount, 1982) and that are common to other yeast ribosomal proteins (Planta et al., 1986; Mager, 1988). The TACTAAC box required for splicing in yeast (Pikielny et al., 1983; Langford et al., 1984) is present in this intron, too.

The presence of introns in the genes for yeast ribosomal proteins is the rule (Fried et al., 1981; Bollen et al., 1982), although exceptions have been reported, such as in the case of proteins L3 (Schultz and Friesen, 1983), L16 (Leer et al., 1984), S24 (Leer et al., 1986b), and S33 (Leer et al., 1983). From this point of view, protein L44' falls in the group of standard ribosomal proteins, whereas proteins L44 and L45 induce the number of exceptions.

The regulation of mRNA splicing seems to be a mechanism for controlling the expression of some ribosomal proteins (Warner et al., 1985). Proteins without introns in their genes would be free from this control and, in principle, able to accumulate in higher amounts. The acidic proteins are present in multiple copies in the yeast ribosome (Kruiswijk et al., 1978; Juan-Vidales et al., 1984); and, in addition, an important pool of acidic ribosomal proteins is found in the cytoplasm. It is interesting to note that the cytoplasmic pool seems to be lower for protein L44' than for proteins L44 and L45, which do not have introns in their genes. However, other ribosomal proteins coded by intron-free genes (proteins L3, L16, S24, and S33) do not seem to accumulate in the cell, indicating that additional mechanisms must control the expression of ribosomal proteins. Moreover, splicing control does not work in all intron-containing genes (Warner et al., 1986).

Ribosomal protein genes are also characterized by the usual presence of two conserved sequences (UASex boxes) which are separated by no more than 40 nucleotides, at higher than 20 nucleotides upstream from the translation initiation site, and which act as transcription activators (Leer et al., 1985; Rothenberg and Woolford, 1986). Using a sequence-fitting computer program, two boxes that fit the UASex consensus sequence (Woudt et al., 1987) can be found in the three genes (Figs. 4-6). The quality of the fit is different for all three cases (high for L44', reasonable for L44, and poor for L45), meaning that, probably in the last case, the sequences do not function as activators. In the L44' gene, the sequences are present in the coding strand; whereas in the L44 and L45 genes, the boxes are in different strands in opposite orientation and, in the case of L45, separated by more than 100 nucleotides.

It seems therefore that, also from the point of view of the UASex boxes, protein L44' is closer to the typical ribosomal protein genes than the other two proteins. This is a very interesting fact since different biochemical results suggest that protein L44' seems to function like a typical ribosomal component, contrary to proteins L44 and L45. Thus, in addition to the different cytoplasmic pools of the three proteins discussed above, we have found that protein L44' requires a higher salt concentration to be removed from the ribosomal particle.

It is interesting to note that the ribosomal protein genes

---

* M. T. Sáenz-Robles, S. Zinker, M. Remacha, M. D. Vilella, and J. P. G. Ballesta, manuscript in preparation.
without introns either lack UAS$_{up}$ boxes (like proteins L3 and S33) or have apparently less efficient ones (like proteins S24, L45, and L44). Proteins L16 is the exception, having standard UAS$_{up}$ regions in the two unspliced gene copies present in the genome. The T-rich region detected (in the ribosomal protein genes) closer to the transcription initiation site than to UAS$_{up}$ is present in the three genes, too (Rotenberg and Woolford, 1986).

The transcription initiation sites have been located by nuclease S1 mapping at about 113, 80, and 13 nucleotides upstream of the translation initiation site in L44' and L45 and L44 genes, respectively. The L44' and L45 initiation sites are in the sequence Pyr-Ado-Ade-Pur, common to other yeast genes; whereas L44 initiates in the sequence AACCAA present in some, but not all, ribosomal protein genes (Planta et al., 1986). The size of the 5'-leader region is noticeably long in the case of the protein L44' gene since the reported extension of this flanking region does not exceed 40 nucleotides in other ribosomal protein genes (Planta et al., 1986). In the first two genes, the 3'-polyadenylation sites have also been identified from the sequence of their cDNA clones and are preceded at adequate distance by the presumed polyadenylation signal AATAAA (Fitzgerald and Shenk, 1981). In spite of having a AATAAA signal, a polyadenine track was not found in the cDNA from the L44 gene, which might reflect a possible mistake in the preparation of the cDNA bank.

Some other features in the nucleotide sequence of the three genes are summarized in Table I. A number of hairpin loops can be formed in the 3' and 5'-flanking regions of the protein L44 gene; and, in one case, the same stretch (nucleotides −381 to −390) can be part of two different stem-loops. These structures are rare, however, in the genes of proteins L44' and L45, although they are found in the intervening region of L44'. It is interesting to note the presence of two sets of direct complementary sequences of 12 and 9 nucleotides each which can stabilize a circular loop in the 5'-leader region of the L44' gene. The biological meaning of all these structures is difficult to assess at this moment, but its implication in transcriptional or translational control should not be disregarded.

The amino acid sequences of each of the three proteins deduced from the gene nucleotide sequence have been compared. The three polypeptides share an almost identical carboxyl terminus, including the last 30 amino acids. The similarity at the amino terminus is also high in proteins L44 and L45, which have the same first 10 amino acid residues. Protein L44' has a totally different amino-terminal end. Protein L45 has practically the same amino acid sequence as ribosomal protein A1, reported by Itoh (1980) from amino acid sequencing studies.

Presently, the amino acid sequences of the acidic ribosomal proteins from five different organisms, including A. salmon (Maassen et al., 1985), D. melanogaster (Qian et al., 1987; Bigboldus, 1986), rat liver (Lin et al., 1985), human liver (Rich and Steitz, 1987), and S. cerevisiae (this report), are known. In the case of S. pombe, only one acidic protein has been reported (Beltrame and Bianchi, 1987). When the amino acid sequences of all of them are compared, they can be divided easily in two groups whose components have, on the average, around 70% similarity in their sequences. The similarity with the components of the other group is about 55%. One group is formed by mammal protein P2, A. salmon eL12, D. melanogaster A1, S. pombe A1, and S. cerevisiae L44 and L45; the other group is formed by mammal P1, A. salmon eL12', D. melanogaster rp21c, and S. cerevisiae L44'. Both protein groups share a highly conserved carboxyl terminus, with the same last 10 amino acids, but differ in the amino terminus, which is conserved among the members of each group. S. cerevisiae protein L44' is an exception since its amino-terminal end does not correspond to the rest of proteins in its group.

S. cerevisiae is unique in having two different proteins, L44 and L45, in the same group. Preliminary data indicate that the yeast Hansenula anomala and the filamentous fungi Geotrichum lactis have, like S. cerevisiae, three ribosomal acidic proteins when checked by isoelectricfocusing, suggesting that this might be a characteristic of lower eukaryotes. Additional evidence is required to confirm these results and to show that, in S. cerevisiae, two out of the three proteins belong to the same structural group, as well as to determine the functional significance of this fact.

The presence of an identical C terminus suggests the existence of a common interacting site in the ribosome for the different polypeptides. If this is so, the eukaryotic acidic proteins would interact with the particle in the opposite way as their bacterial counterparts, which have been shown to bind through their amino terminus (Liljas, 1982). However, this fact might be explained by the existence of a transposition during evolution affecting the ends of the molecule, which postulated by different laboratories (Lin et al., 1982; Otaka et al., 1985). Alternatively, the eukaryotic acidic proteins may interact with the ribosome through their amino terminus in different ribosomal sites. This possibility would be compatible with the reported location of the acidic proteins at different regions in the ribosome (Traut et al., 1986).

The structural similarity between the bacterial and eukaryotic acidic proteins, stressed by their functional interchangeability (Sánchez-Madrid et al., 1981b), has favored the idea that, as in bacteria, the different types of acidic proteins are carrying out an identical function in the eukaryotic ribosome. The data available now on the structure of the acidic proteins in higher organisms cast some doubts on this concept and raise the possibility of a different role for each type of acidic protein.

Additional information is required to clarify the real function of the acidic ribosomal proteins in eukaryotic cells, and the availability of their genes in a genetically accessible organism such as S. cerevisiae will help in that direction.

Acknowledgments—We thank M. C. Fernández Moyano for expert technical assistance, Rosario Vacas for help in performing some experimental work, Dr. S. Zinker for reading and commenting on the manuscript, and Drs. Y. L. Chan and I. G. Wool for supplying the sequence of rat acidic protein P1 prior to publication. We are grateful to Drs. Altman, Trachsel, and Iñáez for their help and for the gift of yeast DNA libraries.

REFERENCES

Ballesta, J. P. G., and Vázquez, D. (1972) FEBS Lett. 28, 337–342
Ballesta, J. P. G., Montejo, V., and Vázquez, D. (1972) FEBS Lett. 19, 79–82
Beltrame, M., and Bianchi, M. E. (1987) Nucleic Acids Res. 15, 9089
Bielka, H. (1982) The Eukaryotic Ribosome, Springer-Verlag, Berlin, and references therein
Bigboldus, J. D. (1987) Nucleic Acids Res. 15, 1064
Bollen, G. H. P. M., Molenaar, C. M. T., Cohen, L. H., van Raamsdonk-Duin, M. M. C., Mager, W. H., and Planta, R. J. (1982) Gene (Amst.) 18, 29–37
Breschamps, J., Hildreth, J. E., Derr, D., and August, J. T. (1985) Anal. Biochem. 147, 451–454
Fitzgerald, M., and Shenk, T. (1981) Cell 24, 251–260
Fried, H. M., Pearson, N. J., Kim, C. H., and Warner, J. R. (1981) J. Biol. Chem. 256, 10176–10183

5 M. Remacha, M. T. Sáenz-Robles, M. D. Vilella, and J. P. G. Ballesta, unpublished results.
6 S. Zinker, personal communication.
The cDNA and genomic libraries in the vector Aqtll (Young and Davis, 1983b) were gifts of Centro de Genetica Molecular, Netherlandsl. DNA ligase, DNA polymerase etc. were obtained from Boehringer Mannheinl and DNA sequencing by the dideoxy chain termination method as described (Sanger et al. 1977).

The amino acid sequences of the different acidic proteins were compared in the Genome project from the Genetic Group software package of the Wisconsin University Biotechnology Center using a gap weight of 0.1 and a gap length weight of 8.0.

Results

Isolation of Phages Expressing Yeast Ribosomal Proteins

The cDNA clones were sequenced from partial cDNA clones and DNA sequence analysis of the inserts was determined to confirm that the inserts were of yeast origin. The inserts were sequenced from both strands and the consensus sequences were then used to design primers for the amplification of the corresponding genes.

Characterization of the Protein

The protein expressed by the recombinant phages was found to have a mobility on SDS-PAGE gels that was consistent with the predicted molecular weight of the protein. The protein was also found to be reactive with antibodies raised against the native protein.

Identification of the Fusogenic Peptides

The site of insertion for the foreign DNA in vplll is located 53 base pairs upstream of the termination codon in the lacZ gene (Young and Davis, 1983). The recombinant phages will therefore generate a protein made of the 5'-flanking nucleotides minus the last 18 amino acids at the carboxy terminal end. Purified to the protein coded by the insert. The fusogenic peptide would be the core of the fusogenic protein in the recombinant phages. Since the intact fusogenic protein in the recombinant phages contains approximately 125,000 dalton, it is expected that the attempt to identify the fusogenic protein in lysates of infected cells using a SDS-PAGE gel is unsuccessful, in spite of the fact that the intact fusogenic protein in the recombinant phages is a monomeric protein of 30,000 daltons. This procedure was therefore repeated using a plasmid (pJH2) isolated from E. coli, which contained the foreign DNA inserted at the SalI site, and the fusogenic protein was then purified to electrophoresis on a 12.5% SDS-PAGE gel. The fusogenic protein in the recombinant phages is a monomeric protein of 30,000 daltons. Protein A is expressed in the staphylococcal system, and protein B is expressed in the E. coli system. Protein A is expressed by the fusion protein in the recombinant phages and it is of interest to note that expression of the fused fusogenic protein takes place as the intact fusogenic protein (Fig. 12C). The fusogenic protein was characterized using antibodies specific for the acidic protein and d-galactosidase. Only in the extracts from the recombinant phages does the anti-acidic protein serum recognize the intact fusogenic protein, which is also recognized by the anti-d-galactosidase serum cross-reacts, as expected, with the tagged fusogenic protein in the recombinant phages but not in the parental phages.

This protein is not detected, however, by the anti-ribosomal protein antibodies (Fig. 12B), suggesting that it may correspond to a degradation product of the fusion protein lacking the acidic protein part of the sequence.
Figure 10. Characterization of the acidic protein expressed by different recombinant plasmids. A rabbit anti-acidic protein serum was absorbed on E. coli strains harboring the plasmids pAF111 (A), pAF3032 (B), pAF3042 (C), and pAF311 (D). The bound immunoglobulins and the unbound supernatant serum were used for treatment of acidic protein electrophoresis gels. The original antisera was included as a control (E). A shows a stained copy of the gel indicating the position of phosphorylated and dephosphorylated forms of the acidic protein.

Figure 11. Detection of the fusion protein expressed by the recombinant plasmids. Extracts from E. coli were separated by SDS-PAGE, stained as in Figure 3, and analyzed for protein expression. The blot was probed with antibodies against the acidic ribosomal proteins. The positions of the acidic ribosomal proteins are indicated on the gel. The original antisera was included as a control (A).