Dever, T. E., Glynias, M. J., & Merrick, W. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1814–1818.
Fowler, A. V., & Zabin, I. (1983) *J. Biol. Chem. 258*, 14354–14358.
Germino, J., & Bastia, D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4692–4696.
Gilman, A. G. (1987) *Annu. Rev. Biochem. 56*, 615–649.
Hershey, J. W. B. (1987) in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology* (Ingraham, J., Low, K. B., Magasanik, B., Shaechter, M., Petersen, H. U., & Neidhart, F. C., Eds.) pp 613–647, American Society of Microbiology, Washington, DC.
Hirel, P. H., Lêvêque, F., Mellot, P., Dardel, F., Panvert, M., Mechulam, Y., & Fayat, G. (1988) *Biochimie 70*, 773–782.
Holbrook, S. R., & Kim, S. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1751–1755.
Jurnak, F. (1985) *Science 230*, 32–36.
Kolakofsky, D., Dewey, K. F., Hershey, J. W. B., & Thach, R. E. (1968) *Proc. Natl. Acad. Sci. U.S.A. 61*, 1066–1071.
La Cour, T. F. M., Nyborg, J., Thirups, S., & Clark, B. F. C. (1985) *EMBO J. 4*, 2385–2388.
Leberman, R., & Egner, U. (1984) *EMBO J. 3*, 339–341.
Lelong, J. C., Grunberg-Manago, M., Dondon, J., Gros, D., & Gros, F. (1970) *Nature 226*, 505–510.
March, P. E., & Inouye, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7500–7504.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035–10038.
Minton, N. P. (1984) *Gene 31*, 268–273.
Morel-Deville, F., Vachon, G., Sacerdot, C., Cozzone, A. J., Grunberg-Manago, M., & Cenatiempo, Y. (1990) *Eur. J. Biochem. 188*, 605–614.
Niu, C. H., Han, K. H., & Roller, P. P. (1989) *Biochem. Biophys. Res. Commun. 160*, 282–288.
Parmeggiani, A., Swart, G. W. M., Mortensen, K. K., Jensen, M., Clark, B. F. C., Dente, L., & Cortese, R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3141–3145.
Pfeuffer, T. P., & Helmreich, E. J. M. (1988) *Curr. Top. Cell. Regul. 29*, 129–217.
Plumbridge, J. A., Deville, F., Sacerdot, C., Petersen, H. U., Cenatiempo, Y., Cozzone, A. J., Grunberg-Manago, M., & Hershey, J. W. B. (1985) *EMBO J. 4*, 223–229.
Pon, C. L., Paci, M., Pawlik, R. T., & Gualerzi, C. (1985) *J. Biol. Chem. 260*, 8918–8924.
Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A., & Grunberg-Manago, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7787–7791.
Shiba, K., Ito, K., Nakamura, Y., Dondon, J., & Grunberg-Manago, M. (1986) *EMBO J. 5*, 3001–3006.
Travers, A. A., Debenham, P. G., & Pongs, O. (1980) *Biochemistry 19*, 1651–1656.
Ullmann, A. (1984) *Gene 29*, 27–31.

Ribosomal Protein L35: Identification in Spinach Chloroplasts and Isolation of a cDNA Clone Encoding Its Cytoplasmic Precursor†

Peter M. Smooker,‡ Theodora Choli, and Alap Raman Subramanian*
Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33, Federal Republic of Germany

Received April 10, 1990; Revised Manuscript Received June 4, 1990

ABSTRACT: We describe the isolation of spinach chloroplast ribosomal protein L35 and characterization of a cDNA clone encoding its cytoplasmic precursor. This protein was only recently identified in ribosomes, but the sequences of four L35 genes have now been reported and confirm its presence in eubacteria, chloroplasts, and cyanelles. Using N-terminal sequence data, oligonucleotides were designed and a cDNA library was screened. The nucleotide sequence of the cDNA clones shows that the spinach L35 protein is encoded as a precursor of 159 residues, comprising a mature protein of 73 residues and a transit peptide of 86 residues. The cleavage site for forming the mature protein is deduced to be Thr-Val-Phe-Ala↓Ala-Lys-Gly-Tyr. The L35 protein in the photosynthetic organelle of the protozoan *Cyanophora paradoxa* is encoded in the organelle DNA [Bryant & Stirewalt (1990) *FEBS Lett. 259*, 273–280]. The corresponding gene has not been found in the chloroplast DNA of a lower plant (liverwort) and two higher plants. Our results demonstrate that the L35 protein in a higher plant (spinach) is encoded in the nucleus. This finding, in light of the endosymbiont hypothesis, suggests an organelle to nucleus transfer of the L35 gene at the evolutionary beginnings of land plants.

The chloroplast ribosome is eubacterial in type (Boynton et al. 1980), but the genes encoding ribosomal proteins (r-proteins) are distributed in two cellular compartments: 20 or 21 r-protein genes have been identified in the chloroplast genomes of three land plants (Shinozaki et al., 1986; Ohyama et al., 1986; Hiratsuka et al., 1989). The remaining r-proteins (≥40) are all assumed to be nuclear-coded and imported into the organelle from the cytoplasm. The endosymbiont hypothesis (Bogorad, 1975; Gray, 1989) proposes that these nuclear genes encoding chloroplast r-proteins were originally located in the prokaryotic endosymbiont and have subsequently transferred to their present location. We are studying the adaptions which the genes encoding chloroplast r-proteins have undergone to ensure efficient and coordinate expression from their nuclear locations [see review: Subramanian et al. (1990)].

The isolation and characterization of cDNA clones for r-proteins L12, L13, L21, and PSrp-1 have been previously reported from our laboratory (Giese & Subramanian, 1989;

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02928.

* Author to whom correspondence should be addressed.
‡ Present address: The Murdoch Institute, Royal Children's Hospital, Parkville, 3054 Victoria, Australia.

0006-2960/90/0429-9733$02.50/0

Phua et al., 1989; Smooker et al., 1990; Johnson et al., 1990). Here we report characterization of both the L35 protein, only recently definitively identified as a component of the *Escherichia coli* ribosome (Wada & Sako, 1987; Kashiwagi & Igarashi, 1987), and a cDNA clone that encodes its cytoplasmic precursor. The significance of the finding that the gene for chloroplast r-protein L35 is located in the nucleus is discussed.

Experimental Procedures

Chloroplast ribosomes were isolated from freshly grown spinach leaves (*Spinacia oleracea*, cultivar Alwaro) and r-proteins separated into pools of similar molecular size (Bartsch et al., 1982; Kamp et al., 1987). Pool G of this experiment [Figure 2, Kamp et al. (1987)] was fractionated by using a 1 × 20 cm column of carboxymethyl-Sephadex C-25, and collected fractions were analyzed by SDS–PAGE. One set of fractions (named pool G-61) consisted predominantly of an 8-kDa polypeptide. This polypeptide was purified to homogeneity by HPLC on a 0.4 × 25 cm column of Nucleosil 3005 C18 (Machery and Nagel) by elution with a gradient of 2-propanol in 0.1% trifluoroacetic acid.

NH_2-terminal sequencing was done on an Applied Biosystems pulsed liquid-phase sequencer (Model 447A) equipped with a Model 120A PTH-amino acid analyzer.

Oligodeoxynucleotides were synthesized on an Applied Biosystems DNA synthesizer (Model 380A). For screening purposes, the oligonucleotides were labeled with ^{32}P by using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (Boehringer Mannheim). The construction of a spinach cDNA library has been described previously (Giese & Subramanian, 1989). The library was screened according to standard procedures, as described (Giese & Subramanian, 1989). The temperature of hybridization was 5 °C below the T_m calculated for the designed oligonucleotide.

DNA manipulations were by standard procedures (Maniatis et al., 1982). λ DNA was prepared (Patterson & Dean, 1987), and cDNA inserts were cleaved by *Eco*RI digestion and subcloned into the plasmid vector pT7/T3-19U (Pharmacia). The nucleotide sequence of the insert was determined by the dideoxy chain termination method (Sanger et al., 1977), modified for use with double-stranded templates (Chen & Seeburg, 1985). Synthetic oligonucleotides, designed from previous rounds of sequencing, were used to prime chain extension using a Pharmacia sequencing kit.

Computer operations were done on a VAX 8600/VMS computer, using the UWGCG (Devereux et al., 1984) suite of programs. For identification of homologous sequences of the NBRF database and the RIBO database (of this Institute) were searched.

Results

Identification of a Homologue of E. coli L35 in Chloroplast Ribosomes. Spinach chloroplast ribosomal proteins were prepared as described (Experimental Procedures), and an individual r-protein was purified by HPLC as shown in Figure 1. The purified material was subjected to automated NH_2-terminal amino acid sequencing and analyzed to 49 cycles with the result:

```
1                   10                  20
A-K-G-Y-K-M-K-T-H-K-A-S-A-K-R-F-R-V-T-G-K-

                    30                  40
G-K-I-V-R-R-( )-A-G-K-Q-H-L-( )-A-K-K-N-T-

                    49
K-( )-K-N-( )-L-( )-K-L-
```

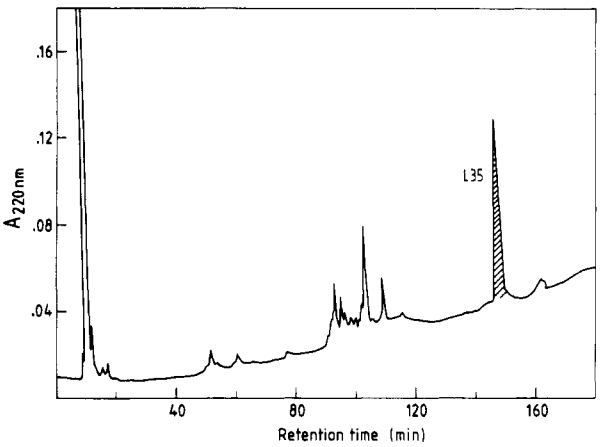

FIGURE 1: Purification of spinach chloroplast ribosomal protein L35. A sample (pool G-61) from a previous fractionation (see Experimental Procedures) was applied to a 0.4 × 25 cm column of Nucleosil 300-5-C18 through a 3-cm precolumn of the same material. Proteins were eluted with a gradient of $H_2O \rightarrow$ 2-propanol (both containing 0.1% trifluoroacetic acid).

The sequence obtained (44 residues) was compared to the protein sequences in the NBRF and RIBO databases: it showed 45% identity with the *E. coli* ribosomal protein A (Wada & Sako, 1987), subsequently named L35 (Pon et al., 1989).

The complete nucleotide sequences of the chloroplast DNA from three plants have been reported, but none of them contains a reading frame encoding a homologue of *E. coli* L35 (Shinozaki et al., 1986; Ohyama et al., 1986; Hiratsuka et al., 1989). Since our protein data pointed to the existence of a L35-like protein in the chloroplast ribosome, we inferred that the gene encoding this chloroplast protein would be nuclear. Therefore, a spinach cDNA library (Giese & Subramanian, 1989) was screened with a mixed oligonucleotide probe designed from the known amino acid sequence.

The oligonucleotide sequence used, and the amino acid sequence from which it was derived, is

```
                    4                          10
AA sequence         Y   K   M   K   T   H   K

Oligonucleotide 3'  ATA TTT TAC TTT TGA GTA TT  5'
                      G   C     C   G   G
                                  T
                                  C
```

The mixed oligonucleotide consists of 64 isomers. The hybridization temperature used was 41 °C (T_m −5 °C), T_m being calculated for the isomer of the lowest G + C content (Suggs et al., 1981).

Of a total of 150000 recombinant phages screened, approximately 20 positive signals were obtained. Individual positive plaques were purified by three rounds of rescreening, and phage DNA was isolated. The nucleotide sequence of the 750 bp insert was determined as described under Experimental Procedures. The complete sequence, obtained from a series of overlapping sequence determinations of the insert, is shown in Figure 2. The insert was found to be 753 nucleotides long, with the last 13 nucleotides comprising the poly(A) tail.

The nucleotide sequence contained a reading frame capable of encoding a polypeptide of 159 amino acid residues. This included a perfect match of the 44 NH_2-terminal amino acid residues previously determined (Figure 2). These 44 amino acids match the reading frame beginning at residue 87. The reading frame thus encodes a precursor form that includes a large N-terminal presequence. The mature L35 protein is predicted to be 73 amino acids in length, with a molecular mass

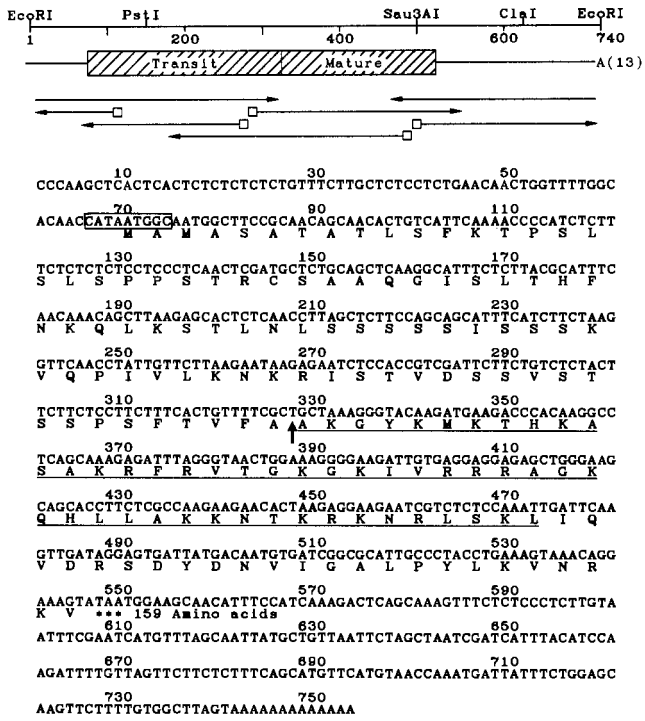

FIGURE 2: Nucleotide sequence of the cDNA clone encoding r-protein L35 and the sequencing strategy. Arrows with rectangles indicate sequencing with synthetic oligonucleotides designed by using data from previous rounds of sequencing. The bold arrow in the protein sequence indicates the point at which the transit peptide is removed. The experimentally determined N-terminal amino acids of the mature L35 protein are underlined, and the context of the putative initiating ATG codon is boxed.

of 8434 Da. The protein would be highly basic, having a net charge of +21/+23 with nonprotonated/protonated histidines.

As the N-terminus of the mature protein is known, the site at which the precursor protein is cleaved is also known. The relatively long presequence of 86 amino acids has an amino acid composition similar to that of known transit peptides of chloroplast r-proteins (Giese & Subramanian, 1989; Phua et al., 1989; Smooker et al., 1990), exhibiting high proportions of serine and threonine, few or no acidic residues, and no tryptophan or tyrosine.

The putative initiating methionine in the cDNA of the r-protein L35 (CATAATGGC, Figure 2) is in the conserved initiation context for plant nuclear genes (Lütcke et al., 1987), with an A at position −3 and G at +4 (the A of the ATG is taken as +1). A second ATG codon only two codons from the first ATG is also found, as in many nuclear-coded chloroplast r-protein genes. It has been recently shown (Giese & Subramanian, 1990) in the case of r-protein L12 that initiation of translation also occurs from the downstream ATG codon.

A typical plant polyadenylation signal (Heidecker & Messing, 1986) is not evident in the 5′ region preceding the poly(A) tail (Figure 2), but a sequence with some variation (AAAUGAU) is present after nucleotide 700. It appears [see Phua et al. (1989) and Heidecker and Messing (1986)] that polyadenylation signals in plants have less stringent requirements than in the animal system, where the signal AAUAAA is almost absolutely conserved.

The alignment of chloroplast L35 with that of *E. coli* (Wada & Sako, 1987), *Bacillus stearothermophilus* (Pon et al., 1989), and *Cyanophora paradoxa* (Bryant & Stirewalt, 1990) is shown in Figure 3. Identity of mature chloroplast L35 with that in the cyanelle of *C. paradoxa* is 46%, and with that of the two eubacterial species is 42%. The important difference is that the spinach L35 protein is synthesized as a precursor molecule with an N-terminal extension. Such a presequence is absent not only in the eubacterial L35 proteins but also in the organelle-encoded L35 protein of the photosynthetic protozoan, *C. paradoxa*.

Discussion

We have described the purification and N-terminal sequencing of ribosomal protein L35 from chloroplast ribosomes of a higher plant (*Spinacia oleracea*; spinach). The isolation and sequencing of a cDNA clone encoding the precursor form of this protein is also reported. The precursor protein (as derived from the nucleotide sequence) is 159 amino acid residues long, in contrast to the mature protein which has only 73 residues (M_r 8434). The cleavage site in the precursor, i.e., TVFA↓AKGY, has been precisely determined from the N-terminal sequence of the purified protein. The 86-residue N-terminal part in the precursor has the characteristic amino acid composition of transit peptides that target cytoplasmically synthesized proteins into chloroplasts. By isolating cDNA clones having a transit peptide coding region and a poly(A) tail, we show that the gene for chloroplast r-protein L35 is located in the nucleus. This result complements the previous finding that a coding sequence for r-protein L35 is absent in the sequenced chloroplast DNA of three land plants [reviewed

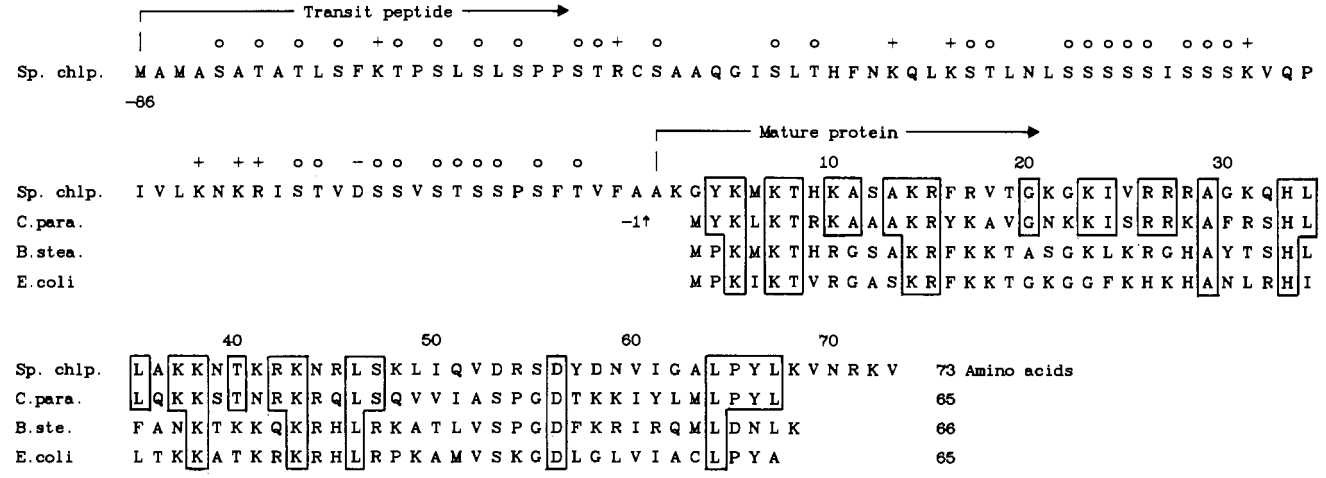

FIGURE 3: Alignment of the L35 protein sequence of spinach chloroplast with that from *C. paradoxa* (C.para.), *B. stearothermophilus* (B.ste.), and *E. coli*. Regions of amino acid identity in the two orgenelle-located proteins (Sp. chlp. and C.para.) or in all four proteins are boxed. Serine, threonine, and the charged amino acids in the presequence are indicated.

in Sugiura (1989)].

The most obvious difference between the L35 cDNA and those for the other chloroplast r-proteins reported so far is in the relative length of the transit peptide. Longer transit peptides are known, e.g., one of 139 amino acids directing the import of 341-residue-long *Euglena gracilis* chloroplast porphobilinogen deaminase (Sharif et al., 1989). However, as a percentage of the length of the precursor protein, the L35 transit (54%) is the highest. The reason for the large size of the L35 transit peptide may be the highly basic nature of L35 protein. With a net charge of +21 (i.e., 30% of the residues are positively charged), L35 is one of the most basic chloroplast ribosomal proteins.

The nucleotide sequence encoding r-protein L35 in the photosynthetic protozoan *C. paradoxa* (Bryant & Stirewalt, 1989; see Figure 3) is located in the organelle DNA (termed cyanelle). The cyanelle DNA is similar in size and in its gene arrangement to the DNA of chloroplasts (Wasman et al., 1987). Therefore, it is significant that L35 is encoded in the organelle DNA in *C. paradoxa* but in the nuclear DNA in spinach as reported here. The three completely sequences chloroplast genomes of land plants [liverwort, tobacco, and rice; see Sugiura (1989) for review] do not contain reading frames encoding r-protein L35. It is therefore likely that L35 is nuclear-coded in these plants as well. This finding therefore suggests that the L35 gene was located in the organelle DNA during the early phase of endosymbiosis but was transferred to the nucleus during the evolution of land plants. A similar situation with respect to the gene for an elongation factor (EF-Tu) has been reported while this paper was in preparation (Baldauf & Palmer, 1990).

In both *E. coli* (Fayat et al., 1983) and *B. stearothermophilus* (Pon et al., 1989), belonging to two diverse groups of eubacteria (Gram negative and Gram positive), the L35 gene is found in an identical gene cluster that includes the genes for an initiation factor (IF-3) and the r-protein L20. In plant cells two of the genes from this cluster (for IF-3 and L35) are nuclear while the third (L20 gene) is chloroplast-located [this work and review by Sugiura (1989)]. This kind of gene redistribution (i.e., 2:1 between the nucleus and the chloroplast) is the general rule for chloroplast r-proteins [reviewed in Subramanian et al. (1990)], but the basis for such a gene allocation remains to be explained.

ACKNOWLEDGMENTS

We thank Dr. B. R. Srinivasa, Gerd Bolken, Stephen Bantz, and Klaus von Knoblauch for participation at different stages of chloroplast r-protein purification.

REFERENCES

Baldauf, S. L., & Palmer, J. D. (1990) *Nature 334*, 262–265.
Bartsch, M., Kimura, M., & Subramanian, A. R. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6871–6875.
Bogorad, L. (1975) *Science 188*, 891–898.
Boynton, J. E., Gillham, N. W., & Lambowitz, A. M. (1980) in *Ribosomes: Structure, Function, and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., & Nomura, M. Eds.) pp 903–950, University Park Press, Baltimore.
Bryant, D. A., & Stirewalt, V. L. (1990) *FEBS Lett. 259*, 273–280.
Chen, E. Y., & Seeburg, P. H. (1985) *DNA 4*, 165–170.
Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.
Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., & Blanquet, S. (1983) *J. Mol. Biol. 171*, 239–261.
Giese, K., & Subramanian, A. R. (1989) *Biochemistry 28*, 3525–3529.
Giese, K., & Subramanian, A. R. (1990) *Biochemistry* (in press).
Gray, M. W. (1989) *Trends Genet. 5*, 294–299.
Heidecker, G., & Messing, J. (1986) *Annu. Rev. Plant Physiol. 37*, 439–466.
Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., & Siguira, M. (1989) *Mol. Gen. Genet. 217*, 185–194.
Johnson, C. J., Kruft, V., & Subramanian, A. R. (1990) *J. Biol. Chem. 265*, 12790–12795.
Kamp, R. M., Srinivasa, B. R., von Knoblauch, K., & Subramanian, A. R. (1987) *Biochemistry 26*, 5866–5870.
Kashiwagi, K., & Igarashi, K. (1987) *Biochem. Biophys. Acta 911*, 180–190.
Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., & Scheele, G. A. (1987) *EMBO J. 6*, 43–48.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, T., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., & Ozeki, H. (1986) *Nature 322*, 572–574.
Patterson, T. A., & Dean, M. (1987) *Nucleic Acids* Res. *15*, 6298.
Phua, S. H., Srinivasa, B. R., & Subramanian, A. R. (1989) *J. Biol. Chem. 264*, 1968–1971.
Pon, C. L., Brombach, M., Thamm, S., & Gualerzi, C. O. (1989) *Mol. Gen. Genet. 218*, 355–357.
Sanger, R., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342–1346.
Sharif, A. L., Smith, A. G., & Abell, C. (1989) *Eur. J. Biochem. 184*, 353–359.
Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., & Suguira, M. (1986) *EMBO J. 5*, 2043–2049.
Smooker, P. M., Kruft, V., & Subramanian A. R. (1990) *J. Biol. Chem.* (in press).
Subramanian, A. R., Smooker, P. M., & Giese, K. (1990) in *Structure, Function, and* Genetics *of Ribosomes* (Hill, W. E., Garrett, R., & Schlessinger, D., Eds.) pp 655–663, ASM Publications, Washington, DC.
Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakura, K., & Wallace, R. B. (1981) in *Developmental Biology Using Purified Genes* (Brown, D. D., Ed.) p 683, Academic Press, New York.
Sugiura, M. (1989) *Annu. Rev. Cell Biol. 5*, 51–70.
Wada, A., & Sako, T. (1987) *J. Biochem. (Tokyo) 101*, 817–820.
Wasmann, C. C., Löffelhardt, W., & Bohnert, H. J. (1987) in *The Cyanobacteria* (Fay, P., & van Baalen, C., Eds.) pp 303–324, Elsevier, Amsterdam.