Volume 15 Number 23 1987 Nucleic Acids Research

Sequence analysis of the EF-1α gene family of *Mucor racemosus*

Paula Sundstrom, L.M.Lira, D.Choi, John E.Linz[1] and Paul S.Sypherd

Department of Microbiology and Molecular Genetics, University of California, Irvine, CA and [1]Department of Food Science and Human Nutrition, 234 B Food Science Building, Michigan State University, East Lansing, MI 48824, USA

Received August 18, 1987; Accepted September 9, 1987

ABSTRACT

Our previous studies have shown that Mucor racemosus possesses three genes (TEF-1, -2 and -3) for EF-1α, and that all three genes are transcribed. However, the level of transcription varies markedly between the three genes, with TEF-1 mRNA levels being approximately two fold higher than TEF-3 and 6 fold higher than TEF-2. We have now completed the DNA sequence of both strands of all three genes and have found that these genes are highly homologous. TEF-2 and TEF-3 are more similar to each other than they are to TEF-1. The TEF-2 and the TEF-3 coding regions differ from TEF-1 at 30 and 37 positions respectively out of 1374 nucleotides. Twenty-six of these nucleotide substitutions were common to both TEF-2 and TEF-3, and the majority of the substitutions were clustered in the 5' region of the coding sequences. While the majority of these changes were silent, TEF-2 and TEF-3 differed from TEF-1 by having a lysine instead of a glutamate at amino acid position 41. In addition, TEF-2 and -3, but not TEF-1, each have an intron located near the 5' end of the coding region, although its size and sequence is not conserved between the two genes. All three genes have a conserved intron near the 3' end of the coding region. The sequence data have been analyzed with respect to the structure and function of EF-1α in protein biosynthesis.

INTRODUCTION

Mucor racemosus is a fungus that displays at least three morphogenetic events; the germination of spores, the formation of spores, and the transitions between yeast or hyphal forms. A variety of biochemical and molecular changes accompany the morphogenetic events, including changes in the translational apparatus. Among the latter events are increased rates of peptide chain elongation (1,2), phosphorylation of ribosomal protein S-6 (3,4) and changes in the level of N-lysine methylation of the translational factor EF-1α (5). EF-1α activity increases six-fold during spore germination (6), paralleling an increase in the rate of peptide chain elongation (1,2). It has been hypothesized that this enhanced EF-1α activity, is a result of a postranslational modification, as suggested by the fact that a concommitant increase in lysine methylation of EF-1α also occurs during this period (5).

The finding of three genes designated TEF-1,-2 and -3 in Mucor racemosus (7) raised the possibility that isomers of EF-1α exist in Mucor, and that differential expression of these

genes could also contribute in some way to the changes in protein synthesis and EF-1α activity seen during morphogenesis. A partial sequence analysis of these genes revealed nonhomologous 5' and 3' flanking regions, for which we prepared gene-specific oligonucleotide probes (8). The presence of different mRNA levels for the three genes was detected by analyzing mRNA isolated from different growth phases with the gene-specific oligonucleotide probes (8). It was found that TEF-1 mRNA levels are highest in all growth phases while the level of TEF-3 mRNA appears to have a morphogenetic pattern of expression. In order to gain more information about the genes which could help explain their different expression levels, we have completely sequenced both strands of TEF-2 and -3. In comparison with the previously sequenced TEF-1 (9), we have found that the coding regions of all three genes are nearly identical, with only two amino acid differences between them. Interestingly, one of the amino acid changes results in the substitution of a glutamate in TEF-1 to a lysine in TEF-2 and TEF-3, which would influence the overall charge on the protein and might add an additional methylation site.

MATERIALS AND METHODS

Organism. DNA used for sequencing was originally derived from M. racemosus (Mucor lusitanicus) ATCC 1216B.

Bacterial Strain and Plasmids. E. coli strains JM109 (10) was the host strain for the bacteriophages M13mp18 and M13mp19, and the plasmids PUC18 and PUC19, which were used to sublone restriction fragments from the EF-1α clones in pUC9 (9). Replicative form and single stranded DNA were purified using the procedures recommended by Bethesda Research Laboratories (11). Plasmid DNA for sequencing was prepared as previously described (12).

DNA sequencing. Overlapping clones for sequencing were prepared by the Exonuclease III deletion procedure of Henikoff (13). Nucleotide sequences were determined by the dideoxy chain termination procedure of Sanger and Coulson (14).

RESULTS

All three genes for EF-1α in Mucor were found to code for nearly identical proteins of 458 amino acids, however, the genetic organization of the genes varied (Figure 1) in that TEF-2 and -3 had an additional intron identically located after codon 18.

As predicted from restriction enzyme analysis, the nucleotide sequences of the exons of the three EF-1α genes were similar (Figure 2), with 37 nucleotide changes between TEF-1 and TEF-3 (97.4 % homology), 30 nucleotide changes between TEF-1 and TEF-2 (97.8 %) homology, and only 13 differences between TEF-2 and TEF-3 (99.5 % homology). Most of these differences were clustered near the 5' end of the genes (Figure 3), and resided in the third nucleotide position with little effect on the amino acid sequences. However, two amino acid changes were noted (Boxed residues in Figure 2). A lysine residue was found at amino acid position 41 In TEF-2 and -3, where TEF-1 had a glutamate. The only other amino acid

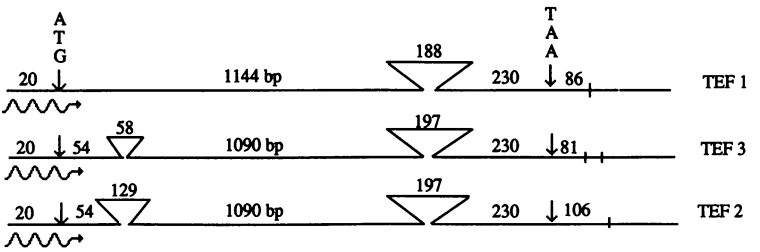

Figure 1. Genetic Organization of TEF 1, 2 and 3. Transcription of all three genes starts approximately 20 base pairs 5' of the translation initiation site (squiggly arrows). Introns are indicated by triangles. The numbers above the bars indicate the length in base pairs of the respective introns and exons. Vertical marks at the 3' ends of the genes indicate the approximate ends of the transcripts. Diagrams are not drawn to scale.

difference between the three genes was found at position 145 where TEF-3 had a valine residue while TEF-1 and -2 had phenylalanine.

The most striking difference between the genes was the presence of a second intron in TEF-2 and -3 which is not present in TEF-1 (9). Unlike the intron near the 3' end, which was present in all three genes and highly conserved, the 5' introns of TEF-2 and -3 were not homologous and differed in length, being only 58 nucleotides long in TEF -3 and 129 nucleotides in TEF-2. The donor splice sites of the 5' introns of TEF-2 and -3 differed from the conventional sequence (15) by 2 nucleotides (TCGTAAGT instead of AGGTAAGT) and differed from the donor site of the conserved 3' intron by 1 nucleotide. The acceptor site of all introns contained the consensus dinucleotide AG. The conserved introns near the 3' end of the genes were slightly less homologous than the coding regions of the genes. In this conserved intron there were 13 nucleotide changes between TEF-1 and -3 (93.5 % homology), 10 changes between TEF-1 and -2 (95 % homology), and 2 changes between TEF-2 and -3 (98.5 % homology). Compared to TEF-1, both TEF -2 and -3 contained a nine base pair insert located four positions prior to the 3' end of the conserved introns. In addition, the conserved 3' intron possessed a possible branch site, TATTAAC, in all three genes which differed by only one nucleotide from the branch consensus sequence, TACTAAC, which is essential for intron cleavage in S. cerevisiae (16). The 5' intron of TEF-2 was found to have an exact TACTAAC sequence while no consensus sequence for splicing could be found in the 5' intron of TEF-3.

The 5' flanking regions of both strands, as shown in Figure 4, were largely non-homologous, as were the 3' flanking regions (8). No TATAA elements or CAAT boxes were found, however some conserved sequence elements were identified as seen in Figure 4. The 11 nucleotide direct repeat shown in the box was found in the flanking regions of TEF-1 and -3 only. Sequencing and analysis of the 3' flanking regions has been previously reported (8).

The frequency of codon usage in each of the three genes was calculated from their

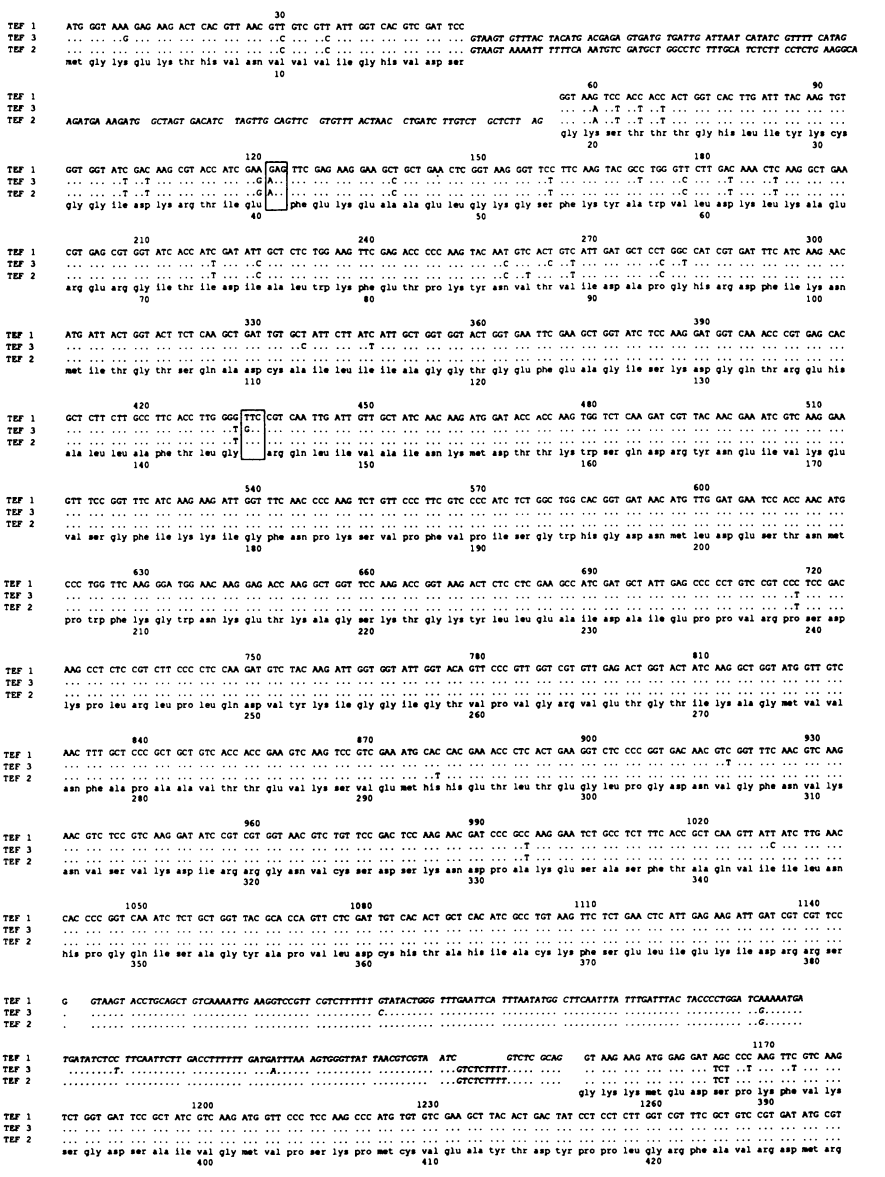

Figure 2. DNA sequence of exons and introns of TEF-1, -2, and -3. Dots are placed in positions where TEF-2 or -3 have the same nucleotide as TEF-1. Boxes indicate amino acid differences. At amino acid position 41, TEF-2 and -3 have a lysine where TEF-1 has a glutamate. At position 145, TEF-3 has a valine where TEF-1 and -2 have phenylalanine. Introns are in italics and are not included in numbering. TEF-1 has only one intron.

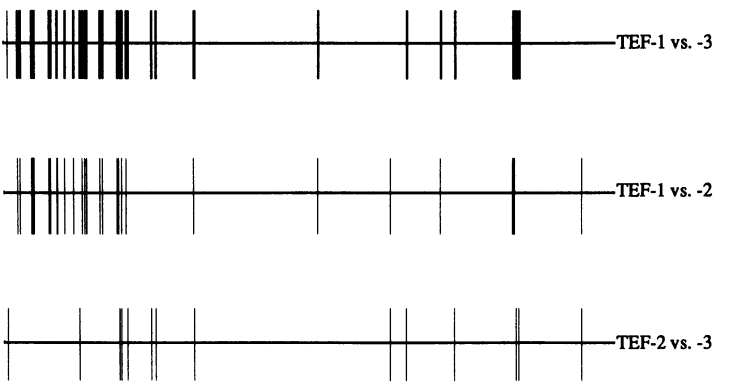

Figure 3. Nucleotide differences between exons of TEF-1, -2 and -3. The vertical lines represent nucleotide substitutions between the indicated genes. Each horizontal line represents a distance of 1374 base pairs.

DNA sequence data with the result seen in Table I. TEF-2 and TEF-3 showed the same codon bias seen with TEF-1 (9).

DISCUSSION

In previous work (8), we reported that there were differences in the levels of mRNA transcribed from the three genes for EF-1α in Mucor. Therefore, it was not surprising that comparison of the three sequences revealed several features of the genes which could be responsible for these differences. These included differences in 5' and 3' regions, as well as intron characteristics. There appear to be only two common features in the 5' regions. First,

```
                                                    TEF 1
GATCCCTCCGTGTTTGATGGCTAAATTTATAGAATTAGATTGTATATATCACTCAATAAAAAGACTATACCATGATGTGGCT
TGTTGCTTGTAGGGACTTGTTGTATCCAAACGTTGACTACAGACGCAATAATTTTTTTCGTCCTTTAAATAGCTTCTCGGAT
TTGGGGATGTGTTGTGTTCCCTAATTTCTATCCACTAAAAGGTTTCATACTATGCATCCATCCATCCATCCATCCATGATAC
TAAACATTGATACTATAGTTATTCAAA......

                                                    TEF 3
GGATCCATCCATGCCACAAATCAGCATAAATGCTATCCATCCATCCATCAAACATACTTACATGTATCATCTTTCATTATAG
TCGCA......

                                                    TEF 2
TTTTTCTCATTGGGAATCCATTGGAATGAAAGGACAAATGCACTCTCGCAATGAGATGCTTTAAATGCTGGCAAATTTGAAG
GATGTACAATCGAAACTTTCCAAATGTCCTCAAACAAGACGGGCTATCCATATGAAAACTAGCAATGTATAGAGACTAAAGC
AACTGATCCTTTGTCCAAGAAGCAGTTAGTGCTGTATCATAAGCCACTAATGGAAGTTTTGCAGCTTTCCTTTATAGTAAAA
CGTCCAATCAATGCATGTAACTGTGACCATCGGGCACAAGTGGCTTTGAGCTGCGTAACAAAAATTTTTTTGTAGTATTTA
AAAACATGGATCTAAAAAAACTTTCTTGTTTTTTTTTTCTCCTTCTATCAATTTACAAGTTTAATCAGTGCGTATACTGCTA
TGTTTTTGTTCTAGTAATAGATCTTGGTTCTAACATGTCTTTTTCTCTATAGCA.........
```

Figure 4. 5' Flanking Regions of TEF 1, 2, and 3. Underlined nucleotides indicate positions of AACAT or ATCAT conserved sequences in the regions predicted for transcription initiation by S1 analysis. Boxed sequences indicate repeated sequences present in TEF 1 and TEF 3.

Table 1. Comparison of frequency of codon usage for the 3 EF-1α genes from Mucor

		TEF1	TEF2	TEF3			TEF1	TEF2	TEF3			TEF1	TEF2	TEF3
Ala	GCT	27	28	27	Gly	GGT	37	38	39	Pro	CCT	5	5	5
	GCC	9	8	9		GGC	2	2	1		CCC	16	16	16
	GCA	1	1	1		GGA	1	1	1		CCA	1	1	1
	GCG	0	0	0		GGG	1	0	0		CCG	0	0	0
Arg	CGT	17	17	17	His	CAT	1	2	1	Ser	TCT	9	12	12
	CGC	0	0	0		CAC	10	9	10		TCC	16	14	14
	CGA	0	0	0							TCA	0	0	0
	CGG	0	0	0	Ile	ATT	15	16	16		TCG	0	0	0
	AGA	0	0	0		ATC	19	18	18		AGT	0	0	0
	AGG	0	0	0		ATA	0	0	0		AGC	1	0	0
Asn	AAT	1	0	0	Leu	TTA	0	0	0	Thr	ACT	12	14	13
	AAC	15	16	16		TTG	5	5	5		ACC	18	16	17
						CTT	6	7	7		ACA	1	1	1
Asp	GAT	19	21	21		CTC	11	10	10		ACG	0	0	0
	GAC	7	5	5		CTA	0	0	0					
						CTG	0	0	0	Trp	TGG	6	6	6
Cys	TGT	6	6	6										
	TGC	0	0	0	Lys	AAA	3	3	3	Tyr	TAT	1	1	1
						AAG	46	47	47		TAC	7	7	7
Gln	CAA	8	8	8										
	CAG	0	0	0	Met	ATG	11	11	11	Val	GTT	15	14	14
											GTC	26	27	28
Glu	GAA	17	16	16	Phe	TTT	1	1	1		GTA	0	0	0
	GAG	12	12	12		TTC	16	16	15		GTG	0	0	0

in the region where the initiation of transcription is predicted to start by S1 analysis (7,8), there is a pentanucleotide sequence AACAT (TEF-1 and -2) and ATCAT in TEF-3 which occurs 25, 23 and 22 nucleotides upstream from the initial ATG codon, in TEF-1, -2 and -3 respectively. In addition AACAT is also present in TEF-3, 37 nucleotides prior to the initial ATG codon. It would be interesting if this pentanucleotide sequence coincided with transcript initiation, since transcription initiation sites are not conserved in general. In E. coli, transcription initiation usually starts with a purine and is often the central nucleotide in the sequence CAT (17). The second notable feature in the 5' region of the Mucor genes is common only to TEF-1 and TEF-3, the genes which have the highest message levels. This is a 12 nucleotide direct repeat CATCCATCCA, present in two copies in each gene, but spaced differently. The two repeats are separated by 25 nucleotides in TEF-3, but are adjacent in TEF-1. This element could be involved in the level of transcription of these genes, based on analogy to Saccharomyces. The EF-1α genes of Saccharomyces have been shown to have RPG sequences, activating sequences for many ribosomal protein genes in Saccharomyces (18,19,20) as well as in Schizosaccharomyes pombe (21), located 200-400 nucleotides upstream of the translation start site. It has been shown that the level of transcription of genes with these sequences is proportional to the number of activating sequences, as well as the to the lack of deviation from an optimal sequence (22). Thus the CATCCATCCA pattern in Mucor could play a similar role. In addition, the Mucor sequence is similar to the RPG seqeunce ACCCATACATCA (20). The observation that TEF-2, which generates the lowest message levels, does not have the 12 nucleotide direct repeat, at least in the 463 nucleotides

which were sequenced 5' of the ATG codon, is consistent with an activating role for this sequence. In addition, at least one protein has been shown to bind to the RPG sequences in Saccharomyces (22,23). In contrast to other eucaryotic activating sequences which are located hundreds of base pairs upstream of the transcriptional initiation site, the CATCCATCCA sequence of Mucor is much closer, being only a few nucleotides upstream from transcription initiation. This feature, along with the conserved pentanucleotide in the vicinity of transcription initiation, is more analogous to transcription initiation in E. coli than in other eucaryotes. Sequencing other coordinately regulated genes such as ribosomal protein genes of Mucor, would determine if this is an important activating element, and would serve to identify precisely the optimal activating sequence. In addition, identifying proteins which bind to these sequences would also suggest an important role for this sequence. Both of these approaches are currently in progress. More precise localization of important regions in regulating transcription by deletion and fusion to marker genes in vivo is needed.

In addition to transcriptional efficiency, other features of the three EF-1α genes which could influence message levels are differences in intron splicing, or message degradation, or differences in secondary structure of the mRNA.

Analysis of the complete sequences of the three EF-1α genes from Mucor predicted that each gene is potentially able to code for a functional protein. The nucleotide sequence revealed only 2 changes in amino acid sequence for the three proteins, and all proteins possessed conserved domains characteristic of GTP binding regions of many proteins such as those of ras proteins and G proteins (24,25). Both amino acid changes occurred in the N terminal GDP-binding domain and potentially altered protein function. For example, if all genes were translated, the substitution of a lysine in TEF-2 and -3 for a glutamate in TEF-1 at position 41 could provide an additional methylation site for the protein, which might make the TEF-2 or TEF-3 gene products more active than the TEF-1 gene product, under conditions where methylation occurs. Although it has been shown that EF-1α and EF-Tu have regions of homology (9), amino acid 41 of the Mucor sequence lies in a stretch of 15 amino acids which is not present in EF-Tu, but is in the vicinity of α helix A of EF-Tu (26). Analysis of the Mucor sequence by Chou Fasman (27) parameters showed that the region around position 41 is highly compatible with an α helix as well, with either a glutamate or a lysine residue in that position. The additional 15 amino acids present in EF-1α could result in an additonal α helix separated from the analogous EF-Tu A helix by a β bend at the glycine residues 32 and 33 (predicted by Chou Fasman analysis). Based on analogy with EF-Tu, these residues are positioned between a region interacting with the aminoacyl group of tRNA (personal communication with F. Jurnak), and α helix A which contributes to the GDP binding region. Thus, methylation of lysine 41 in the TEF 2 and 3 gene products could alter the function of the protein either through its interaction with GDP or tRNA.

The only other amino acid change found among the three proteins was a valine (TEF -2 and -3) substitution for a pheylalanine (TEF-1) at position 145. This position is in a region

homologous to EF-Tu. The crystallized GDP domain of EF-Tu reveals that this position lies in a random coil between α helix C and beta sheet region 5 (26) and is close to a hinge region of the first and second domains of EF-Tu (personal communication with F. Jurnak). Thus, differences in this position could influence intramolecular conformational changes. In addition, since both Artemia and Saccharomyces EF-1α have lysine at position 41 and valine at position 145, it is unlikely that any of the genes would produce nonfunctional proteins.

The codon usage of the three genes for EF-1α in Mucor was quite similar which reflects the close relatedness of the three genes. In addition, the codon usage was highly biased in that 92 % of the codons used represented only 25 out of the 61 possible triplets. A similar degree of codon bias, but for different codons, is found in Saccharomyces for genes which are highly expressed (28). This has been found to reflect the predominant isoacceptor tRNAs present. Since EF-1α is a highly expressed gene, the codon usage seen could reflect the prevalent tRNAs in Mucor.

The new sequence data presented in this paper agree with the previously suggested hyposthesis that the origin of the EF-1α genes of Mucor could have arisen as duplications of a common predecessor gene with two introns (8). Several features of the sequences suggested that TEF-2 and -3 are more related to each other than either of them is to TEF-1. First there were fewer differences in the exons between TEF-2 and -3 than between either of these genes and TEF-1. Also TEF -2 and -3 possessed an additional intron at the same location which was not present in TEF-1. Both TEF -2 and TEF -3 had an identical nine base pair insert close to the 3' end of the conserved intron, as well as a silent codon change (TCT in TEF 2 and 3 insead of AGC in TEF 1) for serine at amino acid position 388. The non-homology of the 5' introns in TEF 2 and 3 compared to the highly conserved 3' introns is difficult to explain. The 3' introns appear to be slightly more divergent among the three genes than is the coding sequence consistent with the idea that all genes came from a common predecessor. Perhaps the 5' introns are vestigial remnants of a precursor fragment, which was differentially deleted subsequent to the divergence of the 2 genes. The presence of repeated sequences (GTAAGT) at the ends of this intron supports this hypothesis.

The reasons for the presence of multiple genes for EF-1α are unknown. It is unlikely that the advantage of multiple genes could be a simple quantitative effect, since Saccharomyces can grow normally with either of its EF-1α genes deleted (29). In addition, overproduction of EF-1α in Saccharomyces may be deleterious. When the copy number of TEF genes is increased from 1 to 3 per cell, increased misreading occurs, as shown by the enhancement of suppression of nonsense mutations, and additional suppression in the presence of paromomycin (personal communications with Susan Liebman and Jae Mahn Song). In Mucor however, it is possible that subtle differences in the protein, such as an additional methylation site, could provide a more active product during growth phases where more rapid protein synthesis is advantageous.

ACKNOWLEDGEMENTS

We wish to recognize Dr. William Fonzi for his many useful discussions during the course of this investigation. This research was supported by grants from the national institute of Allergy and Infectious Disease (AI 16252) and the National Institute of General Medical Science (GM23999). P. Sundstrom was supported by a fellowship from the Biotechnology Research and Education Program of the University of California. This is publication number 136 from the California Colleges and Universities Mycology Research Unit.

REFERENCES

1. Orlowski, M., and Sypherd, P.S. (1978) J. Bacteriol. **134**, 76-83.
2. Orlowski, Ml and Sypherd, P.S (1978) Biochemistry, **17**, 569-575.
3. Larsen, A. and Sypherd, P.S. (1979) Mol. Gen. Genet. **175**, 99-109.
4. Larsen, A. and Sypherd, P.S. (1980) J. Bacteriol. **141**, 20-25.
5. Hiatt, W.R., Garcia, R., Merric, W.C., and Sypherd, P.S., (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3433-3437.
6. Fonzi, W.A., Katayama, C., Leathers, T., and Sypherd, P.S. (1985) Mol. Cell. Biol. **5**, 1100-1103.
7. Linz, J.E., Katayama, C., and Sypherd, P.S. (1986) Mol. and Cell. Biol. **6**, 593-600.
8. Linz, J.E., and Sypherd, P.S., (1987) Mol. Cell. Biol. **7**, 1925-1932.
9. Linz, J.E., Lira, L.M. and Sypherd, P.S. (1986) J. Biol. Chem. **261**, 15022-15029.
10. Yanisch-Perron, C. Vieira, J. and Messing, J. (1985) Gene **33**, 103-119.
11. Bethesda Research Laboratories, Inc. (1985) Instruction Manual. Bethesda Research Laboratories Inc. Bethesda, MD.
12. Chen, E.Y., and Seeburg, P.H. (1985) DNA **4**, 165-170.
13. Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**, 351-359.
14. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. **94**, 441-448.
15. Lewin, B. (1987) Genes pp 473-474. John Wiley and Sons, New York.
16. Lewin, B. (1987) Genes pp 476-477. John Wiley and Sons, New York.
17. Lewin, B. (1987) Genes p 194. John Wiley and Sons, New York.
18. Nagashima, K., Masataka, K., Nagata, S., and Kaziro, Y., (1986) Gene **45**, 265-273.
19. Teem, H.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Wollford, J., Leer, R.J., van Raamsdonk-Duin, J.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D., Fried, H., and Rosbash, M., (1984) Nuc. Acids Res. **12**, 8295-8307.
20. Leer, R.J., Van Raamsdonk-Duin, M.C., Mager, W.H., and Planta, R.J. (1985) Current Genetics **9**, 273-277.
21. Nischt, R., Gross, T., Gatermann, K., Swida, U., and Kaufer, N.F. (1987) **15**, 1477-1492.
22. Huet, J. and Sentenac, A. (1987). Proc. Natl. Acad. Sci. **84**, 36648-3652.
23. Huet,J., Cottrelle, P., Cool, M., Vignais, M.-L., Thiele, D., Marck, C., Buhler,J.M., Sentenac, A. and Romageot, P. (1985) EMBO Journal **4**, 3539-3547.

24. Dever, T.E., Glynias, M.J., and Merrick, W.C. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 1814-1818.
25. Kohno, K., Uchida, T., Ohkubo, H., Nakanish, S., Hakanishi, T., Fukui, T., Ohtsuka, E., Ikehara, M., and Okada, Y. Proc. Natl. Acad. Sci. U.S.A. **83**, 4978-4982.
26. Jurnak, F. (1985) Science. **230**, 32-36.
27. Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. **47**, 251-271.
28. Bennetzen, J.L., and Hall, B.D., (1982) Codon selection in yeast. J. Biol. Chem. **257**, 3026-3031.
29. Cotrelle, P., Thiele, D., Price, V., Memet, Sl., Micoun, J., Marck, C., Buhler, J., Sentnac, A., and Fromageot, P., (1985) J. Biol. Chem **260**, 3090-3096.