JOURNAL OF BACTERIOLOGY, July 1983, p. 8–14
0021-9193/83/070008-07$02.00/0

Vol. 155, No. 1

Nucleotide Sequence of the *tcm1* Gene (Ribosomal Protein L3) of *Saccharomyces cerevisiae*

L. D. SCHULTZ† AND J. D. FRIESEN*

Department of Medical Genetics, University of Toronto, Toronto, Ontario, Canada, M5S 1A8

Received 30 December 1982/Accepted 12 April 1983

The yeast *tcm1* gene, which codes for ribosomal protein L3, has been isolated by using recombinant DNA and genetic complementation. The DNA fragment carrying this gene has been subcloned and we have determined its DNA sequence. The 20 amino acid residues at the amino terminus as inferred from the nucleotide sequence agreed exactly with the amino acid sequence data. The amino acid composition of the encoded protein agreed with that determined for purified ribosomal protein L3. Codon usage in the *tcm1* gene was strongly biased in the direction found for several other abundant *Saccharomyces cerevisiae* proteins. The *tcm1* gene has no introns, which appears to be atypical of ribosomal protein structural genes.

Eucaryotic ribosomes consist of a complex of four RNA species and more than 70 proteins (39). The biosynthesis of these gene products in *Saccharomyces cerevisiae* is regulated in a remarkably coordinated manner (16, 21, 38), yet with one exception (12), no genetic linkage among them has yet been demonstrated (5, 12, 40). The mechanism for this coordinate regulation is not obvious. However, at least in the case of one ribosomal protein (r-protein) gene, it has been suggested that this regulation might be affected by autogenous feedback at the translational level (30). This has beem amply demonstrated in *Escherichia coli* (15, 23, 28).

Among those genes which affect protein synthesis in *S. cerevisiae*, there are several which have an easily identifiable genetic phenotype. These include three drug-resistant genes: tricodermin resistance (*tcm1*; 18), cyclohexamide resistance (*cyh2*; 34), and cryptopleurine resistance (*cry1*; 17). The first two of these code for r-proteins L3 and L29, respectively (13, 14). In addition, there are at least three genetic loci which cause omnipotent suppression, apparently as a result of RNA misreading by the ribosome: *sup35* and *sup45* (19) and SUP46 (25). It is quite possible that these also code for r-proteins. These six genes offer a twofold advantage. First, the phenotype can be used as a possible selection method to clone the structural gene. This approach has been used by Fried and Warner (13, 14) in their isolation of the genes responsible for tricodermin resistance (*tcm1*) and cycloheximide resistance (*cyh2*). Those studies established that the *tcm1* and *cyh2* genes code for r-proteins L3 and L29, respectively (13, 14). Second, one might be able to make use of a selectable phenotype in the isolation of further mutants, for example, those which lie in sequences of regulatory importance.

We have chosen to work with the *tcm1* gene. This codes for the largest r-protein, L3, of *S. cerevisiae* (13), and its biosynthesis appears to be quite strictly regulated (30). We present the nucleotide sequence of this gene and its environs as a basis for detailed genetic studies of its biosynthetic regulation.

MATERIALS AND METHODS

Microbial strains and plasmids. The tricodermin-resistant strain CLP1 (a *leu2 tcm1*; obtained from J. Davies [18]) of *S. cerevisiae* was used as the source of DNA for constructing the clone library. *S. cerevisiae* strain LL20 (α *leu2-3 leu2-112 his3-11 his3-15 can1*; obtained from G. R. Fink) was the recipient for direct isolation of the *tcm1* gene. *E. coli* strain JF1754 (*hsdR lac gal metB leuB hisB*) was used as the bacterial recipient. The cloning vehicle was pJZ1 (constructed by J. Zeissler; Fig. 1); this is a typical yeast shuttle vector expressing *Leu2*$^+$, ampicillin resistance (Ap^r), and tetracycline resistance (Tc^r) (35). Growth and transformation of *S. cerevisiae* and *E. coli* were as previously described (27). In the isolation of the *tcm1* gene, the *S. cerevisiae* genomic clone library carried on pJZ1 in *E. coli* JF1754 was transformed into *S. cervisiae* strain LL20, selecting for *Leu2*$^+$ on minimal medium plus histidine after embedding the spheroplasts in 3% agar. After 18 h at 23°C to allow expression of tricodermin resistance, the thin slab of agar containing the transformants was transferred to the surface of selection agar containing minimal medium

† Present address: Virus and Cell Biology Research, Merck Sharp and Dohme Research Laboratories, West Point, PA 19486.

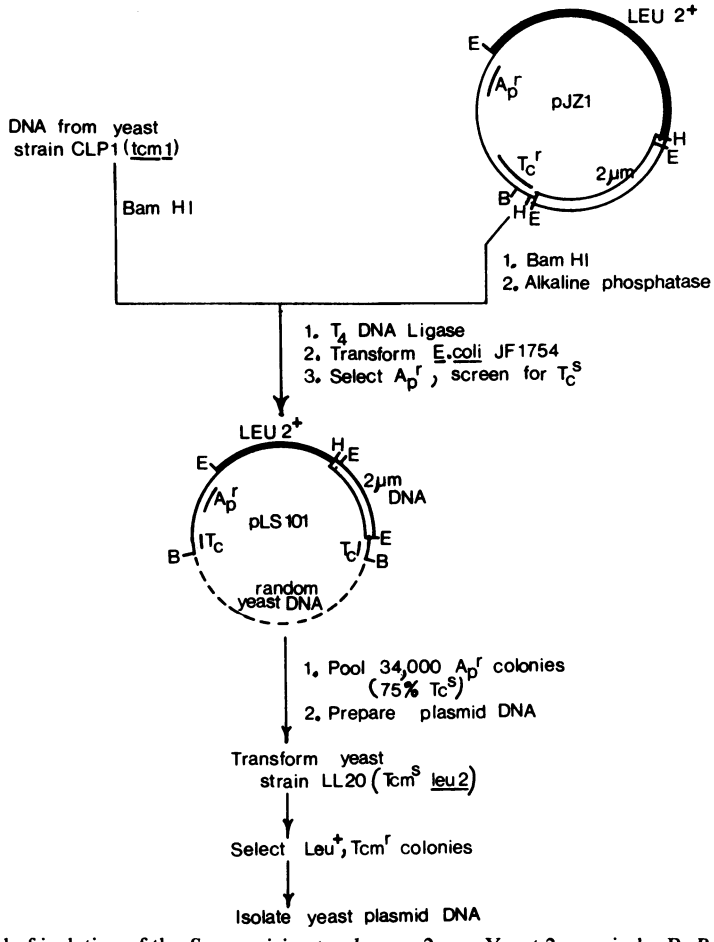

FIG. 1. Method of isolation of the *S. cerevisiae tcm1* gene. 2 μm, Yeast 2-μm circle. B, *Bam*HI; H, *Hin*dIII, and E, *Eco*RI, sites for restriction endonuclease cleavage.

plus histidine and 5 μg of tricodermin per ml. This is a convenient way to transfer spheroplasts from one growth condition to another. The drug concentration used was sufficient for complete growth inhibition of sensitive LL20 yeast cells (data not shown). After 2 days of incubation at 30°C, Leu^+ and tricodermin-resistant (Tcm^r) colonies were picked, purified, and retested. The proportion of Leu^+ colonies which were also Tcm^r was 1 per 1,600.

DNA manipulations and nucleotide sequence analysis. Enzymes and reagents were purchased from Boehringer-Mannheim Canada, Bethesda Research Laboratories, New England Nuclear, and New England Biolabs. Enzyme reactions were carried out according to the procedures supplied by the manufacturers. DNA manipulations were by standard methods (24). DNA sequencing was carried out according to the method of Maxam and Gilbert (26). Fine-structure mapping of restriction sites was carried out by the method of Smith and Birnstiel (33) with DNA fragments labeled at the 3′ end with the Klenow fragment of DNA polymerase I and α-^{32}P-labeled deoxyribonucleotide triphosphates.

RESULTS AND DISCUSSION

Cloning the *tcm1* gene. The scheme for cloning the *tcm1* gene (which is a mutant allele) made use of the fact that drug resistance is semidominant over sensitivity at low concentrations of tricodermin (2 to 5 μg/ml) which completely inhibit sensitive strains (unpublished data). Thus, a partial diploid cell in which tricodermin resistance is carried on a medium-copy replicating plasmid but which bears a chromosomal drug-sensitive allele might be expected to show elevated drug resistance. A clone library in *E. coli* from DNA fragments of a total *Bam*HI digest of DNA from the tricodermin-resistant *S. cerevisiae* strain CLP1 was prepared with pJZ1 as vector (Fig. 1). Transformants of strain LL20 were initially selected for the Leu^+ phenotype. After an expression time of 18 h at 23°C, the thin top agar slab containing the transformants was transferred to a selection agar containing mini-

mal medium plus histidine plus 5 μg of tricodermin per ml to select directly for Leu$^+$ Tcm^r transformants.

Yeast plasmid DNA was prepared from several Leu$^+$ Tcm^r transformants and used to transform *E. coli* strain JF1754 to ampicillin resistance. Plasmid pLS2-II was isolated from one such *E. coli* transformant; this carried a 13.5-kilobase *Bam*HI yeast DNA insert containing the *tcm1* gene. This plasmid DNA retransformed *S. cerevisiae* strain LL20 to both Tcm^r and Leu$^+$ at a high frequency.

Restriction map and nucleotide sequence of the *tcm1* gene. The restriction map of the insert carried on pLS2-II was determined by conventional means (24) (Fig. 2a). This restriction map agrees closely with the map obtained by Fried and Warner (13) for the *Bam*HI fragment carrying the *tcm1* gene. It has been shown by Fried and Warner (13) that the *tcm1* gene is present on the 3.4-kilobase *Bam*HI-*Ava*I region (heavy line in Fig. 2a) of the 13.5-kilobase *Bam*HI fragment. This *Bam*HI-*Ava*I fragment contains the information necessary to transform sensitive yeast cells to tricodermin resistance (13; our unpublished data). To facilitate further analysis, the *Bam*HI-*Ava*I fragment was subcloned onto pBR325 (4), producing a plasmid called pLS5T. With pLS5T, a more detailed restriction map of the *Bam*HI-*Ava*I region was determined (Fig. 2b). Figure 2c shows the tactics used in determining the nucleotide sequence of the region

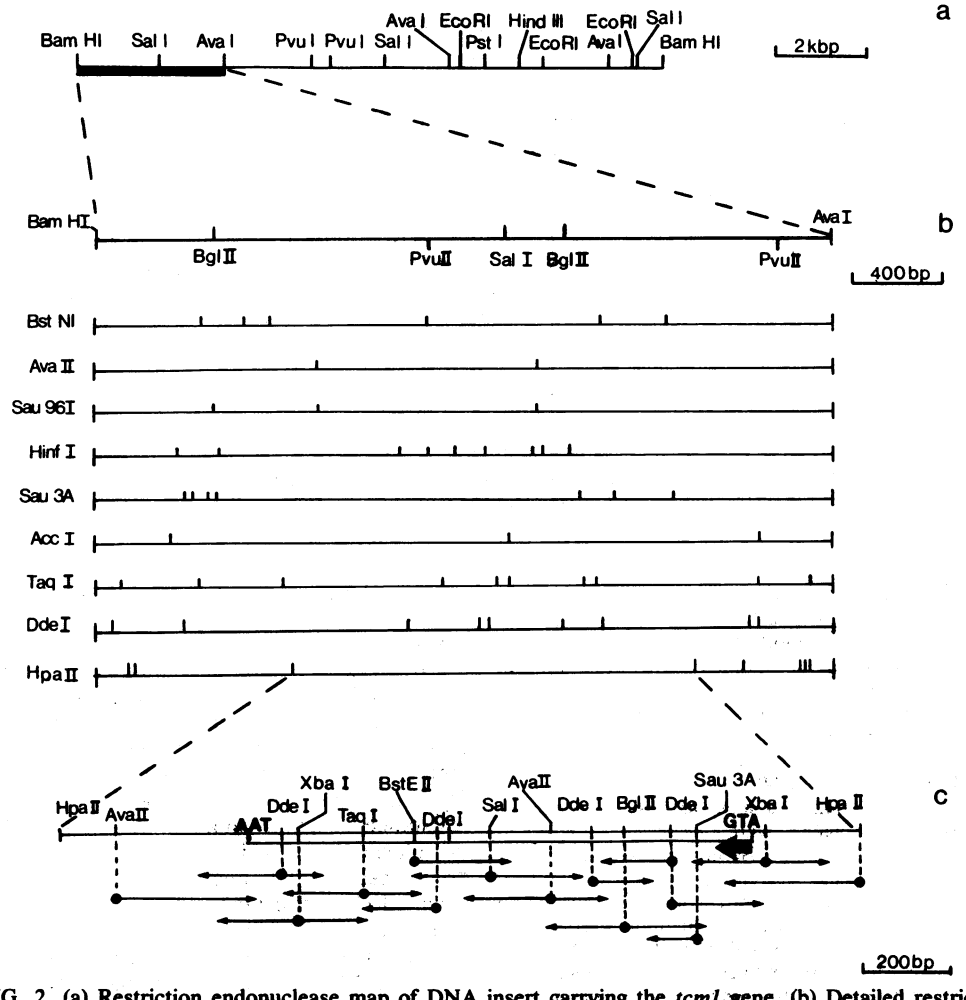

FIG. 2. (a) Restriction endonuclease map of DNA insert carrying the *tcm1* gene. (b) Detailed restriction endonuclease map of the *Bam*HI-*Ava*I portion of the DNA shown in (a), which carries the *tcm1* gene. (c) Tactics for nucleotide sequence determination of the *tcm1* gene. Arrowheads indicate the direction and extent of sequence determination by the method of Maxam and Gilbert (26). The predicted coding region for r-protein L3 (from Fig. 3) is shown by the boxed region between ATG and TAA. The heavy arrow indicates the inferred direction of transcription. bp, Base pairs.

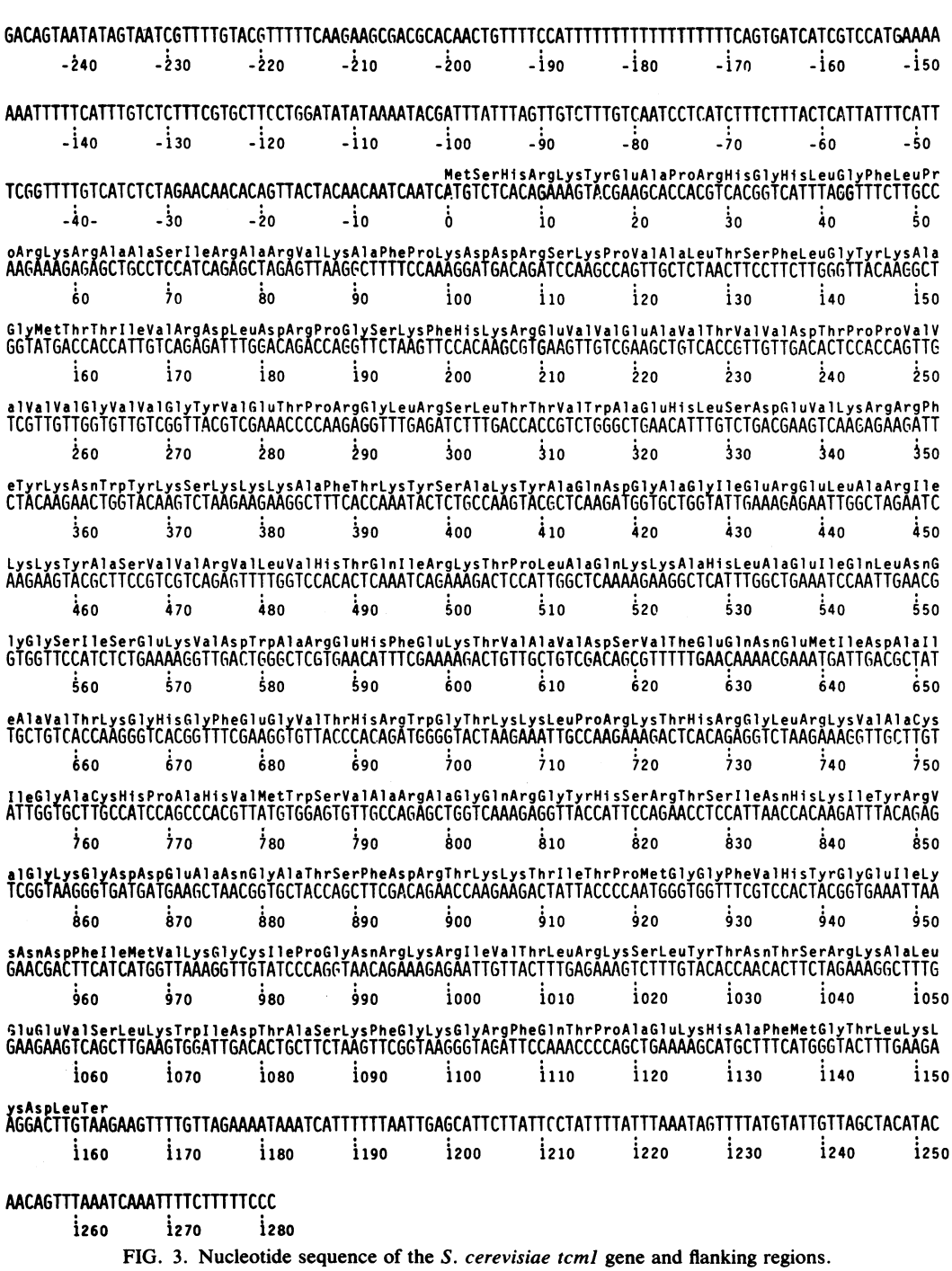

FIG. 3. Nucleotide sequence of the *S. cerevisiae tcm1* gene and flanking regions.

believed to encode the *tcm1* gene (13); over 95% of the sequence was determined for both DNA strands. The sequence itself is shown in Fig. 3.

Features of the nucleotide sequence of the *tcm1* gene. The amino acid sequence inferred from the nucleotide sequence yields a protein of 43.5 kilodaltons, a size similar to that reported by Fried and Warner (13) for r-protein L3. The predicted amino acid composition of this protein agrees with that determined by Itoh et al. (20) for the purified yeast r-protein L3. Furthermore, the sequence of 20 amino acid residues at the amino

		L3	Y			L3	Y			L3	Y			L3	Y
Phe	TTT	2	5	Ser	TCT	10	79	Tyr	TAT	0	2	Cys	TGT	2	20
	TTC	13	63		TCC	7	73		TAC	12	73		TGC	1	2
Leu	TTA	1	14		TCA	0	1	och	TAA	1		umb	TGA	0	
	TTG	18	148		TCG	0	1	amb	TAG	0		Trp	TGG	6	26
Leu	CTT	0	0	Pro	CCT	0	6	His	CAT	7	3	Arg	CGT	3	2
	CTC	0	1		CCC	0	1		CAC	10	56		CGC	0	0
	CTA	2	6		CCA	14	71	Gln	CAA	7	44		CGA	0	0
	CTG	0	0		CCG	0	0		CAG	0	0		CGG	0	0
Ile	ATT	11	58	Thr	ACT	12	49	Asn	AAT	0	3	Ser	AGT	1	1
	ATC	7	66		ACC	15	66		AAC	8	86		AGC	3	1
	ATA	0	1		ACA	0	0	Lys	AAA	3	21	Arg	AGA	30	63
Met	ATG	7	39		ACG	0	0		AAG	40	151		AGG	0	1
Val	GTT	19	112	Ala	GCT	28	182	Asp	GAT	5	37	Gly	GGT	31	164
	GTC	16	93		GCC	4	62		GAC	11	100		GGC	0	11
	GTA	0	3		GCA	1	3	Glu	GAA	20	67		GGA	0	1
	GTG	0	1		GCG	0	4		GAG	0	3		GGG	0	1

FIG. 4. Codon usage for the *S. cerevisiae tcm1* gene and several highly expressed *S. cerevisiae* genes. The values represent the actual number of times each codon is present in the structural gene. The numbers under the columns headed L3 are for the *tcm1* gene, and those under the columns headed Y are summed for two glyceraldehyde-3-phosphate dehydrogenase genes, two enolase genes, alcohol dehydrogenase isozyme I, and alcohol dehydrogenase isozyme II. The data for these six are from Ammerer et al. (1).

terminus as determined by direct analysis of purified r-protein L3 (E. Otaka, personal communication) agrees exactly with that inferred from the nucleotide sequence. Finally, there is no evidence for a noncoding intervening sequence in the *tcm1* gene. Our sequencing results substantiate the results of Fried and Warner (13), who showed on the basis of translation of hybrid-selected mRNA in vitro that the *tcm1* gene codes for r-protein L3. Schwindinger and Warner (personal communication) have shown by S1 mapping that the *tcm1* gene is a rarity among r-protein genes examined thus far in that it does not have an intervening sequence (6, 12); this supports the conclusion drawn from our sequence data.

The data in Fig. 4 show the frequency of codon usage for the *S. cerevisiae tcm1* gene. A summation of the total codon usage for six other yeast genes, two for glyceraldehyde-3-phosphate dehydrogenase, two for enolase, alcohol dehydrogenase isozyme I, and alcohol dehydrogenase isozyme II, is also shown (1, 3). These results clearly show that there is a pronounced bias in the choice of amino acid codons in the *tcm1* gene, similar to the bias in codon usage seen for a number of highly expressed yeast genes (1, 3). These data also show that r-protein L3 is very rich in the basic amino acids lysine and arginine, as is to be expected for an r-protein.

Other features of the DNA sequence include the following. (i) A TATA box, in this case the sequence TATATAAAA, is located 113 nucleotides upstream from the translation start of the *tcm1* gene. This presumably represents an essential feature of the promoter in eucaryotic cells (8). Although we have not determined the transcription start site, by analogy with several other yeast genes (2, 10, 11, 36) one might predict this to lie about 40 to 70 nucleotides on the 3′ side of the TATA box. If this is so, then the untranslated leader of the mRNA would be about 40 to 70 nucleotides long. It is this stretch of RNA that might be implicated in autogenous translational regulation or r-protein gene expression in *S. cerevisiae* (30). The 5′-leader region is very adenine-thymine rich and does not appear to allow for a great deal of secondary structure. This is in contrast to several *E. coli* r-protein leaders with which it has been suggested that the secondary structure is important for binding the autogenous regulator (7, 28, 29).

(ii) Zaret and Sherman (41) observed that the 3′ untranslated regions of many *S. cerevisiae* genes show sequence similarities in a region that is postulated to be involved in transcription termination. Based on a comparison of the 3′ untranslated regions from 15 yeast genes, Zaret and Sherman (41) have proposed the following general sequence as a possible signal for transcription termination: . . . (T rich) . . . TAG . . . TAGT . . . . (AT rich) . . . TTT. The *tcm1* gene sequence shows similarities to this general sequence in the region from +1,217 to +1,262 (Fig. 3); in particular, there is a TAG sequence at +1,228 and a TATGT sequence at +1,234. The 3′-flanking region of the *tcm1* gene is also very adenine-thymine rich. There are several AATAA sequences, which for other genes have been suggested as possible polyadenylic acid addition sites (31).

(iii) A comparison of the nucleotide sequence of *tcm1* in the region of the ATG start codon (CTACAACAATCAATCATGTCTCAC) with the corresponding region of 11 other *S. cerevisiae* genes (1) shows several similarities. (a) There is a scarcity of G residues in the 15 nucleotides of the plus-strand sequence just preceding the ATG. Instead, this region is rich in sequences such as CA, CAA, and CAAA. (b) There appears to be an invariant A residue at position −3. (c) There is a prevalent T residue at position +6.

(iv) Perhaps most striking, the *tcm1* gene has no intron, in contrast to most other yeast r-protein structural genes so far studied (6, 13, 22, 32). Thus, one can conclude that coordinate regulation of r-protein biosynthesis does not depend directly on intron processing. Despite this, r-protein L3 biosynthesis appears to be affected by the *rna2* mutation (16, 37), which is involved with the processing of introns from precursor mRNA (9, 32). This raises the question of whether there is a mechanism of r-protein regulation in which the expression of non-intron-containing r-protein genes is dependent on the expression of other r-protein genes that do contain introns.

In total, five r-protein genes from *S. cerevisiae* have now been sequenced (22; M. Rosbash, J. Warner, and J. Woolford, personal communication). A detailed structural comparison of all five genes will be presented in a joint publication (manuscript in preparation).

ACKNOWLEDGMENTS

We thank Josie Chapman-Smith for typographical assistance, Jon Warner for advance information on S1 mapping of *tcm1*, and E. Otaka for preliminary data on the amino acid sequence of r-protein L3.

This work was supported by the Natural Sciences and Engineering Research Council (grants A5734 and A0627) and the National Cancer Institute of Canada.

LITERATURE CITED

1. **Ammerer, G., R. Hitzeman, F. Hagie, A. Barta, and B. D. Hall.** 1981. The functional expression of mammalian genes in yeast, pp. 185–197. *In* A. G. Walton (ed.), Recombinant DNA. Elsevier Scientific Publishing Co., Amsterdam.
2. **Astell, C. R., L. Ahlstrom-Jonasson, and M. Smith.** 1981. The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*. Cell **27**:15–23.
3. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.
4. **Bolivar, F.** 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *Eco*RI sites for selection of *Eco*RI generated recombinant DNA molecules. Gene **4**:121–136.
5. **Bollen, G. H. P. M., L. H. Cohen, W. H. Mager, A. W. Klaassen, and R. J. Planta.** 1981. Isolation of cloned ribosomal protein genes from the yeast *Saccharomyces carlsbergensis*. Gene **14**:279–287.
6. **Bollen, G. H. P. M., C. M. T. Molenaar, L. H. Cohen, M. M. C. van Raamsdonk-Duin, W. H. Mager, and R. J. Planta.** 1982. Ribosomal protein genes of yeast contain intervening sequences. Gene **18**:29–37.
7. **Branlant, C., A. Krol, A. Machatt, and J.-P. Ebel.** 1981. The secondary structure of the protein L1 binding region of ribosomal 23S RNA. Homologies with putative secondary structure of the L11 mRNA and of a region of mitochondrial 16S rRNA. Nucleic Acids Res. **9**:293–307.
8. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
9. **Bromley, S., L. Hereford, and M. Rosbash.** 1982. Further evidence that the *rna2* mutation of *Saccharomyces cerevisiae* affects mRNA processing. Mol. Cell. Biol. **2**:1205–1211.
10. **Dobson, M. J., M. F. Tuite, N. A. Roberts, A. J. Kingsman, and S. M. Kingsman.** 1982. Conservation of high efficiency promoter sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. **10**:2625–2637.
11. **Donahue, T. F., P. J. Farabaugh, and G. R. Fink.** 1982. The nucleotide sequence of the *HIS4* region of yeast. Gene **18**:47–59.
12. **Fried, H. M., N. J. Pearson, C. H. Kim, and J. R. Warner.** 1981. The genes for fifteen ribosomal proteins of *Saccharomyces cerevisiae*. J. Biol. Chem. **256**:10176–10183.
13. **Fried, H. M., and J. R. Warner.** 1981. Cloning of yeast gene for tricodermin resistance and ribosomal protein L3. Proc. Natl. Acad. Sci. U.S.A. **78**:238–242.
14. **Fried, H. M., and J. R. Warner.** 1982. Molecular cloning and analysis of yeast gene for cycloheximide resistance and ribosomal protein L29. Nucleic Acids Res. **10**:3133–3148.
15. **Friesen, J. D., M. Tropak, and G. An.** 1982. Mutations in the *rplJ* leader of *Escherichia coli* that abolish feedback regulation. Cell **32**:361–369.
16. **Gorenstein, C., and J. R. Warner.** 1976. Coordinate regulation of the synthesis of eukaryotic ribosomal proteins. Proc. Natl. Acad. Sci. U.S.A. **73**:1547–1551.
17. **Grant, P., L. Sanchez, and A. Jiminez.** 1974. Cryptopleurine resistance: genetic locus for a 40S ribosomal component in *Saccharomyces cerevisiae*. J. Bacteriol. **120**:1308–1314.
18. **Grant, P. G., D. Schindler, and J. E. Davies.** 1976. Mapping of tricodermin resistance in *Saccharomyces cerevisiae*: a genetic locus for a component of the 60 S ribosomal subunit. Genetics **83**:667–673.
19. **Hawthorne, D. C., and U. Leupold.** 1974. Suppressor mutations in yeast. Curr. Top. Microbiol. Immunol. **64**:1–47.
20. **Itoh, I., K. Higo, and E. Otaka.** 1979. Isolation and characterization of twenty-three ribosomal proteins from large subunits of yeast. Biochemistry **18**:5787–5793.
21. **Kief, D. R., and J. R. Warner.** 1981. Coordinate control of synthesis of ribosomal ribonucleic acid and ribosomal proteins during nutritional shift-up in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **1**:1007–1015.
22. **Leer, R. J., M. M. C. van Raamsdonk-Duin, C. M. Molenaar, L. H. Cohen, W. H. Mager, and R. J. Planta.** 1982. The structure of the gene coding for the phosphorylated ribosomal protein S10 in yeast. Nucleic Acids Res. **10**:5869–5878.
23. **Lindahl, L., and J. M. Zengel.** 1982. Expression of ribosomal genes in bacteria. Adv. Genet. **21**:53–121.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Masurekar, M., E. Palmer, B.-I. Ono, J. M. Wilhelm, and F. Sherman.** 1981. Misreading of the ribosomal suppressor SUP46 due to an altered 40S subunit in yeast. J. Mol. Biol. **147**:381–390.
26. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499.
27. **McNeil, J. B., and J. D. Friesen.** 1981. Expression of the *Herpes simplex* virus thymidine kinase gene in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **184**:386–393.
28. **Nomura, M., D. Dean, and J. L. Yates.** 1982. Feedback

regulation of ribosomal protein synthesis in *Escherichia coli*. Trends Biochem. Sci. 7:92–95.
29. **Olins, P., and M. Nomura.** 1981. Translational regulation by ribosomal protein S8 in *Escherichia coli*: structural homology between rRNA binding site and feedback target on mRNA. Nucleic Acids Res. 9:1757–1764.
30. **Pearson, N. J., H. M. Fried, and J. R. Warner.** 1982. Yeast uses translational control to compensate for extra copies of a ribosomal protein gene. Cell 29:347–355.
31. **Proudfoot, N. J., and G. G. Brownlee,** 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature (London) 263:211–214.
32. **Rosbash, M., P. K. W. Harris, J. L. Woolford, Jr., and J. L. Teem.** 1981. The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24:679–686.
33. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids Res. 3:2387–2398.
34. **Stocklein, W., and W. Piepersberg.** 1980. Altered ribosomal protein L29 in a cycloheximide-resistant strain of *Saccharomyces cerevisiae*. Curr. Genet. 1:177–183.
35. **Storms, R. K., J. B. McNeil, P. S. Khandekar, G. An, J. Parker, and J. D. Friesen.** Chimeric plasmids for cloning of deoxyribonucleic acid sequences in *Saccharomyces cerevisiae*. J. Bacteriol. 140:73–82.
36. **Struhl, K.** 1982. Regulatory sites for *his*3 gene expression in yeast. Nature (London) 300:284–287.
37. **Warner, J. R., and C. Gorenstein.** 1977. The synthesis of eukaryotic ribosomal proteins *in vitro*. Cell 11:201–212.
38. **Warner, J. R., and C. Gorenstein.** 1978. Yeast has a true stringent response. Nature (London) 275:338–339.
39. **Warner, J. R., R. J. Tushinski, and P. J. Wejksnora.** 1980. Coordination of RNA and proteins in eukaryotic ribosome production, pp. 889–902. *In* G. Chambliss, G. R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (ed.), Ribosomes: structure, function and genetics. University Park Press, Baltimore.
40. **Woolford, J. L., Jr., and M. Rosbash.** 1981. Ribosomal protein genes rp 39(10-78), rp 39(11-40), rp 51, and rp 52 are not contiguous to other ribosomal protein genes in the *Saccharomyces cerevisiae* genome. Nucleic Acids Res. 9:5021–5036.
41. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell 28:563–573.