Volume 15 Number 21 1987 Nucleic Acids Research

Isolation and characterization of a *Neurospora crassa* ribosomal protein gene homologous to *CYH2* of yeast

Carol A.Kreader and Joyce E.Heckman

Department of Chemistry, Indiana University, Bloomington, IN 47405, USA

Received July 24, 1987; Revised and Accepted September 23, 1987

ABSTRACT

We have isolated and characterized a *Neurospora crassa* gene homologous to the yeast *CYH2* gene encoding L29, a cycloheximide sensitivity-conferring protein of the cytoplasmic ribosome. The cloned *Neurospora* gene was isolated by cross-hybridization to *CYH2*. It was sequenced from both cDNA and genomic clones. The coding region is interrupted by seven intervening sequences. Its deduced amino acid sequence shows 70% homology to that of yeast ribosomal protein L29 and 60% homology to that of mammalian ribosomal protein L27', suggesting that the protein has an important role in ribosomal function. The pattern of codon usage is highly biased, consistent with high translation efficiency. There is a single copy of this gene in *N. crassa*, and R. Metzenberg and coworkers have mapped its genetic location to the vicinity of the *cyh-2* locus.

INTRODUCTION

Ribosomes are composed of 50 to 80 different proteins and 2 to 4 different RNA molecules, the exact number of each depending upon the source. Both prokaryotes and eukaryotes regulate the synthesis of ribosomal components under a variety of growth conditions (reviewed in 1, 2). As a consequence, ribosome biogenesis is an excellent system for studying the mechanisms used to regulate a large group of functionally related genes.

The isolation of cloned genes for ribosomal proteins has proven invaluable to the study of gene regulation. Nomura and coworkers used cloned genes to demonstrate that, in most cases, the regulation of ribosomal protein gene expression in *E. coli* is subject to feedback inhibition at the level of translation (reviewed in 3). In eukaryotes, cloned genes have been used to show that ribosomal protein genes are not clustered as in *E. coli*, many are present in more than one copy, and most contain one or more introns (reviewed in 4). Additional studies have delimited transcriptional promoter elements in yeast (5-10) and in mice (11), and have detected post-transcriptional control of ribosomal protein production in yeast (12-15).

In addition to serving as an attractive model system for the study of

gene expression, ribosomal protein genes are of interest from an evolutionary point of view. Ribosomes isolated from diverse organisms show similar subunit structure, and the mechanism for translating nucleic acid sequences into proteins is comparable in all systems analyzed. Therefore, one would expect the functionally important structures of ribosomal components to be conserved from one organism to the next. Indeed, a vast database of rRNA sequences has been compiled, and these molecules do have conserved structures which are important to ribosome function (reviewed in 16). More recently, the sequences of ribosomal protein genes from several sources have been determined, and the deduced amino acid sequences have been compared to reveal some intriguing homologies. In particular, the homologies found between proteins that confer sensitivity to the drugs cryptopleurine (cry) and cycloheximide (cyh) are especially striking. The amino acid sequence of the yeast *CRY1* protein (rp59) is 80% identical to that of ribosomal protein S14 from hamsters and humans, and 37% identical to that of S11 from *E. coli* (17, 9). The sequence of the yeast *CYH2* protein (L29) is 62% identical to that of mouse L27′ (18, 19). The apparently strong evolutionary constraints upon the structures of these proteins suggest that they play key roles in translation.

We have begun to study the genes for proteins from both cytoplasmic and mitochondrial ribosomes of *Neurospora crassa*, to investigate the control of their expression and to study structure-function relationships of the genes and their products. We report the first sequence for a cytoplasmic ribosomal protein gene from *N. crassa*. The cloned gene was isolated by cross-hybridization with *CYH2* from yeast. Its coding region is interrupted by seven intervening sequences, all of which conform fairly well to the consensus pattern for *N. crassa* introns. The deduced amino acid sequence is 70% identical to that of yeast L29 and 60% identical to that of mouse L27′, indicating that this protein has an important role in ribosome function. The pattern of codon usage is highly biased, consistent with the expectation that the mRNA for this gene is translated efficiently. We show that there is one copy of this gene in the *N. crassa* genome, and R. Metzenberg and coworkers have used restriction fragment polymorphism mapping to localize this gene to the area of the *cyh-2* locus on linkage group V.

MATERIALS AND METHODS

Phages, strains, and plasmids

The λJ1:74A genomic library was a gift from M. Orbach and the laboratory of C. Yanofsky (Stanford University). It was constructed by cloning 10 to 23 kb *Sau*3A fragments from a partial digest of *Neurospora crassa* (strain 74A) DNA

between the *BamHI* sites in replacement vector λJ1 (20). λJ1 clones were propagated in *E. coli* strain DG75 (21).

The λgt11 library of *N. crassa* cDNA was constructed by M. Sachs (22), and was a gift from the laboratory of U.L. RajBhandary. λgt11 clones were propagated in *E. coli* strain Y1090 or Y1088 (23).

Fragments of *N. crassa* DNA from phage clones were recloned into pUC18 (24), and propagated in *E. coli* strain JM103 (25).

The cloned gene for yeast *CYH2* (18) was a gift from J. Teem.

Neurospora crassa strain *nuc-1*, a slime mutant lacking a cell wall, was a gift from R. Metzenberg (University of Wisconsin).

Screening and isolation of phage clones

E. coli was pre-infected with phage and plated in top agarose essentially as described by Maniatis *et al.* (26) for plating bacteriophage lambda. λJ1 clones formed plaques on DG75 after 12 to 15 hours at 37°C. λgt11 clones in Y1090 were preincubated at 42°C for one hour, and then formed plaques after 4 to 6 hours at 37°C. Phage plaques were blotted, and their DNA denatured and immobilized onto nitrocellulose filters by the method of Schleif and Wensink (27).

DNA-DNA hybridizations were performed using modified BLOTTO procedures (28). Normal stringency prehybridization and hybridization treatments were in 50% formamide, 5xSSC (0.75 M NaCl, 75 mM sodium citrate), 0.5% non-fat dry milk. Incubations were overnight (15-20 hours) at 42°C. After hybridization, the filters were washed three times for 30 minutes each in 15% formamide, 2xSSC (0.3 M NaCl, 30 mM sodium citrate), 0.1% SDS at 42°C, then briefly rinsed in 2xSSC. Low stringency prehybridization and hybridization treatments were in 10xSSC, 0.5% non-fat dry milk. (Recently, however, we have discovered that prehybridization in a filtered solution of 10xSSC, 5% milk appears to give better blocking and cleaner backgrounds for low stringency hybridization. Filters are then briefly rinsed with 10xSSC, 0.5% milk, 0.05% SDS and treated as usual for hybridization.) Incubations were for 20 to 24 hours at 50°C. Post-hybridization washes were as above, except that the wash solution contained 10xSSC, 0.5% milk, 0.05% SDS, and incubations were at 50°C.

Probes for hybridizations were fragments of cloned ribosomal protein genes, labelled at their 5′ ends and gel purified as described by Maxam and Gilbert (29). Isolated probe fragments were passed through spin columns of Sephadex G-50 resin (30, 26) to remove gel particles. 5×10^5 CPM were used for 137 mm filters and 2×10^5 CPM were used for 82 mm filters.

Preparation and analysis of DNA

Pure recombinant phage from the λgt11 library were propagated in Y1088 in

liquid culture, separated from host nucleic acids by tossing with a slurry of DEAE-cellulose, and subjected to DNA extraction, using methods described by Silhavy (31). Attempts to prepare phage DNA from λJ1 clones by this same procedure failed, probably due to slow phage propagation. Instead, λJ1 phage were propagated on plates. Phage were eluted from plates and purified by passing them through a column of DEAE-cellulose as described by Carlock (32).

N. crassa DNA was prepared from strain *nuc-1*. This wall-less mutant was grown in 7.5% sorbitol, 1.5% sucrose, 1xVogel's salts (33) with gentle agitation at 30°C. Cells were harvested and broken in 15% sucrose, 10 mM Tricine-KOH (pH 7.5), 0.2 mM EDTA, with 15 to 20 strokes of a Potter homogenizer. Nuclei and cell debris were pelleted at 1000 x g for 30 minutes. The pellet was resuspended in 20 mM Tris-HCl (pH 7.5), 2 mM EDTA, and membranes were lysed by the addition of Sarkosyl to 2%. 0.7664 g CsCl/ml of final volume were added. This solution was centrifuged at 20,000 RPM in an SW28 rotor for 2 hours, and the protein/lipid clot was removed from the surface and discarded. CsCl density gradients were established by centrifugation, and DNA was recovered from the gradients as described by Maniatis *et al.* (26).

N. crassa DNA was digested with restriction enzymes and the fragments were fractionated by electrophoresis on 1% agarose gels in Tris-acetate buffer using procedures described by Maniatis *et al.* (26). Restriction patterns were transferred to nitrocellulose by the Southern (34) transfer method, and hybridized with 5′ end-labelled probes as described above for phage blots.

Plasmid DNA was prepared from stationary cultures using the boiling procedure described by Maniatis *et al.* (26), and twice purified on CsCl gradients.

The nucleotide sequence of both genomic and cDNA clones in pUC18 was determined by the Maxam and Gilbert method (29).

RESULTS AND DISCUSSION

Isolation of clones and determination of nucleotide sequences

The *CYH2* gene, which encodes cytoplasmic ribosomal protein L29 of *Saccharomyces cerevisiae*, was used as a hybridization probe to isolate a clone for the analogous gene from *Neurospora crassa*. The 600 bp *Xho*I-*Bgl*II fragment of *CYH2*, which spans the 3′ half of its single intron and two-thirds of the coding region (18), was used to screen a genomic library of *N. crassa* DNA in λJ1. A single phage clone was isolated using low stringency hybridization. Phage DNA was prepared, digested with restriction enzymes, and hybridized with the *CYH2* probe to identify fragments which contained the *N. crassa* gene. A 4

kb *EcoRI* fragment was identified, recloned into pUC8, and partially sequenced. Nucleotide sequence homology between the yeast gene and the *N. crassa* DNA was used to establish a reading frame for the *N. crassa* cytoplasmic ribosomal protein gene (*crp-1*). Additional short stretches of sequence, which encoded multiple translational stop signals, were found between blocks of homology in *crp-1* but were absent from the yeast gene. The latter suggested that *crp-1* contains several closely spaced introns. In addition, this first genomic clone encoded a polypeptide corresponding to only the amino-terminal one-fourth of L29, and ended in what was later identified as intron 5 of *crp-1*.

To determine, unequivocally, the protein coding sequence of *crp-1*, cDNA clones were isolated and sequenced. The 180 bp *Xho*I-*Pst*I genomic fragment (indicated by a "c" in figure 1B), which contains exons 3 and 4 and introns 2 and 3, was labelled and used to screen a λgt11 library of *N. crassa* cDNA using normal stringency hybridization. Fifteen positive clones were identified among roughly 3000 plaques. DNA was isolated from eight, and the three longest cDNA inserts were recloned into pUC18 for sequence determination.

To complete the genomic DNA sequence of *crp-1*, additional λJ1 clones were

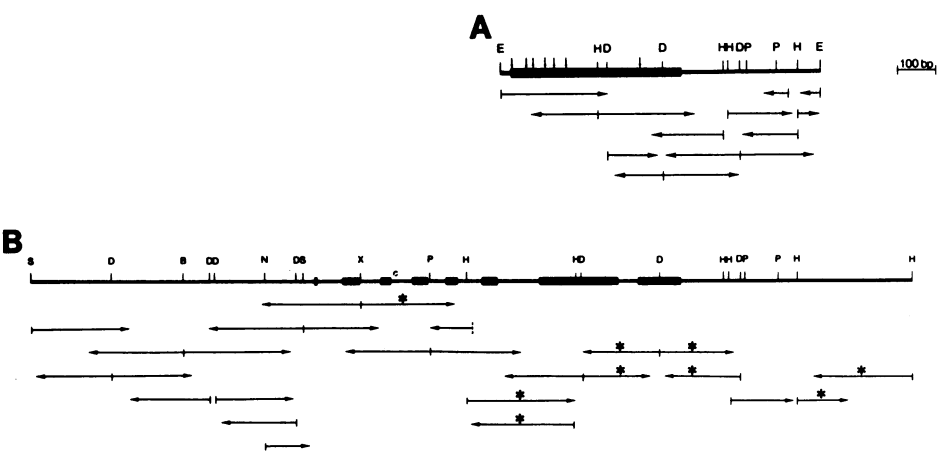
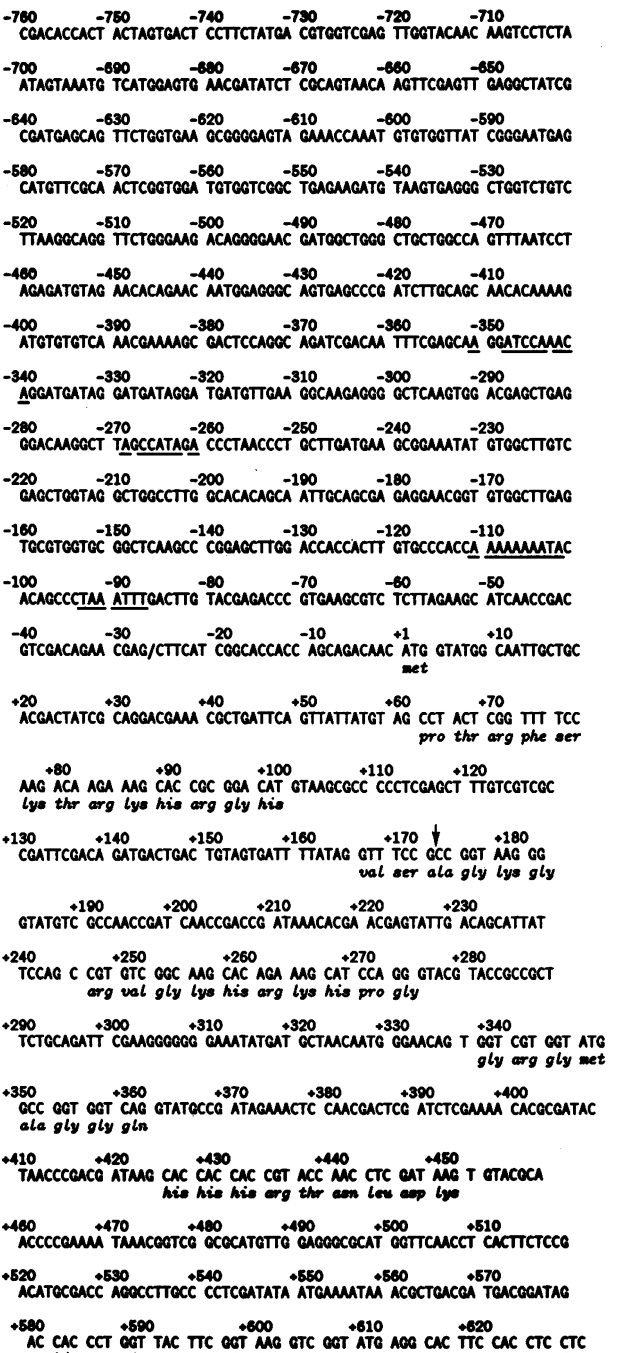


Figure 1. Positions of restriction endonuclease sites and strategy for sequencing of DNA. Horizontal arrows indicate the direction and extent of sequence determination from cDNA (A) and genomic DNA (B) clones. The vertical dotted line indicates the position of the λJ1 polylinker at the 3′ end of the first genomic clone. Only restriction sites used for sequencing are shown. B, *BamHI*; D, *Dde*I; E, *EcoRI*; H, *Hinf*I; N, *Nci*I; P, *Pst*I; S, *Sal*I; X, *Xho*I. Positions of coding sequences are indicated with thickened lines. Intron positions are shown on the cDNA map with vertical arrows. The "c" on line B indicates the 180 bp *Xho*I-*Pst*I fragment used as a probe to isolate cDNA clones. Asterisks indicate fragments used in the mixed probe for the genomic southern hybridization (figure 6).


-760 -750 -740 -730 -720 -710
CGACACCACT ACTAGTGACT CCTTCTATGA CGTGGTCGAG TTGGTACAAC AAGTCCTCTA
-700 -690 -680 -670 -660 -650
ATAGTAAATG TCATGGAGTG AACGATATCT CGCAGTAACA AGTTCGAGTT GAGGCTATCG
-640 -630 -620 -610 -600 -590
CGATGAGCAG TTCTGGTGAA GCGGGGAGTA GAAACCAAAT GTGTGGTTAT CGGGAATGAG
-580 -570 -560 -550 -540 -530
CATGTTCGCA ACTCGGTGGA TGTGGTCGGC TGAGAAGATG TAAGTGAGGG CTGGTCTGTC
-520 -510 -500 -490 -480 -470
TTAAGGCAGG TTCTGGGAAG ACAGGGGAAC GATGGCTGGG CTGCTGGCCA GTTTAATCCT
-460 -450 -440 -430 -420 -410
AGAGATGTAG AACACAGAAC AATGGAGGGC AGTGAGCCCG ATCTTGCAGC AACACAAAAG
-400 -390 -380 -370 -360 -350
ATGTGTGTCA AACGAAAAGC GACTCCAGGC AGATCGACAA TTTCGAGCAA GGATCCAAAC
-340 -330 -320 -310 -300 -290
AGGATGATAG GATGATAGGA TGATGTTGAA GGCAAGAGGG GCTCAAGTGG ACGAGCTGAG
-280 -270 -260 -250 -240 -230
GGACAAGGCT TAGCCATAGA CCCTAACCCT GCTTGATGAA GCGGAAATAT GTGGCTTGTC
-220 -210 -200 -190 -180 -170
GAGCTGGTAG GCTGGCCTTG GCACACAGCA ATTGCAGCGA GAGGAACGGT GTGGCTTGAG
-160 -150 -140 -130 -120 -110
TGCGTGGTGC GGCTCAAGCC CGGAGCTTGG ACCACCACTT GTGCCCACCA AAAAAAATAC
-100 -90 -80 -70 -60 -50
ACAGCCCTAA ATTTGACTTG TACGAGACCC GTGAAGCGTC TCTTAGAAGC ATCAACCGAC
-40 -30 -20 -10 +1 +10
GTCGACAGAA CGAG/CTTCAT CGGCACCACC AGCAGACAAC ATG GTATGG CAATTGCTGC
met
+20 +30 +40 +50 +60 +70
ACGACTATCG CAGGACGAAA CGCTGATTCA GTTATTATGT AG CCT ACT CGG TTT TCC
pro thr arg phe ser
+80 +90 +100 +110 +120
AAG ACA AGA AAG CAC CGC GGA CAT GTAAGCGCC CCCTCGAGCT TTGTCGTCGC
lys thr arg lys his arg gly his
+130 +140 +150 +160 +170 +180
CGATTCGACA GATGACTGAC TGTAGTGATT TTATAG GTT TCC GCC GGT AAG GG
val ser ala gly lys gly
+190 +200 +210 +220 +230
GTATGTC GCCAACCGAT CAACCGACCG ATAAACACGA ACGAGTATTG ACAGCATTAT
+240 +250 +260 +270 +280
TCCAG C CGT GTC GGC AAG CAC AGA AAG CAT CCA GG GTACG TACCGCCGCT
arg val gly lys his arg lys his pro gly
+290 +300 +310 +320 +330 +340
TCTGCAGATT CGAAGGGGGG GAAATATGAT GCTAACAATG GGAACAG T GGT CGT GGT ATG
gly arg gly met
+350 +360 +370 +380 +390 +400
GCC GGT GGT CAG GTATGCCG ATAGAAACTC CAACGACTCG ATCTCGAAAA CACGCGATAC
ala gly gly gln
+410 +420 +430 +440 +450
TAACCCGACG ATAAG CAC CAC CAC CGT ACC AAC CTC GAT AAG T GTACGCA
his his his arg thr asn leu asp lys
+460 +470 +480 +490 +500 +510
ACCCCGAAAA TAAACGGTCG GCGCATGTTG GAGGGCGCAT GGTTCAACCT CACTTCTCCG
+520 +530 +540 +550 +560 +570
ACATGCGACC AGGCCTTGCC CCTCGATATA ATGAAAATAA ACGCTGACGA TGACGGATAG
+580 +590 +600 +610 +620
AC CAC CCT GGT TAC TTC GGT AAG GTC GGT ATG AGG CAC TTC CAC CTC CTC
tyr his pro gly tyr phe gly lys val gly met arg his phe his leu leu

```
+630        +640        +650        +660        +670        +680
 CGC AAC CAC CAG TGG GCC CCC ATC CTC AAC ATT GAG AAG CTC TGG ACT CTC
 arg asn his gln trp ala pro ile leu asn ile glu lys leu trp thr leu

          +690        +700        +710        +720        +730
GTC CCC GCT GAG GCC CGC GAG AAG TAC GTC TCT GGT GCC GCC ACC GAG ACT
val pro ala glu ala arg glu lys tyr val ser gly ala ala thr glu thr

        +740        +750        +760        +770        +780
GCT CCC GTC ATC GAC CTC CTT TCC CAC GGC TAT GCC AAG CTC CTC GGC AAG G
ala pro val ile asp leu leu ser his gly tyr ala lys leu leu gly lys

      +790       +800       +810       +820       +830       +840
GTAAGA ACAAGCAATT GCAAAGAAAC ACCCCCGAGA AGGCTCGTCT AACAACACAT TCTAG GC
                                                              gly

   +850        +860        +870        +880        +890
CGT CTC CCC CAG GTC CCC ATC GTC GTC CGT GCC CGC TAC GTC TCC GCC GAG
arg leu pro gln val pro ile val val arg ala arg tyr val ser ala glu

 +900        +910        +920        +930        +940
GCT GAG CGC AAG ATC AAG GAG GCC GGT GGT GTC ATT GAG CTC GTT GCT TAA A
ala glu arg lys ile lys glu ala gly gly val ile glu leu val ala ***

+950       +960       +970       +980       +990       +1000
TTCTGAATGC GGCATGATAT ATTCAAGGAC TGGCGATTTG GCATCACGGT ACGAGGGAAA

+1010      +1020      +1030      +1040      +1050      +1060
CCGGAAACGA ACCCGATGGT CATGGATCGA ACAGTCGGAC AGACGAACAA CGACAAGATT

+1070      +1080      +1090      +1100      +1110      +1120
CGGTCCCGGA GTCCTGGATG GGTCAGAGGT GGTTGAAAAT CTCAGCGCCT CGTCTGCAGC

+1130      +1140      +1150      +1160      +1170      +1180
ATTTAGGGAA ATGGCAGTTC ATTTGCATCG CAACATTGGG TCCGGGATTT CGTTCACAAC

+1190      +1200      +1210      +1220      +1230      +1240
CGTGCCATGC TGCAGACTGG CGAACTCGGG TGGTCCCGAG GGGTAACCCG GGCCGCGGCT

+1250      +1260      +1270      +1280      +1290      +1300
GCGTCGCAAA CTGCTGATTC AGGGGCAAGT AACGAATAAA TGAAGGCCTA TGCCCGACGG

+1310      +1320      +1330      +1340      +1350      +1360
TCGGTTGATC TCA/TCCATGA GCATTTCCTT GGCCATTGTG AAGTTGAACA TTCGTCCAGA

+1370      +1380      +1390      +1400      +1410      +1420
AGCCCACTTT CAGCACGTTG GATTTGTTTG TCCTGTCGTC ATGTCATGAG AACCCCAGTG

+1430      +1440      +1450
TGAGGCCGTT ATATGCCTTT GTCTG
```

Figure 2. DNA sequence of the *crp-1* gene and deduced amino acid sequence of the encoded ribosomal protein. Slashes mark the beginning and end of the longest cDNA sequence. Sequences homologous to transcriptional control elements from yeast ribosomal protein genes are underlined. An arrow marks the position of the intron in yeast *CYH2*.

isolated and pieces recloned as described above, using fragments from the 3′ ends of the cDNA clones as hybridization probes. This second group of genomic clones had the same 5′ end as the first clone, but continued 6 kb farther in the 3′ direction. A 750 bp *Sal*I fragment of upstream sequence and a 2.6 kb *Sal*I fragment of coding plus downstream sequence were recloned into pUC18 for sequence determination.

The nucleotide sequence of *crp-1* was determined from four of the independently isolated, overlapping clones--two of cDNA and two of genomic DNA. Both strands were analyzed for the entire coding region, as well as for

760 nucleotides of upstream and 500 nucleotides of downstream sequence (figures 1 and 2).

Upstream homologies

Analysis of the upstream sequence from *crp-1* revealed short stretches homologous to conserved elements found in yeast. Sequence comparisons of yeast ribosomal protein genes have identified three blocks of homology, in defined order--HOMOLI (consensus, A A C A T C C/T G/A T A/G C A), RPG (consensus, A C C C A T A C A T T/C T/A), and a thymine-rich region--120 to 450 nucleotides upstream from the transcriptional start site in most of these genes (35, 36, 6). HOMOLI and RPG are unique to ribosomal protein genes; however, the T-rich region may be related to poly dA-dT stretches (6), which have been shown to function as promoter elements for constitutive transcription of "housekeeping" genes in yeast (37). A computer-facilitated search for similar elements in *crp-1* detected a block of sequence at nucleotides -351 to -340 that matches HOMOLI in 9 of 12 positions, and a second block at nucleotides -269 to -261 that matches RPG in 7 of 12 positions (underlined in figure 2). In addition, poly dA-dT stretches were found at nucleotides -111 to -102 and -93 to -87.

Deletion analyses have demonstrated that the conserved ribosomal protein gene elements are essential for high level expression in yeast (6-9). In addition, binding assays indicate that a single protein factor from yeast extracts interacts with both HOMOLI and RPG sequences (5, 10). Analogous studies in *Neurospora* are essential to determine what role, if any, the homologous elements from *crp-1* play in transcription.

Intron sequences

The coding region of *crp-1* is interrupted by seven intervening sequences, as deduced by comparing cDNA and genomic DNA sequences. These show general agreement with the consensus sequences for *Neurospora* introns at the 5′ and 3′ junctions and for an internal sequence, as illustrated in figure 3. I5 is one notable exception, in that its 3′ junction is TAAG rather than pyrimidine-AG.

It can be noted that the internal conserved sequence for yeast introns is one of the eight possibilities for *N. crassa* introns (bottom of figure 3). Although it has not been demonstrated yet in *N. crassa*, the invariant A in the internal conserved sequence presumably serves as the branch site for lariat formation during intron excision, as seen for type II introns of yeast and mammals. Thus, the splicing machinery of *N. crassa* must require less stringent adherence to primary structure than that of yeast. However, it is stricter than that of higher eukaryotes, in which no internal conserved sequence has been identified.

I1	GTATGG----GCTGATT----11----TAG	47 bp
I2	GTAAGC----ACTGACT----12----TAG	50 bp
I3	GTATGT----ATTGACA---- 9----CAG	62 bp
I4	GTACGT----GCTAACA---- 7----CAG	62 bp
I5	GTATGC----ACTAACC---- 6---TAAG	63 bp
I6	GTACGC----GCTGACG---- 8----TAG	126 bp
I7	GTAAGA----TCTAACA---- 8----TAG	61 bp

N. crassa consensus
A A A T C
GTA GT---- CT AC ---9-16--- AG
C G G A T

S. cerevisiae consensus GTATGT---TACTAACA--18-53--PyAG

higher eukaryote consensus GTAAGT----------CAG

Figure 3. Comparison of *crp-1* introns with consensus sequences from *Neurospora crassa*, *Saccharomyces cerevisiae*, and higher eukaryotes (20). 5′ and 3′ splice junction and internal conserved sequences are presented. The internal numbers indicate nucleotides separating internal conserved sequences from the 3′ splice signals. Lengths of *crp-1* introns are given to the right of each in base pairs (bp).

The positions and lengths of the introns in *crp-1* seem to be typical as well. I1 through 5 are clustered within the first one-fourth of the coding region in *crp-1* (figures 1 and 2). The only other *N. crassa* gene reported to have nearly as many introns, the β-tubulin gene, has five of its six introns similarly located in the first one-fourth of the coding region (20). This structural similarity may be a coincidence however, since in other *N. crassa* genes that contain only one or two introns, these introns are not necessarily restricted to the 5′ end. Notably, the gene for histone H3 has a single intron in the middle of its coding region (38), and the gene for the proteolipid subunit of mitochondrial ATP synthase is interrupted by its second intron immediately before the last codon (39). In yeast, most ribosomal protein genes have a single intron near their 5′ ends. Interestingly, the position of the intron in *CYH2* (indicated by an arrow in figure 2) does not correspond to any of the seven introns in *crp-1*. In addition, all seven introns in *crp-1* are quite short, which seems to be typical for *N. crassa*. The introns in *crp-1* range in size from 47 to 126 nucleotides (figure 3). The longest intron reported for *N. crassa*, I1 of β-tubulin (20), is 240 nucleotides in length. In contrast, introns of yeast range in size from 300

to 500 nucleotides, and introns of higher eukaryotes can be several kilobases in length.

Amino acid homology and codon usage

The deduced amino acid sequence is highly conserved between *Neurospora crassa crp-1* and *Saccharomyces cerevisiae CYH2*. In addition, the gene sequence for mouse L27′ was reported recently, and it is homologous to those of the fungal proteins (19). The fungal genes encode proteins of 149 amino acids, or approximately 16,500 mw, whereas the mouse gene encodes a protein of 148 amino acids. 50% of the amino acids are identical in all three proteins, as illustrated in figure 4. The *N. crassa crp-1* protein is 70% homologous to yeast L29 and 60% homologous to mouse L27′. The homology is most extensive in the amino-terminal one-third of the sequences, where 80% of the residues are identical in all three proteins. Eight proline and seventeen glycine residues are conserved. Mouse L27′ has an additional proline and there are four glycine residues that are not aligned, but all are in the carboxyl halves of their respective proteins. Therefore, the overall shapes of the amino-terminal portions of these proteins must be quite similar. Mouse L27′ contains two cysteine residues, one in the middle and one near the carboxy-terminus, which are absent from the fungal proteins. If these cysteines form a disulfide bond, the conformation of the mouse protein might be different from those of the fungal proteins in the carboxyl end.

It is of interest that the amino-terminal portions of these proteins are so highly conserved, since this region surrounds the position of two mutations giving rise to cycloheximide resistance in yeast. In the yeast mutants, glutamine 37 (indicated by an arrow in figure 4) is replaced by either glutamate or lysine (40). This position is normally occupied by methionine in mouse L27′. Therefore, for normal protein structure, this position apparently must be filled by a residue which is large and polar, but not charged. In addition, this position is followed immediately by three histidines and an arginine, reminiscent of an enzyme active site. Since this region of the protein is conserved over a long evolutionary distance--from fungi to a mammal--there must be strong constraints upon its structure, indicating that it serves an important function in protein synthesis. Indeed, cycloheximide inhibits the EF-2 mediated translocation of peptidyl-tRNA on eukaryotic ribosomes (41, 42). The cycloheximide-sensitivity conferring protein may be involved in this function directly, or indirectly by interacting with another ribosomal protein that is involved.

The cryptopleurine-sensitivity conferring protein presents a picture similar to that for cycloheximide. Cryptopleurine and cycloheximide block the

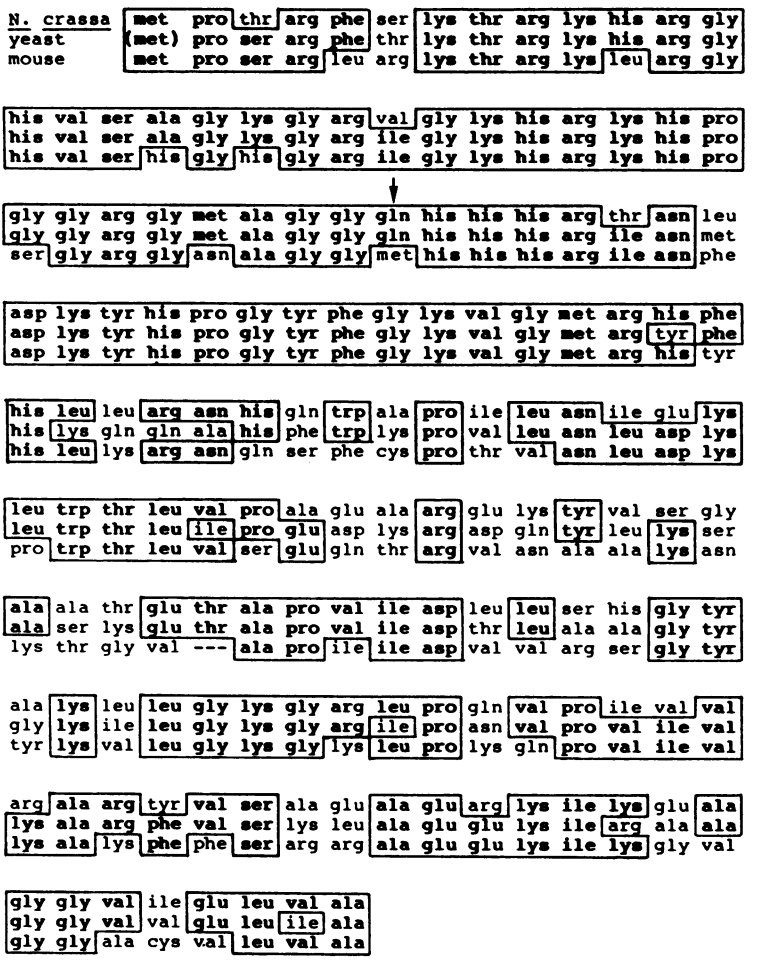


Figure 4. Comparison of the deduced amino acid sequences for the *crp-1* protein and homologous ribosomal proteins. The amino acid sequences of the *N. crassa crp-1* protein, *S. cerevisiae* L29 (18), and mouse L27′ (19) are shown. Identical residues are enclosed in boxes. Glutamine 37, the site of mutations giving rise to cycloheximide resistance in yeast, is marked with an arrow.

same step in protein synthesis (42). However, whereas cycloheximide acts upon the large ribosomal subunit, cryptopleurine interacts with the small subunit. As might be expected, the ribosomal protein affected by cryptopleurine is highly conserved. The sequence of the *CRY1* protein of yeast (rp59) is 80% identical to that of L14 of hamsters and humans, and 37% identical to that of S11 from *E. coli* (17, 9).

Although the deduced amino acid sequences of *N. crassa crp-1* and yeast

	Nc	Sc		Nc	Sc		Nc	Sc
phe UUU	1	0	pro CCU	2	1	lys AAA	0	6
UUC	2	5	CCC	5	1	AAG	13	12
	3	5	CCA	1	6		13	18
			CCG	0	0			
leu UUA	0	0		8	8	asp GAU	1	1
UUG	0	9				GAC	1	4
CUU	1	0	thr ACU	3	4		2	5
CUC	11	0	ACC	2	0			
CUA	0	0	ACA	1	1	glu GAA	0	5
CUG	0	0	ACG	0	0	GAG	8	0
	12	9		6	5		8	5
ile AUU	2	3	ala GCU	4	8	cys UGU	0	0
AUC	4	6	GCC	10	3	UGC	0	0
AUA	0	0	GCA	0	1		0	0
	6	9	GCG	0	0			
				14	12	trp UGG	2	2
met AUG	3	4						
			tyr UAU	1	1	arg CGU	5	1
val GUU	2	6	UAC	4	4	CGC	5	0
GUC	10	4		5	5	CGA	0	0
GUA	0	0				CGG	1	0
GUG	0	0	his CAU	2	3	AGA	2	11
	12	10	CAC	10	7	AGG	1	0
				12	10		14	12
ser AGU	0	0						
AGC	0	0	gln CAA	0	4	gly GGU	12	18
UCU	1	2	CAG	3	0	GGC	5	0
UCC	4	2		3	4	GGA	1	0
UCA	0	1				GGG	0	0
UCG	0	0	asn AAU	0	1		18	18
	5	5	AAC	3	2			
				3	3			

Figure 5. Comparison of codon usage patterns for the *N. crassa crp-1* (Nc) and *S. cerevisiae CYH2* (Sc) ribosomal protein genes.

CYH2 are homologous, their patterns of codon usage differ and are biased towards the patterns for highly expressed genes in their respective organisms. As mentioned above, all eight proline residues are in conserved locations; however, the nucleotide sequences encoding them differ (figure 5). *Neurospora* preferentially uses CCC for proline, whereas yeast uses CCA. Highly expressed genes in *Neurospora* show a bias toward codons ending in pyrimidines rather than purines. In the case of two-codon families where a purine is the only choice in the third position, guanine is preferred and adenine is avoided. *Crp-1* uses only five codons ending in adenine, and completely avoids A-ending codons for several two-codon families (gln, glu, and lys). As might be expected, the degree of bias seen in the *crp-1* gene is similar to that of *tub-2*, the gene for β-tubulin (20), but slightly less stringent than that of the more abundantly expressed histone genes (38).

Gene copy number and chromosomal location

In eukaryotes, ribosomal protein genes can be represented at varying copy numbers in the genome. Duplication of ribosomal protein genes is common in yeast. In those cases studied, investigators have shown that both copies are expressed, but at different levels (43, 44). In addition, yeast must have

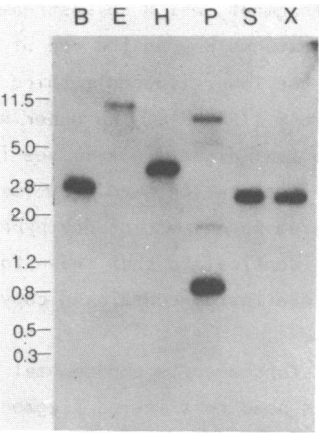


Figure 6. Determination of copy number for the *crp-1* gene. 5 μg of *N. crassa* DNA from strain *nuc-1* was digested with *BamHI* (B), *EcoRI* (E), *HindIII* (H), *PstI* (P), *SalI* (S), or *XhoI* (X). DNA fragments were separated by electrophoresis on a 1% agarose gel, transferred to nitrocellulose, and hybridized with a mixture of 5′ end labeled fragments from genomic clones using normal stringency conditions. The probe contained an approximately equal number of counts of fragments marked with asterisks in figure 1B, which were prepared for determination of the genomic sequence. A total of 2×10^5 CPM were used.

some mechanism to compensate for unequal copy number, since the proteins encoded in the duplicated genes do not accumulate above the level of proteins having single copy genes (45). In mice, ribosomal protein genes are reiterated from 8 to more than 20 times. However, most copies are processed pseudogenes and only one or a few intron-containing genes are functional (46, 47, 48). Since *crp-1* is the first cytoplasmic ribosomal protein gene from *Neurospora* to be analyzed, we did not know what type of representation to expect.

To determine the number of copies for *crp-1* present in the *Neurospora* genome, restriction patterns of *Neurospora* DNA were probed with pieces of *crp-1* genomic clones using normal stringency hybridization. The restriction fragments identified (figure 6) are consistent with the digestion pattern predicted from the restriction map of genomic clones. The *crp-1* probe hybridized to single *BamHI*, *EcoRI*, *HindIII*, *SalI*, and *XhoI* fragments, indicating that there is only one copy of *crp-1*. The *EcoRI* fragment is significantly larger than these other hybridizing fragments (10 kb versus 2-4 kb), resulting in a less efficient transfer from gel to nitrocellulose, and thus a weaker signal. Three *PstI* fragments of *crp-1* were represented in the

probe - an 800 bp coding fragment, a 1.8 kb upstream fragment, and a downstream fragment which extends beyond the end of the genomic clones (>3.5 kb). These correspond to the lower (fourth), third, and presumably, second hybridizing bands in Figure 6 (lane P). The upper hybridizing band is probably the product of partial digestion. To see if there are any copies of *crp-1* lacking introns, an identical pattern of *N. crassa* DNA was probed with fragments of *crp-1* cDNA using normal stringency hybridization. The hybridization pattern was identical to that shown for the genomic probe in figure 6. Thus, there is one intron-containing copy and no pseudogenes of *crp-1* in *Neurospora crassa*.

To characterize *crp-1* further, its chromosomal location was determined. R. Metzenberg and coworkers used restriction fragment length polymorphism analysis to map *crp-1* to the area of the *cyh-2* locus of linkage group V from *Neurospora crassa* (R. Metzenberg, personal communication). *Cyh-2* is one of at least three loci where resistance to cycloheximide has been mapped in *Neurospora* (49, 50). Thus, *cyh-2* of *Neurospora crassa* and *CYH2* of *Saccharomyces cerevisiae* apparently encode homologous ribosomal proteins. Additional experiments are in progress to determine the definitive map position of this gene.

ACKNOWLEDGEMENTS

We thank M. Orbach, C. Yanofsky, and M. Sachs for *N. crassa* DNA libraries, and J. Teem for the cloned *CYH2* gene. In addition, we thank R. Metzenberg and coworkers for mapping the chromosomal location of *crp-1*, and J. Preer for computer software. This work was supported by NIH grant GM29478-04A1 and by NIH Biomedical Research Support Grant PHS S07-RR7031H. C.K. was partially supported by a fellowship from the Lubrizol Corporation.

REFERENCES

1. Lindahl, L. and Zengel, J.M. (1982) Adv. Genet. 21, 53-121.
2. Warner, J.R., Tushinski, R.J., and Wejksnora, P.J. (1980) In Chamblis, G., Craven, G.R., Davies, J., Davis, K., Kahan, L., and Nomura, M. (eds), Ribosomes: Structure, Function, and Genetics, University Park Press, Baltimore, pp. 889-902.
3. Nomura, M., Gourse, R., and Baughman, G. (1984) Ann. Rev. Biochem. 53, 75-117.
4. Fried, H.M., and Warner, J.R. (1984) In Stein, G.S. and Stein, J.L. (eds), Recombinant DNA and Cell Proliferation, Academic Press, New York, pp. 169-192.
5. Huet, J., Cottrelle, P., Cool, M., Vignais, M.L., Thiele, D., Marck, C., Buhler, J.M., Sentenac, A., and Fromageot, P. (1985) EMBO J. 4, 3539-3547.

6. Rotenberg, M.O., and Woolford, J.L., Jr. (1986) Mol. Cell. Biol. **6**, 674-687.
7. Woudt, L.P., Smit, A.B., Mager, W.H., and Planta, R.J. (1986) EMBO J. **5**, 1037-1040.
8. Schwindinger, W.F. and Warner, J.R. (1987) J. Biol. Chem. **262**, 5690-5695.
9. Larkin, J.C., Thompson, J.R., and Woolford, J.L., Jr. (1987) Mol. Cell. Biol. **7**, 1764-1775.
10. Vignais, M.L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A., and Planta, R.J. (1987) EMBO J. **6**, 1451-1457.
11. Dudov, K.P., and Perry, R.P. (1986) Proc. Natl. Acad. Sci. USA **83**, 8545-8549.
12. Pearson, N.J., Fried, H.M., and Warner, J.R. (1982) Cell **29**, 347-355.
13. Warner, J.R., Mitra, G., Schwindinger, W.F., Studeny, M., and Fried, H.M. (1985) Mol. Cell. Biol. **5**, 1512-1521.
14. Gritz, L., Abovich, N., Teem, J.L., and Rosbash, M. (1985) Mol. Cell. Biol. **5**, 3436-3442.
15. ElBaradi, T.T.A.L., van der Sande, C.A.F.M., Mager, W.H., Raue, H.A., and Planta, R.J. (1986) Curr. Genet. **10**, 733-739.
16. Noller, H.F., Asire, M., Barta, A., Douthwaite, S., Goldstein, T., Gutell, R.R., Moazed, D., Normanly, J., Prince, J.B., Stern, S., Triman, K., Turner, S., Van Stolk, B., Wheaton, V., Weiser, B., and Woese, C.R. (1986) In Hardesty, B. and Kramer, G. (eds), Structure, Function, and Genetics of Ribosomes, Springer-Verlag, New York, pp. 143-163.
17. Chen, I.T., Dixit, A., Rhoads, D.D., and Roufa, D.J. (1986) Proc. Natl. Acad. Sci. USA **83**, 6907-6911.
18. Kaufer, N.F., Fried, H.M., Schwindinger, W.F., Jasin, M., and Warner, J.R. (1983) Nucleic Acids Res. **11**, 3123-3135.
19. Belhumeur, P., Paterno, G.D., Boileau, G., Claverie, J.M., and Skup, D. (1987) Nucleic Acids Res. **15**, 1019-1029.
20. Orbach, M.J., Porro, E.B., and Yanofsky, C. (1986) Mol. Cell. Biol. **6**, 2452-2461.
21. O'Farrell, P.H., Polisky, B., and Gelfand, D.H. (1978) J. Bacteriol. **134**, 645-654.
22. Sachs, M.S., David, M., Werner, S., and RajBhandary, U.L. (1986) J. Biol. Chem. **261**, 869-873.
23. Young, R.A., and Davis, R.W. (1983) Science **222**, 778-782.
24. Norrander, J., Kempe, T., and Messing, J. (1983) Gene **26**, 101-106.
25. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucleic Acids Res. **9**, 309-321.
26. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982), Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor.
27. Schlief, R.F., and Wensink, P.C. (1981) Practical Methods in Molecular Biology, p. 148, Springer-Verlag, New York.
28. Johnson, D.A., Gautsch, J.W., Sportsman, J.R., and Elder, J.H. (1984) Gene. Anal. Techn. **1**, 3-8.
29. Maxam, A.M., and Gilbert, W. (1980) Methods Enzymol. **65** (part I), 499-560.
30. Penefsky, H.S. (1977) J. Biol. Chem. **252**, 2891-2899.
31. Silhavy, T.J. (1984) Experiments with Gene Fusions, pp. 140-143, Cold Spring Harbor Laboratory, Cold Spring Harbor.
32. Carlock, L.R. (1986) Focus (BRL) **8** (2), 6-8.
33. Vogel, H.J. (1964) Am. Naturalist **98**, 435-446.
34. Southern, E.M. (1975) J. Mol. Biol. **98**, 503-517.
35. Teem, J.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., van Raamsdonk-Duin, M.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D., Fried, H., and Rosbash, M. (1984) Nucleic Acids Res. **12**, 8295-8312.

36. Leer, R.J., van Raamsdonk-Duin, M.M.C., Mager, W.H., and Planta, R.J. (1985) Curr. Genet. 9, 273-277.
37. Struhl, K. (1984) Proc. Natl. Acad. Sci. USA 81, 7865-7869.
38. Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H., and Planta, R.J. (1983) Nucleic Acids Res. 11, 5347-5360.
39. Sebald, W. and Kruse, B. (1984) In Papa, S., Altendorf, L., Ernster, L., and Packer, L. (eds), H^+-ATPase (ATP Synthase): Structure, Function, Biogenesis: the F_OF_1 Complex of Coupling Membranes, Adriatica Editrice, Bari, pp. 67-75.
40. Stocklein, W., Piepersberg, W., and Bock, A. (1981) FEBS Letters 136, 265-268.
41. McKeehan, W. and Hardesty, B. (1969) Biochem. Biophys. Res. Commun. 36, 625-630.
42. Barbacid, M., Fesno, M., and Vazquez, D. (1975) J. Antibiot. 28, 453-462.
43. Abovich, N., and Rosbash, M. (1984) Mol. Cell. Biol. 4, 1871-1879.
44. Leer, R.J., van Raamsdonk-Duin, M.M.C., Molenaar, C.M.T., Witsenboer, H.M.A., Mager, W.H., and Planta, R.J. (1985) Nucleic Acids Res. 13, 5027-5039.
45. Kim, C.H., and Warner, J.R. (1983) J. Mol. Biol. 165, 79-89.
46. Dudov, K.P., and Perry, R.P. (1984) Cell 37, 457-468.
47. Wiedemann, L.M., and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518-2528.
48. Wagner, M., and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560-3576.
49. Hsu, K.S. (1963) J. Gen. Microbiol. 32, 341-347
50. Vomvoyanni, V.E. and Argyrakis, M.P. (1979) J. Bacteriol. 139, 620-624.