Two complex regions, including a TATA sequence, are required for transcription by RNA polymerase I in *Neurospora crassa*

Brett M.Tyler
Department of Plant Pathology, University of California at Davis, Davis, CA 95616, USA

Received November 27, 1989; Revised and Accepted March 5, 1990

ABSTRACT

In order to define the RNA polymerase I transcriptional apparatus and how it might interact with regulatory signals, the DNA sequences necessary for 40S rRNA transcription in *Neurospora crassa* were determined. A systematic set of deletion, substitution and insertion mutations were assayed in a homologous *in vitro* system. The sequences required for transcription of the gene consist of two large domains (I and II) from −113 to −37, and −29 to +4, respectively. Complete deletion of either domain abolished transcription. Upstream sequences confer a small stimulation of transcription. Domain II includes a TATA sequence at −5 which is sensitive to a small (2 bp) substitution and which is conserved among the large rRNA genes of many organisms. Domain I includes a sequence, termed the 'Ribo box', which is also required for transcription of the *Neurospora* 5S rRNA genes (1), and which occurs in the 5′ region of a *Neurospora* ribosomal protein gene. The 5S and 40S Ribo boxes are shown to be functionally interchangeable.

INTRODUCTION

RNA polymerase I transcription accounts for a large proportion of the RNA synthesis of a eukaryotic cell (approximately 30% in rapidly growing *Neurospora crassa* mycelia (2)). Including the synthesis of the ribosomal proteins, which comprise up to 37% of the cell protein (3), the synthesis of ribosomal components constitutes a major metabolic investment by the cell. In several lower eukaryotes, the regulation of ribosomal RNA and protein syntheses are directly correlated under a variety of conditions (4, 5). However, the mechanisms for co-ordinating expression are not understood.

Transcription of vertebrate large rRNA genes by RNA polymerase I has been well characterized (reviewed in 6). The rRNA promoters of human (7), mouse (8−10), rat (11), *Xenopus* (12, 13) and yeast (14) have been defined using *in vivo* and *in vitro* transcription assays. A variety of protein fractions containing transcription factors which bind RNA polymerase I promoter elements and/or which are required for transcription *in vitro* have been isolated (reviewed in (6)). Some factors have been purified. These factors form stable complexes which can direct specific initiation by RNA polymerase I *in vitro* (reviewed in 6). In mammalian cells, RNA transcription is closely tied to growth rate. Nutrient starvation, glucocorticoids, cycloheximide or serum withdrawal, which inhibit growth, cause a rapid shutoff of RNA polymerase I transcription. In mouse cells, these regulatory events appear to be mediated by a protein tightly associated with RNA polymerase I which is required for specific transcription (15−17). This protein appears to become modified when cell growth is rapidly slowed. A similar event has been demonstrated during *Acanthamoeba* sporulation (18). However, the mechanism, if any, by which RNA polymerase I transcription is linked to 5S rRNA and ribosomal protein synthesis has not been defined in any of these organisms.

In *Neurospora crassa*, the rate of accumulation of the large rRNAs and the levels of ribosomal proteins relative to the total cell protein increase in direct proportion to the growth rate (3). During a shift-up from acetate to glucose medium, there is a co-ordinate early increase in the *de novo* synthesis of the large rRNAs and of the r-proteins 30 min following the shift (19). During a shift to poorer medium, there is a rapid cessation of new rRNA synthesis for approximately 2.5 hours before a slower rate of synthesis recommences (20). Phosphorus deprivation also leads to a rapid cessation of large rRNA synthesis (2). Ribosomal RNA synthesis is also under developmental control—during conidial germination, there is a burst of large rRNA and tRNA transcription (21).

A highly efficient RNA polymerase I *in vitro* transcription system has previously been developed for *Neurospora crassa* (22). This was used to locate the startpoint of transcription to a site 650 bp upstream of the 17S rRNA gene and to show that the region from −96 to +95 was sufficient for transcription (22). This region contains a sequence (−65 to −48), termed here the 'Ribo box', which is homologous to a sequence required for transcription of *Neurospora* 5S rRNA genes *in vitro* (1), and to three sequences in the 5′ flanking region of a *Neurospora* ribosomal protein gene (B.T., manuscript in preparation). In this study, the sequences required for RNA polymerase I transcription in *Neurospora crassa* are defined using a systematic set of deletion, substitution and insertion mutations assayed *in vitro*. Two complex regions, −113 to −37 and −28 to +4, are shown to be necessary for full transcription. These contain the 'Ribo box' and a conserved TATA sequence, respectively.

MATERIALS and METHODS

Construction of Mutant Plasmids. The 5′ and 3′ deletion mutants described here were constructed using convenient restriction sites. In some cases, e.g mutants 549 (5′ Δ −108) and 465 (3′Δ +4), the deletion endpoint was modified by choosing replacement sequences that matched the deletion for a distance of several nucleotides. Substitution mutations were created using restriction fragment replacement (1): restriction sites were used to delete short internal sequences and to replace them with unrelated sequences, usually from the polylinker of pUC18 or pUC13 (23). No substitutions except 550 (3 bp Δ) and 527 (1 bp Δ) altered the spacing of flanking sequences. Substitution mutations between −70 and −15 were introduced into parent constructs (484 or 440) which contained a 5′ deletion to the DraI site at −96, in order to facilitate construction of the mutations. Although this deletion reduced transcription an additional 2.5 fold in normal extracts, no interactions were observed between internal substitution mutations and the −96 deletion in the cases that were directly compared. Three substitutions (476, 458 and 475) were introduced by oligonucleotide mutagenesis (see below). Restriction sites created by the oligonucleotide mutations were used to create additional larger substitutions. All 5′ deletions and substitution mutations carried 310 bp or 560 bp of 3′ flanking sequences. All 3′ deletions, and several substitution mutations, carried 1300 bp of 5′ flanking sequences. To standardize the assay of 3′ deletions, these templates carried downstream sequences from +95 to +310. Due to complete cleavage of primary transcripts at +120 *in vitro* (22), these 3′ deletion templates generate the same downstream run-off transcripts as wild type templates. All mutations were verified by sequencing.

The Ribo box exchange mutants described in Figure 5 were created by oligonucleotide mutagenesis. The parental wild-type plasmids were pFS415, which contains the 5S gene on a *Sal*I (−230) to *Pst*I (+160) fragment and pRR524 which carries the 40S promoter on a *Hin*dIII (−1050) to *Eco*RI (+560) fragment. Both templates give maximal transcription *in vitro*. Both mutations were verified by sequencing.

DNA Manipulation. Routine manipulations of DNA, such as restriction enzyme digestions, DNA ligation and *E.coli* transformation, were carried out essentially as described by Maniatis *et al* (24). For small scale and large scale plasmid DNA extractions the method of (25) was used. Template plasmid DNA was purified further by CsCl centrifugation until no further RNA remained. Extraction of *Neurospora* DNA, and DNA blot hybridizations (Southerns) were according to Geever (26). Mutant plasmids were sequenced by dideoxy sequencing of alkali-denatured double-stranded plasmid DNA.

Oligonucleotide mutagenesis was carried out according to Nisbet and Beilharz (27). The plasmid vector in each case was pTZ18R (Pharmacia) which carries an f1 phage origin of replication. Single-stranded plasmid DNA for the mutagenesis was produced by super-infecting the host cells with M13K07 (28). Enrichment for mutant plasmids was achieved by incorporating uridine residues into the template DNA (29).

In vitro Transcription Assays. Transcription extracts were prepared as described from *Neurospora crassa* strain 105C (22, 30). 40S rRNA transcription assays (25 μl) contained 4μl of extract, 10 mM KHepes pH 7.9, 5 mM K_2 ethylene glycol-bis (β-aminoethylether) N,N,N′,N′ tetraacetate (EGTA), 10 mM MgOAc, 2.5% glycerol, 2.5 mM dithiothreitol, 400 μM each ATP, CTP and GTP, 50 μM UTP, 1.0 μCi [α-^{32}P] UTP, 4 mM K-phosphoenolpyruvate, and 1.2×10^{-14} M (≈30 ng) linear template DNA cleaved with EcoR1 or Pst1. RNA polymerase III (5S rRNA) transcription assays had the same composition, with the addition of 103 mM potassium acetate and 570 ng (3.2×10^{-13} M) pUC13 carrier DNA. Also the template DNA was circular. The DNA concentrations of the working stocks (6μg/ml) of the template DNAs were confirmed by measuring fluorescence in the presence of the dye bisbenzamide. Reactions were started by addition of extract, incubated for 20 min at 30 °C, and terminated by phenol/chloroform extraction of the RNA. Purified RNAs were then fractionated on a 6% or 8% polyacrylamide, 7M urea gel. The gel was dried and the transcripts detected by autoradiography.

To quantitate transcription, the radioactivities of particular transcripts were determined either by cutting the transcripts from the dried gel and counting them in a scintillation counter, or the whole dried gel was scanned for radioactivity using an AMBIS Radioanalytic Imaging System (San Diego, CA). Controls for non-specific transcription were determined by counting a segment of the lane contiguous to each specific transcript. All transcription levels reported were averaged from the results of at least three independent experiments except where indicated. Standard errors were usually less than 10% of the mean, though in some cases, especially mutants with low transcription levels, the standard error was up to 20% of the mean.

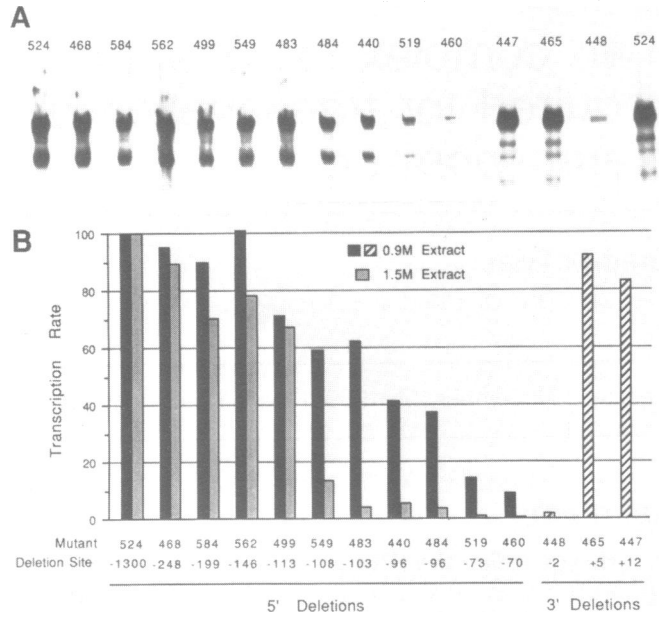

Figure 1. Deletion mapping of the 40S rRNA promoter. A. Transcription of the deletion mutants with 0.9M $(NH_4)_2SO_4$ extracts. The lower band (400 nt) results from some additional processing of the larger 440 nt transcript. Both bands were included in the quantitation. A 120 nt RNA, which is quantitatively cleaved from the 5′ end of all *in vitro* transcripts (22), does not appear on the gels and was not included in the quantitation. B. Transcription levels of the mutants. The 5′ deletion mutants were assayed with 0.9M $(NH_4)_2SO_4$ (filled bars) and 1.5M $(NH_4)_2SO_4$ extracts (open bars). 3′ deletions were assayed with 0.9M $(NH_4)_2SO_4$ extracts only (striped bars). Each data point is an average of at least three experiments. Standard errors were less than 15% of the mean in all cases (most were less than 10%). 5′ mutants 440 and 484 (−96) contain different plasmid sequences at the deletion site. 5′ deletion mutants 524, 468, 584, 499, 549, 483 and 519 carried plasmid sequences similar to 484 at the deletion site, while 562 and 460 carried sequences similar to 440. The sequences of the deletion sites are shown in figure 2.

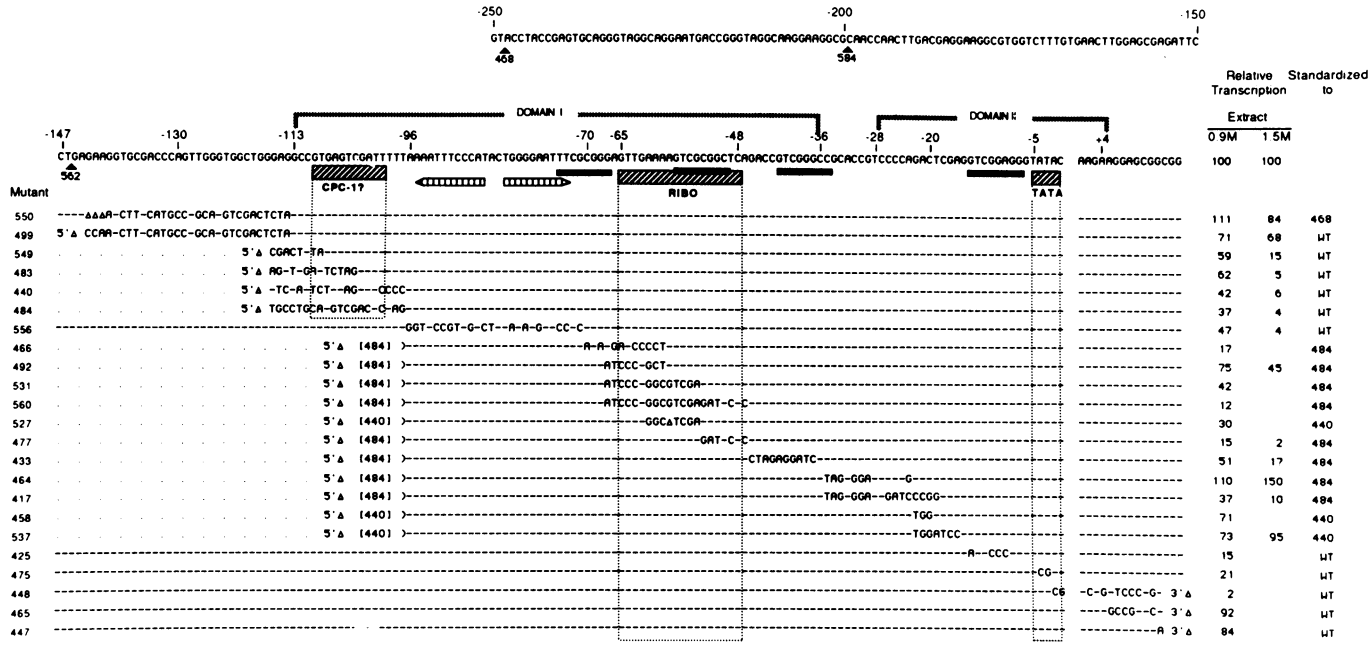

Figure 2. Substitution and deletion mutagenesis of the promoter. The wildtype sequence from −250 to +15 is shown. In the sequences of the mutants, dashes (−) indicate homology with wildtype, Δ indicates single base pair deletions, 5′Δ and 3′Δ indicate 5′ and 3′ deletions respectively, [484] or [440] indicate that the plasmid sequences at the 5′ deletion sites match those of 5′ deletion mutants 484 or 440 respectively. Symbols along the wildtype sequence: numbered arrowheads (e.g. 468) indicate the deletion sites of various 5′ deletion mutants described in figure 1; hatched boxes marked CPC-1?, RIBO and TATA indicate potential transcriptional elements described in the text; solid boxes indicate GC elements (see text) and the double arrowhead box indicates an inverted repeat sequence; domain I and domain II indicate the two major regions required for transcription. Assay data: all mutants were assayed using 0.9M $(NH_4)_2SO_4$ transcription extracts, only those for which data is shown were assayed using 1.5M $(NH_4)_2SO_4$ extracts. All data from 0.9M extracts are the means of at least three experiments; the standard errors were less than 10% of the mean for all mutants, except 466, 464, 417 and 425 where it was less than 20% of the mean. The data from the 1.5M $(NH_4)_2SO_4$ extracts represents means from at least two experiments; in each case the standard error was less than 20% of the mean. In the cases where a substitution mutant also contained a 5′ deletion, the assay data were standardized to the parent 5′ deletion as indicated in the final column.

RESULTS

Delimiting the Sequences Required for Transcription. 5′ and 3′ deletions, constructed using convenient restriction sites, were used to define the boundaries of the region required for transcription. Figure 1 shows the transcription levels of 5′ and 3′ deletion mutants. Two types of transcription extracts were used to assay the mutants. Standard extracts were prepared by treating the initial whole-cell homogenate with 0.9M $(NH_4)_2SO_4$ (22). In addition, extracts prepared by treating the initial homogenate with 1.5M $(NH_4)_2SO_4$ were used. Extracts prepared using the higher salt concentration displayed levels of transcriptional activity comparable to the 0.9M $(NH_4)_2SO_4$ extracts when wild-type templates were used. However, many mutants had much more severe effects on transcription when assayed with high salt extracts. No mutants had significantly milder effects on transcription using high salt extracts. Presumably the increased levels of non-specific DNA binding proteins in the 1.5M extracts interfered more strongly with the binding of transcription factors to mutant templates than to wild type templates (31). The 1.5M $(NH_4)_2SO_4$ extraction may also have increased the yield of some transcription factors. In addition, the 1.5M extracts lacked RNA polymerase II and RNA polymerase III transcription activity (30).

When assayed with transcription extracts prepared by the standard 0.9M $(NH_4)_2SO_4$ extraction, 5′ deletions to −146 retained at least 90% of wild type (1300 bp 5′ flanking sequences) activity. A further deletion to −113 reduced transcription to 70% activity, suggesting that sequences between −146 and −113 have a small but significant effect on transcription. Further deletions to −103 retained at least 63% activity, but from −103, transcription dropped progressively to approximately 40% at −96 and to 10% at −70. These results suggest an additional sequence between −103 and −70 is important for transcription. When the 5′ deletions were assayed with a 1.5 M $(NH_4)_2SO_4$ extract, slightly different sequences appeared important. A deletion to −199 reduced activity to 70%, and but further deletions to −146 or −113 had little additional effect, suggesting a primary influence from the region −248 to −199. Further deletions to −108 or −103 substantially reduced transcription, to 15% or 5% respectively (Figures 1 and 2). These results indicate that in the high salt extracts, important sequences for transcription occur immediately downstream of −113. Interestingly, a potential binding sequence for the amino acid regulatory protein *cpc*-1 (D.J. Ebbole, J.L. Paluh, M. Plamann, M.S.Sachs and C. Yanofsky, submitted to Genes Devel., 1990) lies from −110 to −100. At the 3′ end, deletions to +5 allow greater than 90% transcription but further deletions to −2 almost completely abolish transcription, placing the 3′ boundary of the promoter region between −2 and +4. Similar results were obtained with the 3′ deletions using either type of extract.

Substitutions Define Two Sequence Domains between −113 and +4 Required for Transcription. To identify sequences between −113 and +4 required for transcription, substitution mutations were constructed spanning the region. The mutations were constructed primarily by restriction fragment replacement; some were introduced or facilitated by oligonucleotide mutagenesis.

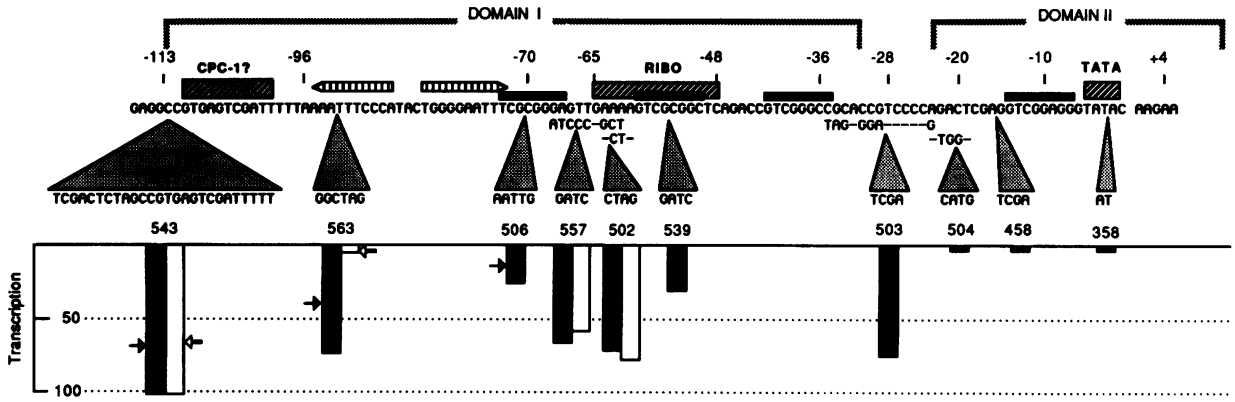

Figure 3. Insertional mutagenesis. The wildtype sequence from −116 to +5 is shown. Symbols along the sequence are the same as for figure 2. Filled bars indicate assays with 0.9M $(NH_4)_2SO_4$ extracts and open bars with 1.5M extracts. Arrows adjacent to the bars indicate the transcription efficiency of 5′ deletions with endpoints at the insertion site (black, 0.9M extracts; outline, 1.5M extracts). Several insertion mutants were derived from substitution mutants. In those cases, the sequence of the parent substitution is shown and the assay data shown have been standardized against that of the parental mutant (shown in figure 2). The parent of insertion 502 was transcribed at 93% (0.9M) or 65% (1.5M) of wild type (not shown in Fig. 2). The data from 0.9M extracts are the means of at least three experiments. The data from the 1.5M $(NH_4)_2SO_4$ extracts represents means from at least two experiments. In each case the standard error was less than 20% of the mean.

All but two of the mutations did not alter the spacing of surrounding sequences. Many of the mutations were combined with a 5′ deletion to −96 to aid their construction. Therefore, in some cases, the observed effects of the mutations may include interactions with the 5′ deletion. The sequences and transcription activities of the substitution mutations are shown in Figure 2.

The results of the 5′ deletions suggested that a sequence between −146 and −113 contributed 30% to transcription. However, a substitution mutation (550) from −144 to −114 actually stimulated transcription slightly. Therefore, there may be interactions between −144/-114 and sequences upstream of −146. This would be consistent with the results with the 1.5M $(NH_4)_2SO_4$ extract which suggested sequences upstream of −199 were important for transcription. All other substitutions between −96 and +1 measurably reduced transcription, except for a single substitution (464) from −36 to −28. This non-essential segment divides the region from −113 to +4 into two domains, designated domain I (−113 to −37) and domain II (−27 to +4).

Domain I contains a potential *cpc*-1 binding site (see above) (−110 to −100), a near perfect 23 bp inverted repeat (−95 to −73), the Ribo box (−65 to −48) (see Introduction) and three copies of an 8 bp GC rich sequence (the GC element, consensus GTCG$^{C}_{G}$G$^{GG}_{CC}$). Domain II contains one GC element and a TATA sequence at −5. Deletions of the *cpc*-1 site (484 and 440) reduced transcription to about 40% in 0.9M $(NH_4)_2SO_4$ extracts and to 5% in 1.5 M $(NH_4)_2SO_4$ extracts (see above). A substitution (556) eliminating the inverted repeat reduced transcription to 47% in 0.9 M extracts and to 4% in 1.5M extracts (figure 2). Deletions which remove this repeat (519 and 460, figure 1) reduced transcription to around 30% in 0.9M extracts, compared to penultimate deletions (440 and 484) which do not alter the repeat (figure 1). The Ribo box (−65/−48) overlaps one GC element (−57/−50) and abuts a second (−74/−67) (fig 2). A substitution which eliminates the 5′ end of the Ribo box reduced transcription only slightly, to 75% of wildtype, while substitutions which also eliminate all or part of a GC element (e.g. 466, 560) reduced transcription to 12−42%, suggesting that the GC element is more important to transcription than the entire Ribo box. Mutations which eliminate the other two GC elements at −42 and −14 (433 and 425) reduced transcription to 51% and 15% respectively. The TATA sequence at −5 is conserved among the RNA polymerase I promoters of many organisms (22). A two base pair substitution within the *Neurospora* TATA sequence reduced transcription to 21%, while a 3′ deletion into the TATA sequence reduced transcription to 2%.

Insertional Mutagenesis of Domain I and Domain II. To further define the individual sequence elements which make up domains I and II, and to begin to examine the interactions between the transcription factors which presumably bind to them, a series of insertion mutations were constructed. Most of the insertions were 4 bp to 6 bp in size; an insertion within the TATA sequence was 2 bp, and a 26 bp duplication was placed at −113. Several insertion mutations were derived from substitution mutations. The effect of these mutations on transcription was quantitated relative to the parent substitution. The insertion mutations and their effects on transcription are shown in figure 3.

Insertion mutations at −113 and −96 had effects on transcription comparable to 5′ deletions to the same site (indicated by arrows in figure 3), though in each case the insertion mutations were less severe. For example, mutant 563 (insertion −96) showed 70% transcription in 0.9 M $(NH_4)_2SO_4$ extracts (or 5% in 1.5M extracts) while mutants 440 and 484 (5′ Δ −96) showed 40% transcription (or 4−6% in1.5M extracts). These results suggest that the sequences upstream of these insertion sites contribute little to transcription when spaced incorrectly from downstream sequences. The poor transcription of mutant 563 (insertion −96) in 1.5M $(NH_4)_2SO_4$ extracts supports the importance of the sequences from −113 to −96 as determined from the 5′ deletion analysis.

In contrast to the insertions at −96 and −70, 4 bp insertions at −64 and −60 (mutants 557 and 502) in the 5′ end of the Ribo box reduced transcription to only 60−70%, in both 0.9M and 1.5M $(NH_4)_2SO_4$ extracts. These results suggest that sequences located either side of −64/−60 constitute separate binding sites for transcription factors. Since the two insertions were placed at sites of substitution mutations, no conclusion can be drawn as to whether the insertions directly disrupt a binding site for a transcriptional element. A 4 bp insertion (539) directly into

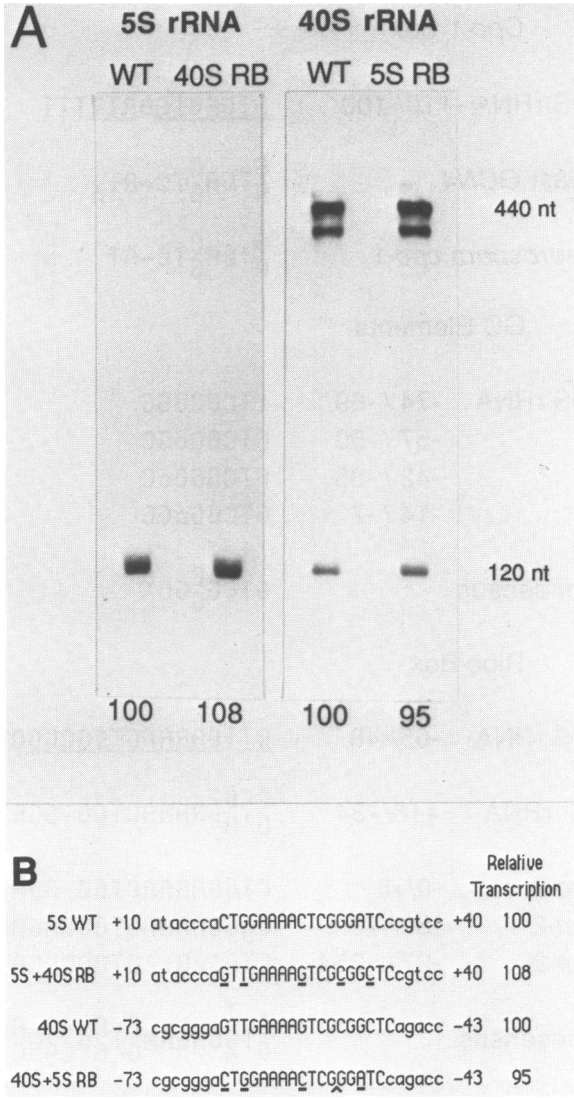

Figure 4. Exchange of the 5S and 40S Ribo Boxes does not Affect Transcription. A. Transcripts from wild type 5S and 40S templates are compared to those from mutants in which the 40S ribo box has been placed in the 5S gene and vice versa. Transcription efficiencies are shown at the bottom of the figure and derive from three replicates. For an explanation of the 40S rRNA transcripts, see Figure 1. B. Sequences the wild type and Ribo box exchange mutants of the 5S and 40S promoters. The Ribo box sequences are capitalized. The differences between the sequences are underlined. The 40S ribo box contains an additional base compared to the 5S ribo box.

the GC element in the 3′ end of the ribo box reduced transcription to 30%, a similar effect to substitution mutations (527, 477 in figure 2) which disrupt this sequence.

A 4 bp substitution at −29 between domains I and II reduced transcription to only 70%, indicating that the spacing requirement between these two domains is not strict. However all three insertions within domain II, including a 2 bp insertion into the TATA sequence severely reduced transcription, to less than 5% of the respective parent templates.

Ribo Box Sequences from 5S and 40S rRNA Genes are Interchangeable

Based on the above results, the sequence termed the Ribo box (−65 to −48) either constitutes or overlaps a sequence required for 40S rRNA transcription. A sequence in the 5S rRNA gene (+18 to +34) is similar to the 40S Ribo box and is essential for 5S transcription *in vitro* (1). Therefore, it appeared possible that a common transcription factor, binding to the Ribo box, could be involved in each case. However, since there are 5 differences between the 5S and 40S Ribo box sequences, including an additional base pair within the 40S sequence (Fig. 4B), it was also possible that two distinct factors were involved. To test the latter possibility, oligonucleotide mutagenesis was used to alter the 5S Ribo box sequence to match the 40S sequence precisely, and the 40S Ribo box was altered to match the 5S sequence precisely. If replacing the 5S Ribo box with the 40S Ribo box reduced 5S rRNA transcription, and vice versa, then one could conclude that different transcription factors bound the Ribo boxes of the 5S and 40S rRNA genes. As shown in Fig. 4A, the relative transcription of the 5S gene containing the 40S Ribo box was 108%—virtually identical to wild-type. Similarly, transcription from the 40S rRNA promoter containing the 5S Ribo box was 95% of the wild-type transcription rate. These results are fully consistent with the involvement a common Ribo box binding transcription factor. However they also do not rule out the involvement of two separate transcription factors.

DISCUSSION

A systematic set of deletion, substitution and insertion mutations has been used to define the DNA sequences required for transcription *in vitro* by RNA polymerase I in *Neurospora crassa*. Two long regions, one from −113 to −36 and one from −28 to +4, called domains I and II, were necessary for full transcription; complete deletion of either domain abolished transcription. Sequences from −250 to −113 stimulate transcription an additional 30−40%. All substitution mutations within domains I and II significantly reduced transcription, depending on the location and size of the mutation, suggesting that most of domain I (77 bp) and domain II (32 bp) contain important binding sites for transcription factors. This arrangement of transcriptional sequences is very similar to that described for *Xenopus laevis* (13) in which all sequences between −140 and +10 except for −82 to −77 were essential for full transcription. It is also similar to the arrangement in yeast, where there are two domains (−70 to −51 and −28 to +8) essential for transcription plus an upstream domain (−46 to −76) which provides further stimulation (14, 32). In humans, a core element from −45 to +7 and an upstream stimulatory domain from −156 to −107 have been described (7). Both bind similar or identical transcription factors (7). Core and upstream elements similar to that of humans have been described for the mouse (8, 9, 10), rat (11), and *Xenopus* (12), while core elements alone have been described for *Acanthamoeba* (33) and *Drosophila* (34). The core elements of these organisms corresponds closely to domain II of the *Neurospora crassa* 40S rRNA promoter (−28 to +4). Indeed, the sequence of domain II from −15 to −2 shows about 70% homology to the same region of the core elements of the fungi *Saccharomyces carlsbergensis* and *Dictyostelium discoideum*, and also of human, mouse, rat and *Xenopus* (22). From −5 to +10, the *Neurospora* sequence shows homology to *Saccharomyces carlsbergensis*, *Schizosaccharomyces pombe*, *Xenopus borealis*, *Drosophila melanogaster*, *Tetrahymena pyriformis*, *Dictyostelium discoideum* and *Acanthamoeba castellani* (22, 33, 35). The region −15 to +20, encompassing these two conserved regions, is directly contacted by RNA

polymerase I and appears to be important for the actual initiation process in *Acanthamoeba* (36). Therefore the sequence conservation may relate to a conserved function of RNA polymerase I.

Although the data do not enable firm conclusions to be drawn as to the role of individual sequences within domains I and II, a variety of sequence comparisons and some of the insertional mutagenesis results suggest some speculations regarding possible transcriptional elements, as diagrammed in figures 2 and 3.

One such element is the TATA sequence at −5. This sequence is partially or completely conserved at −5 or −3 in the rRNA promoters of most eukaryotes (excepting *Xenopus*, and possibly mouse) (22). A two base pair substitution (TATA → TcgA) in the TATA sequence reduced transcription of the *Neurospora* gene 5 fold, while a single base change in the corresponding mouse sequence (−3 GGTA → GGgA) reduced transcription 10 fold. In *Acanthamoeba*, the TATA sequence is contacted non-specifically by RNA polymerase I and is essential for efficient initiation (36). The effects of these mutations, and the conservation of the TATA sequence close to the initiation site strongly suggest that this sequence is involved in the initiation of RNA synthesis by RNA polymerase I in eukaryotes. A TATA sequence is also required in many cases for initiation by RNA polymerase II (where it is the binding site for a separate transcription factor, TFIID), for efficient and specific initiation by RNA polymerase III on the 5S rRNA genes of *Neurospora crassa* (30, 37) and possibly insects (38−40), and for transcription initiation by *E.coli* RNA polymerase (containing σ^{70}) (41). Therefore a TATA sequence appears to be nearly universal in its role in transcription initiation, and may reflect a highly conserved mechanism of action of eukaryotic and prokaryotic RNA polymerases.

At the 5′ end of domain I, from −110 to −100, there is a sequence with homology to the binding site for the amino acid biosynthesis regulatory proteins of *Neurospora* , (*cpc*-1) (42; D.J. Ebbole, J.L. Paluh, M. Plamann, M.S.Sachs and C. Yanofsky, submitted to Genes Devel., 1990) and of yeast (*GCN4*) (43). The homology is shown in figure 4A. Deletion of this sequence, or insertions downstream of it reduce transcription 30−60% in 0.9M $(NH_4)_2SO_4$ extracts and 95% in 1.5M $(NH_4)_2SO_4$ extracts. However, these sequences are still required for full transcription in 1.5M $(NH_4)_2SO_4$ extracts prepared from a *cpc*-1 disruption mutant (Y. Shi and B. Tyler, unpublished). We are currently examining the regulation of rRNA synthesis in *cpc*-1 mutants *in vivo*.

The 3′ half of domain I contains three copies of a GC rich sequence designated the GC element, and there is an additional GC element at −14 in domain II (figures 2 and 5B). Substitutions and insertions which disrupt any of these GC elements reduce transcription to 15%−50%. Hence these elements are also candidate binding sites for transcription factors. In vertebrates, the binding sites for several RNA polymerase I transcription factors, including the cloned factor UBF (44) are very GC rich (7, 45). Additional GC elements with the opposite orientation occur at −140 and −203. However, since these are located outside domains I and II, their functional significance is unclear.

The sequence from −65 to −48, which we have termed the 'Ribo box', shows homology to a sequence which is essential for 5S rRNA transcription (30), and which occurs three times in the 5′ flanking region of a *Neurospora* ribosomal protein gene, *crp*-2 (B.T., manuscript in preparation). Figure 4C shows the homology among the 5S rRNA, 40S rRNA and *crp*-2 Ribo boxes.

A. *Cpc*-1 Box

40S rRNA -110/-100		GTGAGTCGATTTTT
Yeast *GCN4*		G/A TGA C/G TC-AT$_n$
Neurospora cpc-1		G/A TGA C/G TC-AT

B. GC Elements

40S rRNA	-74 / -69	tTCGCGGG
	-57 / -50	GTCGCGGC
	-42 / -35	GTCGGGcC
	-14 / -7	GTCGGaGG
Consensus		GTCG C/G GG G/C

C. Ribo Box

40S rRNA	-65/-48	GTTGAAAAGTCGCGGCTC
5S rRNA	+18/+34	C/G T G/A GAAAACTCG-GGATC
Crp-2	-9/+8	CTGGAAAACTGG-AGATC
Crp-2	-89/-72	GgGGAAAAGTGGCAGAaa
Crp-2	-271/-254	CTGGcAtcGTCGCAGATC
Consensus		C/G T/A TGGAAAA C/G T C/G G-G C/A A/C TC

Figure 5. Potential Transcriptional Elements in the 40S Promoter. A. Homology with the binding sites for the amino acid regulatory proteins of *N.crassa* (*cpc*-1; J. Paluh and C. Yanofsky, submitted to Genes Devel., 1990) and of yeast (*GCN4*; 43). Underlining indicates matching bases. The dash (−) in the binding site sequences indicates the position of an extra base pair in the *N.crassa* 40S sequence. B. Homology of the four copies of the GC element within domains I and II (see text). Lower case letters indicate mismatches with the consensus. C. Homology with Ribo box sequences from *N.crassa* 5S rRNA genes and a ribosomal protein gene (*crp*-2). underlining indicates matches to the 5S and crp-2 Ribo boxes. The dash (−) in two of the sequences indicates the position of an extra base pair in the other sequences. The consensus sequence is derived from the 5S rRNA, 40S rRNA and *crp*-2 −9 Ribo boxes (unpublished). Lower case letters in the −89 and −271 *crp*-2 Ribo boxes indicates mismatch with the consensus.

In the 40S gene, the 3′ half of the Ribo box is coincident with a GC element and mutations there reduce transcription up to 15 fold. Substitutions and insertions within the 5′ end of the Ribo box have smaller effects. Therefore, either the Ribo box itself is a binding site for a transcription factor required for RNA polymerase I transcription, or it overlaps the site for a transcription factor which has specificity for another sequence such as the GC element. Either interpretation raises the possibility that the Ribo box is involved in coordinating ribosomal RNA and protein gene expression in *Neurospora crassa*. In the first interpretation, the Ribo box could be the binding site for a

transcription factor common to 5S and 40S rRNA genes, as well as possibly ribosomal protein genes. This would provide a very direct mechanism for coordinating the expression of these genes. The fact that the 5S and 40S Ribo boxes are functionally interchangeable, despite their sequence differences (Fig. 4), is consistent with this hypothesis. If the 40S Ribo box simply overlaps a transcription factor binding site (e.g. the GC element), then coordination of ribosomal RNA and protein gene expression potentially could be achieved by a negative regulatory protein which acts to interfere with the binding of transcription factors specific to each gene system.

ACKNOWLEDGEMENTS

I thank D.J. Ebbole, J.L. Paluh, M. Plamann, M.S. Sachs and C. Yanofsky for communicating data regarding *cpc*-1 prior to publication, Karin Harrison for technical assistance, Yuguang Shi and Layne Huiet for stimulating discussions, and Michael Holland, Yuguang Shi, Eva Steinberger, Tom Cujec and Julie Billington for criticizing the manuscript. This work was supported in part by National Institutes of Health grant GM42178 (to B.M.T.) and by the Research School of Biological Sciences, Australian National University.

REFERENCES

1. Tyler, B.M. (1987) *J. Mol. Biol.* **196,** 801–811.
2. Stellwag, E.J. & Metzenberg, R.L. (1984) *Mol. Gen. Genet.* **194,** 105–110.
3. Alberghina, F.A.M., Sturani, E. & Gohlke, J.R. (1975) *J. Biol. Chem.* **250,** 4381–4388.
4. Alberghina, L., Sturani, E., Costantini, M.G., Martegani, E. & Zippel, R. (1978) In Burnett, J.H. & Trinci, A.P.J. (e.d.), Fungal Walls and Cell Growth. Cambridge University Press, London, pp. 295–318.
5. Warner, J.R. (1989) *Microbiol. Rev.* In press.
6. Sollner-Webb, B. & Tower, J. (1986) *Ann. Rev. Biochem.* **55,** 801–830.
7. Learned, R.M., Learned, T.K., Haltiner, M.M. & Tjian, R.T. (1986) *Cell* **45,** 847–857.
8. Grummt, I. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6908–6911.
9. Yamamoto, O., Takakusa, N., Mishima, Y., Kominani, R. & Muramatsu, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 299–303.
10. Miller, K.G., Tower, J. & Sollner-Webb, B. (1985) *Mol.Cell. Biol.* **5,** 554–562.
11. Cassidy, B.G., Haglund, R. & Rothblum, L.I. (1987) *Bioch. Bioph. Acta* **909,** 133–144.
12. Windle, J.J. & Sollner-Webb, B. (1986) *Mol.Cell. Biol.* **6,** 4585–4593.
13. Reeder, R.H., Pennock, D., McStay, B., Roan, J., Tolentino, E. & Walker, P. (1987) *Nucl. Acids Res.* **15,** 7429–7441.
14. Musters, W., Knol, J., Maas, P., Dekker, A.F., van Heerikhuizen, H. & Planta, R.J. (1990) *Nucl. Acids Res.* in press.
15. Buttgereit, D., Plugfelder, G. & Grummt, I. (1985) *Nucl. Acids Res.* **13,** 8165–8180.
16. Gokal, P.K., Cavanaugh, A.H. & Thompson, E.A. (1986) *J. Biol. Chem.* **261,** 2536–2541.
17. Tower, J. & Sollner-Webb, B. (1987) *Cell* **50,** 873–883.
18. Bateman, E. & Paule, M.R. (1986) *Cell* **47,** 445–450.
19. Sturani, E., Costantini, M.G., Zippel, R. & Alberghina, F.A.M. (1976) *Exp. Cell Res.* **99,** 245–252.
20. Sturani, E., Magnani, F. & Alberghina, F.A.M. (1973) *Bioch. Bioph. A.* **319,** 153–164.
21. Mirkes, P.E. (1977) *Exp. Mycol.* **1,** 271–279.
22. Tyler, B.M. & Giles, N.H. (1985) *Nucl. Acids Res.* **13,** 4311–4332.
23. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26,** 101–106.
24. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. Birnboim, H.C. & Doly, J. (1979) *Nucl. Acids Res* **7,** 1513–1523.
26. Geever, R.F., Case, M.E., Tyler, B.M., Buxton, F. & Giles, N.H. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7298–7302.
27. Nisbet, I.T. & Beilharz, M.W. (1985) *Gene Anal. Technol.* **2,** 23–29.
28. Vieira, J. & Messing, J. (1987) *Meths. Enzymology* **153,** 3–11.
29. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
30. Tyler, B.M. & Giles, N.H. (1984) *Nucl. Acids Res.* **12,** 5737–5757.
31. Wilson, E.T., Larson, D., Young, L.S. & Sprague, K.U. (1985) *J. Mol. Biol.* **183,** 153–163.
32. Kempers-Veenstra, A.E., Musters, W., Dekker, A.F., Klootwijk, J. & Planta, R.J. (1985) *Curr. Genet.* **10,** 253–260.
33. Kownin, P., Iida, C.T., Brown-Shimer, S. & Paule, M.R. (1985) *Nucl. Acids Res.* **13,** 6237–6247.
34. Kohorn, B.D. & Rae, P.M.M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3265–3268.
35. Balzi, E., Di Pietro, A., Goffeau, A., van Heerikhuizen, H. & Klootwijk, J. (1985) *Gene* **39,** 165–172.
36. Kownin, P., Bateman, E. & Paule, M.R. (1987) *Cell* **50,** 693–699.
37. Selker, E.U., Morzycka–Wroblewska, E., Stevens, J.N. & Metzenberg, R.L. (1986) *Mol. Gen. Genet.* **205,** 189–192.
38. Indik, Z. & Tartof, K.D. (1982) *Nucl. Acids Res.* **10,** 4159–4172.
39. Rubacha, A., Sumner III, A., Richter, W. & Beckingham, K. (1984) *Nucl. Acids Res.* **12,** 8193–8207.
40. Morton, D.G. & Sprague, K.U. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5519–5522.
41. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
42. Barthelmess, I.B. (1982) *Genet. Res.* **39,** 169–185.
43. Hill, D.E., Hope, I.A., Macke, J.P. & Struhl, K. (1986) *Science* **234,** 451–457.
44. Bell, S.P., Pikaard, C.S., Reeder, R.H. & Tjian, R. (1989) *Cell* **59,** 489–497.
45. Tower, J., Culotte, V.C. & Sollner-Webb, B. (1986) *Mol.Cell. Biol.* **6,** 3451–3462.