Genes expressed during conidiation in *Neurospora crassa*: characterization of *con-8*

Anne N.Roberts⁺ and Charles Yanofsky*

Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA

Received September 19, 1988; Revised and Accepted November 28, 1988

Accession no. X07040

ABSTRACT

The filamentous fungus *Neurospora crassa*, by a series of defined changes, differentiates from a mycelium composed of branching hyphae to form dormant spores, called conidia. Several genes of unknown function (*con* genes) that are preferentially expressed during this period have been cloned (1). Transcription of these genes has been examined in conidiation-defective mutants, and the results obtained revealed that *con-6*, *con-8*, *con-10*, *con-11* and *con-13* are most likely to play a unique role during conidiation. *con-8* is expressed early during conidial differentiation. Genomic and cDNA sequence analyses with *con-8* clones identified one open reading frame, interrupted by two introns, which encodes a weakly acidic 18.4 kDa polypeptide containing 176 amino acid residues. *con-8* is unusual in that it is transcribed as two mRNA species, 1.0 and 1.25 kb in length. S1 nuclease mapping and primer extension analyses identify one major initiation site, one major polyadenylation site, and demonstrated the existence of heterogeneity at the messenger's 5' and 3' ends.

INTRODUCTION

During its life cycle the filamentous fungus *Neurospora crassa* undergoes an asexual developmental process known as conidiation. Vegetatively growing mycelium, in response to external signals, primarily nutrient deprivation (2), produce aerial hyphae which undergo multiple changes to form dormant spores, macroconidia (termed conidia throughout this paper). Conidia are airborne agents of dispersal.

One of our aims in studying this developmental process is to determine the regulatory mechanisms *Neurospora* employs to achieve temporal gene expression. Our initial approach involved the isolation of a series of plasmid clones (the pCon series) that contain genes encoding mRNA species found predominantly or exclusively during conidiation (1). In this study we have analyzed expression of these *con* genes in a number of conidiation-defective mutants and from these results selected the most promising genes for further study of differential expression during conidiation. We present the molecular characterization of one of these genes, *con-8*.

The series of plasmids containing genomic DNA that hybridizes with mRNA species expressed during conidiation have been named pCon-1 to pCon-12. The gene (or genes) encoded on these plasmids have been named con-1, con-2 ... con-n. The gene corresponding to the most abundant mRNA transcript detected by each plasmid has been given the same number as the plasmid. Thus the plasmid pCon-1 encodes the con-1 gene. Additional genes encoded by some of the pCon plasmids have been given a con- designation as necessary. Of relevance to this paper is the fact that pCon-10a encodes two genes, con-10 and con-13.

MATERIALS AND METHODS

N. crassa strains

The wild type strain used was 74-OR23-1A. Mutant strains used (3) were aconidiate-2 (acon-2), aconidiate-3 (acon-3), conidial separation-1 (csp-1), conidial separation-2 (csp-2), easily wettable (eas) and fluffy (fl). Growth conditions and media are described in Berlin and Yanofsky (4). Mycelial cultures were grown in liquid medium for approximately 20 h with aeration until a concentration of 0.7-0.9 mg dry weight/ml had been reached. Conidiation was induced by filtering 200 ml aliquots of this material onto Whatman No. 1 paper discs. These discs were placed on a layer of glass beads in Petri dishes containing medium and incubated in a sterile hood under fluorescent light (4). Conidial samples were harvested 16-24 h later. For the experiment described in Figure 2 growth of the mycelial culture was continued in liquid medium for a total of 44 h.

General molecular biological techniques

Methods used for the isolation of nucleic acids, preparation of radio-labelled probes, and Southern analyses, are described in Roberts *et al.* (3). The cosmid library of Vollmer and Yanofsky (5) was screened as described previously. The method used for the isolation of cDNA clones from a conidial cDNA library has been described (3).

Northern analyses

Northern blots were carried out as described by Berlin and Yanofsky (1). Equal loading of total RNA (20-25 µg) in each lane was verified by ethidium bromide staining of the gel. Further confirmation of the integrity of the RNA preparations was obtained by stripping the blots and reprobing with the cDNA clone encoding the gene crp-1 which codes for a cytoplasmic ribosomal protein (6). The intensity of bands obtained on autoradiograms was quantitated by densitometry.

DNA sequencing

The 2.3 kb SalI fragment that contains the con-8 gene was sequenced by the M13 dideoxynucleotide method of Sanger *et al.* (7). A library of random clones derived from this fragment was obtained by sonication followed by subcloning into the SmaI site of M13mp18 (8) as described by Bankier and Barrell (9). 110 clones were randomly picked and sequenced to cover both strands at an average of 9.0 readings per bp. Additional sequence at the 5' end of con-8 was obtained by subcloning appropriate restriction fragments. Sequence assembly used the computer programs of Staden (10, 11). Analysis of the sequence was performed using the programs of the University of Wisconsin Genetics Computer Group (12).

S1 nuclease mapping and primer extension

Internally labelled probes of single stranded DNA used for S1 mapping and primer extension were prepared on appropriate M13 templates as described by Hudson and Davidson (13). S1 mapping and primer extension were performed by published procedures (13), with the following modifications. For S1 mapping, 50,000 Cerenkov cpm/probe was hybridized with 20 μ g total RNA overnight. In early experiments annealing was performed at 50° C, but the temperature was later reduced to 45° C after we found that some probes did not anneal at the higher temperature. The only double stranded probe used was denatured at 80° C for 15 min before use. Primer extension was performed with 20,000 Cerenkov cpm/probe hybridized with 20 μ g total RNA annealed at 45° C.

RESULTS AND DISCUSSION

Timing of expression of con genes

Transcription of the con genes was examined by Northern blot hybridization of RNA prepared from a selection of N. crassa mutants, to radiolabelled probes prepared from various pCon plasmids. The mutants analyzed were either aconidial or formed defective conidia, and are listed in MATERIALS AND METHODS. Our results, based on at least two separate Northern blots for each probe, are compiled in Table 1. A typical example of a Northern blot obtained with con-8 is shown in Figure 1. The same filter, reprobed with a control gene (crp-1), is displayed underneath. We have confined our analysis to the major transcript(s) detected by each plasmid. Their sizes are indicated in Table 1. The levels of mRNA from con-6 and con-8, two genes whose transcripts accumulated early during conidiation (0 - 4 h), were only slightly altered in the mutants; the levels of these mRNAs were the same in most mutants as in wild type (25-100% of wild type), apart from a reduction

Table 1
Levels of con transcripts in various mutants of N. crassa

Gene	transcript level				
	<u>con-6</u>	<u>con-8</u>	<u>con-10</u>	<u>con-13</u>	<u>con-11</u>
Size (kb)	0.66	1.0,1.25	0.66	1.35	1.75
Conidiation induction	0-2 h	2-4 h	4-8 h	4-8 h	4-8 h
WT mycelia (20 h)	+/-	-	•	-	-
WT conidia ^a	++	++	++	++	++
<u>csp-1</u>	++	++	++	+	+
<u>csp-2</u>	++	++	++	+	++
<u>eas</u>	++	++	+	++	+
<u>fl</u>	+	++	•	-	-
<u>acon-3</u>	++	++	•	-	•
34°C <u>acon-2</u>	++	++	•	-	•
25°C <u>acon-2</u>	++	++	•	-	+

^a RNA levels are defined as 100%

++ estimated at 25-100% wild type conidial levels

+ estimated at 5-25% wild type conidial levels

• estimated at less than 5-25% wild type conidial levels

+/- present in some samples

- not detected

of con-6 mRNA in fl. The levels of mRNA for genes whose transcripts accumulate later during conidiation (4 - 12 h) in wild type, con-10, con-11, and con-13, were more severely affected. In general, these mRNAs were very low or absent in RNA isolated from fl, acon-2 and acon-3 cultures and reduced in eas and in some cases csp-1 and csp-2 cultures. Transcripts corresponding to pCon-3 and pCon-5 showed a slight increase in comparisons of conidial RNA and mycelial RNA in wild type, whereas no significant increase was observed using pCon-1, pCon-2, and pCon-4a. Transcripts corresponding to pCon-1, pCon-2, pCon-3, pCon-4a and pCon-5 were found in all the mutants in both mycelial and conidial RNA and showed only minor differences from the pattern of expression found in wild type (data not shown). Transcription of the genes contained in pCon-7 and pCon-9 was not examined.

Transcription of the genes encoded within the pCon clones was also examined in cultures that had been grown in liquid medium for 44 h. These

Figure 1. Northern blot of RNA from *N. crassa* conidial mutants probed with a 2.3 kb *SalI* fragment containing *con-8*. RNA was extracted from a mycelial (m) and conidial (c) culture of each mutant and run in adjacent lanes. The time in parentheses (below) refers to the time after induction of conidiation at which RNA was prepared. Panel 1, wild type (24 h); panel 2, *csp-1* (16 h); panel 3, *csp-2* (16 h); panel 4, *eas* (16 h); panel 5, *fl* (16 h); panel 6, *acon-3* (16 h); panel 7, *acon-2* at 34° C (16 h and 28 h); panel 8 *acon-2* at 25° C (30 h). Underneath are shown the results obtained after reprobing with the *crp-1* gene. The transcripts corresponding to *con-8* and *crp-1* are indicated by arrows.

cultures were therefore 24 h older than our normal mycelial preparations which were usually analyzed after 20 h, when growth was believed to be most vigorous. In contrast, by 44 h, the culture may be senescent. Figure 2 shows Northern blots of RNA prepared from such cultures probed with pCon-6, pCon-8, pCon-10a and pCon-11a. Transcripts corresponding to *con-10* and *con-13* (encoded on pCon-10a), *con-6* and *con-11* were not found, but *con-8* mRNA was detected. (A constitutively expressed transcript of 1.1 kb encoded by pCon-11a (1) was also found. This gene is not thought to be involved with conidiation (1)).

These initial studies permit us to suggest the order in which the *con* genes and genetically defined conidiation-defective loci are expressed during conidiation (Figure 3). Our evidence indicates that the genes encoded by

Figure 2. Northern blots of RNA from *N. crassa* 74-OR23-1A grown under different conditions. RNA was extracted from a 20 h old mycelial culture (m), 44 h old mycelial culture (m44), and 24 h conidial culture (c). The plasmids used as probes are indicated beside each set of lanes. The relevant transcripts are indicated by their sizes in kb.

pCon-1, pCon-2, and pCon-4a may not be preferentially expressed during conidiation over the time period examined. These findings differ slightly from those of Berlin and Yanofsky (1) who observed an increase in these messages during conidiation, although in the case of pCon-1 and pCon-4a significant induction was only apparent at 40 h, beyond the time of this analysis. These observed differences may be due to our examination of total RNA, rather than poly(A)⁺ RNA as was used previously (1). The genes

Figure 3. Diagram representing the proposed order of *con* gene expression, and the stages at which various mutants are blocked in the conidiation pathway.

encoded by pCon-3 and pCon-5a probably do not play a unique role in conidiation since in wild type their mRNAs accumulated only slightly in conidial cultures compared to mycelial cultures. In contrast, the transcripts corresponding to con-10, con-11 and con-13 accumulated only in the later stages of conidiation (3), and were not found or were found at severely reduced levels in the aconidial mutants acon-2, acon-3 and fl. Furthermore, their mRNAs were not present in 44 h old mycelial cultures, implying that their respective gene products have a more specific function during conidiation (Figure 2). The levels of mRNA for con-6 and con-8, which accumulated early in conidiation in wild type, were only slightly altered in the mutants. The presence of con-6 mRNA in mycelial preparations was variable, in both mutants and wild type. This may be due to slight variations in growth conditions. The finding that con-8 transcripts accumulated during conidiation and in 44 h old mycelial preparations implies that the con-8 gene product functions at both of these times.

From these results we conclude that con-8, con-10, con-11, and con-13 are bona fide conidiation-specific genes, and are attractive candidates for studies on differential gene expression in N. crassa.

The coding region of con-8

The con-8 gene was selected for further study since elucidation of its structure and expression could contribute to our understanding of both conidiation and culture aging. We sequenced the 2.3 kb SalI fragment that hybridized with both con-8 transcripts. From the sequence it was apparent that this SalI fragment contained only a short segment of 5' flanking DNA. To detect additional upstream DNA, we probed an ordered cosmid library (5) with the 2.3 kb SalI fragment and identified a single positive clone, 24:8:6. An additional 400 bp of 5' flanking DNA was subcloned from this cosmid and sequenced. A restriction map of the con-8 gene and surrounding DNA is shown in Figure 4 together with the boundaries of useful clones related to con-8.

Isolation of cDNA clones of con-8

Approximately 50,000 plaques of a conidial cDNA library (3) were screened with a radiolabelled sample of the 2.3 kb SalI fragment. 108 positive plaques were identified of which 10 were purified. The two longest clones, designated cDNA4.8 and cDNA11.8, were selected for further study. From a combination of sequence analysis and Northern blotting it was apparent that cDNA4.8 and cDNA11.8 corresponded to con-8. We were surprised to observe that cDNA4.8 and cDNA11.8 each hybridized with both of the transcripts (1.0

Figure 4. Restriction map of *con-8* gene and flanking regions. The top line represents *N. crassa* genomic DNA. The sequenced portion is indicated by a hatched bar. Arrows indicate direction of transcription. Underneath, the positions of relevant subclones are displayed; vector DNA is shown as a broken line. Their insert sizes, cloning sites, and names are indicated. Appropriate restriction sites are shown in pCon-8 to indicate the insert orientation. The restriction sites are as follows: A = *Apal*; B = *BamHI*; E = *EcoRI*; H = *HpaI*; K = *KpnI*; N = *NcoI*; P = *PstI*; Pv = *PvuII*; Sc = *SacI*; Sl = *SalI*; S = *SphI*; Su = *Sau3A*. The *EcoRI* and *BamHI* sites on the left-hand side originate from the polylinker of the cosmid pSV50. The precise order of the *SacI* and *PstI* sites at 4 kb and *PstI* and *Apal* sites at 5 kb is not known. The parent vectors are pRK9 for pCon-8 (37), pSV50 for 24:8:G (5) and Bluescript KS M13+ for pRPCon-8 (38).

kb and 1.25 kb) identified by pCon-8. This was the first indication that pCon-8 contained one gene that was transcribed into two mRNA species.

DNA sequence of *con-8* and its predicted gene product

The DNA sequence of *con-8* and flanking regions is shown in Figure 5. This segment corresponds to the 2.6 kb *EcoRI-SalI* fragment that is shaded in Figure 4. The organization of this gene is shown diagrammatically in Figure 6. Sequence analysis of the two cDNA clones located the positions of two introns. Within the cDNA sequence, only one sizable open reading frame was found. This open reading frame is translated above the nucleotide sequence in Figure 5.

In most respects, *con-8* displays features characteristic of a *N. crassa* gene. The G+C content of the whole sequence is 49%, slightly lower than the value of 54% for *N. crassa* genomic DNA (14). However it is noteworthy that the G+C content of the three exons (63, 53, and 68%, respectively) is significantly higher than the G+C content of the introns (41 and 42%,

respectively) or 200 bp flanking the 5' (44%) and 3' (38%) ends of the coding region. The two introns of con-8 are 170 bp and 72 bp in length, respectively, typical sizes for N. crassa. The splice sequences and internal element involved in lariat formation of each intron (intron 1, GTATGT.....GCTAACA....19....TAG; intron 2, GTACGT.....GCTAACT.... 13TAG) closely match the N. crassa consensus sequences (3). These sequences are boxed in Figure 5. A striking feature of intron 1 is the stretches of repetitive sequence. In the central portion of intron 1, the motif TTGTGGG is repeated six times. In addition, near the beginning of intron 1, a highly T-rich region is found interspersed by C residues. Part of this sequence is shared by intron 2 and also by a region just distal to the stop codon for the con-8 polypeptide (bp 1642-1652). These sequences are indicated by broken underlining in Figure 5. Examination of the sequence for stem and loop structures located a number of potential candidates at the 3' end of the sequence. Two of these are indicated by horizontal arrows in Figure 5.

The codon usage pattern displayed in the first part of exon 1 of this gene follows the general rules for a N. crassa gene, namely a preference for codons terminating with pyrimidine residues rather than purines in the four and six codon families and displaying a bias for codons terminating with a G or C residue over those terminating with A and T residues throughout the entire sequence (15,16). However, the second part of exon 1 and all of exon 2 do not follow the typical N. crassa codon usage profile. Most noticeably, codons terminating with C residues are found at a lower frequency, 18%, than the normally observed 48% (compiled from codon usage data for ten N. crassa genes).

The con-8 gene product

The first methionine codon of the con-8 open reading frame occurs at position 852 (Figure 5). The bases immediately adjacent to this codon (ATAACC ATG G) conform to sequences found around the ATG initiation codon (ATCAAA ATG G) in N. crassa (17) and higher eukaryotes (GCCACC ATG G; 18). These sequences are believed to be important for ensuring correct initiation of protein synthesis (19). We have no direct evidence that this is the initiator methionine. There is a second methionine codon 148 bp further downstream, but it does not possess a good translation initiation consensus sequence.

The deduced con-8 polypeptide is weakly acidic, contains 176 amino acid residues, and a mass of 18.4 kDa. Examination of the secondary structure of this polypeptide by the algorithm of Chou and Fasman (20) indicates some

potential to form α -helices and very little tendency to form β -pleated sheets. The hydropathy plot of Kyte and Doolittle (21) reveals that the proposed con-8 polypeptide is predominantly hydrophilic; there are no hydrophobic potential membrane-spanning regions. It is noteworthy that 11% of all residues are proline.

Comparison of the deduced amino acid sequence of the con-8 product with

the polypeptide sequences contained in the NBRF Protein Identification Resource (22) by the FASTP program (23) revealed only small pockets of homology with known protein sequences. The most extensive of these was with the adenovirus 55 kDa protein encoded by the early E1b mRNA (24). The program ALIGN (25) run using a matrix bias parameter of 6 and a break penalty of 6 showed that amino acids 90-148 of the con-8 polypeptide shares 19 out of 59 (32%) identities with part of the adenovirus 55 kDa protein. This homology yielded an alignment score of approximately 6, which is considered significant (25). We also reexamined the con-10 polypeptide sequence using the same procedure and found a previously unrecognized homology with a segment of actin. Exon 2 of con-10 shares 11 out of 29 (38%) amino acid identities with Dictyostelium discoideum actin 3-sub 2 (26). This region of homology is shared by other actins including that of yeast (27).

S1 nuclease mapping

Northern blots performed with the cDNA clones as probes had shown that the con-8 gene is represented by two RNA species. S1 nuclease mapping was performed to identify the origin of each of these RNA species. Nine single-stranded M13 probes (named 1-9) were prepared spanning approximately a 1 kb region at the 5' end of the gene (Figure 6). One major mRNA initiation site, labelled A1, was found with probes 3, 4 and 5 extending from bp 631-666. This site was always represented by a cluster of four bands (Figure 7a). The average positions of these bands (from four experiments) indicate S1 cleavage at bp 631-633, 635-638, 653-656 and 659-666 (Figure 5). A primer extension experiment (data not shown) using an M13 single-stranded probe extending from nt 812 to 722 located the same cluster of four bands. Their positions were a few bases further upstream than those determined by S1 mapping analysis and are shown by vertical arrows in Figure 5. The region between the A1 mRNA

Figure 5. The nucleotide sequence of con-8 and flanking regions. The mRNA-equivalent strand is shown with the predicted amino acid sequence in the one-letter code above. The sequence is numbered from the EcoRI site. A1 and A2 designate the major and minor RNA initiation sites respectively. The exact bases are indicated by (↓). The potential major (B1) and minor (B2) polyadenylation sites are indicated (*). Introns are labelled. The splice sequence and internal element of each intron are boxed. The con-8 translation stop codon (*) and relevant restriction sites are indicated. A potential Goldberg-Hogness sequence (TATAAA) and CCAAT element also are boxed. Two potential capping sequences are underlined. A poly(A)⁺ recognition element is overlined. Other interesting features are indicated by broken underlining: the TTGTGGG repeated motif in intron 1, and a short sequence conserved between introns 1 and 2 and the 3' untranslated region. Two potential stem-loop structures in the 3' segment are indicated by horizontal arrows.

Figure 6. Organization of *con-8*. The sequenced DNA extends from the *EcoRI* site (E) to the second *SalI* site (S). Stippled boxes represent exons. The boundaries of the exons in the genomic DNA were defined by S1 mapping and primer extension. A1 and A2 indicate the major and minor RNA initiation sites respectively. B1 and B2 mark the positions of the proposed major and minor polyadenylation sites. The two cDNA clones (11.8 and 4.8) are displayed underneath. The translation initiation and termination codons are shown. The numbered bars represent uniformly labelled single-stranded probes used for S1 nuclease mapping.

initiation sites (bp 636 and 644) contains pyrimidine residues exclusively, which often characterizes transcription initiation sites of *N. crassa* (15). The 5' end of the longer cDNA clone, cDNA11.8, also extends to this region (bp 660). cDNA4.8 is a little shorter (5' end at bp 771). A weak mRNA initiation site (A2) was detected with probes 5, 6, 7 and 9 as a doublet at bp 503 and 514. A2 was at most 10% the intensity of A1 (Figure 7a). There was also an indication of minor RNA transcripts extending further 5' of A2 (Figure 7a). The smallest DNA fragment protected in this series of experiments was 150 nt. After allowing for this limitation, the probes used had the resolution to detect mRNA initiation sites extending from bp 1 to 900. No other initiation sites were found. Therefore, our analyses show that transcription of *con-8* begins predominantly at A1 and infrequently at A2. Noteworthy features of the sequence preceding A1 include a Goldberg - Hogness box (TATAAA) located 46 bp upstream of A1 (28), and a CCAAT box found 207 bp upstream of A1 (29). This CCAAT location is more upstream than usual. In addition, two sequences resembling the capping consensus (30) are found near

Figure 7. S1 nuclease mapping of *con-8* mRNA with uniformly labelled M13 probes. The lanes are: (1) probe without S1 nuclease treatment; (2) probe with 20 μ g mycelial RNA; (3) probe with 1 μ g conidial RNA; (4) probe with 10 μ g conidial RNA; (5) probe with 20 μ g conidial RNA; (5') probe with 20 μ g conidial RNA (different preparation); (6) probe with 20 μ g dormant conidia RNA; (M) end-labelled *Hpa*II digest of pBR322. The positions of the probes are shown in Figure 6. (a) Probe 5; (b) Probe 16; (c) Probe 14; (d) Probe 11. The markers in nt for (a) are 404, 309, 242, 238, 217, 201, 190, 180, 160, 147, 122. The markers for (b), (c) and (d) in nt are 622, 527, 404, 309, 242, 238, 217, 201, 190, 180, 160, 147, 122. The RNA initiation and termination sites are indicated by labelled arrows.

the 5' end of the transcript (Figure 5).

At the 3' end of *con-8*, S1 nuclease mapping identified one major polyadenylation site (B1) at bp 2019 (Probes 11, 12, 13, 14 and 15, Figure 6; Figure 7c and 7d) and a minor polyadenylation site (B2) at bp 2167 (Probes 10, 11, 12 and 13, Figure 6; Figure 7d). There was some evidence of minor transcripts extending beyond B2 (not shown). Extensive mapping in this region with seven probes (Figure 6) that had the capacity to detect any polyadenylation sites from bp 1765 to 2307 did not reveal any other major species. However, a number of probes (12, 13, 14, 15 and 16; Figure 6) detected a zone of minor mRNA species corresponding most likely to 3' ends in the region 1800 to 2019 bp, which were visible simply as a smear (Figure 7c) or as a group of more discrete bands (Figure 7b). The 3' end of cDNA11.8 (bp 1936) and cDNA4.8 (bp 1860) are included in this zone. These ends do not correspond to B1. Extensive runs of A residues are present at the end of

each cDNA clone. A potential polyadenylation recognition element (AATGAA) is located at an appropriate distance upstream from B1 (31). Close matches with this sequence are not found near B2 nor the 3' ends of the two cDNA clones.

Identifying two mRNA transcripts for *con-8*

The two mRNA transcripts corresponding to *con-8* are approximately 1.0 and 1.25 kb in length. Attempts to identify the 5' and 3' end of each message have so far been inconclusive. The length of RNA between the predominant sites of transcription initiation (A1) at bp 630 and polyadenylation (B1) is 1148 bp which could correspond to one of the two mRNA transcripts seen in Northern blots. A2 lies approximately 200 bp 5' of A1 whereas B2 lies approximately 200 bp 3' of B1. Thus RNA species may exist that extend from A1 to B2 and from A2 to B1 which together represent the larger RNA transcript. An experiment designed to detect transcription initiating from A2 using a uniformly labelled M13 probe corresponding to nt 504-569 (Figure 5), that is between A2 and A1, did not detect any species on a Northern blot. (A positive control was included.) Most probably the minor RNA species found by the more sensitive S1 mapping technique are beyond the sensitivity of Northern blotting. Therefore it is unlikely that transcription from A2 could explain the existence of a mRNA species that is approximately 200 nt larger than a message transcribed from A1. The possibility that the larger mRNA is spliced differently from the smaller mRNA and includes the intron 1 sequence was ruled out by demonstrating that a probe spanning intron 1 (bp 1335-1223) did not hybridize with either *con-8* transcript. In fact this probe hybridized strongly with a small RNA species of < 300 nt that is presumably derived from the spliced out intron (data not shown). The trivial explanation that another gene exists on the opposite strand of DNA was excluded by showing that a single-stranded probe complementary to the proposed *con-8* mRNA hybridized with both mRNA species whereas one in the opposite orientation hybridized with neither (data not shown). In addition, genomic Southern analyses (1, data not shown) have given no indication of the existence of two *con-8* genes.

Our evidence from S1 mapping studies indicates that extensive heterogeneity at the 3' end of the *con-8* mRNA exists. This is therefore the most likely source of the additional RNA transcript found on Northern blots. RNA extending from the 5' end at A1 to the 3' end at B1 can clearly account for one of the mRNA bands, probably the larger. The cDNA clones share the same exons and are 1035 bp and 848 bp in length (Figure 6). Each cDNA identifies a genuine polyadenylation site inside of B1 which falls within the zone of

Table 2
Upstream sequence homologies between con-8 and con-10

Gene	Sequence ^a	Homology	Location ^b
1. <u>con-8</u>	77 5' CTGGAAGGTC [→] CCGAAA 3' 92	14/16	-553
	x x		
<u>con-10</u>	1248 5' CTGGAAGCTG [→] CCGAAA 3' 1263		-408
2. <u>con-8</u>	158 5' GACTTCGAAATCTCAATGAATC 3' 179	9/10-	-472
	x x x x xx x		
<u>con-10</u>	1316 5' CATTGGAAATGTCAACAATTC 3' 1295	15/21	-361
3. <u>con-8</u>	183 5' GGGAGCTTA 3' 191	9/9	-447
<u>con-10</u>	1141 5' GGGAGCTTA 3' 1149		-515
4. <u>con-8</u>	187 5' GCTTACACA 3' 195	9/9	-443
<u>con-10</u>	1166 5' GCTTACACA 3' 1174		-490
5. <u>con-8</u>	224 5' CAAACACTATTG [→] CCA 3' 239	9/10-	-406
	x x xx x		
<u>con-10</u>	1533 5' GAGATGCTATCTG [→] CCA 3' 1518	11/15	-138
6. <u>con-8</u>	377 5' GAGGTACATCCATCT 3' 391	8/8-	-253
	x x xx		
<u>con-10</u>	1364 5' GGGTTCGATCCATCT 3' 1350	11/15	-306
7. <u>con-8</u>	485 5' TGGCCCTGCGCTA 3' 497	12/12	-145
	x		
<u>con-10</u>	1193 5' CGGCCCTGCGCTA 3' 1181		-475

^a The numbers for con-8 sequences are taken from Figure 5. The numbers for con-10 sequences are from Figure 3 of Roberts et al. (3). Overline and underline arrows identify dyad symmetries. Non-identical matched nucleotides are marked x.

^b Furthest distance upstream of the transcription initiation site.

3'ends detected. It is possible that some of the minor RNA species run together as a single band on the Northern gel.

Relationship between con-8 and con-10

Since we have upstream sequence data for con-8 and con-10, we were interested in determining if these sequences have any features in common. A comparison of approximately 600 bp of sequence upstream of the coding region of these two genes on both strands of DNA has identified several perfect

matches between the two sequences and many imperfect matches located at equivalent distances from the mRNA initiation sites (A1 for con-8). The perfect matches are displayed as examples 3, 4 and 7 in Table 2. Further analysis of the same regions identified a number of palindromic sequences, of which some occurred in regions in common between con-8 and con-10. Five of these are shown as examples 1, 2, 5, 6 and 7 in Table 2. As a control for these analyses, the fourteen sequences displayed in Table 2 were compared in both orientations against 600 bp sequences upstream of the major S1 mapped mRNA initiation site of the N. crassa plasma membrane H⁺/ATPase gene (32) and ga-4 gene (33) under the homology parameters listed in Table 2. Four matches with the upstream region of the H⁺/ATPase gene were obtained (example 2, con-8, three times; example 6, con-8, once) and one match with ga-4 was obtained (example 5, con-10). Functional tests are required to evaluate the significance, if any, of the common sequences listed in Table 2. Although mRNAs for con-8 and con-10 are found at the same time during most of conidiation there are also times when con-8 mRNA is present and con-10 mRNA is not, indicating that each gene must be controlled by unique regulatory sites in addition to any common ones.

Using yeast as a model, we can predict that analogs of upstream activator sequences (UAS) could be located 600 bp or more from the mRNA initiation site (34). Specific sequences important for gene regulation in yeast vary in length between 6-24 bp; some of these have been shown to function in either orientation and some display rotational symmetry (34,35,36). It is possible that genes expressed at the same time in the life cycle of N. crassa, such as con-8 and con-10, might belong to the same family and share similar upstream sequences.

The con-8 gene is an example of a conidiation specific gene that is differentially regulated. The mRNA appears early in conidiation and late in mycelial growth. con-8 is unusual in that it is transcribed into two related RNA species of similar intensity. Future efforts will be directed toward understanding the factors responsible for the differential regulation of this gene and at identifying the function of its gene product.

ACKNOWLEDGEMENTS

We thank our laboratory colleagues for their support and advice and Tony Day for help with computing. The patient assistance of Susan Lacoste during the preparation of this manuscript is greatly appreciated. This work is supported by Public Health Service Grant AG-05568. C.Y. is a Career Investigator of the American Heart Association.

*To whom correspondence should be addressed

[†]Present address: MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

REFERENCES:

1. Berlin, V. and Yanofsky, C. (1985) *Mol. Cell. Biol.* **5**, 849-855.
2. Nelson, R.E., Selitrennikoff, C.P. and Siegel, R.W. (1975) In Reinert, J. and Holzer, H. (eds), *Results and problems in cell differentiation*, Springer-Verlag, Berlin, pp. 291-310.
3. Roberts, A.N., Berlin, V., Hager, K.M. and Yanofsky, C. (1988) *Mol. Cell. Biol.* **8**, 2411-2418.
4. Berlin, V. and Yanofsky, C. (1985) *Mol. Cell. Biol.* **5**, 839-848.
5. Vollmer, S.J. and Yanofsky, C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4869-4873.
6. Kreader, C.A. and Heckman, J.E. (1987) *Nucl. Acids Res.* **15**, 9027-9042.
7. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* **143**, 161-178.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103-119.
9. Bankier, A.T. and Barrell, B.G. (1983) In Flavell, R.A. (ed), *Techniques in the Life Sciences*, Elsevier, Shannon, Vol. B508, pp. 1-34.
10. Staden, R. (1982) *Nucl. Acids Res.* **10**, 4731-4751.
11. Staden, R. (1984) *Nucl. Acids Res.* **12**, 499-503.
12. Devereux, J., Haeblerli, P. and Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387-395.
13. Hudson, G.S. and Davidson, B.E. (1984) *J. Mol. Biol.* **180**, 1023-1051.
14. Villa, V.D. and Storck, R. (1968) *J. Bacteriol.* **96**, 184-190.
15. Ballance, D.J. (1986) *Yeast* **2**, 229-236.
16. Orbach, M.J., Porro, E.B. and Yanofsky, C. (1986) *Mol. Cell. Biol.* **6**, 2452-2461.
17. Paluh, J.L., Orbach, M.J., Legerton, T.L. and Yanofsky, C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3728-3732.
18. Kozak, M. (1987) *Nucl. Acids Res.* **15**, 8125-8148.
19. Kozak, M. (1986) *Cell* **44**, 283-292.
20. Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enz.* **47**, 45-148.
21. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* **157**, 105-132.
22. George, D.G., Barker, W.C. and Hunt, L.T. (1986) *Nucl. Acids Res.* **14**, 11-15.
23. Lipman, D.J. and Pearson, W.R. (1985) *Science* **227**, 1435-1441.
24. Dijkema, R., Dekker, B.M.M. and van Ormondt, H. (1982) *Gene* **18**, 143-156.
25. Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) *Methods Enzymol.* **91**, 524-545.
26. Romans, P. and Firtel, R.A. (1985) *J. Mol. Biol.* **186**, 321-335.
27. Gallwitz, D. and Sures, I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2546-2550.
28. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349-383.
29. Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) *Cell* **50**, 863-872.
30. Sures, I., Lowry, J. and Kedes, L.H. (1978) *Cell* **15**, 1033-1044.
31. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* **263**, 211-214.
32. Hager, K.M., Mandala, S.M., Davenport, J.W., Speicher, D.W., Benz, E.J., Jr. and Slayman, C.W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7693-7697.

33. Rutledge, B.J. (1984) *Gene* 32, 275-287.
34. Struhl, K. (1987) *Cell* 49, 295-297.
35. Bram, R.J., Lue, N.F. and Kornberg, R.D. (1986) *EMBO J.* 5, 603-608.
36. Pfeifer, K., Prezant, T. and Guarente, L. (1987) *Cell* 49, 19-27.
37. Schechtman, M.G. and Yanofsky, C. (1983) *J. Mol. Appl. Genet.* 2, 83-99.
38. *Strategene Catalog*, 1988.