Journal of Nematology 29(2):144–152. 1997.

Promoter Structure of the RNA Polymerase II Large Subunit Gene in *Caenorhabditis elegans* and *C. briggsae*

D. McK. Bird,[1] I. Kaloshian,[2] and S. Molinari[3]

Abstract: The 5′-end of the *Caenorhabditis elegans ama-1* gene transcript, which encodes the largest subunit of RNA polymerase II, was cloned. Sequencing revealed that the message is trans-spliced. To characterize the *Ce-ama-1* promoter, DNA sequence spanning 3 kb upstream from the initiation codon was determined. Typical elements, such as TATA and Sp1 sites, were absent. The homologue of *ama-1* in *C. briggsae, Cb-ama-1,* was isolated and its 5′ flanking sequence compared with that of *Ce-ama-1,* revealing only limited similarity, although both sequences included a potential initiator-class transcriptional regulator and phased repeats of an AT_3C motif. The latter elements are postulated to facilitate DNA bending and may play a role in transcription regulation.

Key words: ama-1, Caenorhabditis briggsae, Caenorhabditis elegans, DNA sequencing, molecular nematology, promoter elements, RACE PCR, TATA-box, trans-splicing.

Regulation of eukaryotic gene expression is governed to a large measure by specific interactions between DNA sequence elements and complexes of accessory proteins termed "transcription factors" (Conaway and Conaway, 1993; Johnson and McKnight, 1989). For most genes, the primary event in transcription initiation is recognition by TATA-binding protein (TBP) of a short AT-rich sequence in the gene promoter (the TATA-element). TBP-binding, which results in bending of the DNA through approximately 80° (Kim et al., 1993a; Kim et al., 1993b), is necessary and sufficient to direct RNA polymerase II (RNAP II) to initiate basal transcription and synthesize mRNA from TATA-containing promoters (Martinez et al., 1995). High-level, specific transcription is modulated by interactions with further accessory proteins termed "specific transcription factors" (Johnson and McKnight, 1989). However, although biochemical analyses have elucidated the subunit structure of RNAP II from a range of organisms (reviewed by Young, 1991), and many of the accessory proteins required for transcription initiation have been defined (Conaway and Conaway, 1993; Johnson and McKnight, 1989), the nature of interactions between these components (and also with the DNA template) is only beginning to be elucidated (Dynan and Gilman, 1993).

Genetic studies of the *Saccharomyces cerevisiae* (yeast) gene encoding the largest RNAP II subunit revealed that it participates in transcriptional initiation (Hekmatpanah and Young, 1991) and elongation (Archambault et al., 1992), and direct biochemical evidence has been obtained for the interaction of a suite of proteins with a domain at the carboxyl terminus of this subunit (Chao et al., 1996). Isolation of mutants resistant to α-amanitin, a toxin that binds RNAP II and inhibits RNA chain elongation (Cochet-Mielhac and Chambon, 1974), has permitted genetic analyses of this enzyme in a number of other organisms, including *Caenorhabditis elegans* (Sanford et al., 1983; Rogalski and Riddle, 1988). In *C. elegans,* the gene encoding the largest RNAP II subunit (*Ce-ama-1*) has been mapped to chromosome IV, 0.05 map units to the right of *dpy-13* (Rogalski and Riddle, 1988). The 9.9-kb coding portion of the wild type gene (Bird and Riddle, 1989) and also the α-amanitin-resistant allele *Ce-ama-1(m118),* which carries a G-to-A change at nucleotide 3,891 (Bird, Driscoll and Riddle, unpubl.), has been cloned and sequenced. A physical map of this region is shown in Fig. 1.

A large collection of *Ce-ama-1* alleles, many of which exhibit developmental phenotypes, have been analyzed (Rogalski et al.,

Received for publication 13 September 1996.

[1] Department of Plant Pathology, Box 7616, North Carolina State University, Raleigh, NC 27695.

[2] Department of Nematology, University of California, Davis, CA 95616.

[3] Instituto di Nematologia Applicata ai Vegetali, C.N.R., 70126 Bari, Italy.

E-mail: david_bird@ncsu.edu

Nematodes were obtained from the *Caenorhabditis* Genetics Center, which is supported by the National Institutes of Health, National Center for Research Resources.

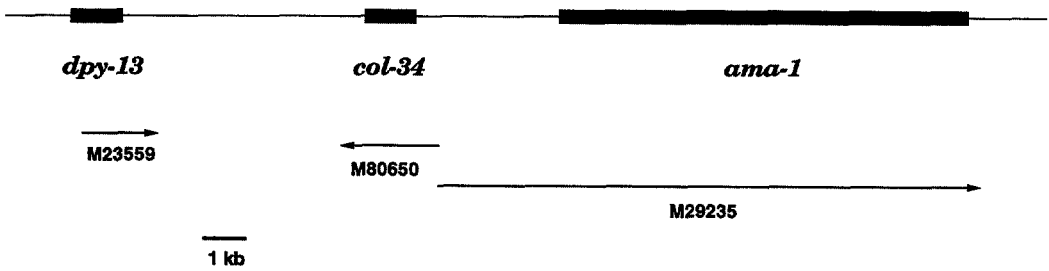

FIG. 1. Physical map around *ama-1* on *Caenorhabditis elegans* chromosome IV. Coding regions of *ama-1, dpy-13,* and *col-34* are shown as black boxes with regions sequenced (Bird, 1992; Bird and Riddle, 1989; von Mende et al., 1988; this paper) shown as thin arrows (which also indicate transcription direction) with GenBank accession numbers below.

1988; Bullerjahn and Riddle, 1988). Null alleles of *Ce-ama-1* are embryo-lethal, and the presence of the dominant-negative allele *Ce-ama-1(m118m252)* gives a late embryonic-arrest phenotype, indicating that expression of this gene occurs during embryogenesis (Rogalski et al., 1988). Using mutants with various gene doses of *Ce-ama-1,* Dalley et al. (1993) showed that expression of this gene is post-transcriptionally regulated and speculated that sequences responsible for translational regulation might reside in the 5′ untranslated region (UTR). However, despite the crucial cellular role of RNAP II, little is known about the mechanisms regulating expression of RNAP II subunits. The work presented here was aimed at understanding these processes.

In particular, we were interested in identifying sites responsible for translational regulation in the 5′ UTR, and also targets for transcription factors in the *Ce-ama-1* promoter. Using a PCR approach, we cloned the 5′-end of the *Ce-ama-1* transcript. Sequencing revealed that the *Ce-ama-1* transcript is transpliced, thereby unambiguously defining the 5′-end of the message but providing no information about the transcription start site. Simple examination of the short 5′ UTR failed to reveal potential sites of post-transcriptional regulation. Similarly, examination of genomic sequences upstream of the *Ce-ama-1* coding region also failed to reveal typical promoter elements. One approach that has proved powerful for identification of functionally important sequences has been to compare analogous gene regions from related species. Functionally important sequences such as coding regions and promoter elements are likely to be conserved, whereas non-functional, intergenic regions are likely to have diverged. Although morphologically similar to *C. elegans, C. briggsae* is sufficiently diverged (Fitch et al., 1995), and this approach has been used to identify regulatory elements of various *C. elegans* genes, including members of the vitellogenin family (Zucker-Aprison and Blumenthal, 1989). Consequently, we isolated the *Ce-ama-1* homologue from *C. briggsae* (*Cb-ama-1*) and compared genomic sequences. Based on these results, it appears that the *Ce-ama-1* promoter belongs to the Inr (initiator) class and lacks a TATA element. In contrast, the *Cb-ama-1* promoter has a recognizable TATA-motif. Additionally, both genes share repeats of a motif with the canonical sequence AT_3C phased in a manner consistent with them playing a role in DNA bending.

MATERIALS AND METHODS

Nematode cultures: Wild type *C. elegans* (N2) and *C. briggsae* were grown on *Escherichia coli* OP50 (Brenner, 1974), either on plates or in liquid S medium (Sulston and Brenner, 1974), and were harvested for nucleic acid isolation immediately upon depletion of the food supply; RNA and DNA were isolated as described (Bird and Riddle, 1989). Molecular clones and genes were named according to Bird and Riddle (1994).

cDNA cloning: Based on the previously de-

termined genomic and cDNA sequences of the *Ce-ama-1* coding region (Bird and Riddle, 1989), a set of oligonucleotides (Fig. 2) were designed to prime cDNA synthesis and subsequent PCR amplification of the *Ce-ama-1* 5′-end using the RACE procedure (Froman et al., 1988). Twenty nmols of a primer (P44: 5′-AGAAGCCGTCCGCAG-TAG-3′), which is complementary to the *Ce-ama-1* mRNA beginning 418 nucleotides 3′ from the initiation codon (i.e., +418), was annealed to 1 µg of *C. elegans* total RNA and double-stranded cDNA synthesized using a 5′ RACE System kit (BRL, Gaithersburg, MD). An aliquot was then subjected to 35 rounds of PCR amplification using the 5′ RACE System kit anchor primer (P1) and a primer (P34: 5′-CTTTGGCTTTCC-GTTCTC-3′) complementary to the *Ce-ama-1* transcript at +224. An aliquot was re-amplified by 35 rounds of PCR using the 5′ RACE System kit universal amplification primer (P2) and a primer complementary to the *Ce-ama-1* transcript at +49 (P12: 5′-CGTGAGACAATGCGCAGCGGCGCT-3′). This final amplification product was digested with Sal I, ligated into the Sal I site of the phagemid vector pUC119, and transformed into electrocompetent *E. coli* DH12S cells (Dower et al., 1988).

Genomic cloning: A 4-kb Bam HI-Mlu I fragment from the *C. elegans* genomic cosmid clone CB#B0016 was isolated and subcloned into the vector pGem7Zf+ (Promega, Madison, WI) to generate DB#10. CB#B0016 had previously been shown to span the 5′-end of *Ce-ama-1* (Bird and Riddle, 1989) and was kindly provided by A. Coulson (Coulson et al., 1986).

Phage DB#1 was isolated from a *C. briggsae* Charon 4 genomic library (constructed by T. Snutch and kindly supplied by D. Baillie) by 3 rounds of plaque purification and screening with ^{32}P-labeled insert from DB#16 at a final stringency of 330 mM Na^+/ 42 °C. The DB#16 probe is a 2.1-kb Eco RI fragment from the near-full-length *Ce-ama-1* cDNA clone DB#2 (Bird and Riddle, 1989) cloned into the Eco RI site of pGem7Zf+ and spans the *Ce-ama-1* initiation codon. The genomic library was re-screened with subclones from DB#1 to identify the overlapping clone DB#45.

DNA sequencing: Templates were generated for sequencing from phagemid subclones using random sonication (Deininger, 1983) and *Exo*III-*Exo*VII nested deletion approaches. Single-stranded DNA was prepared following rescue with the helper phage M13K07 and subjected to sequencing using the dideoxy chain-termination method (Sanger et al., 1977). Formamide (25%) was included in some sequencing gels to enable resolution of compressions. Both strands were completely sequenced. *C. elegans* and *C. briggsae* sequences were compared with software from the University of Wisconsin Genetics Computer Group (Devereux et al., 1984).

Results

The Ce-ama-1 mRNA is trans-spliced: To experimentally determine the transcription

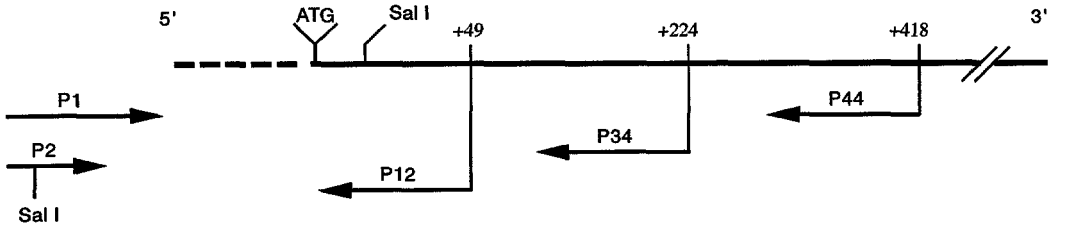

FIG. 2. Oligonucleotide primers used for cloning the 5′-end of the *Caenorhabditis elegans ama-1* transcript. Previously determined *ama-1* mRNA sequence (Bird and Riddle, 1989) is represented as a solid line; unknown sequences upstream of the initiation codon (ATG) are shown dashed. Oligonucleotide primers P12, P34, and P44 are represented as arrows (indicating 5′-3′ direction); the position of the *ama-1* residue complementary to the 5′-end of each primer is numbered. Overlapping primers P1 and P2 were from the BRL RACE System kit. Sal I sites used to clone the *ama-1* 5′-end are shown.

start site, we cloned the 5′-end of the *Ce-ama-1* transcript by 5′ RACE PCR. DNA from 27 independent recombinants from two independent cDNA syntheses was sequenced. Except for the length of the homopolymeric tail added during the RACE procedure, each insert was identical and established the length of *Ce-ama-1* 5′ UTR at 84 nucleotides (Fig. 3). Alignment of cDNA and genomic sequences revealed a 56 base pair intron, immediately 5′ of the ATG (i.e., the A of the initiation codon is the first nucleotide of exon 2). The final 22 base pairs of each clone distal to the homopolymeric tail (corresponding to the 5′ end of the *Ce-ama-1* transcript) were identical to the *C. elegans* trans-spliced leader, SL-1 (Krause and Hirsh, 1987). The splice acceptor is indicated in Fig. 3.

Isolation of the Ce-ama-1 5′-flanking sequence: To identify potential *Ce-ama-1* promoter elements, a 4-kb Bam HI-Mlu I fragment spanning the *Ce-ama-1* ATG and extending into the adjacent, divergently transcribed *col-34* gene (Bird, 1992) was subcloned, and 2,963 nucleotides of *ama-1* 5′ flanking sequence determined. This sequence was appended to that already obtained for *ama-1* (Bird and Riddle, 1989) to generate GenBank accession number M29235. M29235 abuts with M80650 (Bird, 1992) and defines 15,224 of contiguous genomic sequence (Fig. 1).

Inspection of DNA sequence 5′ from the *Ce-ama-1* ATG failed to reveal any of the typical promoter elements associated with genes transcribed by RNA polymerase II, including TATA and CAAT motifs (Johnson and McKnight, 1989), or targets for specific transcription factors (e.g., SP-1) common to other housekeeping genes (Reynolds et al., 1984; Ye et al., 1993). Inspection did, however, reveal four copies of the motif AT_3C (two have 4 T residues) spaced 10 base pairs (or 11 base pairs) apart in the *Ce-ama-1* 5′ flanking region (Fig. 4). The most distal of these elements begins at position −192 (with respect to the ATG).

```
5′              ▼
aatttaccgttttcag ATCGGGTTCACCAAAAAATCGCCGGAACACTATCGGATA
ATCGGGGAGAAAAACCGTAAAAC gtaagttttaaaagtcgaaaatcaatttttttt
aaatcccatttttttaaatttttcag ATG ...
                              3′
```

FIG. 3. The *ama-1* 5′-end. The arrowhead indicates the site in the transcript of trans-spliced leader (SL-1) addition. Underlined, lowercase sequences are postulated to be present in the primary *ama-1* transcript, which is proposed to begin with the first residue shown. Uppercase letters represent sequence of the mature *ama-1* 5′-UTR, ending with the initiation codon (ATG). Intron 1 is shown in lowercase.

Isolation of Cb-ama-1: To identify additional *ama-1* promoter elements, a *C. briggsae* genomic clone (DB#1) was isolated by screening with a *Ce-ama-1* probe. By hybridizing Southern blots of DB#1 restriction digests with *Ce-ama-1* sequences, it was ascertained that DB#1 shared homology with the entire *ama-1* coding region (not shown). Subclones spanning DB#1 were constructed, and the 778 base pair insert from one, DB#40, which hybridized with the 5′-end of *Ce-ama-1* in a Southern blot (not shown), was used to re-screen the *C. briggsae* genomic library. One of the recombinants thus isolated, DB#45, extended farther 5′ and, based on Southern blotting with a *C. elegans* collagen gene probe, includes the presumed *C. briggsae* orthologue of *col-34* (Fig. 1).

The DNA sequence of the DB#40 insert was determined. Analysis of open reading frames implicated a coding region at one end of the clone. Alignment of this deduced protein sequence revealed a high degree of similarity with the amino terminus of the *Ce-ama-1* product; 59 of 64 amino acids are identical (Fig. 4). Introns (experimentally confirmed for *Ce-ama-1* and deduced for *Cb-ama-1*) disrupt each coding region at identical points, although intron length is not conserved (Fig. 4). These results confirmed that the genomic clones DB#1 and DB#45 together encode the *C. briggsae ama-1* gene, named *Cb-ama-1*. The 778 nucleotide sequence of DB#40 has been entered into GenBank, with the accession number L23763.

Reduced stringency Southern blotting (not shown) suggested that the 5′ flanking region of the *Ce-ama-1* gene shared sequences with *Cb-ama-1*, but that this homology extended no farther than 500 base pairs 5′ from the ATG. The DB#40 sequence,

A.

```
Cb-ama-1:  ATG GCT CTC GTC GGA GTC GAC TTT CAG GCT CCT CTC CGA ACC
           M   A   L   V   G   V   D   F   Q   A   P   L   R   T      14
Ce-ama-1:  ATG GCT CTC GTC GGT GTC GAC TTC CAA GCG CCG CTG CGC ATT
                                                               I

Cb-ama-1:  GTC TGC CGT GTG CAG TTC GGT TTT TTG GGA CCC GAG GAG ATT
           V   C   R   V   Q   F   G   F   L   G   P   E   E   I      28
Ce-ama-1:  GTC TCA CGC GTT CAG TTT GGA ATT CTC GGC CCG GAG GAG ATT
               S                   I

Cb-ama-1:  gtaagtggctgaaaatggctgaaatctcgagatttttttttcag        AAA
                                                               K      29
Ce-ama-1:  gtaggcggaaacccccgatttaacccgaaatttatgatttcttattcag AAA

Cb-ama-1:  CGA ATG TCA GTG GCT CAT GTC GAG TTT CCC GAG GTC TAC GAG
           R   M   S   V   A   H   V   E   F   P   E   V   Y   E      43
Ce-ama-1:  CGC ATG TCA GTG GCT CAT GTC GAG TTT CCA GAA GTC TAC GAG

Cb-ama-1:  AAC GGA AAA CCC AAA ATG GGC GGT CTG ATG GAC CCG CGC CAG
           N   G   K   P   K   M   G   G   L   M   D   P   R   Q      57
Ce-ama-1:  AAC GGA AAG CCA AAG TTG GGC GGT CTC ATG GAT CCA AGA CAA
                               L

Cb-ama-1:  GGA GTG ATT GGA CGT CGT GGA AG gttagga
           G   V   I   G   R   R   G                                  64
Ce-ama-1:  GGA GTC ATA GAT CGT CGT GGA AG gtgcggt
                   D
```

B.

```
                    -188
                    ATTTTTTC GCATTTCATGGTTATAAACAGCTGAAAATTCAGTTT
                    **** * *   *****  *  * **      * *********  ***
AATTTTCGTGGATTTTCGCAATATTTCT CTCAATTTCTTCCTGATTTTTT C GAAAATTCAATTTACCGTTTTCAGATC
-192                                                                    ▲
```

FIG. 4. Alignment of *Caenorhabditis elegans* (*Ce-ama-1*) and *C. briggsae ama-1* sequences. A) Predicted coding sequence from the *C. briggsae* genomic subclone DB#40 (*Cb-ama-1*), beginning with the presumed ATG start codon, aligned with the previously determined *C. elegans ama-1* genomic sequence (*Ce-ama-1*) (Bird and Riddle, 1989). Predicted intron sequences are shown in lowercase; a gap has been introduced in the first *Cb-ama-1* intron shown to retain coding region alignment. The deduced *Cb-ama-1* amino acid sequence is shown below the DNA sequence in italic, uppercase, one-letter code, and numbered on the right. Deduced *Ce-ama-1* amino acids that differ from those at equivalent positions in the *Cb-ama-1* product are shown in bold italic, uppercase, one-letter code. B) *Cb-ama-1* (top) and *Ce-ama-1* 5′ flanking regions, beginning 188 base pairs and 192 base pairs 5′ from the respective initiation codons, are shown aligned. Single nucleotide gaps have been introduced into both sequences to maximize similarity. Asterisks indicate identical residues. AT_3C motifs (and derivatives) are underlined, and the Inr consensus is double-underlined. The arrowhead indicates the site in the *Ce-ama-1* transcript of SL-1 addition. The putative *C. briggsae* TATA box is shown in bold.

which extends 534 base pairs 5′ from the *Cb-ama-1* initiation codon, was therefore compared with the equivalent region of *Ce-ama-1,* revealing only a short, but none-theless significant, region of similarity (Fig. 4). The *Cb-ama-1* sequence includes two forms of the AT_3C motif, aligned with two such motifs in *Ce-ama-1,* and also a potential TATA box. Farther 3′, an additional region of similarity was observed between the two

genes (Fig. 4). Significantly, these sequences correspond to the consensus for the basal transcription control element, "initiator" (Inr), which is YYANWYY (Corden et al., 1980).

Discussion

Transcription regulation of housekeeping genes, including those encoding RNA polymerase subunits, is poorly understood. Because the *C. elegans ama-1* gene has been subjected to extensive genetic scrutiny, we initiated a molecular study of its regulation. Previous analysis of genomic and cDNA clones had defined the *Ce-ama-1* transcript from the initiation codon to the poly-A tail, but no information was available for the 5′ UTR, and it had been speculated that the sequence immediately upstream of the ATG might function as a trans-splice acceptor (Bird and Riddle, 1989). To test this, we isolated and sequenced the 5′-end of the mature transcript, revealing that the proposed splice acceptor at −1 is, in fact, a cis-splicing acceptor. Thus, *Ce-ama-1* has 13 exons, not 12 as previously reported (Bird and Riddle, 1989). However, an additional splice acceptor was identified at −116 and, based on sequence obtained from 27 *Ce-ama-1* cDNA clones, this site is employed for addition of the trans-spliced leader SL-1. These recombinants were randomly picked from two independent cDNA syntheses and thus must have been derived from at least two independent transcripts. Additionally, different recombinants showed variation in the length of the homopolymeric tail added prior to the PCR steps of the RACE process, consistent with the notion that all 27 recombinants were derived from independent transcripts. Thus, it seems likely that all stable *Ce-ama-1* mRNAs are trans-spliced with SL-1.

The presence of SL-1, which is capped, unambiguously defines the 5′-end of the mature *Ce-ama-1* transcript, and establishes the length of 5′ UTR as being 84 nucleotides (including the 22 nucleotides of SL-1). This is substantially shorter than for the homologous transcripts in *Drosophila* (436 nucleotides: Biggs et al., 1985), mouse (407 nucleotides: Ahearn et al., 1987), and yeast (166 nucleotides: Allison et al., 1985). Relatively short 5′ UTRs are typical for *C. elegans* genes.

Dalley et al. (1993) used an elegant genetic and biochemical approach to demonstrate post-transcriptional regulation of *Ce-ama-1,* and speculated that sites for such regulation might reside in the 5′ UTR. The finding that bacterial RNAP large subunits also are subject to post-transcriptional regulation (Dennis, 1984) suggests that regulation at this level might be a general feature of RNAP subunit genes. If this is the case, then the sites of this regulation might exhibit conservation of primary or secondary structure. However, sequence comparisons of *Ce-ama-1* and the 5′ UTRs of the mouse, fly, and yeast RNAP II large subunit genes using the BESTFIT algorithm (Devereux et al., 1984) failed to reveal similarity judged to be significant, nor was similarity found between the *Ce-ama-1* 5′ UTR and the translational regulator sequences of the *E. coli* β and β′ subunits (Dennis, 1984). A search for potential secondary structure in the *ama-1* 5′ UTR also failed to reveal any meaningful (i.e., with a negative free energy) stem-loop structures, although more complex structures involving the entire *ama-1* transcript cannot be ruled out. Thus, the existence of potential post-transcriptional regulation sites in the *ama-1* 5′ UTR remains unresolved.

The presence of the trans-spliced leader precludes inference of the transcription start site simply from analysis of the mature *ama-1* transcript, although it does place it at or 5′ to the trans-splice acceptor. To identify core promoter elements (such as the TATA element), which typically are upstream of the transcription site, we determined the sequence of approximately 3 kb of DNA extending 5′ from the *Ce-ama-1* coding region. This sequence extended to the promoter of the adjacent, divergently transcribed *col-34* gene (Bird, 1992); it is unlikely that the *Ce-ama-1* transcript could originate beyond this point. We isolated the homologue of *Ce-ama-1* from the related nematode *C. briggsae*

and determined the first two coding exons, plus sequences extending approximately 500 base pairs upstream. Based on their high degree of coding region homology, it was straightforward to align the *Ce-ama-1* and *Cb-ama-1* genomic sequences. Beyond the coding regions the sequences diverged, and because we have no data from the *Cb-ama-1* message sequence, we are unable reliably to infer the intron/exon structure upstream of the ATG. Nevertheless, manual and computer-assisted inspection of the *Ce-ama-1* and *Cb-ama-1* 5′-flanking sequences revealed what appears to be the core promoter for each of these genes. Both genes possess a sequence motif that corresponds to an Inr element, which is the simplest class of element known to direct accurate basal transcription. Both postulated elements (*Ce-ama-1:* TCAATTT; *Cb-ama-1:* TCAGTTT) are strikingly similar to the strong Inr (TCATTCT) of the human terminal deoxynucleotidyl transferase gene, especially at positions two (C), three (A), and five (T), which are the most important for function (Javahery et al., 1994). Complete identity between the two nematode genes continues 5′ for a short region, perhaps revealing an expanded template target for nematode Inr-binding proteins. Inr elements span the transcription start point, and based on this assignment the transcription start point of *Ce-ama-1* is postulated to be at position −134, and that of *Cb-ama-1* at position −149. For the *C. elegans* gene, this is 16 nucleotides upstream from the addition point of SL-1 and is consistent with the known structure of the mature transcript. In addition to an Inr element, the *C. briggsae* gene also contains a TATA motif; no such element is apparent in *Ce-ama-1*. In this regard, *Cb-ama-1* more closely resembles its homologues in mouse (Ahearn et al., 1987), yeast (Allison et al., 1985), and *Drosophila* (Biggs et al., 1985), which all have TATA elements. The putative *Cb-ama-1* TATA is at −20 with respect to the postulated, Inr-specified transcription start site and is thus consistent with the position of TATA elements with known function.

Based on biochemical approaches, Martinez et al. (1995) have proposed a model for Inr function. In simplified form, Inr serves as a direct target for part of the TFIID complex. The TFIID complex is an incompletely defined suite of transcription factors, which includes the TATA-binding protein (TBP). If a TATA element is present upstream of the Inr (as in *Cb-ama-1*), TBP is able to bind, thus further stimulating transcription. Components of the TFIID complex also interact with specific transcription factors bound to DNA targets (typically farther upstream). Targets for such factors are apparent in the *ama-1* homologues from other species. The *Drosophila* gene has a CAAT motif, and the mouse gene has several potential targets for the transcription factor SP-1 (Dynan and Tjian, 1983), a feature common to housekeeping genes (Reynolds et al., 1984; Ye et al., 1993). However, careful examination of sequences upstream of *Cb-ama-1* and *Ce-ama-1* failed to reveal previously described transcription factor targets. Rather, multiple copies (four for *Ce-ama-1* and two for *Cb-ama-1*) of a motif with the canonical sequence AT_3C were observed (two of the motifs in *Ce-ama-1* were AT_4C, and one in *Cb-ama-1* was AT_6C). The spacing of the AT_3C motifs places them at equivalent positions on adjacent turns of the DNA helix, suggesting that they might serve as targets for regulatory proteins. The transcription regulator Pit-1 (Ingraham et al., 1990), for example, binds cooperatively to paired motifs that are typically 8–12 base pairs apart.

An intriguing alternative is that rather than being targets for DNA-binding proteins, the AT_3C elements play a role in DNA bending. Bending has been observed upstream of a number of genes, including the adenovirus E1A enhancer, where two of the responsible elements correspond to AT_4C (Ohyama and Hashimoto, 1989). Computer modeling showed that the basis for this type of DNA bending is a series of oligo T_{3-6} tracts spaced with an average periodicity of 10.2 base pairs (Van Wye et al., 1991). The *ama-1* AT_3C elements are spaced (5′-3′) with a periodicity of 10, 11, and 10 base pairs, and thus are predicted by this model to effect bending. Significantly, direct experi-

mental evidence has been obtained to show that *Ce-ama-1* (as naked DNA) is strongly bent throughout a region extending 700 base pairs upstream from the ATG (VanWye et al., 1991).

The biological role of promoter bending is unclear, but a strong correlation between gene activity and bending has been observed. Most striking is the 80° DNA bend caused when TBP binds the TATA element (Kim et al., 1993a; Kim et al., 1993b), and a recent structural analysis of the *Drosophila* TFIID (Xie et al., 1996) indicates that the entire TFIID complex might adopt a conformation similar to that of the histone octamer, around which the DNA strand is wound twice. These data are suggestive of the importance of bending at the promoter. It is likely that TFIID complex binds to the *C. elegans* and *C. briggsae ama-1* promoters, recognizing the Inr element upstream of each gene. Whether this interaction results in any DNA bending is unknown. Presumably, the TBP component of TFIID also recognizes the TATA of *Cb-ama-1,* resulting in the characteristic 80° bend. Experiments in which a TATA element was added to a promoter containing only an Inr element resulted in a much stronger promoter (Martinez et al., 1995). Perhaps the additional (compared to *Cb-ama-1*) AT_3C elements upstream of *Ce-ama-1* ameliorate for the absence of a TATA element and increase the strength of the *Ce-ama-1* promoter by potentiating bending. If the AT_3C elements function as targets for transcription factors, the TFIID presumably also interacts with those proteins (perhaps also causing bending). However, it seems likely that the AT_3C elements serve as bending sites alone, possibly permitting a degree of basal transcription in the absence of transcription factors. This might be an important promoter attribute for a function as central to cell viability as proper expression of an RNA polymerase subunit gene.

Literature Cited

Ahearn, J. M., M. S. Bartolomei, M. L. West, L. J. Cisek, and J. L. Corden. 1987. Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. Journal of Biological Chemistry 262:10695–10705.

Allison, L. A., M. Moyle, M. Shales, and C. J. Ingles. 1985. Extensive homology amongst the large subunits of eukaryotic and prokaryotic RNA polymerases. Cell 42:599–610.

Archambault, J., F. Lacroute, A. Ruet, and J. D. Friesen. 1992. Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. Molecular and Cellular Biology 12:4142–4152.

Biggs, J., L. L. Searles, and A. L. Greenleaf. 1985. Structure of the eukaryotic transcription apparatus: Features of the gene for the largest subunit of *Drosophila* RNA polymerase II. Cell 42:611–621.

Bird, D. McK. 1992. Sequence comparison of the *Caenorhabditis elegans dpy-13* and *col-34* genes, and their deduced collagen products. Gene 120:261–266.

Bird, D. McK., and D. L. Riddle. 1989. Molecular cloning and sequence of *ama-1,* the gene encoding the largest subunit of *Caenorhabditis elegans* RNA polymerase II. Molecular and Cellular Biology 9:4119–4130.

Bird, D. McK., and D. L. Riddle. 1994. A genetic nomenclature for parasitic nematodes. Journal of Nematology 26:138–143.

Brenner, S. 1974. The genetics of *C. elegans.* Genetics 77:71–94.

Bullerjahn, A. M. E., and D. L. Riddle. 1988. Fine-structure genetics of *ama-1,* an essential gene encoding the large subunit of RNA polymerase II in *Caenorhabditis elegans.* Genetics 120:423–434.

Chao, D. M., E. L. Gadbois, P. J. Murray, S. F. Anderson, M. S. Sonu, J. D. Parvin, and R. A. Young. 1996. A mammalian SRB protein associated with an RNA polymerase II holoenzyme. Nature 380:82–85.

Cochet-Mielhac, M., and P. Chambon. 1974. Animal DNA-dependent RNA polymerases. II. Mechanism of the inhibition of RNA polymerase B by amatoxins. Biochemica et Biophysica Acta 353:160–184.

Conaway, R. C., and J. W. Conaway. 1993. General initiation factors for RNA polymerase II. Annual Review of Biochemistry 62:161–190.

Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon. 1980. Promoter sequences of eukaryotic protein-coding genes. Science 209:1405–1414.

Coulson, A., J. Sulston, S. Brenner, and J. Karn. 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans.* Proceedings of the National Academy of Science (USA) 83:7821–7825.

Dalley, B. K., T. M. Rogalski, G. E. Tullis, D. L. Riddle, and M. Golomb. 1993. Post-transcriptional regulation of RNA polymerase II levels in *Caenorhabditis elegans.* Genetics 133:237–245.

Deininger, P. L. 1983. Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. Analytical Biochemistry 129:216–223.

Dennis, P. P. 1984. Site-specific deletions of regulatory sequences in a ribosomal protein-RNA polymerase operon in *Escherichia coli.* Journal of Biological Chemistry 259:3202–3209.

Devereux, J., P. Haeberli, and O. W. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Research 10:387–395.

Dower, W. J., J. F. Miller, and C. W. Ragsdale. 1988.

High-efficiency transformation of *E. coli* by high-voltage electroporation. Nucleic Acids Research 16:6127–6145.

Dynan, W. S., and M. Z. Gilman. 1993. Transcriptional control and signal transduction: Of soap-operas and reductionism. Trends in Genetics 9:154–156.

Dynan, W. S., and R. Tjian. 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32: 669–680.

Fitch, D. H. A., B. Bugaj-Gaweda, and S. W. Emmons. 1995. 18S ribosomal RNA gene phylogeny for some Rhabditidae related to *Caenorhabditis*. Molecular Biology and Evolution 12:346–358.

Froman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proceedings of the National Academy of Science (USA) 85:8998–9002.

Hekmatpanah, D. S., and R. A. Young. 1991. Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. Molecular and Cellular Biology 11:5781–5791.

Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S. Kapiloff, L. Wilson, and M. G. Rosenfeld. 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high-affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. Cell 61:1021–1033.

Javahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory, and S. T. Smale. 1994. DNA sequence requirements for transcription initiator activity in mammalian cells. Molecular and Cellular Biology 14:116–127.

Johnson, P. F., and S. L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. Annual Review of Biochemistry 58:799–839.

Kim, J. L., D. B. Nikolov, and S. K. Burley, 1993a. Co-crystal structure of a TBP recognizing the minor groove of a TATA element. Nature 365:520–527.

Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler. 1993b. Crystal structure of a yeast TBP/TATA-box complex. Nature 365:512–520.

Krause, M., and D. Hirsh. 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49:735–761.

Martinez, E., Q. Zhou, N. D. L'Etoile, T. Oelgeschläger, A. J. Berk, and R. G. Roeder. 1995. Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. Proceedings of the National Academy of Science (USA) 92:11864–11868.

Ohyama, T., and S. Hashimoto. 1989. Upstream half of adenovirus type 2 enhancer adopts a curved DNA conformation. Nucleic Acids Research 17:3845–3853.

Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey. 1984. HMG CoA reductase: A negatively regulated gene with an unusual promoter and 5′ untranslated regions. Cell 38:275–285.

Rogalski, T. M., A. M. E. Bullerjahn, and D. L. Riddle. 1988. Lethal and amanitin-resistance mutations in the *Caenorhabditis elegans ama-1* and *ama-2* genes. Genetics 120:409–422.

Rogalski, T. M., and D. L. Riddle. 1988. A *Caenorhabditis elegans* RNA polymerase II gene, *ama-1 IV,* and nearby essential genes. Genetics 118:61–74.

Sanford, T., M. Golomb, and D. L. Riddle. 1983. RNA Polymerase II from wild type and α-amanitin-resistant strains of *Caenorhabditis elegans*. Journal of Biological Chemistry 258:12804–12809.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science (USA) 74: 5463–5467.

Sulston, J., and S. Brenner. 1974. The DNA of *Caenorhabditis elegans*. Genetics 77:95–104.

VanWye, J. D., E. C. Bronson, and J. N. Anderson. 1991. Species-specific patterns of DNA bending and sequence. Nucleic Acids Research 19:5253–5261.

von Mende, N., D. McK. Bird, P. S. Albert, and D. L. Riddle. 1988. *dpy-13:* A nematode collagen gene that affects body shape. Cell 55:567–576.

Xie, X., T. Kobudo, S. L. Cohen, U. A. Mirza, A. Hoffman, B. T. Chait, R. G. Roeder, Y. Nakatani, and S. K. Burley. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. Nature 380:316–322.

Ye, K., C. A. Dinarello, and B. D. Clark. 1993. Identification of the promoter region of human interleukin 1 type I receptor gene: Multiple initiation sites, high G + C content, and constituitive expression. Proceedings of the National Academy of Science (USA) 90:2295–2299.

Young, R. A. 1991. RNA polymerase II. Annual Review of Biochemistry 60:689–715.

Zucker-Aprison, E., and T. Blumenthal. 1989. Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequences comparison. Journal of Molecular Evolution 28:487–496.