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 AT3C 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.,
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 transspliced, 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$_3$C 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'-AGAAGCCGTCCGCAGTAG-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'-CTTTGGCTTTGGTCTC-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'-CGTGAGACAATGCAGCGGCCT-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. Bailie) by 3 rounds of plaque purification and screening with 32P-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 ExoIII-ExoVII 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
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 Barn HI-Mlu I fragment spanning the Ce-ama-1 ATG and extending into the adjacent, divergently transcribed coi-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-i 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 Mc Knight, 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_\text{C} (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).

Isolation of Cb-ama-l: 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.

\[
\text{Cb-ama-1: } \text{ATG GCT CTC GTC GGA GTC GAC TTT CAG GCT CTT CTC CAA GAG CAG GAC ATT} \\
\text{Ce-ama-1: } \text{ATG GCT CTC GTC GGT GTC GAC TTC CAA GCG CCG CTG GGC ATT} \\
\]

\[
\text{Cb-ama-1: } \text{GTC TCC GGT CAG TCC GGT TTT TIG GGA CAG GAG ATT} \\
\text{Ce-ama-1: } \text{GTC TCA GGC GGT CAG TTT GGA ATT CTC GGC CCG GAG GAC ATT} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GGTTGTTTAC} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

\[
\text{Cb-ama-1: } \text{gtaagtggctggaaatggctgaaatctcgagattttttttttcag} \\
\text{Ce-ama-1: } \text{gtaggcggaaaccccccagatatttaaccccgaaatttatgatttcttatttcag} \\
\]

\[
\text{Cb-ama-1: } \text{GTTATGTCGAGTTCG} \\
\text{Ce-ama-1: } \text{GTTATGTCGAGTTCG} \\
\]

\[
\text{Cb-ama-1: } \text{AAC AAA CLA AAA} \\
\text{Ce-ama-1: } \text{AAC GGA AAG CUA AAG} \\
\]

\[
\text{Cb-ama-1: } \text{G V I G R R G} \\
\text{Ce-ama-1: } \text{G V I G R R G} \\
\]

B.

**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 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. ATsC 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-the-less significant, region of similarity (Fig. 4). The Cb-ama-1 sequence includes two forms of the ATsC 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 YANWWY (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 deoxyribonucleotidyl 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 AT₃C were observed (two of the motifs in Ce-ama-1 were AT₅C, and one in Cb-ama-1 was AT₆C). The spacing of the AT₃C 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 AT₃C elements play a role in DNA bending. Bending has been observed upstream of a number of genes, including the adenovirus EIA enhancer, where two of the responsible elements correspond to AT₄C (Ohyama and Hashimoto, 1989). Computer modeling showed that the basis for this type of DNA bending is a series of oligo T₅₆ tracts spaced with an average periodicity of 10.2 base pairs (Van Wye et al., 1991). The ama-1 AT₃C 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 Inn 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 Ch-ama-1, resulting in the characteristic 80° bend. Experiments in which a TATA element was added to a promoter containing only an Inn element resulted in a much stronger promoter (Martinez et al., 1995). Perhaps the additional (compared to Ch-ama-1) AT₃C 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 AT₃C 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 AT₃C 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**


