

***Caenorhabditis elegans* mRNAs that encode a protein similar to ADARs derive from an operon containing six genes**

Ronald F. Hough, Arunth T. Lingam and Brenda L. Bass*

Department of Biochemistry and Howard Hughes Medical Institute, University of Utah, 50 North Medical Drive, Salt Lake City, UT 84132, USA

Received June 8, 1999; Revised and Accepted July 15, 1999

DDBJ/EMBL/GenBank accession nos[†]

ABSTRACT

The *Caenorhabditis elegans* T20H4.4 open reading frame (GenBank accession no. U00037) predicted by Genefinder encodes a 367 amino acid protein that is 32–35% identical to the C-terminal domain of adenosine deaminases that act on RNA. We show that T20H4.4 cDNAs (GenBank accession no. AF051275) encode a larger 495 amino acid protein that is extended at its N-terminus to include a single double-stranded RNA-binding motif, and that T20H4.4 occupies the second position in a six-gene operon (5'-T20H4.5, T20H4.4, R151.8A, R151.8B, R151.7, R151.6-3'). Ten different spliced-leader (SL) sequences were found attached to T20H4.4 mRNAs, including SL1, SL2 and eight SL2-like leaders that include two new variants. Characterization of cDNAs derived from all six genes confirmed the essential features of *C.elegans* operons: intergenic distances in the range of 104–257 nt between the upstream polyadenylation sites and the downstream *trans*-splice sites; SL2, or SL2-like leaders, attached to the downstream mRNAs. Polycistronic mRNA fragments revealed a 5'-untranslated region (5'-UTR) >705 nt. The 5'-UTR is removed in mature mRNAs from the first gene (T20H4.5) and replaced primarily by SL1, and to a lesser extent by SL2. Our study provides new information regarding operons and how they are processed.

INTRODUCTION

The *Caenorhabditis elegans* T20H4.4 open reading frame (ORF) encodes a protein that is remarkably similar to the C-terminal domain of the adenosine deaminases that act on RNA (ADARs) (1,2). ADARs require base-paired substrates and were first discovered by their ability to modify adenosines to inosines within double-stranded RNA (dsRNA) (3,4). *In vivo* the enzymes act as RNA editing enzymes to deaminate

adenosines within base-paired regions of cellular pre-mRNAs and viral RNAs (reviewed in 5). In addition to the C-terminal domain, which contains the catalytic active site (2,6,7), ADARs contain variable numbers of an amino acid sequence known as the dsRNA binding motif (dsRBM) (8,9).

We isolated several T20H4.4 clones from a *C.elegans* cDNA library. The cDNAs included two exons in addition to those identified as part of the T20H4.4 ORF by Genefinder (10), and encoded a larger protein (55.3 kDa) that contained a single dsRBM. While one of the newly identified exons was created by conventional *cis*-splicing, the second derived from a *trans*-splicing event since the 5' ends of the three longest clones contained non-genomic spliced-leader (SL) sequences. We show that the 5' ends of T20H4.4 mRNAs are *trans*-spliced to at least 10 distinct SLs, mainly SL2-like leaders (11–13) which are found in mature mRNAs that derive from downstream regions of polycistronic pre-mRNAs (reviewed in 14). Northern analyses demonstrated that the T20H4.4 mRNAs are expressed as a single species (~1.7 kb).

Toward our goal of understanding the expression and function of ADAR-like T20H4.4, we set out to determine whether T20H4.4 mRNAs are co-transcribed with nearby genes, and to define the genes of the operon. T20H4.4 is the second ORF in a gene cluster located near the center of chromosome III, as described by the overlapping ends of cosmids T20H4 and R151. Six mature transcripts from this region were characterized, all of which were *trans*-spliced and contained atypical polyadenylation signals. The predicted ORF for the first gene, T20H4.5, was confirmed, while the primary structures of the ORFs for the five downstream genes were corrected. cDNAs corresponding to partially processed fragments of polycistronic precursor RNAs were also isolated. These revealed an untranslated region >705 nt upstream of T20H4.5 and unprocessed intergenic regions between the first three ORFs. The 5'-untranslated region (5'-UTR) is removed from T20H4.5 mRNAs and replaced primarily by SL1. However, approximately one in six of the leaders were SL2. This finding suggests that mRNAs derived from the first gene in an operon may, under certain conditions, be *trans*-spliced with leaders other than SL1.

*To whom correspondence should be addressed. Tel: +1 801 581 4884; Fax: +1 801 581 5379; Email: bbass@howard.genetics.utah.edu

[†]AF140272, AF051275 and AF143147–AF143152

MATERIALS AND METHODS

Accession numbers

The GenBank database accession numbers for *C.elegans* cosmids and cDNAs are: cosmid T20H4 (U00037); cosmid R151 (U00036); T20H4.5 cDNA (AF140272); T20H4.4 cDNA (AF051275); R151.8A cDNA (AF143147); R151.8B cDNAs (5', AF143148; 3', AF143149); R151.7 cDNAs (5', AF143150; 3', AF143151); R151.6 cDNA (AF143152).

RNA preparation and cloning of T20H4.4 cDNAs

Cultures of *C.elegans* (Bristol N2) were prepared as described (15). Total RNA was prepared according to standard protocols (16). Poly(A)⁺ RNA was selected from this total RNA using an Oligotex-dT kit (Qiagen). To generate a 343 bp T20H4.4 fragment (Table 1; primers C and D), RT-PCR was performed with superscript II reverse transcriptase (Life Technologies) and AmpliTaqI polymerase (Perkin Elmer-Cetus) followed by cloning into pCRII (Invitrogen). The 343 bp insert was excised by *EcoRI* digestion and used for random-priming (BMB kit) utilizing [α -³²P]dATP (DuPont NEN; >3000 Ci/mmol). These random-primed probes were used to screen a Uni-Zap XR, *EcoRI*-*XhoI*, *C.elegans* cDNA library (Stratagene). To characterize the 5' sequence more carefully the library was rescreened with random-primed probes prepared from a 410 bp, *EcoRI*-*BstBI* fragment, corresponding to the 5' end of a full-length clone.

Rapid amplification of cDNA ends (RACE)

First strand T20H4.4 cDNA was synthesized from poly(A)⁺ RNA, in two independent preparations, primed with RACE-3 or RACE-4 (Table 1). After polyadenylation of the 5' ends catalyzed by terminal deoxynucleotidyltransferase (Life Technologies) in the presence of dATP, two-step nested PCR amplification of the 5' ends was carried out exactly as described by Frohman (17). First round PCR amplification of the cDNA templates was done with the 5' primers [Q_T + Q₀] combined with the 3' primers RACE-2 (RACE-3 RT preparation) or RACE-3 (RACE-4 RT preparation). The second round was done with Q₁ and RACE-1 in both reactions. The amplified DNAs (~220 bp) in each preparation were gel purified and cloned into pCRII (Invitrogen). T20H4.4 inserts were identified in plasmids from positive transformants by restriction analysis and sequenced.

Isolation of six full-length cDNAs and characterization of the operon

The 5' ends of mRNAs corresponding to all six genes were characterized by RACE as described above for T20H4.4 using 3' gene-specific nested primers for RT-PCR. In addition, primers corresponding to SL1, SL2, SL2* (Table 1) and 5' gene-specific primers (Table 3) were substituted for Q₁ during the second round of amplification to determine if *trans*-splicing occurred. SL2* is a 4-fold degenerate primer complementary to eight of the 11 SL2-like leaders (SLa, SLc, SLd, SLf, SLg, SLh, SLi and SLk; Table 2). Relative densities of ethidium bromide stained SL1-T20H4.5, and SL2-T20H4.5 PCR products, were estimated using an Eagle Eye II (Stratagene) gel documentation system.

After cloning and sequencing the 3' end PCR products, full-length cDNAs were obtained by long-distance PCR (Gene

AMP XL PCR; Perkin Elmer-Cetus) using gene-specific 5' primers complementary to the 5' cDNA sequences immediately downstream of the *trans*-splice sites (Table 3). These PCR products were cloned and sequenced to verify the ORFs and 3'-UTRs, and to complete other regions not covered by the expressed sequence tag database (18).

An extensive RT-PCR screen designed to amplify possible polycistronic transcripts was performed. Gene-specific primers were used to prime poly (A)⁺ RNA and total RNA during the RT step. During the PCR step, control reactions contained templates from mock RT reactions incubated without reverse transcriptase (no-RT control), and RT products obtained from RNA that was pretreated with DNase prior to the RT step (no DNA control). PCR reactions were also done with several 5' and 3' primers complementary to possible exonic sequences upstream of T20H4.5 in a effort to reveal a previously unrecognized ORF or identify a portion of the 5'-UTR. Oligo(dT)-primed cDNA was used as a template with these upstream 5' primers and Q_T, or (dT)₁₇, to search for a fragment of a polyadenylated mRNA.

Table 1. Deoxyoligonucleotides

T20H4.4 Cloning	
A (antisense)	CTTCCACAGTTTCCTTCACAA
B (antisense)	ACTGGAGTGTTCCTGTCATA
C (sense)	TTGTCATGCTGAAATTCCTCGC
D (antisense)	CATTGTCTCATTCGTTCTCC
5' End cDNA Amplification	
5' Primers	
Q _T	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC(T) ₁₇
Q ₀	CCAGTGAGCAGAGTGACG
Q ₁	GAGGACTCGAGCTCAAGC
SL1	GGTTTAATTACCCAAGTTTGAG
SL2	GGTTTAACCCAGTTACTCAAG
SL2*	GGTTT(A/T)(A/T)ACCCAGTTAACCAAG
5' UTR	AACATCATGTGCGCAACA
T20H4.5 3' Primers	
4.5-RACE-1	TTGAACTGAGTGGTCCCTTTTCG
4.5-RACE-2	GACGAAGAGCGTGTTCCTCTCG
4.5-RACE-3	TCACCAGACGGGTATCGACG
T20H4.4 3' Primers	
4.4-RACE-1	ACTGGAGTGTTCCTGTCATA
4.4-RACE-2	TGCCTCTTCTTTTGGATTTT
4.4-RACE-3	GTTTCATAAAAATTCGCGTATGTATG
4.4-RACE-4	TGAGAGCTGTTCATCGTTTAC

Northern analyses and *in vitro* translation of T20H4.5 and T20H4.4

Northern analyses were done as described previously (2). Random primed ³²P-labeled deoxyoligonucleotides were prepared from the 343 bp DNA fragment (Table 1, primers C and D) to identify T20H4.4 mRNAs. PCR products corresponding to the 705 bp 5'-UTR and the first 308 bp of the T20H4.5 ORF were used to prepare random primed ³²P-labeled deoxyoligonucleotides specific for T20H4.5 mRNAs. Blots hybridized with T20H4.4 probes were stripped according to standard protocols (19) and re-probed with deoxyoligonucleotides specific for the 5'-UTR or the 5' end of the T20H4.5 ORF. Coupled *in vitro*

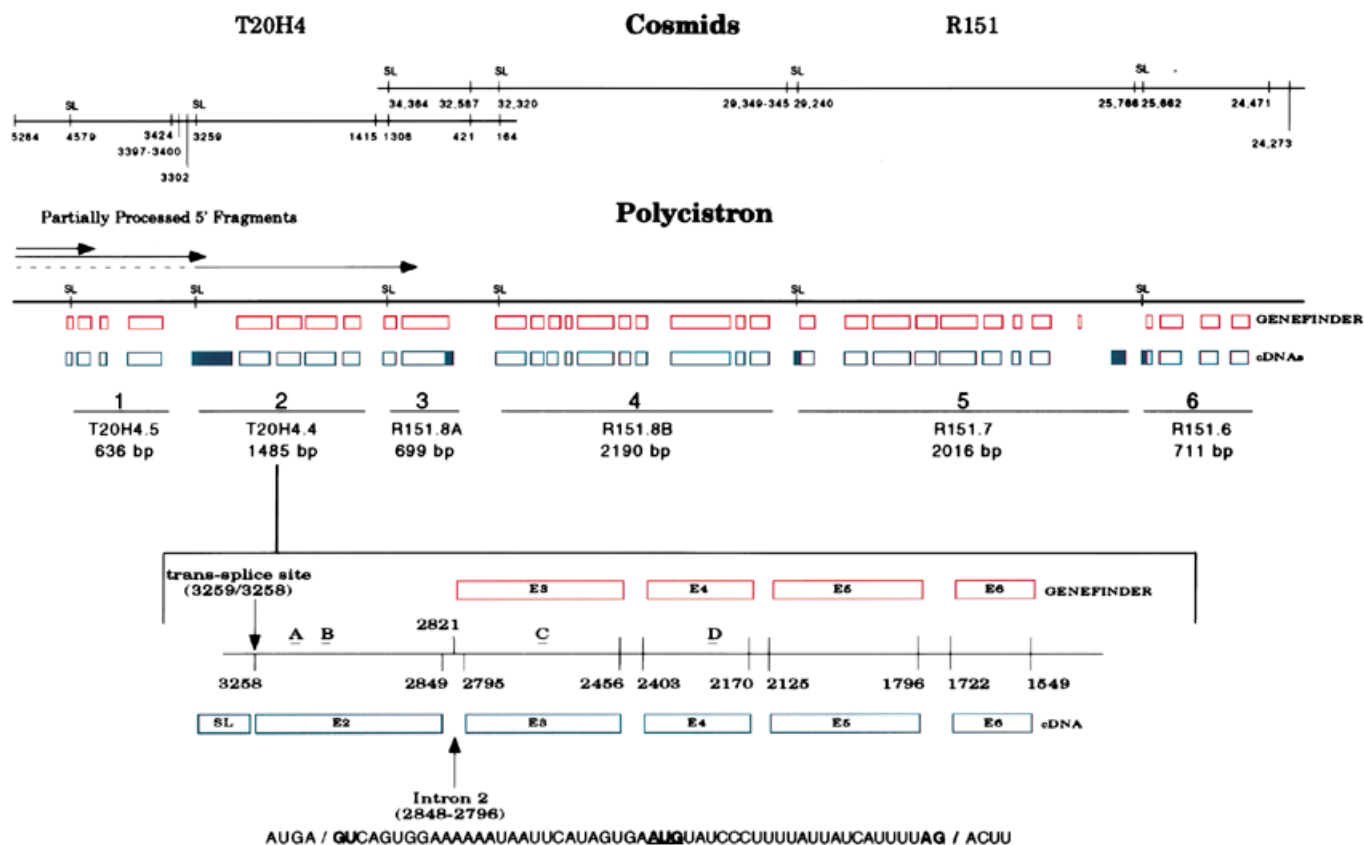


Figure 1. A six-gene operon in *C. elegans*. A cluster of genes located near the center of chromosome III is represented schematically. Six genes are encoded in the region where cosmids T20H4 and R151 overlap. Numbers indicate the nucleotide positions in the cosmid sequences that correspond to polyadenylation sites and 3' SL acceptor sites (SL). Partially processed cDNA fragments corresponding to the 5' ends of putative polycistronic transcripts are indicated by arrows. Boxes denote exons chosen by Genefinder (red) and corrections to the *cis*-splicing pattern found in the corresponding cDNAs (blue). Filled boxes (blue) denote extensions of ORFs revealed by the presence of *trans*-splice sites (SL), an additional 5' exon (T20H4.4) and a different 3' exon (R151.7). The numbered regions 1–6 are the actual ORFs using the nomenclature established by the genome sequencing project. (Inset) An enlarged comparison of the T20H4.4 cDNA sequence (GenBank accession no. AF051275) with the T20H4.4 genomic sequence (GenBank accession no. U00037). The *trans*-splicing site and newly characterized intron are indicated. Positions of four primers used in these studies for RT-PCR and cloning of cDNAs from the phage library are shown (A–D) (Table 1). The AUG (bold and underlined) within intron 2 is the translation start site selected incorrectly by Genefinder.

transcription/translation reactions were done in wheat germ extracts (TNT; Promega) in the presence of [³⁵S]methionine with linearized plasmids (1 μg pCR II) containing full-length inserts of T20H4.5 and T20H4.4.

RESULTS

Cloning of *C. elegans* T20H4.4 cDNAs

Ten T20H4.4 cDNAs were isolated from a *C. elegans* cDNA library, three of which contained full-length ORFs (see Materials and Methods and Fig. 1). All of the cDNAs coded for identical proteins and contained identical 3'-UTRs of 137 nt. The longest, 1662 nt cDNA (Fig. 2), encoded a 495 amino acid protein with a calculated mass of 55 326 Da and an isoelectric point of 8.0. The predicted protein sequence was 128 amino acids longer than that encoded by the T20H4.4 ORF, and included a dsRBM that fit well with the previously established consensus for this motif (8,9,20). An atypical AAUUA polyadenylation signal, located 13 nt upstream of the poly(A) tail, was found in all cDNAs (Fig. 2).

cDNA sequences revealed new RNA processing sites

When compared to the genomic sequence, the cDNAs revealed a *trans*-splice site and an additional *cis*-spliced intron that were not selected by Genefinder when the T20H4.4 ORF was chosen (Fig. 1). The *trans*-splice site (. . . AAUUCAG/AUGUCC . . .) was at nucleotides 3259/3258 in the cosmid sequence, which corresponds to nucleotides 11/12 in the cDNA sequence (Fig. 2). The *cis*-spliced intron was located between positions 2849 and 2795 in the cosmid sequence (Fig. 1), which correspond to the exon2/exon3 boundary at nucleotides 421/422 in the cDNA sequence (Fig. 2). The 53 nt intron released from this junction was similar in length and composition (74% A+U) to most *C. elegans* introns (14).

Identification of 10 spliced leaders at the 5' ends of T20H4.4 mRNAs

The 5' end of the longest T20H4.4 cDNA contained 11 nt (GTTTAACCAAG) that were not found in the genomic sequence (Fig. 2). This sequence is identical to the 3' end of a previously described 23 nt SL2-like spliced leader sequence

```

GTTTAAACCAAGATGTCCTCGAAGAAGGTATGGAAGTTGAGTTAAAGTCAGAGAAAATGGATTTGGATGACAATATCCCGACTTTGTGAAGGAACTGT 100
SL M S V E E G M E V E L K S E K M D L D D N I P D F V K E T V
GGAAAGAACCGGAAAGAATCCAATGTCATTATTCTCAGAGCTTTATGTTTCATATGACAGCGCAACACTCCAGTTTTTGATTCTATAACGAAACCAACCC 200
E R S G K N P M S L F S E L Y V H M T G N T P V F D F Y N R N Q P
dsRBM
AACGGATCCATGAAATTTATATGTGTGTAATTTTAAATGGAGAACGAATCGAGGGAAATGTAATAATCCAAAAAGAAAGAGGCGAAAGGTGTCGTGCTCAT 300
N G S M K F I C V V I L N G E R I E G N V K S K K K E A K V S C S
dsRBM
TGAAAGGACTGGAAGTCGTGCTCAAGCATGTTAGCGAATACGTTGTGCTGCTGTCAAATTTGAGGAGAAAACCAATTTTTCGAGATGCTCAGAGAACA 400
L K G L E V V L K H V S E Y V V P A V K F E E K T T F F E M L R E H
dsRBM*
TACATACGCGAAATTTTGAACCTTTGTAAGAACAACGCCTTAATCTATGGATTGCAAAAAGTGATCGCTTCTGTTTTCTGAAAATCAATGAAACCTC 500
T Y A K F Y E L C K N N A L I Y G F E K V I A S V F L K I N G N L
CAAATCATGCTCTCCACTGGAACAAGGGATTACGCGGAGACAAAATGTAACAGTGAACAGCTCTCATAGATTGTCATGCTGAAATTTCTCGCAA 600
Q I I A L S T G N K G L R G D K I V N D G T A L I D C H A E I L A
GAAGAGGCTTCTGAGATTTTGTATCCGAAGTCTAAAGTCTCTACAGAGCCACCAATCAATTTTTACAAAAGGAAAGAACGCATTTGGTGTGAA 700
R R G L L R F L Y S E V L K F S T E P P N S I F T K G K N A L V L K
GCCAGGAATTTCTTCCATTTTTCATAAACACCGCACCATGTGGAGTTGCGAGAATAGATAAAAAGCTAAAACCGGAACATCAGATGACCTCAAAAT 800
P G I S F H L F I N T A P C G V A R I D K K L K P G T S D D L Q N
TCTTCTCGTCTCGATCAAATGATAAAGGTATGGCCACAGTCTCGGTGGTCTTCTGAGTTTGAAGCTCCCAAAACATTTGATGGAATCATGATGG 900
S S R L R F K I D K G M G T V L G G A S E F E A P Q T F D G I M M
GAGAACGAATGAGGACAATGCTTGTCTGATAAATGCTTCGAGCAATGTACTGGGTGCAAGGAGCAATATTATCACATTTTCATCGATCCAATCTA 1000
G E R M R T M S C S D K L L R A N V L G V Q G A I L S H F I D P I Y
TTACTCGTCAATCGCGTTGCCGAGTTGAACAATGCAGATCGTCTCCGTAAGCAGTTTACAGCAGAGCAGCCACCTTCAAACCTCCAGCTCCATTCAT 1100
Y S S I A V A E L N N A D R L R K A V Y S R A A T F K P P F H
GTTCAGAGCTGGAATTCGAGAGTGTCAAGTGGAGATACTGAGCAATCAACTTCTGAGCAGCAGCATCCAGATAGTTCGATGAATTTGGAAGTGG 1200
V Q D V E I G E C Q V E D T E Q S T S A A A R S T I S S M N W N L
CTGATGGAATACAGAAGTTGTACGAACATCGGATGGAATGGTTTCATGATAAGGATATGAGCGGGGTGACATCAGACTCCATCCAGATTTGTAAAAA 1300
A D G N T E V V R T S D G M V H D K D M S G A D I T T P S R L C K K
AAATATGGCAGAATGATGATTACTATCTGCACGTTGACAAAGACTTCTGGGATTATCCAATCTCATACGAAGAAGCTCAAAGCTGGATCTCAAGAATAC 1400
N M A E L M I T I C T L T K T S V D Y P I S Y E E L K A G S Q E Y
GCTGCTCAAAAAAGTCGTTTACCTGGCTGGCAGAGAAGGATCTCGAAATTTGGCAACGAAAGCCACGAGAATTTCAAATGTTTACTATAAATTAAT 1500
A A A K K S F I T W L R Q K D L G I W Q R K P R E F Q M F T I N
CAATGCAATTTTACCTTTATGTTATCAAAAAGTTCAATGTGGTTCCACTTTTAAAAAATCTCATGCAGCATCGTGTATTCATCCATTTTACGTTAACTC 1600
TTAATTTCTGCATTAATTAATGTTAATCTTTTAAAAAAAAAAAAAAAAAAAAAAAAA
    
```

Figure 2. T20H4.4 cDNA sequence. A complete 1662 nt cDNA sequence (GenBank accession no. AF051275) is shown with the encoded amino acid sequence indicated below. The SL fragment and the single dsRBM are marked with bold underlines. The 11 nucleotides at the extreme 5' end are identical to the 3' end of the 23 nt spliced leader sequence, SLb. The newly identified exon 2-exon 3 splice site junction (A₄₂₁-A₄₂₂), denoted by an asterisk, extends the Genefinder predicted ORF by 384 nt (128 amino acids). Deoxyoligonucleotides (RACE 1-4; Table 1) used for 5' end cDNA amplification of T20H4.4 mRNAs (Table 2) are indicated in bold type.

named either SL2A (11) or SLb (13). To further characterize the spliced leader sequences on T20H4.4 mRNAs, we synthesized cDNA copies of the 5' ends by reverse transcription of poly(A)⁺ RNA, and amplified these cDNAs by two rounds of nested PCR (RACE; see Materials and Methods). Subsequent cloning and sequencing of the RACE products demonstrated that six previously described SLs with CCAAG 3' ends (SLa, SLb, SLc, SLd, SLf and SLi) were attached to T20H4.4 mRNAs (Table 2). In addition, four other SLs were identified: the classic SL1 (21) and SL2 (22) leaders plus two new variants, SLj and SLk, also related to SL2 (Table 2). The SLj variant differed from SL2D (11) by a single A→C change at position -5 and was a major species spliced to T20H4.4 (Table 2). Only a single copy of the SLk variant was identified. It deviated significantly at the 3' end as compared to the other SL2-like sequences, and was identical to the 3' end (. . . TTGAG) of SL1. The SL sequences at the 5' ends of putative SL2-like genes recently identified in cosmids r13h9 and f36h12 are identical to the SLk leader sequence (23).

Characterization of a six-gene operon

An additional exon in the coding region extended the 5' end of the T20H4.4 ORF significantly, placing it much closer to the upstream T20H4.5 ORF (Fig. 1). Combined with the presence of multiple SL2-like leaders at the 5' ends of T20H4.4 mRNAs

Table 2. *Caenorhabditis elegans* T20H4.4 mRNA SL sequences

Spliced Leader Names			Sequence	No.
(Ref 13)	(Ref 14)	(Ref 15)		
		SL1	GGTTTAAATACCC AGTT---TGAG	1
SL2	SL2	SL2	GGTTTAA--ACCC -AGTT- ACTCAAG	2
		SLa	GGTTTAT--ACCC -AGTT- AACCAAG	2
SL2A		SLb	GGTTTAA--ACCC -AGTT AACCAAG	2
SL2B		SLc	GGTTTAA--ACCC -AGTT-- ACCAAG	3
SL2E	B	SLd	GGTTTAA--ACCC -AGTT- AACCAAG	5
SL2D		SLe	GGTTTAAA-ACCC -AGTT-- AACAAG	0
SL2C		SLf	GGTTTAA--ACCC -AGTT- AACCAAG	8
		SLg	GGTTTAA--ACC --AGTT- AACTAAG	0
		SLh*	GGTTTAA--ACCC -ATAT- AACCAAG	0
		SLi*	GGTTTAA--ACCCAAGTT - AACCAAG	2
		SLj*	GGTTTAAA-ACCC -AGTT-- ACCAAG	5
		SLk*	GGTTTAA--ACCC -AGTT- AATTGAG	1

*SLh and SLi derived from the SL4 and SL5 (or related genes) respectively (15).

*The new sequence, SLj, is a variant of SLe (position -5). The new SLk sequence is identical to the 5' SL donors of two putative SL genes identified in cosmids f36h12 and r13h9 (21).

(Table 2), these data strongly suggested that the mature T20H4.4 mRNAs were produced from polycistronic transcripts initiated from a promoter upstream of the T20H4.5 gene (14). To determine if these genes and the predicted genes

Table 3. Summary of polycistronic mRNA processing sites

cDNA	Spliced Leaders			Trans-splice Junctions / and Translation Start Sites
	SL1	SL2	SL2*	
1. T20H4.5	+	+	(-)	TTTTCAG / <u>TAACCTAACCACGCTGTCAA</u> ATG CGCATGAAAAGTGTG...
2. T20H4.4	+	+	+	AATTCAG / <u>ATGTC</u> CGTCGAAGAAG GTATGGA AGTTGAGTTAAAGTCA...
3. R151.8A	(+)	+	(+)	TTTCCAG / <u>ATG</u> CCTGCATCAGT GACCC ATGTGATTTTCGATTTTGACG...
4. R151.8B	(+)	(+)	+	GTTCAG / <u>AAATG</u> CAATCAGATGAGCCATCAACGTCATCTGCCAGTTC...
5. R151.7	(-)	(+)	+	TTATTAG / <u>ATGTC</u> CGCTCCCGGTACGGATGTTAACTGGCAGGATCATG...
6. R151.6	(+)	+	+	GTTCAG / <u>ATGA</u> ATGGAGTAGTTGCTGCTCTGGAAGAA ATG CCACCG...

cDNA	Polyadenylation Signals	Cleavage Sites & Intercistronic Distances	
1. T20H4.5 (short)	TTTTCTGATAACTTCCGAATTGT / (A) _n	+ 11 nts	165 nts
T20H4.5 (major)	ATTGTAGATGAAAAACAAAATATGTTATA / (A) _n	+ 9-16nts	141 nts
T20H4.5 (major)	ATTGTAGATGAAAAACAAAATATGTTATAT / (A) _n	+ 11-18nts	140 nts
T20H4.5 (major)	ATTGTAGATGAAAAACAAAATATGTTATATAC / (A) _n	+ 13-20nts	138 nts
T20H4.5 (long)	TCTCAAGGTAAAACCTGTTACACATT / (A) _n	+ 13 nts	43 nts
2. T20H4.4	TGCATTAATTAATTGTTAATCTTTT / (A) _n	+ 13 nts	107 nts
3. R151.8A	TTTTTGAATATATTTTTGCAACAGTTT / (A) _n	+ 15 nts	257 nts
4. R151.8B	TTGCTTCATAAAAATTTTTTATACTG / (A) _n	+ 14 nts	109 nts
R151.8B	TTGCTTCATAAAAATTTTTTATACTGTTTT / (A) _n	+ 18 nts	105 nts
5. R151.7	TTTTCCCATACATTTCGTTGATTTTT / (A) _n	+ 13 nts	104 nts
6. R151.6 (short)	TTACCTAATTAATTATTTTCTCTGCTCTT / (A) _n	+ 16 nts	NA
R151.6 (long)	TCGTTTAGTAAAGCTCATTAACT / (A) _n	+ 11 nts	NA

(±) PCR only, unconfirmed by sequencing.

Underlined regions designate 5' end deoxyoligonucleotide primers; bold in-frame ATG codons, putative polyadenylation signals

downstream of T20H4.4 are co-expressed from an operon, we first characterized the 5' and 3' ends of cDNAs corresponding to the individual mRNAs, then cloned six full-length cDNAs (see Materials and Methods). The results of these experiments are summarized in Figure 1 and Table 3.

A total of six ORFs were identified, rather than the five predicted by Genefinder, since two regions of the R151.8 locus were expressed as separate mature mRNAs (Fig. 1; R151.8A and R151.8B). We were unable to detect a single R151.8 mRNA by RT-PCR across the putative exon 2–exon 3 boundary predicted by Genefinder. In addition, two classes of polyadenylated cDNAs corresponding to R151.8A and R151.8B were isolated (Table 3).

cDNAs derived from all five downstream genes contained SL2 or SL2-like leader sequences, as shown by the isolation of PCR products using SL1, SL2 or SL2* primers (Table 1) in the second round of RACE amplification (Table 3; Fig. 3). With the exception of R151.7, the four other downstream cDNAs also contained SL1 to a lesser extent. The presence of either SL2 or SL2-like leaders was confirmed by sequencing a sufficient number of RACE clones, amplified with Q₁ (Table 1) instead of SL primers, to obtain full-length sequences with complete leaders. The following spliced-leaders were confirmed for other mRNAs downstream of T20H4.4: SL2 (R151.8A); SLd (R151.8B); SLb (R151.7); SL2 and SLi spliced to R151.6 mRNAs.

PCR products corresponding to the 5' end of T20H4.5 contained SL1 primarily, but also a small fraction of SL2 (Fig. 3). The SL1:SL2 ratio was estimated to be ~5, based upon the

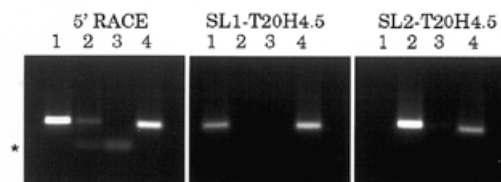


Figure 3. SL1 and SL2 leaders attached to T20H4.5 mRNAs. 5' RACE: the 5' ends of T20H4.5 mRNAs were amplified using 2 µl of a 1:20 dilution of the first round PCR reaction in 100 µl second round reactions (see Materials and Methods). 5' primers were specific to: lane 1, SL1 (328 bp); lane 2, SL2 (328 bp); lane 3, SL2* (130 bp); lane 4, T20H4.5–5' end primer (306 bp). Aliquots (20 µl) of the reactions were resolved by electrophoresis in 1.5% agarose and stained with ethidium bromide. Control reactions using cDNAs verified the SL1 and SL2 primers were specific in reactions that contained ~2 × 10⁷ copies of the 328 bp SL1-T20H4.5 and SL2-T20H4.5 cloned inserts in pCRII (Invitrogen). After a 72°C hot start, thermocycling (Perkin Elmer 480) was carried out for 30 cycles of 94°C for 1 min, 53°C for 1 min, 72°C for 1 min. The asterisk denotes an artifact that arose by template switching (40) during the reaction with SL2*, emphasizing the need to confirm the identity of PCR products by sequencing. The 5' end of this PCR product was identical to 92 nt of the K01G5.8 ORF (accession no. Z92803; nucleotides 130–222, cosmid K01G5) and the 3' end derived from 47 nt of T20H4.5 (GenBank accession no. U00037; nucleotides 4151–4198, cosmid T20H4). The region of overlap consisted of nine identical nucleotides.

relative fluorescence in each PCR product under the conditions described (see Materials and Methods). Similar results were obtained over a range of annealing temperatures, 53–60°C, but

the estimate of one SL2 and five SL1 leaders should be viewed as a qualitative estimate. Most of the 5'-T20H4.5 RACE clones were incomplete, ending prior to the SL sequence (48 of 50). However, two clones contained SL1 *trans*-spliced to a classic 3' splice junction (. . . TTTTCAG/; Table 3). To verify that SL2 was also spliced to T20H4.5 mRNAs, RACE cDNAs amplified during the second round with SL1 and SL2 (Fig. 3) were cloned and sequenced. Both SL1 and SL2 were found at the splice junction (Fig. 1; position 4579 of cosmid T20H4), 22 nt upstream of the putative AUG start codon (Table 3). The SL2 products were not generated by mispriming on SL1 in a region where nine of 11 nucleotides at the 3' end of SL2 match the SL1 sequence [Table 1; 5'- . . . A(G/A)TTAC(T/C)CAAG-3'], since the predicted 3' end of SL1 (. . . TTTGAG-3') was not found 3' of the SL2 sequence. Control experiments demonstrated that SL1 and SL2 specific PCR products were obtained from the corresponding SL1-T20H4.5 and SL2-T20H4.5 cloned cDNAs (Fig. 3). We conclude that T20H4.5 mRNAs are *trans*-spliced predominantly with SL1, but a small fraction contain SL2.

Although *trans*-splicing produces trimethylguanosine-capped mature mRNAs from polycistronic precursors, the precise function of the 21–23 nt leaders is not known (reviewed in 14 and 24). *Trans*-splice sites frequently occur adjacent to, or near, the AUG start codon. This suggests SLs have a direct role in translation initiation by removing upstream AUG codons, or by providing an optimal initiation context (14,25,26). Indeed, affixing SLs adjacent to the first AUG codon of four downstream mRNAs (T20H4.4, R151.8A, R151.7 and R151.6) improves translational context (27) by replacing the genomically encoded pyrimidines with purines at the –3 position (Table 3). The –3 position of T20H4.5 and R151.8B AUG codons was unchanged after *trans*-splicing. Consequently, the first in-frame AUG codons of all six mRNAs are the likely translation initiation sites (Table 3).

All of the polyadenylation signals identified in the cDNAs deviated from the canonical mammalian AAUAAA hexamer (Table 3), consistent with previous surveys which revealed >37% of *C.elegans* cDNAs contained degenerate cleavage and polyadenylation signals (14). The intercistronic distances also fit the bimodal distribution described previously for other *C.elegans* operons (14). We found no evidence for cleavage

and polyadenylation occurring at, or very close to, the downstream *trans*-splice sites as observed in some operons (14,28).

Polycistronic RNA fragments

Mature mRNAs derived from the first gene in more than 30 operons surveyed thus far contain either a 5'-UTR, or the SL1 leader exclusively (14). Our finding of SL2 attached to T20H4.5 mRNAs raised the possibility of an upstream ORF undetected by Genefinder. The next predicted ORF is located ~4.8 kb upstream of T20H4.5 and is oriented in the opposite direction. Despite an extensive screen covering ~4.6 kb of the genomic sequence upstream, no polyadenylated cDNAs, or cDNAs corresponding to another ORF were isolated. In contrast, partially processed RNA fragments, containing some introns but not others, were isolated and found to contain 705 nt of a 5'-UTR extending into the 5' end of the T20H4.4 ORF (Fig. 1 and Table 4). In addition, partially spliced RNAs containing the entire T20H4.4 ORF and extending beyond the first *cis*-splice junction of R151.8A were recovered. Since these cDNAs were partially processed they could not have derived from contaminating genomic DNA (see Materials and Methods). These data, together with the presence of SL2, or SL2-like leaders attached to all downstream mRNAs (Tables 2 and 3), strongly suggest that the mature mRNAs are produced by *trans*-splicing of polycistronic precursor RNAs, beginning with T20H4.5. With the exception of a single clone, the cDNAs suggest there is a bias (3' to 5') to the apparent order of *cis*-splicing (Table 4).

Northern analyses and *in vitro* translation

Northern analyses were performed to determine the size of mRNAs expressed from the first two genes of the operon. Radiolabeled probes specific to each ORF hybridized to single RNAs corresponding to ~0.9 and ~1.7 kb for T20H4.5 and T20H4.4, respectively (Fig. 4A), consistent with the predicted sizes for *trans*-spliced mRNAs from our cDNA analyses. The fact that only a single RNA (~0.9 kb) hybridized with the T20H4.5 ORF probes, and no RNAs were detected with probes specific for the 5'-UTR (data not shown), indicated that the mature T20H4.5 mRNAs were *trans*-spliced and did not contain a 5'-UTR. Thus, RNAs containing the 5'-UTR (Fig. 1;

Table 4. Polycistronic cDNA fragments

5' Transcription -----> 3'												
Gene	1 T20H4.5				2 T20H4.4					3 R151.8A		No. Clones
	Trans-1	Cis-2	Cis-3	Cis-4	Trans-1	Cis-2	Cis-3	Cis-4	Cis-5	Trans-1	Cis-2	
—	+	+										2
—	—	+										2
—	—	+	+	—								1
—	—	—	—	—								3
						+	+	+	+	—	+	3
						+	—	+	+	—	+	1*
						—	—	+	+	—	+	1
						—	—	—	—	—	+	2
						—	—	—	—	—	—	3

+, *cis*-splicing completed; —, unspliced (introns present); *, 5'←3' processing bias (except for single clone indicated).

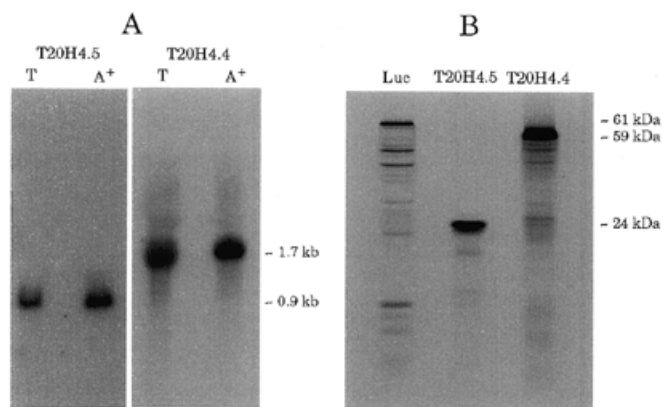


Figure 4. Northern analysis and *in vitro* translation of T20H4.5 and T20H4.4. (A) Blots containing *C.elegans* total RNA (T; 50 μ g) and poly(A)⁺ RNA (A⁺; 10 μ g) were hybridized with a mixture of random-primed ³²P-labeled deoxy-oligonucleotides specific for either T20H4.5 or T20H4.4 (see Materials and Methods). The single labeled RNAs, ~0.9 and ~1.7 kb, agree with the predicted sizes of the *trans*-spliced cDNAs. (B) Equal portions of the ³⁵S-labeled translation mixtures (see Materials and Methods) were resolved by 13% SDS-PAGE, followed by autoradiography. The major protein products migrated with apparent molecular masses of ~24 kDa (T20H4.5) and ~59 kDa (T20H4.4) relative to the migration of marker proteins not shown (NEB broad range protein markers). The positive control lane (Luc) contained luciferase (~61 kDa).

Table 4) were minor species, revealed only by sensitive RT-PCR amplification methods.

To verify that the 636 nt T20H4.5 ORF and 1485 nt T20H4.4 ORF within our cDNAs (Fig. 1) coded for the predicted proteins, RNAs were transcribed from the full-length cDNAs and translated in wheat germ extracts. Proteins were labeled by including [³⁵S]methionine in the reaction, and the labeled products were separated by SDS-PAGE (Fig. 4B). Migration of the ~24 kDa radiolabeled T20H4.5 polypeptide corresponded to the predicted size of 23.8 kDa, whereas the apparent mass (~59 kDa) of the radiolabeled polypeptide in the T20H4.4 lane was slightly larger than the native size of 55.3 kDa predicted by the cDNAs.

DISCUSSION

Spliced leader sequences

In 1987, Krause and Hirsch identified the first *C.elegans* spliced leader, SL1, at the 5' ends of actin mRNAs (21,29). One year later, a different sequence, SL2, was identified on mRNAs derived from a glyceraldehyde-3-phosphate dehydrogenase gene (22). Subsequently, a third class of RNAs, the SL2-like leaders, were found attached to mRNAs for *tra-2* (11), protein kinase C1A (12), and the β subunit of casein kinase II (13,30). Here we show that SL2-like leaders are also attached to T20H4.4 transcripts. In previous reports, identical SL2-like leaders were given different names (Table 2). Using the nomenclature of Ross *et al.* (13) we have updated the list of SL2-like leaders to include *trans*-spliced RNAs predicted from the SL4 and SL5 genes (SLh and SLi respectively), and the two

new variants, SLj and SLk, identified among the T20H4.4 mRNAs (Table 2).

Genes for SL1, SL2 and the SL2-like leaders have been identified. The SL1 genes are tandemly repeated ~100 times on chromosome V (29,31), whereas multiple copies of the SL2 and SL2-like genes are distributed throughout the genome (13,22,28). The transcripts encoded by the spliced leader RNA genes are 95 nt (SL1) or 100–114 nt (SL2 and SL2-like) in length and contain the 21–23 nt SL donor segments at their 5' ends.

Recent estimates suggest the majority of nematode mRNAs (>70%) contain spliced leaders (27,32). In *C.elegans* it is estimated that ~57% of the mRNAs are *trans*-spliced to SL1, ~16% are non-SL1, and ~30% may not contain spliced leaders (32). For the non-SL1 sequences, Ross *et al.* (13) estimate that ~43% derive from SL2 RNAs and ~57% from the SL2-like RNAs (i.e., SL3, SL4 and SL5 genes; 13). Hence, ~10% of *C.elegans* mRNAs may contain SL2-like leaders, and we estimate that \geq 90% of the SLs appended to T20H4.4 mRNAs are of this minor class (Table 2).

T20H4.4 is the second gene in a six-gene operon

Our results showing SL2, or SL2-like leaders on cDNAs derived from T20H4.4 (Table 2) and four other downstream genes (Table 3; R151.8A; R151.8B; R151.7; R151.6), indicate that the mature, SL2-spliced mRNAs, are processed from polycistronic precursors (26,28,32). The partially processed cDNA fragments isolated from the gene cluster we describe here (Fig. 1; Table 4) provide further evidence that these genes are co-transcribed, and augment general information about *C.elegans* operons.

SL2 *trans*-splicing of an mRNA is linked to cleavage and polyadenylation of the mRNA immediately upstream (14,28,33). SL1-splicing at SL2 *trans*-splice sites is commonly observed and may reflect either infrequent transcription initiation at the downstream gene, or an alteration of cleavage and polyadenylation of the upstream mRNA that somehow favors SL1 splicing of the downstream mRNA (14,33). For genes of some operons, the polyadenylation site and the downstream *trans*-splice site are identical, or separated by only a few nucleotides. In these situations, a different mechanism is probably involved that results in splicing of only SL1 to the downstream mRNAs (14,34).

Identification of SL2 spliced mRNAs from the first gene of an operon, as we observed here, is unprecedented. Typically the 5' ends of mature mRNAs from the first gene in an operon contain either a 5'-UTR or SL1 exclusively (35–37). However, SL2 can substitute for SL1 and rescue embryonic lethal *rrs-1* worms lacking SL1-RNAs, suggesting that discrimination involves competition between SL1- and SL2-RNAs (38). Possibly the T20H4.5 *trans*-splice site favors SL2-RNA selection more frequently than observed in other operons.

Although the formal possibility remains that another ORF may exist upstream of T20H4.5, our data strongly suggest T20H4.5 is the first gene of this operon. Northern analyses confirmed that the mRNA population from the first gene is *trans*-spliced and does not contain the 705 nt 5'-UTR (Fig. 4) identified in the precursor polycistronic RNA fragments (Table 4). Based on the length of the 5'-UTR, the promoter for transcription initiation of polycistronic RNA probably resides

Table 5. Proteins encoded by a six-gene operon in *C.elegans*

Gene	AAs	Mr	pI	Function*
1. T20H4.5	212	23,872	5.8	Mitochondrial respiratory chain (NADH complex I)
2. T20H4.4	495	55,326	8.0	Adenosine deaminase that acts on RNA (ADAR)?
3. R151.8A	233	26,609	5.4	GS1-like protein, unknown function (phosphatase?)
4. R151.8B	730	81,668	9.3	Unknown, serine/threonine rich protein
5. R151.7	671	76,471	7.9	Heat shock protein (<i>hsp 75</i>), <i>hsp 90</i> family
6. R151.6	237	28,036	5.2	Mitochondrial respiratory chain (NADH complex I)?

*BLASTP 2.0.4 or TBLASTN 2.0.8 (41) most significant similarities (GenBank accession nos U00036 and U00037).

```

..... A...L.....A...S.....D.....GT.L...A..ILARR..L.....E....ST.... Consensus
141 NNAL IYGFKEKVI ASVF K I N G N L Q - I I A L - S T G N K G L R G K I V N - D G T A I I D C H A E I L A R R G L R F L Y S I V L K F S I - E P P T20H4.4
638 TEYYGI PPSNLF A A F L V N A Q E E K R V L M G S I Q I Y I V E P I T L S G A N G I S L H L D A I I L A R R A M K A F I H L - - - S T V D S E H15N14.1
. S I F . K . . . . K . A L . . . . K P . . . . L . . . . N . . . . P C . . . . A . . . . D . . . . K K L . . . . T . . . . L . . . . . . . . . . V L . . . . . . . . . . Consensus
217 N S T E I K G - - - N A I V L K P G I S F H F I L - T A P C G V A R I D - - - K K L K P G I S D D I O N S S R L R F K I D K G M G T I L G G A S E F E A P O T20H4.4
715 C S I E K K E E G A A L - - - P N L R L V Y S Y S P P C I H M V D D A A T K K L S Y V I P T N I T C V P - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - H15N14.1
T . . . . I . . . . R . . . . D K L . . . . N . L G . G G A . L S . . . . P I . . . . I . . . . E . . . . D . . . . L . . . . A . . . . R . . . . . . . . . . Consensus
290 I F D G I M M G E R M T M S C S D K I L R A N V M G V I G A I I S H F I D P I Y Y S S I A V A - I L N N A D R - I R K A V Y S A A T F K P P A P F H V Q D V T20H4.4
773 I Y E Q I K E T K S L S V H C T A D K I F K W N I L G I D G A I I S N V L H P I I F I D N I F F G S I A P V S I E S I S Y A L G G R L G - - - - - - - - - - - - - - - - - - - - - - - - - H15N14.1
E . . . . E . . . . V E . . . . . . . . . . . . . . . . S . . . . W . . . . . . . . . . E . . . . . . . . . . G . . . . . . . . . . K . . . . . . . . . . P S R . C K . . . . E . . . . . . . . . . Consensus
368 I I G C O F E D T E Q S T S A A A R S T I S M N N L A D G N T I V V R T S D I M V H D I D M S G A D I T T P S R L K K N M A L M I T I C T L T K T S V T20H4.4
844 R - S I I I F S M P V Q M R M H M G I - - - S H L H R G V D S V I T L D Y N T - R T S K G - - - - - - - - - - - - - - - - - - - - - - - - - S P S R V C K A E I F I A Y R K L N G V D Q A - - H15N14.1
. . . . . Y . . . . K . . . . E Y . . . . K K . F . . . . L . . . . L G . W Q . K P . E . . . . F T . . . . . . . . . . Consensus
448 D Y P I S T E E L A G S O I A A A K S I T W I R O K D C I W C R K P F F Q M I N T20H4.4
910 - - V V N Y A K A E M A S F Q Y E K K V I Y E K I E A A G G K W G T K P A L V D S E I L A A F D H15N14.1

```

Figure 5. Alignment of *C.elegans* T20H4.4 and H15N14.1. The predicted C-terminal amino acids 141–495 of T20H4.4 (Fig. 2; GenBank accession no. AF051275) were aligned with the predicted C-terminal amino acids 638–959 of H15N14.1 (accession no. Z96100) as described previously (2). The consensus of the sequences is shown; H15N14.1 contains 32 (bold type) of the 58 identical amino acids shown previously to be conserved in the C-termini of ADARs and putative ADARs, including T20H4.4 (2).

>705 nt upstream of T20H4.5 (Fig. 1; nucleotide 5284, cosmid T20H4).

Common functional roles for six co-expressed proteins?

There are unambiguous examples of co-expressed gene products that function together in *C.elegans*, but the driving force for clustering genes in many other operons is not clear (14,32). For some *C.elegans* operons, identical gene arrangements have been found in *Caenorhabditis briggsae* (14). Recently, SL2 splicing and gene clusters similar to the clusters found in *C.elegans*, were identified in a distantly related nematode in the genus *Dolichorhabditis*, suggesting that the arrangement of operons may have been determined early in evolution (23).

cDNAs corresponding to the six genes of the operon described here (Fig. 1) encode the proteins listed in Table 5. Co-expression of five genes downstream of T20H4.5, which encodes a mitochondrial protein, raises the possibility that the other encoded proteins also are involved with mitochondrial function in some way (e.g., respiratory chain subunits, protein chaperones, protein modification and RNA editing; Table 5). Conceivably, the ADAR-like T20H4.4 is a mitochondrial

RNA editing enzyme. Alternatively, the genes of the operon may all have ‘housekeeping’ functions, and are clustered together because transcription of all genes from the same promoter is more economical (14).

T20H4.4 is related to ADARs

While *in vitro* studies clearly demonstrate the presence of ADAR activity in *C.elegans* extracts (M.Krause, B.Bass and D.Morse, unpublished data), the gene products responsible for the observed activity have not been characterized. With the completion of the *C.elegans* genome project, the best ADAR candidates are T20H4.4 and the H15N14.1 ORF (Fig. 5). Like previously characterized ADARs, H15N14.1 contains multiple dsRBMs (5). Possibly proteins encoded by both T20H4.4 and H15N14.1 ORFs contribute to the ADAR activity detected in worm extracts.

H15N14.1 mRNAs are transcribed from a single gene and are *trans*-spliced with SL1 only (L.Tonkin, R.Hough and B.Bass, unpublished data). We have shown here that T20H4.4 mRNAs are produced by cleavage and SL2 *trans*-splicing of precursor polycistronic RNAs. These data provide the

framework and essential tools necessary to characterize worms defective at the T20H4.4 locus, since expression of four downstream mRNAs may be affected by alteration of the upstream gene sequence. Concurrent with the recent discovery of several ADAR substrates in *C.elegans* (39), experiments aimed at determining the consequences of A→I editing, and the functions of T20H4.4 and H15N14.1 in *C.elegans* are now underway.

ACKNOWLEDGEMENTS

We thank D. Morse and L. Tonkin for helpful discussions and assistance, M. Robertson, E. Lawrence and J. Barrow for their support with DNA sequencing (University of Utah Health Sciences Center Sequencing Facility supported by the National Cancer Institute Grant 5P30CA42014) and E. Meenen for synthesis of deoxyoligonucleotides (Howard Hughes Medical Institute Synthesis Facility supported by Department of Energy Grant DE-FG03-94ER61817). This work was supported by funds to B.L.B. from the National Institute of General Medical Sciences, National Institutes of Health (GM 44073). B.L.B. is a Howard Hughes Medical Institute Associate Investigator.

REFERENCES

- Kim, U., Wang, Y., Sanford, T., Zeng, Y. and Nishikura, K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 11457–11461.
- Hough, R.F. and Bass, B.L. (1997) *RNA*, **3**, 356–370.
- Bass, B.L. and Weintraub, H. (1988) *Cell*, **55**, 1089–1098.
- Wagner, R.W., Smith, J.E., Cooperman, B.S. and Nishikura, K. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2647–2651.
- Bass, B.L. (1997) *Trends Biochem. Sci.*, **22**, 157–162.
- Lai, F., Drakas, R. and Nishikura, K. (1995) *J. Biol. Chem.*, **270**, 17098–17105.
- Maas, S., Melcher, T., Herb, A., Seeburg, P.H., Keller, W., Krause, S., Higuchi, M. and O'Connell, M.A. (1996) *J. Biol. Chem.*, **271**, 12221–12226.
- St Johnston, D., Brown, N.H., Gall, J.G. and Jantsch, M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 10979–10983.
- Bass, B.L., Hurst, S.R. and Singer, J.D. (1994) *Curr. Biol.*, **4**, 301–314.
- Wilson, R. *et al.* (1994) *Nature*, **368**, 32–38.
- Kuwabara, P.E., Okkema, P.G. and Kimble, J. (1992) *Mol. Biol. Cell*, **3**, 461–473.
- Land, M., Islas-Trejo, A. and Rubin, C.S. (1994) *J. Biol. Chem.*, **269**, 14820–14827.
- Ross, L.H., Freedman, J.H. and Rubin, C.S. (1995) *J. Biol. Chem.*, **270**, 22066–22075.
- Blumenthal, T. and Steward, K. (1997) In Riddle, D.L., Blumenthal, T., Meyer, B.J. and Priess, J.R. (eds), *C.elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY, pp. 117–145.
- Caldicott, I.M., Larsen, P.L. and Riddle, D.L. (1994) In Celis, J.E. (ed.), *Cell Biology. A Laboratory Handbook*. Academic Press, San Diego, CA, pp. 389–397.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (eds) (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY.
- Frohman, M.A. (1994) In Mullis, K.B., Ferre, F. and Gibbs, R.A. (eds), *The Polymerase Chain Reaction*. Birkhauser, Boston, MA, pp. 14–37.
- Kohara, Y. (1997) *Tanpakushitsu Kakusan Koso*, **42**, 2907–2913 (www.ddbj.nig.ac.jp).
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. Green Publishing Associates/Wiley-Interscience, New York, NY.
- Kharrat, A., Macias, M.J., Gibson, T.J., Nilges, M. and Pastore, A. (1995) *EMBO J.*, **14**, 3572–3584.
- Krause, M. and Hirsch, D. (1987) *Cell*, **49**, 753–761.
- Huang, X.-Y. and Hirsch, D. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8640–8644.
- Evans, D., Zorio, D., MacMorris, M., Winter, C.E., Lea, K. and Blumenthal, T. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9751–9756.
- Blaxter, M. and Liu, L. (1996) *Int. J. Parasitol.*, **26**, 1025–1033.
- Blumenthal, T. (1995) *Trends Genet.*, **11**, 132–136.
- Maroney, P.A., Denker, J.A., Darzynkiewicz, E., Laneve, R. and Nilsen, T.W. (1995) *RNA*, **1**, 714–723.
- Kozak, M. (1991) *J. Cell. Biol.*, **115**, 887–903.
- Spieth, J., Brooke, G., Kirsten, S., Lea, K. and Blumenthal, T. (1993) *Cell*, **73**, 521–532.
- Bektesh, S., van Doren, K. and Hirsch, D. (1988) *Genes Dev.*, **2**, 1277–1283.
- Hu, E. and Rubin, C.S. (1991) *J. Biol. Chem.*, **266**, 19796–19802.
- Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990) *Nucleic Acids Res.*, **18**, 2633–2642.
- Zorio, D.A., Cheng, N.N., Blumenthal, T. and Spieth, J. (1994) *Nature*, **372**, 270–272.
- Kuersten, S., Lea, K., MacMorris, M., Spieth, J. and Blumenthal, T. (1997) *RNA*, **3**, 269–278.
- Williams, C., Xu, L. and Blumenthal, T. (1999) *Mol. Cell. Biol.*, **19**, 376–383.
- Conrad, R., Thomas, J., Spieth, J. and Blumenthal, T. (1991) *Mol. Cell. Biol.*, **11**, 1921–1926.
- Conrad, R., Liou, R.F. and Blumenthal, T. (1993) *EMBO J.*, **12**, 1249–1255.
- Conrad, R., Lea, K. and Blumenthal, T. (1995) *RNA*, **1**, 164–170.
- Ferguson, K.C., Heid, P.J. and Rothman, J.H. (1996) *Genes Dev.*, **10**, 1543–1556.
- Morse, D.P. and Bass, B.L. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 6048–6053.
- Odelberg, S.J., Weiss, R.B., Hata, A. and White, R. (1995) *Nucleic Acids Res.*, **23**, 2049–2057.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Miller, W. and Lipman, D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.