Published online 9 September 2009

Nucleic Acids Research, 2009, Vol. 37, No. 20 ***6723–6736***
doi:10.1093/nar/gkp744

Regulation of transcription termination in the nematode *Caenorhabditis elegans*

Simon Haenni, Helen E. Sharpe, Maria Gravato Nobre, Kerstin Zechner, Cathy Browne, Jonathan Hodgkin and André Furger*

Genetics Unit Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Received June 23, 2009; Revised August 20, 2009; Accepted August 24, 2009

ABSTRACT

The current predicted mechanisms that describe RNA polymerase II (pol II) transcription termination downstream of protein expressing genes fail to adequately explain, how premature termination is prevented in eukaryotes that possess operon-like structures. Here we address this issue by analysing transcription termination at the end of single protein expressing genes and genes located within operons in the nematode *Caenorhabditis elegans*. By using a combination of RT-PCR and ChIP analysis we found that pol II generally transcribes up to 1 kb past the poly(A) sites into the 3′ flanking regions of the nematode genes before it terminates. We also show that pol II does not terminate after transcription of internal poly(A) sites in operons. We provide experimental evidence that five randomly chosen *C. elegans* operons are transcribed as polycistronic pre-mRNAs. Furthermore, we show that *cis*-splicing of the first intron located in downstream positioned genes in these polycistronic pre-mRNAs is critical for their expression and may play a role in preventing premature pol II transcription termination.

INTRODUCTION

Gene expression requires that the information stored in the DNA is transcribed into an RNA copy by the transcription machinery. The transcription cycle begins with the recognition of promoter sequences and the subsequent assembly of the transcription machinery, which then elongates, and transcribes protein encoding genes into a pre-mRNA molecule. The cycle concludes at the end of the transcription unit, when the large protein complex disengages from the DNA template and its components are recycled for a new round of transcription. The impact of transcription initiation and elongation on gene expression is obvious; the importance of transcription termination however, has long been understated. Termination of RNA polymerase at the end of a transcription unit is a critical process to avoid interference with downstream positioned genes (1). This may be particularly important in organisms with compact genomes such as *Caenorhabditis elegans*. Furthermore, termination of the transcription machinery may also be critical to enable efficient recycling of the polymerase.

The molecular mechanisms that direct termination at the end of protein expressing genes are still only poorly understood. The mapping of termination sites revealed that pol II transcription proceeds up to 700 nucleotides beyond the poly(A) site in the fly *Chironomus tentans* (2) and more than 1.5 kb in human genes (3,4). This is in strong contrast to *Saccharomyces cerevisiae*, where termination is proposed in close proximity of the actual poly(A) cleavage site (5–7). Two major models addressing the mechanism of termination have been proposed in the late 1980s and experimental data supporting both models has since been presented.

The anti-terminator model/allosteric model introduced by Logan (8) describes transcription termination as an event that is triggered by the dissociation of 'anti-termination' factors from the polymerase once a functional poly(A) site is recognized (9).

The torpedo model emphasises that the cleavage of the pre-mRNA at the poly(A) site generates an unprotected 5′-end. This allows a 5′–3′ exoribonuclease to access and subsequently degrade the remaining transcript attached to pol II, which somehow results in termination (10,11).

Since strong support for both models has been found over the years it is highly likely that *in vivo* both mechanisms are interlinked and contribute to the termination process in eukaryotes (9,12–15).

A number of eukaryotes including nematodes, such as *C. elegans* and protozoan parasites, for example Trypanosomes, have the ability to transcribe some of their genes (15% in *C. elegans*) from a single promoter

*To whom correspondence should be addressed. Tel: 01865 613261; Fax: 01865 613276; Email: andre.furger@bioch.ox.ac.uk

into polycistronic pre-mRNAs, similar to bacterial operons (16). These polycistronic primary transcripts have to undergo complex processing reactions in order to generate individual mature mRNAs. The nematode *C. elegans* has been, since the first discovery of operon-like structures in 1993 (17), an important organism for the study of operon expression in multi-cellular eukaryotes. The intergenic regions within nematode operons contain crucial *cis*-elements for the maturation of these individual transcripts. These elements direct two linked reactions that involve cleavage and subsequent polyadenylation at the 3′-end of the upstream transcript, and *trans*-splicing of a small leader sequence from a splice leader snRNP (SL snRNP) onto the 5′-end of the downstream transcript (17). *Trans*-splicing is dependent on a spliceosome containing the same U snRNPs that direct *cis*-splicing with the exception that the U1 snRNP is not believed to be part of the *trans*-splicing machinery (18). In *C. elegans*, monocistronic *trans*-spliced genes and the first gene in operons are *trans*-spliced by SL1 snRNP. In contrast, the second major splice leader, SL2 snRNA, is thought to be *trans*-spliced exclusively to downstream positioned genes in operons. Interestingly, the poly(A) cleavage at the 3′-end of the upstream positioned gene and *trans*-splicing of SL2 are connected. This is evidenced both by mutational analysis (19–22) and by the discovery of a complex containing the SL2 snRNP and the essential poly(A) cleavage stimulation factor subunit *Ce*CstF-64 (16,23).

As outlined above, the polycistronic precursors generated in the nematode are subjected to cleavage and polyadenylation at the end of each individual gene and *trans*-splicing of a short leader sequence at the beginning of each downstream positioned gene. Given the close relationship between the transcription of a functional poly(A) site and subsequent transcription termination, it is not known how multiple functional poly(A) sites can be transcribed by pol II without triggering premature termination. Hence, in order to complete the transcription of all the genes present in a given operon pol II must be prevented from terminating prematurely. This is likely to involve *trans*-splicing in order to prevent exoribonuclease access as suggested (22) and mechanisms that prevent remodelling of the polymerase after recognition of the poly(A) site.

We employed RT-PCR and ChIP protocols that allowed us to investigate operon expression and transcription termination using mixed stage hermaphrodites. We show that pol II at the end of three genes terminates around 1 kb downstream of the poly(A) site. We demonstrate that premature transcription termination is prevented downstream of internal poly(A) sites in operons. In addition, we present evidence that five randomly chosen predicted operons in *C. elegans* are truly transcribed as polycistronic precursors and that splicing of introns located in the downstream positioned genes, can occur prior to both poly(A) cleavage and *trans*-splicing at the upstream intergenic regions. Moreover, we demonstrate that *cis*-splicing of the first intron at a downstream positioned gene is critical for its expression levels and may be involved in preventing premature transcription termination.

MATERIALS AND METHODS

RNA analysis

Isolation of total RNA. All work was carried out using the *C. elegans* strain Bristol N2. Nematodes were grown in liquid cultures of *Escherichia coli* HB101. After three days in culture, nematodes were harvested and washed followed by sucrose flotation (24). Total RNA was isolated using the hot-phenol method (25).

Reverse transcriptase polymerase chain reaction (RT–PCR). Four to eight micrograms of total RNA was used as template for reverse transcription with SuperScript^{TM} III (Invitrogen). The cDNA was amplified by Taq DNA polymerase (28 cycles of PCR). A 'no RT' control was always included. Individual bands were purified, re-amplified and sequenced. Radioactive PCRs were performed with one-tenth of the cDNAs for 22–25 cycles, separated by PAGE and quantitated using a PhosphorImager. For rRNA analysis 1 ng of template was used and the cDNA was further diluted 1:10.

Chromatin immunoprecipitation (ChIP) and real-time PCR

Mixed stage Bristol (N2) worms from a liquid culture or from two large (14 cm Ø) worm egg plates were harvested in M9. The worms were cleared from bacteria by washing in M9 and by sucrose floatation if required. Digestion of intestinal bacteria was performed by 1 hr incubation in M9 at 20°C. The subsequent ChIP procedure was essentially performed as described in (26): 500 µl worm slurry was fixed for 30 min at 20°C in 45 ml buffer containing 1% formaldehyde. The reaction was blocked with 125 mM glycine for 5 min at room temperature. The fixed worms were washed three times in PBS and the final pellet resuspended in 500 µl FA 150 buffer [0.1% SDS, 1% Triton X-100, 10 mM Hepes pH 7.0, 0.1% deoxycholate, 150 mM NaCl, Complete Mini protease inhibitors (Roche Biochemicals)]. The worms were homogenized using acid washed glass beads (SIGMA) and a MagNA Lyser instrument (Roche Applied Science; two runs for 1 min at 4°C with 6000 rpm). The homogenate was sonicated in a Bioruptor TOS-UCD-200TM-EX (Cosmo Bio) twice for 15 min at 4°C with 1 min on per 20 s off cycles (medium power: 160 W ultrasonic wave output). After centrifugation for 15 min at 12 000g at 4°C, the sonicated chromatin sample (SCS) in the supernatant was collected and analysed directly or stored at −70°C. To check for efficient DNA fragmentation to an average length of 300–400 bp, an aliquot of each SCS was reverse cross-linked (see below) and loaded onto a 1.5% agarose gel. For the IP reactions, 50 µl SCS was mixed with 150 µl FA 150 buffer and 8 µl 8WG16 antibodies (Abcam, ab817, preferentially recognising unmodified CTD of pol II) and incubated on a rotating wheel overnight at 4°C. A negative control without antibodies (NO) was included. To collect antibody-bound complexes, 40 µl blocked protein A sepharose CL-4B beads (GE Healthcare; equilibrated as a 1 : 3 slurry in and blocked with 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 100 µg/ml BSA, 50 µg/ml herring sperm DNA) were added and the samples put on a

rotating wheel for 1 h at room temperature. The beads were washed sequentially for 3 min on a rotating wheel with 350 µl TSE-150 (1% Triton-X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris−HCl pH 8.0, 150 mM NaCl), TSE-500 (as TSE-150 but with 500 mM NaCl) and LiCl (0.25 M LiCl, 1% NP-40, 1% dioxycholate, 1 mM EDTA, 10 mM Tris−HCl pH 8.0) buffers. This was followed by two quick washes in TE pH 8.0 and elution of the antibody-bound complexes with 100 µl elution buffer (1% SDS, 0.1 M $NaHCO_3$) at 65°C for 30 min. The beads were pelleted at 16 000g for 1 min and the eluate transferred to a fresh tube. Reverse cross-linking was performed by addition of 7 µl NaCl (5 M) and incubation at 65°C for 4–5 h. A total input control sample (TOT) containing 25 µl SCS and 75 µl FA 150 buffer was included at this step. RNA was digested for 1 h at 37°C with 1 µl RNase A (10 mg/ml). After an overnight incubation at 65°C with 1 µl proteinase K (Roche), the samples were purified through QIAgen PCR purification columns. IP and NO samples were eluted from the columns with 200 µl ddH_2O, the TOT samples with 300 µl ddH_2O.

For real-time PCR we used the RotorGene 3000 system (Corbett Life Science) and the SensiMix DNA kit (Quantace). Each sample was run in triplicate (a 1 : 10 dilution of the TOT sample was used). Primers were used at final concentrations of 200 or 500 nM. Threshold values were calculated with the built-in software and percentage input signals calculated as $100 \times [(IP - NO)/3]/(TOT \times 10)]$, taking into account the different dilution factors and elution volumes of the samples.

RNA interference

RNA interference (RNAi) by feeding on plates or in 150 ml liquid cultures was essentially performed as described by (27). RNAi was performed for 4 days at 20°C in the presence of IPTG (1 mM on plates, 0.4 mM in liquid), 15 µg/ml tetracycline and 50 µg/ml ampicillin. The L4440-rnp7 plasmid used to knockdown *rnp-7* (encoding the U1-70K protein homolog; wormbase annotation *K04G7.10*) contains a 449 nt *rnp-7* specific fragment. RNAi against *unc-22* (*ZK617.1*) served as a control.

Reporter gene constructs and transgenic animals

The p*vha-1::RFP*; *vha-2::GFP::vrk-1* and p*vha-1::RFP*; *vha-2::GFP::vrk-1mut** vectors were constructed by standard cloning procedures. The *vha-1* and *vrk-1* sequences were amplified from genomic DNA and fused in frame to RFP and GFP fragments respectively in the pUC-18 backbone.

Transgenic worms were generated by injecting young hermaphrodite gonads of *unc-119(ed3)* with both wild-type and mutant constructs at 30 ng/µl along with the rescuing plasmid pPD#MM016B at 80 ng/µl. For microscopy, nematodes were mounted onto 2% agarose pads and anaesthetised in 1% 1-phenoxypropan-2-ol in M9 buffer. Specimens were viewed under Normarski optics and epifluorescence using a Zeiss Axioplan 2 Imaging microscope. Wild-type non-unc animals exhibiting RFP signal were further inspected for the GFP signal under 20× objective. Images were captured using an AxioCam digital camera and the following exposure times: GFP: 1.55 s, RFP: 524 ms and DIC: 56 ms. Several transgenic lines were generated carrying extra-chromosomal arrays.

Nematode strains and primers

The following transgenic strains were generated: CB6690 *unc-119(ed3) eEx644* [p*vha-1*::RFP *vha-2*::GFP::*vrk-1*; *unc-119* (+)], CB6691 *unc-119(ed3) eEx645* [p*vha-1*::RFP *vha-2*::GFP::*vrk-1mut**; *unc-119* (+)], CB6692 *unc-119(ed3) eEx646* [p*vha-1*::RFP *vha-2*::GFP::*vrk-1mut**; *unc-119* (+)] and CB6702 *unc-119(ed3) eEx647* [p*vha-1*::RFP *vha-2*::GFP::*vrk-1mut**; *unc-119* (+)]. Primers are available upon request.

RESULTS

The majority of pol II terminates around 1 kb downstream of poly(A) sites at the end of mono- and poly-cistronic transcription units

We developed a RNA pol II ChIP protocol for use with whole nematodes and employed it to analyse how transcription termination is controlled at the end of individual genes or at the end of polycistronic transcription units.

We analysed the single gene *vit-2* by pol II specific ChIP using primer pairs specific for a region immediately upstream of the transcriptional start site (V1), two probes within the *vit-2* coding region (V2, V3) and four probes (V4–V7) specific to sequences downstream of the *vit-2* poly(A) site (Figure 1A). Chromatin preparations from mixed stage hermaphrodite cultures were immunoprecipitated using the 8WG16 antibody raised against the largest subunit of RNA pol II and precipitates were subsequently analysed by real-time PCR. As can be seen in Figure 1B, similarly high levels of polymerases, which are significantly above the rDNA control, are reproducibly detected with probes located over the coding region of *vit-2* (V2, V3) and 400 nucleotides downstream of the *vit-2* poly(A) site (V4). Slightly lower levels of pol II compared to the coding regions, but which are clearly above rDNA background levels, were detected over the promoter region (V1) and a region positioned 700 nucleotides downstream of the poly(A) site (V4b). A considerable drop of polymerase densities can be observed 1 kb or more downstream of the *vit-2* poly(A) site (V5, V7, Figure 1B) indicating that a significant number of polymerases have terminated around 1 kb downstream of the poly(A) site.

We next analysed transcription termination in *vrk-1*, the last of seven genes in the operon CEOP2232 (Figure 1C). Reproducibly high levels of polymerases were detected over the coding region (P2 and P2b, Figure 1D) and regions positioned up to 600 nucleotides downstream of the *vrk-1* poly(A) site (P3 and P3b, Figure 1D). A significant drop of polymerase densities is again observed at 1 kb and further downstream of the poly(A) site (compare P4, P4b, P5 and P6, Figure 1D).

We conclude from these results, that pol II terminates in the region 1 kb downstream of functional poly(A) sites in

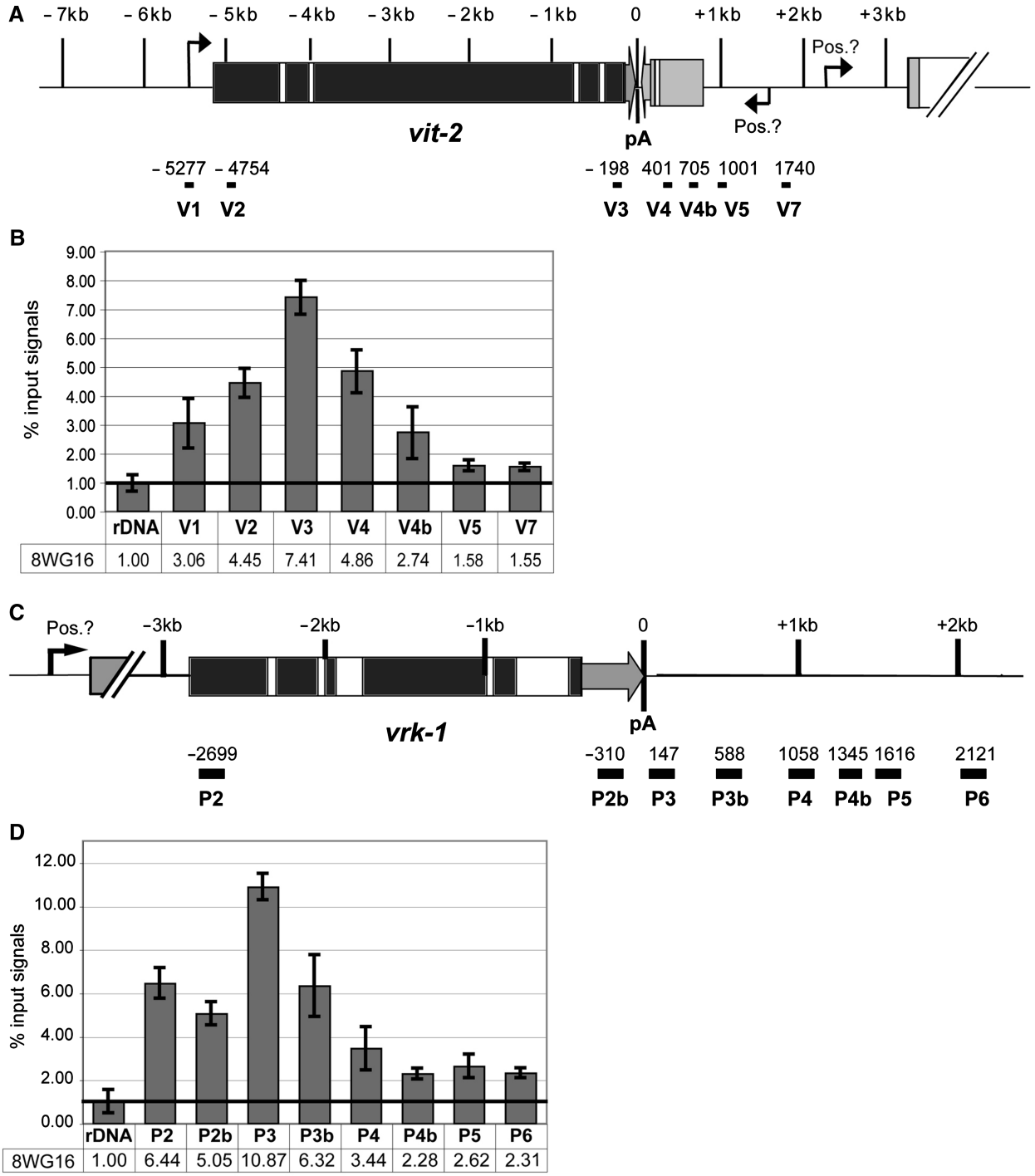

Figure 1. pol II ChIP profile on *vit-2* and *vrk-1* at the end of CEOP2232. (**A**) The *vit-2* (C42D8.2) locus and the positions of the real-time PCR probes are shown. The figure is drawn to scale and probes are about 100 nt long. The numbering starts at the poly(A) site of *vit-2* (pA) and numbers above the probes indicate the middle of the PCR amplicons. Negative numbers indicate positions upstream and positive numbers positions downstream of the poly(A) site. Black boxes: exons of *vit-2*. Grey boxes: exons of flanking genes C42D8.9 and C42D8.1. White boxes: introns. Grey boxed arrows indicate the directionality and 3′-UTR regions of the genes. Black arrows refer to transcription start sites ('Pos?', the precise transcription start sites are not known). (**B**) Average of four 8WG16 mediated pol II ChIP profiles on *vit-2* (three independent chromatin samples, one used twice for two independent IPs). Prior to averaging, a normalization of the individual ChIP experiments was performed against the sum of the signals of a series. The average of the normalized rDNA signals was put to 1. The bars indicate standard errors. (**C**) The *vrk-1* (*F28B12.3*) gene and the positions of the real-time PCR probes. The figure is drawn to scale and the numbering starts at the poly(A) site of *vrk-1*. Labelling as in Figure 1A. Dark grey box represents the first gene in the operon (*F59G1.4*). (**D**) Average of 4 pol II ChIP profiles on *vrk-1* (two independent chromatin samples, each used twice for two independent IPs). Data presented as in Figure 1B.

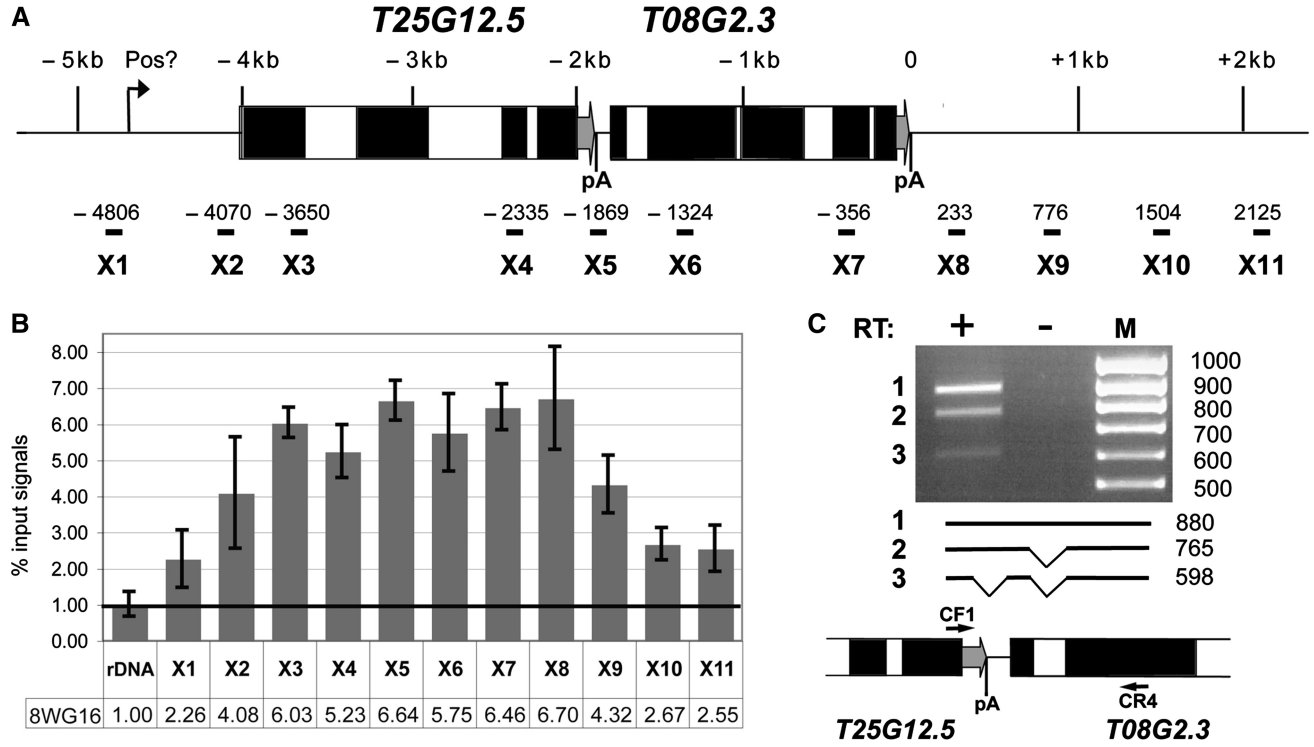

Figure 2. pol II ChIP profile on and RT-PCR analysis in CEOPX144. (**A**) The CEOPX144 operon with the two genes *T25G12.5* and *T08G2.3* and the positions of the real-time PCR probes. The figure is drawn to scale and the numbering starts at the poly(A) site of *T08G2.3*. Labelling as in Figure 1. (**B**) Average of 4 pol II ChIP profiles on CEOPX144 (two independent chromatin samples, each used twice for two independent IPs). The data is presented as described in Figure 1. (**C**) Detection of bicistronic pre-mRNAs in CEOPX144 by RT–PCR analysis. The region of analysis and the position of the RT-PCR primers (CF1 and CR4) are outlined in the graph below. + and – refer to the RT-PCR with and without reverse transcriptase, respectively. Each fragment was gel extracted and sequenced. Band 1 represents the full-length bicistronic pre-mRNA, band 2 has the first intron of *T08G2.3* spliced and band 3 is an alternative bicistronic transcript.

both *vit-2* and *vrk-1,* indicating that there is no difference regarding where pol II terminates at the end of single transcription units or at the end of operons.

The recognition of internal poly(A) sites does not trigger pol II termination in operons

Given the tight connection between the recognition of a functional poly(A) site and transcription termination it is unclear how pol II is able to transcribe entire operons without prematurely terminating downstream of internal poly(A) sites. In order to overcome this hurdle, polymerases that transcribe operons must either be prevented from prematurely terminating downstream of internal poly(A) sites or the expression of downstream genes in operons relies solely on polymerases that escape termination.

From our analysis described above, it appeared that the majority of pol II in both the *vit-2* and *vrk-1* genes terminates around 1 kb downstream of the poly(A) site. However, in both *vit-2* and *vrk-1* we consistently detected low levels of pol II up to 2 kb downstream of the poly(A) sites that were above the rDNA background levels (compare V7 and P6 with 'rDNA' in Figure 1B and D). We could therefore not rule out that the expression of polycistronic pre-mRNAs in the nematode simply relies on these small numbers of polymerases that fail to terminate. If this were the case, no specific mechanism would be required to prevent premature termination in polycistronic transcription units. To address this issue, we used a number of primers complementary to sequences in several critical positions to determine the polymerase densities over the annotated operon CEOPX144 (Figure 2A). Probe X1 covers a region that is 800 nucleotides upstream of the SL1 *trans*-splice site and is likely to be upstream of the transcription start site (note that the precise start site cannot be mapped due to *trans*-splicing). The probes X2–X4 target two sequences in the first gene (*T25G12.5)* of the operon, X5 represents the intergenic region, probes X6 and X7 cover the coding region of the second gene (*T08G2.3)* and probes X8–X11 target sequences in the 3′ flank of the operon. As can be seen in Figure 2B, no significant drop of polymerase densities is observed over the entire operon and up to about 800 nucleotides downstream of the last poly(A) site (compare probes X2 to X9, Figure 2B). In accordance with the results obtained in *vit-2* and *vrk-1* the majority of polymerases have terminated at a distance 1.5 kb downstream of the *T08G2.3* poly(A) site (X10 and X11, Figure 2B). Most importantly, no significant reduction of pol II densities was observed with primers complementary to sequences located 1.5 kb downstream of the internal poly(A) site (X7, Figure 2B). This result suggests that most polymerases that initiate

transcription at the CEOPX144 operon upstream of *T25G12.5* may transcribe the entire operon and terminate around 1.5 kb downstream of the last poly(A) site. It is therefore, unlikely that expression of downstream positioned genes in operons relies on a subset of polymerases that escape termination after internal poly(A) sites.

In order to substantiate this result it was necessary to prove that both genes allocated to CEOPX144 are indeed transcribed into a polycistronic pre-mRNA. This was important because despite the widely accepted proposition that up to a fifth of *C. elegans* genes are arranged in operons, direct experimental evidence for the existence of polycistronic pre-mRNAs is limited. So far, the best evidence for the presence of true operons in *C. elegans* is extrapolated from a genome wide analysis where SL2 snRNP *trans*-splicing to putative downstream operon genes was found to correlate strongly with gene clustering (28).

However, we noticed in earlier unrelated experiments that RNA isolation from whole mixed stage hermaphrodites using the hot-phenol method allows the detection of pre-mRNAs that are not fully processed (data not shown). We subsequently used the same approach to isolate total RNA from mixed stage hermaphrodites and subjected it to reverse transcription using a reverse primer located in the second exon of the downstream positioned *T08G2.3* gene to test whether we could detect polycistronic transcripts. PCR of the resulting cDNA was then performed using the same reverse primer in combination with a forward primer complementary to sequences in the terminal exon of the upstream positioned gene *T25G12.5*. As can be seen in Figure 2C, RT-PCR resulted in two distinct products, which in length corresponded to polycistronic pre-mRNAs. To confirm that the two detected bands represent polycistronic pre-mRNAs from CEOPX144, they were excised, re-amplified and sequenced (data not shown). The longer product 1, represents a polycistronic pre-mRNA complementary to the genomic sequence, whereas the smaller product 2, represents a polycistronic pre-mRNA that was identical to the first, except that the first intron of the downstream positioned *T08G2.3* gene was excised. Sequencing of a third faint product revealed an alternatively spliced bicistronic variant.

We conclude from this data that *T25G12.5* and *T08G2.3* are part of an operon and that they are transcribed into a polycistronic pre-mRNA. Furthermore, transcription of a functional internal poly(A) site in this operon does not result in premature termination of a significant number of polymerases.

Cis-splicing of introns in genes that are located at downstream positions of operons can occur prior to 3′-end processing and *trans*-splicing at the upstream intergenic region

The above analysis showed that *cis*-splicing in the polycistronic pre-mRNAs originating from CEOPX144 can occur prior to 3′-end formation and *trans*-splicing at the intergenic region. To confirm this observation, we analyzed primary transcripts from four additional randomly chosen annotated operons (Figure 3).

We aimed to detect and analyse polycistronic RNA containing *frh-1* and *ptp-2*, the fifth and sixth gene of CEOP2232, pre-mRNA containing *ebp-2* and *adbp-1,* part of CEOP2536, *vha-1* and *vha-2*, part of CEOP3620, and finally pre-mRNA that includes *lin-53* and *rba-1,* from CEOP1552. Total RNA was reverse transcribed with a reverse primer complementary to sequences located in exons of the downstream positioned genes. The resulting cDNAs were subsequently amplified using forward primers complementary to sequences in the terminal exon of the upstream positioned genes (Figure 3A–D, top panels). As can be seen (Figure 3A–D, bottom panels), in all of the operons analysed, polycistronic pre-mRNAs can readily be detected, in which the first intron of the downstream positioned gene is either present (*E1* + *i1* + *E2*) or spliced (*E1* + *E2*) (confirmed by sequencing). Additionally, in CEOP2232, pre-mRNAs can be detected where both intron 1 and intron 2 are spliced prior to poly(A) cleavage and *trans*-splicing at the intergenic processing sites.

Next, we addressed whether *cis*-splicing of the first downstream intron can occur after *trans*-splicing at the intergenic region has been completed. If *trans*-splicing predominantly occurs prior to *cis*-splicing in the analysed operons, then SL2 *trans*-spliced RNAs retaining the first intron should be, like the above observed polycistronic species, readily detectable. To that end, we repeated the RT reactions for CEOP3620 and CEOPX144 and used equivalent amounts of cDNAs with forward primers specific for the amplification of either polycistronic or SL2 *trans*-spliced transcripts. As can be seen (Figure 3E, upper panels), PCRs using SL2 specific forward primers and reverse primers, specific for either of the two operons, did not reveal products at sizes that would correspond to *trans*-spliced transcripts that still retained the first intron of the downstream gene (indicated by * in Figure 3, upper panels, lanes: SL2). In contrast, the same cDNA, when subjected to PCR using forward primers specific for the detection of bicistronic pre-mRNAs, produced clearly visible bands of the correct sizes (Figure 3, upper panels, lanes bicis.). In the case of CEOP3620, an additional band was observed but sequencing of this product revealed that it was the result of a miss-priming by the reverse primer v2R in the gene *K10B3.7/8* (Figure 3, upper left panel). This transcript was SL2 *trans*-spliced and the first intron was spliced. Repeats of the PCR using radioactively labelled nucleotides, revealed several weak additional bands (data not shown). However, none of these bands migrated at the expected lengths that would correspond to transcripts that had completed SL2 *trans*-splicing and still retained the intron. These results suggest that it is unlikely for *trans*-splicing in the analysed operons to be generally completed prior to *cis*-splicing of the downstream first intron.

The same PCR based approach was also used to determine the ratio between polycistronic pre-mRNAs and their respective mRNAs from the two operons. The quantitation using RT-PCR reactions performed on

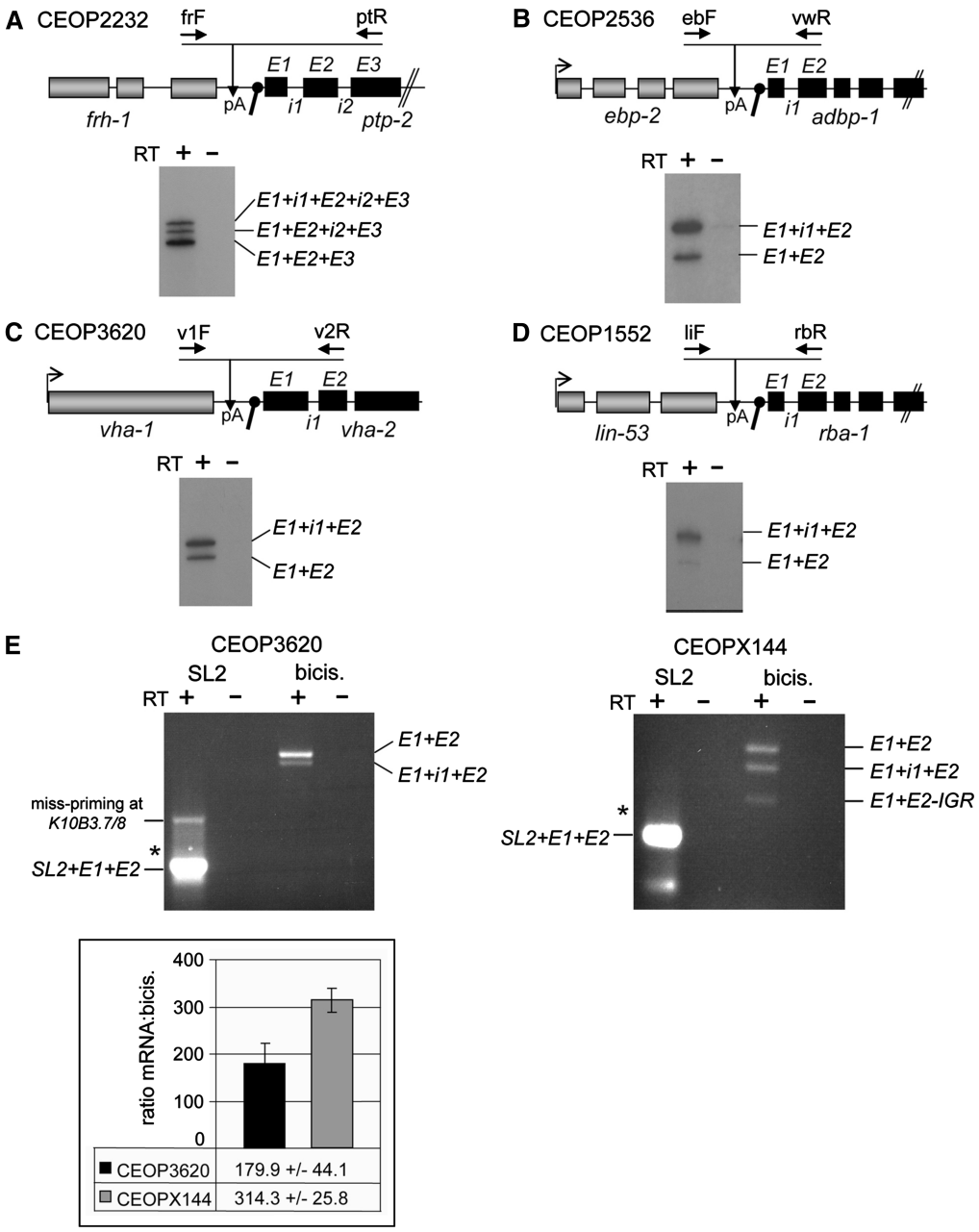

Figure 3. *Cis*-splicing of the downstream positioned gene in a polycistronic pre-mRNA can occur prior to cleavage at the poly(A) site of the upstream gene. RT-PCR analysis of polycistronic transcription units using RNA extracted from mixed stage nematodes. '*i1*' and '*i2*' represent the first and second introns, and '*E1*' and '*E2*' represent the first and second exons of the downstream transcription unit. (**A**) Three CEOP2232 transcripts containing the *frh-1*/*ptp-2* intergenic region are detected with the primer ptR, when amplified with the forward primer frF. The transcripts are with or without the introns *i1* and *i2*. (**B**) Two CEOP2536 polycistronic transcripts, with and without the intron *i1*, are detected with the primer vwR, when amplified with the forward primer ebF. (**C**) Two CEOP3620 polycistronic transcripts, with and without the intron *i1,* are detected with the primers v2R and v1F. (**D**) Spliced and non-spliced polycistronic precursors containing *lin-53* and *rba-1* (CEOP1552) are detected using the primers liF and rbR. (**E**) Analysis of CEOP3620 and CEOPX144 to detect SL2 *trans*-spliced (SL2) and bicistronic (bicis.) transcripts. Upper panels: Representative EtBr-stained gels of the individual RT-PCR reactions (25 µl of 50 µl, 28 cycles). RT-PCR primers as in Figures 2C and 3C. All bands were sequenced. Asterisk indicate expected band of a *trans*-spliced transcript containing intron 1 (*SL2* + *E1* + *i1* + *E2*). Miss-priming of the reverse primer in *K10B3.7/8* is indicated (miss-priming *K10B3.7/8*), representing an SL2 *trans*-spliced and first intron *cis*-spliced transcript. The chart shows the ratios of mRNA levels of the downstream gene to levels of bicistronic transcripts (corrected for radiolabel content). Standard errors were calculated from 5 (CEOP3620) or 3 (CEOPX144) independent experiments.

three independent RNA batches shows that the mRNAs encoded in the downstream genes in the two operons are about 200–300 times more abundant than their respective polycistronic precursors (Figure 3E, bottom panel).

From these experiments we conclude that *cis*-splicing at the downstream positioned genes in operons can occur prior to 3′-end processing and *trans*-splicing at the intergenic region.

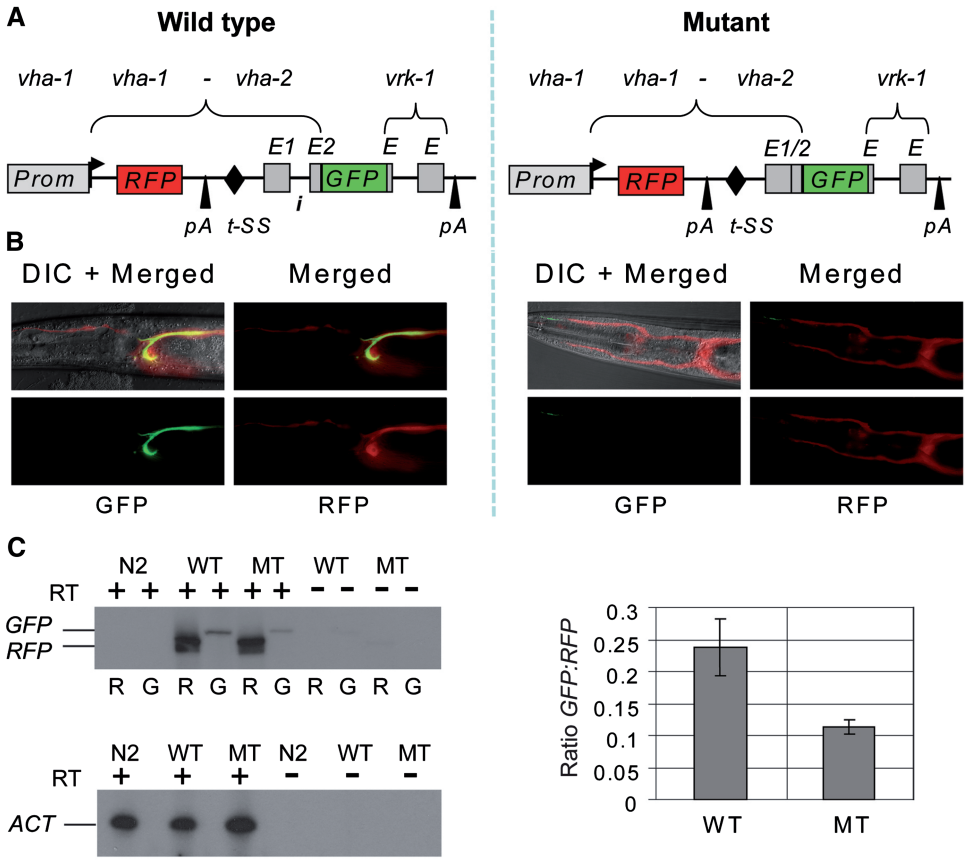

Figure 4. Expression levels from a plasmid bearing a RFP-GFP operon vary depending on the presence or absence of the intergenic proximal intron. (**A**) Depiction of the wild-type (left) and mutant (right) operon expression construct. *pA*, *t-ss* indicate the cleavage and polyadenylation site and the *trans*-splice site, respectively. The *vha-1* promoter (*Prom*) drives transcription of the polycistronic pre-mRNA containing both the ORF of the green fluorescent protein (*GFP*) and the red fluorescent protein (*RFP*). The 'in frame' fusion of exon 1 (*E1*) and exon 2 (*E2*) in the mutant construct is indicated (*E1/2*). '*i*' represents the intergenic proximal downstream intron in the wild type construct. '*E*' states the penultimate and terminal exon of the *vrk-1* gene. The origins of sequences in the operon construct are indicated above the diagrams: *vha-1*, *vha-2* and *vrk-1*. (**B**) Comparison of GFP and RFP expression in the excretory cell in both wild-type (left panels) and mutant (right panels) transgenic animals. (**C**) RNA from transgenic animals expressing the wild-type (WT) and mutant (MT) plasmids and from non-transgenic N2 animals was reverse transcribed using an oligo-dT primer. cDNAs were then amplified with *RFP*- (R) and *GFP*-specific primers (G) (upper left panel). cDNAs were also amplified with *ACT-5*-specific primers (lower left panel). The graph on the right shows the ratio of *GFP*- to *RFP*-RT-PCR products for both WT and MT animals (average of two independent experiments; standard deviation bars are shown). It has to be noted that high levels of aberrant GFP and RFP expression is observed in the vulva of the transgenic animals which may have affected the quantitation. A more detailed RNA analysis and quantitation can be found in Figure S1.

Cis-splicing of the first intron of a downstream positioned gene affects its expression levels

As described earlier, it appears that in polycistronic pre-mRNAs *cis*-splicing of introns located in the downstream positioned genes in all five analysed operons can occur before poly(A) cleavage and *trans*-splicing at the intergenic region. We therefore next questioned whether this observed *cis*-splicing of the first intron in the downstream positioned gene has any functional consequence for the expression of genes located in operons.

We designed a plasmid containing a polycistronic transcription unit, where the first ORF encodes the red fluorescent protein (RFP) followed by the *vha-1*/*vha-2* intergenic region, the *vha-2* 5′-UTR, exon 1, the first intron and parts of exon 2, fused in frame to the green fluorescent protein (GFP) ORF, followed by the penultimate and terminal exon and poly(A) site of the *vrk-1* gene (Figure 4A, wild-type construct). A second 'mutant' plasmid was designed consisting of identical gene fragments with the sole exception that the *vha-2* exon 1 is directly fused in frame to the *vha-2*/*gfp* fragment, omitting the intron sequence (Figure 4A). The resulting *GFP* and *RFP* mRNAs from both constructs are identical but the maturation of the two pre-mRNAs differs in that only the wild-type primary transcript undergoes *cis*-splicing at close proximity to the intergenic region. Both wild-type and mutant transcripts, undergo *cis*-splicing at the terminal intron. Transcription in both plasmids is under the control of the *vha-1* promoter and expression of the reporter genes is expected in the 'H' shaped excretory cell and canal (29). Both plasmids were injected into *C. elegans* hermaphrodites and stable lines inheriting extra chromosomal arrays were obtained. We then compared the expression levels of RFP and GFP by microscopy. As can be seen in Figure 4B, RFP is generally well

expressed in the excretory cell and canal of both wild-type (left panel) and mutant transgenics (right panel). GFP expression, however, is at much lower levels in worms containing the mutant plasmid compared to transgenics containing the wild-type plasmid (Figure 4B). Out of 148 RFP positive wild-type transgenic animals, 79% clearly showed GFP expression in the excretory cell under the 20x objective. In comparison, out of 50 mutant worms, none showed detectable GFP expression under 20x magnification, but half of those showed weak GFP expression under higher magnification. The analysis of two additional transgenic mutant lines and one additional wild-type confirmed the above described observations (data not shown). The analysis of RNA isolated from the transgenic animals showed that the expression of the RNAs from the injected plasmids follows the expected pattern. In the wild-type transgenic worms two pre-mRNA species are detected that are not cleaved or *trans*-spliced at the intergenic poly(A) site and either contain (*E1* + *i1* + *E2*) or lack (*E1* + *E2*) the first intron. In addition, GFP mRNA levels in the mutant compared to wild-type nematodes were modestly reduced (Figure 4C and Figure S1).

This analysis suggests that downstream *cis*-splicing close to the intergenic region can affect expression of the downstream positioned genes.

RNAi knockdown of the U1 snRNP specific 70 kDa protein reduces polycistronic pre-mRNA levels and causes a drop in pol II levels at downstream positions

The observed loss of GFP expression in the mutant transgenic worms described above is likely to be the result of multiple effects. These effects may include reduced nuclear cytoplasmic export, mRNA localisation and reduced translation efficiency (30). However, since it has been reported that in retroviruses the recognition of a downstream positioned 5′ splice site by the U1 snRNP prevents the recognition of a functional upstream poly(A) site (31), we considered whether splicing of the intergenic proximal intron in downstream operon genes could also play an additional role in preventing premature transcription termination by regulating internal poly(A) site use. To address this possibility we decided to employ an RNAi based approach to inhibit *cis*-splicing since site-directed mutagenesis of *cis*-splice sites is not feasible in worms. We cloned a fragment (32) targeting sequences of the U1-70K gene *(rnp-7*, K04G7.10) into the plasmid L4440, which allows the production of dsRNA in the host bacteria upon induction by IPTG. We then confirmed that nematodes feeding on host bacteria containing this plasmid for four days have reduced *rnp-7* mRNA levels and show reduced levels of *cis*-splicing in the non *trans*-spliced *vit-2* gene compared to control worms (Figure S2). Interestingly, knock down of the basic splicing factor U1 70 kDa protein resulted in a significant accumulation of unspliced pre-mRNA in the nematode, which to the best of our knowledge, has previously not been observed in *C. elegans*. Worms were subsequently grown in liquid cultures feeding on induced host bacteria expressing either the *rnp-7* specific or control dsRNAs targeting the non-essential *unc-22* mRNA. After 4 days of feeding on the induced host bacteria, cultures were analysed for phenotypic RNAi effects (high embryonic lethality for *rnp-7* targeted worms) and if positive, subjected to total RNA and chromatin isolation. As can be seen in Figure 5B, we observed a clear knockdown at the whole worm RNA level for *rnp-7*. We analysed the RNA preparation further by RT-PCR using primers that are specific for the detection of transcripts from CEOP3620 (Figure 5A, RT-PCR probes). We used this particular operon because the first gene in the operon, *vha-1*, is intronless and hence its expression is not affected by *cis*-splicing. We compared the levels of *vha-2* mRNA from the *rnp-7* knockdown culture, normalised to *vha-1* levels, with the *unc-22* control. This comparison showed that *vha-2* RNA levels were significantly reduced in the *rnp-7* knockdown culture (Figure 5B, *vha-2*). Interestingly, the levels of polycistronic pre-mRNAs in the *rnp-7* RNAi culture, compared to the *unc-22* control and normalised to *vha-1,* were dramatically reduced. A drop in polycistronic pre-mRNA levels has also been seen in additional RNA preparations, where *rnp-7* knockdown resulted in severe phenotypic effects and a significant reduction in *rnp-7* mRNA levels (data not shown). The reduction in levels of polycistronic pre-mRNAs in the *rnp-7* knockdown cultures could be explained if inhibition of the recognition of a downstream 5′ *cis*-splice site would affect *trans*-splicing, the stability of the polycistronic pre-mRNAs and/or if *cis*-splicing somehow counteracts premature transcription termination. To verify whether inhibition of 5′ *cis*-splice site recognition affects the latter, we used the isolated chromatin from the RNAi cultures for pol II ChIP analysis. The analysis shows that when *rnp-7* is targeted by RNAi there are significantly lower levels of polymerases detectable at positions 1 and 1.5 kb downstream of the internal poly(A) site compared to the *unc-22* control (Figure 5C, compare probes VH5 and VH6). Interestingly, the drop in pol II levels is observed 1–1.5 kb downstream of the internal poly(A) site, which corresponds to the distance where termination is most prominent downstream of terminal poly(A) sites (Figures 1 and 2). In contrast, polymerase densities in the *rnp-7* and control knockdowns are very similar over probes corresponding to regions upstream, immediately around the internal poly(A) site and 500 nucleotides downstream of the internal processing site (Figure 5C, compare probes VH2, VH3 and VH4).

From these results we conclude that inhibition of the recognition of a 5′ *cis*-splice site reduces pol II levels 1–1.5 kb downstream of the internal poly(A) site and causes a drop in polycistronic pre-mRNA levels. This would be consistent with a role of *cis*-splicing in preventing premature transcription termination in *C. elegans* operons.

Most genes in operons that are located at downstream positions contain introns

If *cis*-splicing of the downstream proximal intron is critical for efficient expression of downstream genes in

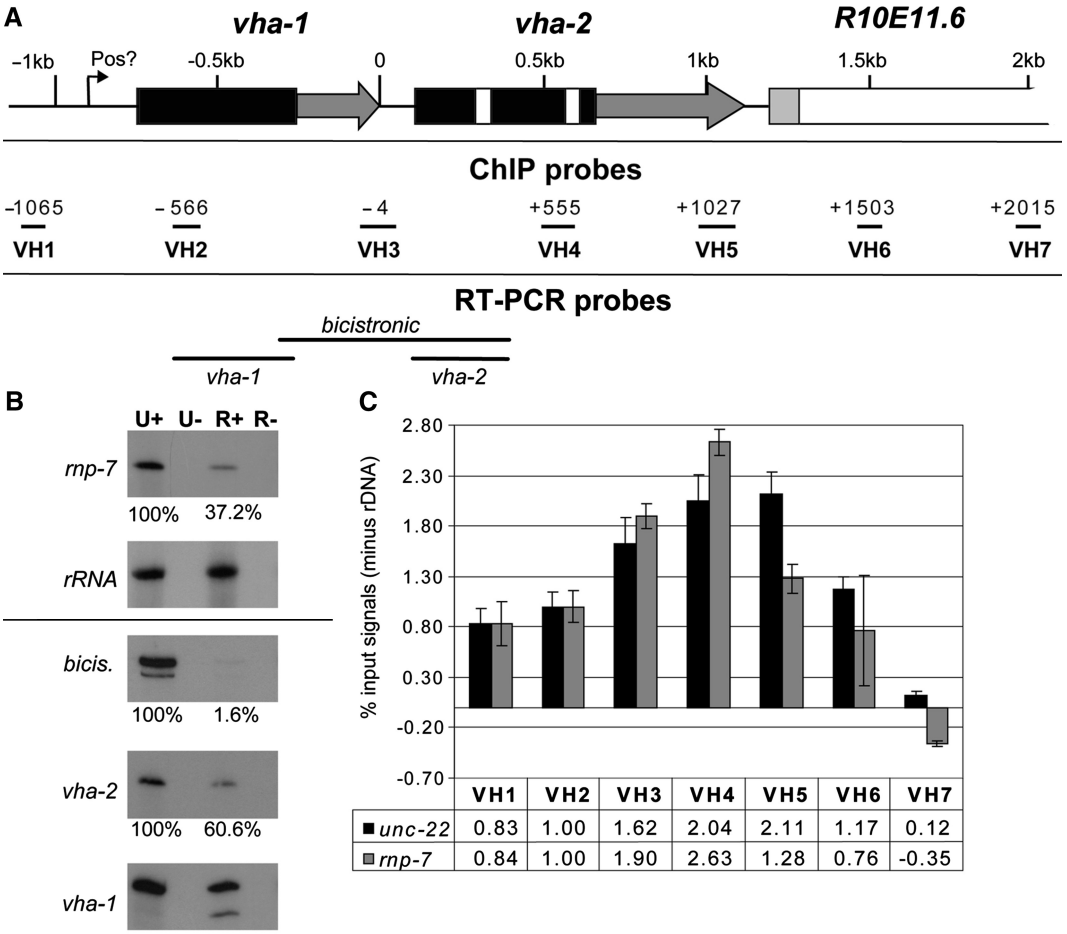

Figure 5. RNAi against *rnp-7* correlates with premature pol II termination within operons. (**A**) The CEOP3620 operon and the positions of the real-time PCR (for ChIP) and RT-PCR probes are indicated. The diagram is drawn to scale and the numbering starts at the poly(A) site of the intronless *vha-1* gene (R10E11.8). The labelling is as in Figure 1. The flanking gene R10E11.6 may not be part of the operon but is shown for completeness. (**B**) Representative hot RT-PCR reactions for *rnp-7* (R) and *unc-22* (U) knockdown samples with (+) and without (−) reverse transcriptase (RT). Gene-specific primers were used as indicated. *Unc-22* levels were set to 100%. Upper panel: *rnp-7* mRNA levels normalized to 18S rRNA. Lower panel: levels of bicistronic transcripts and *vha-2* mRNAs were normalised to *vha-1* mRNA. (**C**) Pol II density profiles with chromatin samples from the RNAi experiment shown in panel A. The data is presented as described in Figure 1, but with rDNA signals deducted. Numbers were obtained from three independent IPs for both the *rnp-7* and *unc-22* knockdowns.

polycistronic transcription units, it would demand that most downstream genes in operons must contain introns and undergo obligatory *cis*-splicing. We therefore, analysed whether the distribution of intronless genes regarding their positions in operons in the nematode genome is biased. There are currently 596 annotated intronless ORFs in the *C. elegans* genome. Only 23 of these intronless genes can be found within operons, 12 of the 23 represent the first gene of the respective operons and hence do not have upstream poly(A) sites. Three of the remaining 11 genes contain non-coding exons and therefore, essentially qualify as spliced genes. This leaves eight genes located in operons that may contradict the possible requirement of *cis*-splicing at downstream positioned genes. There are one thousand operons in the nematode genome encompassing ~2600 genes (28). Random distribution would result in 48 intronless genes at downstream positions in operons and hence is notably higher than the eight intron-free ORFs we have identified.

DISCUSSION

Pol II termination at the end of mono- and polycistronic transcription units

We employed a pol II ChIP protocol and RT-PCR to study how pol II termination is regulated in the nematode *C. elegans* and how premature transcription termination may be prevented in its operons.

From our ChIP analysis it is evident that the positions relative to poly(A) sites, where transcription termination occurs at the end of a monocistronic and polycistronic transcription units are very similar (Figures 1 and 2). This suggests that transcription termination at the end of operons is regulated in the same way as at the end of monocistronic genes. The ChIP analysis of CEOPX144, described in Figure 2, also argues against the possibility that downstream positioned genes in operons could be transcribed by a subset of pol II that simply escapes termination at internal poly(A) sites since no loss of

pol II 1–1.5 kb downstream of the internal 3′-end processing site can be detected. We therefore believe that termination after transcription of functional internal poly(A) sites in operons is likely to be actively prevented.

Interestingly, our results also show that despite the compact *C. elegans* genome, pol II termination in the nematode occurs at similar distances as has been observed in fly genes and it appears that there is no strong correlation between the density of a genome and how far downstream of poly(A) sites pol II terminates.

Detection and analysis of polycistronic pre-mRNAs

Since it has so far been difficult to demonstrate that putative *C. elegans* operons are truly transcribed into polycistronic pre-mRNAs, we needed experimental proof for the existence of these polycistronic transcription units. Our initial RT-PCR approach confirmed that the genes in CEOPX144 are transcribed into a polycistronic pre-mRNA, but also provided further data suggesting that *cis*-splicing may take place before cleavage at the upstream poly(A) site and *trans*-splicing occur (Figure 2).

This initial observation was confirmed, when we analysed more predicted operons (Figure 3). Interestingly, the observation that *cis*-splicing of introns in downstream positioned genes can occur prior to poly(A) cleavage at upstream positioned genes, is not limited to *C. elegans*. This phenomenon has also been found in the nematode *Brugia malayi* (33) and in other operon-containing eukaryotic organisms including *Oikopleura dioica* (a tunicate) (34), the flatworm *Schistosoma mansoni* (35) and *Trypanosoma cruzi* (36). As with any other pre-mRNAs detected *in vivo*, we cannot entirely rule out that these products are dead end transcripts but it would be surprising that this would happen so frequently in various different organisms. In addition, the fact that they are partially processed (downstream introns spliced) makes it less likely that they are dead end products. Furthermore, if *cis*-splicing in *C. elegans* would generally occur after processing at the intergenic region, intermediate RNAs that have completed *trans*-splicing but still retained the first intron should be detectable. However, our RT-PCR analyses demonstrate that this is not the case (Figure 3).

The intergenic proximal downstream intron is critical for the expression of downstream genes in operons

The analysis of the transgenic nematodes indicates that the lack of *cis*-splicing in the second gene (*GFP*) of an operon close to an internal poly(A) site causes a drop in *GFP* expression. Lack of *cis*-splicing has previously been shown to affect gene expression at multiple levels. *Cis*-splicing can affect transcription (37,38), nuclear cytoplasmic transport (39), RNA localization (40) and translation efficiency (41). In addition, we cannot rule out that sequences located in the first intron of *GFP* (derived from the endogenous *vha*-2 gene) are required for *trans*-splicing. Such regulatory sequences have recently been found in introns of some monocistronic *trans*-spliced genes in the nematode *Brugia malayi* (42). However, we did not find any similar sequences in the first intron of the *vha-2* gene by sequence comparison. Furthermore, our analysis of the polycistronic pre-mRNAs from this operon suggests that *cis*-splicing may occur before *trans*-splicing, which inevitably would result in the loss of such an enhancer. It is also unlikely that transcription initiation and re-initiation are affected because the upstream gene *RFP,* transcribed from the same promoter, is expressed well in both mutant and wild-type transgenic worms.

Does the recognition of a 5′ splice site play a role in preventing premature transcription termination in operons?

It is very likely that some of the above discussed phenomena contribute to the overall observed dramatic loss of GFP in the mutant transgenic nematodes. However, our results presented in Figure 5 indicate an additional function that is associated with the recognition of the introns located downstream of the intergenic regions in operons, which could contribute to the loss of *GFP* expression in mutant transgenic nematodes. We show that targeting U1 snRNP specific 70 kDa mRNA by RNAi results both in a drop of polycistronic steady state pre-mRNA levels in CEOP3620 and most importantly in a drop of pol II levels at positions located 1–1.5 kb downstream of the internal poly(A) site. Although we cannot exclude a general effect on transcription elongation, it is striking that a drop in pol II levels at such positions would be consistent with premature termination of pol II downstream of the internal poly(A) sites. This raises two intriguing possible functions for the 70 kDa protein in regulating operon gene expression:

First, although U1 snRNP is not believed to be involved in *trans*-splicing in *Ascaris*, we cannot entirely rule out that the U1-70 kDa protein could still play a role in SL2 *trans*-splicing in *C. elegans*. Furthermore, it is possible that the 5′ splice site recognition and the assembly of the *cis*-splicing complex at the downstream exon/intron border affects SL2 specific *trans*-splicing. It is plausible that cross-exon protein protein interactions could strengthen assembly of splice factors at 3′ '*trans*-splice' sites. Inhibition of *cis*-splicing, caused by knocking down the U1-70 K protein, could therefore affect the efficiency of SL2 *trans*-splicing. Since cleavage and polyadenylation at the upstream gene can happen in the absence of *trans*-splicing (20), this would result in an unprotected 5′-end, allowing exoribonuclease access and promote premature transcription termination downstream of the internal poly(A) site. This would therefore represent experimental evidence that *trans*-splicing is required to prevent premature transcription termination in operons via the torpedo model, but would not suffice to explain why termination is prevented through the allosteric pathway and how the poly(A) factors are rearranged or re-recruited after cleavage at internal poly(A) sites (see below). In addition, although this interpretation would be consistent with most of our results, it would not explain why a drop in polycistronic pre-mRNAs, as observed in our RNAi experiments, would occur. Furthermore, it would somewhat disagree with the recent findings that inhibition of SL2 *trans*-splicing results in the accumulation of

pre-mRNAs that are cleaved at the upstream poly(A) site but are not *trans*-spliced (22). It has been suggested that 5′–3′ exoribonuclease activity is blocked by a so far unidentified protein that associates with the pre-mRNA at U-rich sequences located between internal poly(A) sites and *trans*-splice sites.

We therefore suggest a second possible interpretation of our results. It is possible that U1 snRNP association with the 5′ splice site at the exon/intron border close to the intergenic region may also regulate the recognition of internal poly(A) sites by the pol II associated poly(A) factors.

The molecular mechanism, by which U1 snRNP could inhibit internal poly(A) site recognition by the poly(A) factors associated with the transcription machinery may be similar to that described for the promoter proximal poly(A) site inhibition in HIV-1 (43,44). Due to the replication mechanism in retroviruses, poly(A) sites are present both immediately downstream of the viral promoter and towards the end of the viral genomic DNA. To successfully express any viral transcripts the promoter proximal poly(A) site must be inactivated. This regulation is achieved via the interaction of the U1 snRNP with the major 5′ splice site located downstream of the promoter proximal poly(A) site (45). The 70 kDa protein of the U1 snRNP inhibits poly(A) site recognition and hence prevents cleavage at this premature 3′ processing site in the viral pre-mRNA (43).

Expression of a polycistronic transcription unit in *C. elegans* results in a comparable situation. We therefore, propose a novel speculative model described in Figure 6 how premature transcription termination may be prevented in operons. In this model, the recognition of a 5′ *cis*-splice site at the intergenic-proximal downstream positioned intron in operons inhibits the recognition of the internal poly(A) site by the polymerase associated poly(A) factors (Figure 6A). It is important to note that this process would not demand that *cis*-splicing at the downstream gene is completed prior to *trans*-splicing, although our results indicate that this may sometimes be the case (Figure 3).

However, unlike in HIV-1, the internal poly(A) sites in the nematode genes will eventually have to be recognised to generate mature cleaved individual mRNAs from the polycistronic primary transcript. It is possible that the recognition of a downstream *cis*-splice site simply modulates the assembly of the pol II associated factors into functional complexes but this could present a problem for rearrangement or re-recruitment of poly(A) factors downstream of internal processing sites (see below). We therefore speculate that the internal poly(A) site is recognised by a poly(A) complex that is not associated with RNA pol II. We propose that the intergenic poly(A) sites are recognised by at least some poly(A) factors that are associated with the *trans*-splicing machinery (Figure 6B, see Introduction) rather than the transcription apparatus. The suggestion that internal poly(A) sites could be recognised by such a complex, possibly independently of RNA polymerase II, but in conjunction with the assembly of the *trans*-splicing machinery, is reasonable since SL2 snRNP can be co-purified with the essential poly(A)

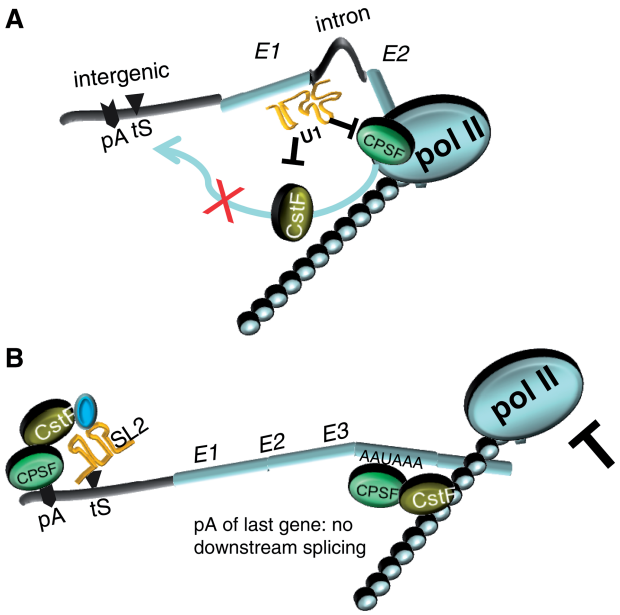

Figure 6. Model. (**A**) The association of the U1 snRNP at the 5′ splice site in exon 1 (*E*1) prevents interaction of polymerase associated poly(A) factors (CPSF, CstF) with the intergenic poly(A) sequences (pA). (**B**) Intergenic poly(A) sites and *trans*-splice sites are recognised independently of pol II by a complex consisting of the SL2 splice leader RNA and the poly(A) factors. Due to a lack of a functional downstream 5′ splice site at the end of the transcription unit the terminal poly(A) site is recognised by the polymerase associated 3′-end processing factors and subsequent cleavage and polyadenylation instigates transcription termination.

factor *Ce*CstF-64 (23). As mentioned above, the association of U1 snRNP at the downstream 5′ splice site may not only prevent the recognition of the internal poly(A) site by the pol II associated factors, but may also promote the interaction of splicing factors at the upstream 3′ (*trans*-) splice site via cross exon interactions.

The poly(A) cleavage machinery associated with SL2 snRNP could then interact with the intergenic region both through the recognition of the poly(A) site by CPSF, CstF and the *trans*-splice site by SL2 snRNP (23). It has been shown that cleavage and polyadenylation can occur when downstream *trans*-splice sites are mutated (20), but the combined recognition of the internal poly(A) sites and the *trans*-splice site in normal circumstances may still enhance efficient tethering of the SL2-poly(A) complex to the pre-mRNA independent of the transcription machinery. Interestingly, it has recently been suggested that both *trans*-splicing and cleavage and polyadenylation of some bicistronic pre-mRNAs from the protozoan *Trypanosoma cruzi* can be completely uncoupled from transcription (36).

The presence or absence of a nearby downstream 5′ *cis*-splice site in this model would be critical to determine which or how the complex will associate with the different poly(A) sites on pre-mRNAs. If there is no functional downstream 5′ *cis*-splice site, such as at the end of transcription units the polymerase associated factors will recognise the poly(A) sites and transcription termination will be instigated (Figure 6B). However, if the poly(A) site

is located in the intergenic region within an operon, recognition of a downstream 5′ *cis*-splice site modulates the association of polymerase bound poly(A) factors with the pre-mRNA (Figure 6A) but instead will promote the recognition of the poly(A) site by the alternative poly(A) and *trans*-splicing complex (Figure 6B). This would lead to delayed processing at the internal sites, equipping the upstream mRNA with a poly(A) tail and the downstream nascent RNA with a cap structure. The latter will permanently avert 5′−3′ exoribonuclease access, preventing premature termination via the torpedo model, as suggested by (22).

The above described pol II independent processing at internal poly(A) sites would ensure that the polymerase associated poly(A) factors would remain linked with the transcription complex in the same way until the recognition of the last poly(A) signal at the end of every polycistronic transcription unit. This mechanism would also avoid the possible need of a rearrangement of the interactions between the cleavage and polyadenylation complex and the polymerase. It has been suggested that CPSF is bound to the body of the polymerase until it engages with a poly(A) hexamer. The interaction of CstF with the RNA causes CPSF to disengage from the body of the polymerase and its association with the polymerase is subsequently mediated, via the interaction of CstF and additional poly(A) factors with the pol II CTD (46). In addition, since there is growing evidence that at least some of the poly(A) factors are recruited to the transcription machinery at the promoter (4,47–49) the suggested alternative poly(A) site recognition would also avoid the potential need for re-recruitment of poly(A) factors such as CPSF that may be lost after 3′-end processing at internal sites. It has been shown that polyadenylation may be initiated, while the transcript is still associated with the gene but may be completed after release and during the transport through the nucleoplasm (2).

It is important to note that we cannot rule out that the observed effect on transcription termination downstream of internal poly(A) sites is caused by a direct effect of the experimental procedure. Knocking down of the 70 kDa protein may result in the formation of an incomplete spliceosome, which remains associated with the pre-mRNA. Splicing factors, if not removed from the pre-mRNA by splicing, could trigger transcription termination as part of a surveillance mechanism. This is plausible, since it has recently been demonstrated that the 3′-ends of transcripts are tethered to the CTD of pol II by the exon definition complex consisting of splice factors and cleavage and polyadenylation factors (50). If splicing is stalled by the inactivation of downstream splicing in a polycistronic pre-mRNA, an aberrant complex could be formed between the poly(A) factors assembled at the upstream internal poly(A) site and the stalled spliceosome at the downstream intron, which may trigger premature termination in the operon.

However, the potential role of *cis*-splicing in the above presented speculative model is indirectly further supported by the finding that intronless genes are underrepresented at downstream positions in operons (data not shown). Finally, it is striking to note that the *C. elegans* genome has the lowest number of intronless protein encoding genes when compared to 21 other eukaryotic genomes (51). Taken together, our study indicates that the recognition of a 5′ *cis*-splice site downstream of operon-internal processing sites is important for expression of downstream positioned genes. This may be at least partly achieved via modulating how upstream poly(A) and/or *trans*-splice sites are recognized and this may be critical to prevent premature transcription termination.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank all the members of the Furger lab and Nick Proudfoot for his great support throughout this study and for the critical reading of this manuscript. They are very grateful to Jane Mellor and her group for invaluable support in the development of the ChIP protocol.

FUNDING

Wellcome Trust [081083/Z/06/Z to A.F.]; the Medical Research Council [AL2445 to H.S., G06000238, ID 77674 to J.H.]; the E.P.A. Cephalosporin Fund [CF 143 to A.F. and S.H.]; and the Swiss National Science Foundation [PBBEA-115061 to S.H., Austrian Academy of Sciences (DOC-fFORTE-fellowship) to K.Z.]. Funding for open access charge: Wellcome Trust [081083/Z/06/Z]; Medical Research Council [G06000238, ID 77674].

Conflict of interest statement. None declared.

REFERENCES

1. Greger,I.H., Demarchi,F., Giacca,M. and Proudfoot,N.J. (1998) Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Res.*, **26**, 1294–1301.
2. Bauren,G., Belikov,S. and Wieslander,L. (1998) Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3′-end formation and excision of the 3′-terminal intron. *Genes Dev.*, **12**, 2759–2769.
3. Dye,M.J. and Proudfoot,N.J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol. Cell*, **3**, 371–378.
4. Glover-Cutter,K., Kim,S., Espinosa,J. and Bentley,D.L. (2008) RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.*, **15**, 71–78.
5. Birse,C.E., Minvielle-Sebastia,L., Lee,B.A., Keller,W. and Proudfoot,N.J. (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science*, **280**, 298–301.
6. Kim,M., Ahn,S.H., Krogan,N.J., Greenblatt,J.F. and Buratowski,S. (2004) Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J.*, **23**, 354–364.
7. Birse,C.E., Lee,B.A., Hansen,K. and Proudfoot,N.J. (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J.*, **16**, 3633–3643.
8. Logan,J., Falck-Pedersen,E., Darnell,J.E. Jr. and Shenk,T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase

II in the mouse beta maj-globin gene. *Proc. Natl Acad. Sci. USA*, **84**, 8306–8310.
9. Rosonina,E., Kaneko,S. and Manley,J.L. (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev.*, **20**, 1050–1056.
10. Connelly,S. and Manley,J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.*, **2**, 440–452.
11. Whitelaw,E. and Proudfoot,N. (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human alpha 2 globin gene. *EMBO J.*, **5**, 2915–2922.
12. Proudfoot,N. (2004) New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.*, **16**, 272–278.
13. Luo,W., Johnson,A.W. and Bentley,D.L. (2006) The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.*, **20**, 954–965.
14. Kaneko,S., Rozenblatt-Rosen,O., Meyerson,M. and Manley,J.L. (2007) The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes Dev.*, **21**, 1779–1789.
15. West,S., Proudfoot,N.J. and Dye,M.J. (2008) Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol. Cell*, **29**, 600–610.
16. Blumenthal,T. (2004) Operons in eukaryotes. *Brief Funct. Genomic Proteomic*, **3**, 199–211.
17. Spieth,J., Brooke,G., Kuersten,S., Lea,K. and Blumenthal,T. (1993) Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell*, **73**, 521–532.
18. Hannon,G.J., Maroney,P.A. and Nilsen,T.W. (1991) U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. *J. Biol. Chem.*, **266**, 22792–22795.
19. Huang,T., Kuersten,S., Deshpande,A.M., Spieth,J., MacMorris,M. and Blumenthal,T. (2001) Intercistronic region required for polycistronic pre-mRNA processing in *Caenorhabditis elegans*. *Mol. Cell Biol.*, **21**, 1111–1120.
20. Kuersten,S., Lea,K., MacMorris,M., Spieth,J. and Blumenthal,T. (1997) Relationship between 3′-end formation and SL2-specific trans-splicing in polycistronic Caenorhabditis elegans pre-mRNA processing. *RNA*, **3**, 269–278.
21. Liu,Y., Huang,T., MacMorris,M. and Blumenthal,T. (2001) Interplay between AAUAAA and the trans-splice site in processing of a Caenorhabditis elegans operon pre-mRNA. *RNA*, **7**, 176–181.
22. Liu,Y., Kuersten,S., Huang,T., Larsen,A., MacMorris,M. and Blumenthal,T. (2003) An uncapped RNA suggests a model for Caenorhabditis elegans polycistronic pre-mRNA processing. *RNA*, **9**, 677–687.
23. Evans,D., Perez,I., MacMorris,M., Leake,D., Wilusz,C.J. and Blumenthal,T. (2001) A complex containing CstF-64 and the SL2 snRNP connects mRNA 3′ end formation and trans-splicing in C. elegans operons. *Genes Dev.*, **15**, 2562–2571.
24. Johnstone,I.L. (1999) In Hope,I.A. (ed.), *C. elegans, A Practical Approach*. Oxford University Press, Oxford, Chapter 11, pp. 201–225.
25. Dalziel,M., Nunes,N.M. and Furger,A. (2007) Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3′ end processing. *Mol. Cell Biol.*, **27**, 1568–1580.
26. Morillon,A., Karabetsou,N., Nair,A. and Mellor,J. (2005) Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol. Cell*, **18**, 723–734.
27. Kamath,R.S., Martinez-Campos,M., Zipperlen,P., Fraser,A.G. and Ahringer,J. (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome Biol.*, **2**, RESEARCH0002.
28. Blumenthal,T., Evans,D., Link,C.D., Guffanti,A., Lawson,D., Thierry-Mieg,J., Thierry-Mieg,D., Chiu,W.L., Duke,K., Kiraly,M. *et al.* (2002) A global analysis of Caenorhabditis elegans operons. *Nature*, **417**, 851–854.
29. Oka,T., Yamamoto,R. and Futai,M. (1997) Three vha genes encode proteolipids of Caenorhabditis elegans vacuolar-type ATPase. Gene structures and preferential expression in an H-shaped excretory cell and rectal cells. *J. Biol. Chem.*, **272**, 24387–24392.
30. Le Hir,H., Nott,A. and Moore,M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.*, **28**, 215–220.
31. Ashe,M.P., Griffin,P., James,W. and Proudfoot,N.J. (1995) Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.*, **9**, 3008–3025.
32. Sonnichsen,B., Koski,L.B., Walsh,A., Marschall,P., Neumann,B., Brehm,M., Alleaume,A.M., Artelt,J., Bettencourt,P., Cassin,E. *et al.* (2005) Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature*, **434**, 462–469.
33. Guiliano,D.B. and Blaxter,M.L. (2006) Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genet.*, **2**, e198.
34. Ganot,P., Kallesoe,T., Reinhardt,R., Chourrout,D. and Thompson,E.M. (2004) Spliced-leader RNA trans splicing in a chordate, Oikopleura dioica, with a compact genome. *Mol. Cell Biol.*, **24**, 7795–7805.
35. Davis,R.E. and Hodgson,S. (1997) Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in Schistosoma mansoni. *Mol. Biochem. Parasitol.*, **89**, 25–39.
36. Jager,A.V., De Gaudenzi,J.G., Cassola,A., D'Orso,I. and Frasch,A.C. (2007) mRNA maturation by two-step trans-splicing/polyadenylation processing in trypanosomes. *Proc. Natl Acad. Sci. USA*, **104**, 2035–2042.
37. Fong,Y.W. and Zhou,Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929–933.
38. Furger,A., O'Sullivan,J.M., Binnie,A., Lee,B.A. and Proudfoot,N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.*, **16**, 2792–2799.
39. Tange,T.O., Nott,A. and Moore,M.J. (2004) The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.*, **16**, 279–284.
40. Hachet,O. and Ephrussi,A. (2004) Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature*, **428**, 959–963.
41. Nott,A., Le Hir,H. and Moore,M.J. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.*, **18**, 210–222.
42. Liu,C., de Oliveira,A., Higazi,T.B., Ghedin,E., DePasse,J. and Unnasch,T.R. (2007) Sequences necessary for trans-splicing in transiently transfected Brugia malayi. *Mol. Biochem. Parasitol.*, **156**, 62–73.
43. Ashe,M.P., Furger,A. and Proudfoot,N.J. (2000) Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA*, **6**, 170–177.
44. Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
45. Ashe,M.P., Pearson,L.H. and Proudfoot,N.J. (1997) The HIV-1 5′ LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J.*, **16**, 5752–5763.
46. Nag,A., Narsinh,K. and Martinson,H.G. (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat. Struct. Mol. Biol.*, **14**, 662–669.
47. Dantonel,J.C., Murthy,K.G., Manley,J.L. and Tora,L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature*, **389**, 399–402.
48. Calvo,O. and Manley,J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol. Cell*, **7**, 1013–1023.
49. Venkataraman,K., Brown,K.M. and Gilmartin,G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.*, **19**, 1315–1327.
50. Rigo,F. and Martinson,H.G. (2009) Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. *RNA*, **15**, 823–836.
51. Lin,K. and Zhang,D.Y. (2005) The excess of 5′ introns in eukaryotic genomes. *Nucleic Acids Res.*, **33**, 6522–6527.