The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced

Christine N. Tennyson[1,2], Henry J. Klamut[1] & Ronald G. Worton[1,2]

The largest known gene is the human dystrophin gene, which has 79 exons spanning at least 2,300 kilobases (kb). Transcript accumulation was monitored from four regions of the gene following induction of expression in muscle cell cultures. Quantitative reverse transcription-polymerase chain reaction (RT-PCR) results indicate that approximately 12 h are required for transcription of 1,770 kb (at an average elongation rate of 2.4 kb min^{-1}), extrapolating to a transcription time of 16 h for the complete gene. Accumulation profiles for spliced and total transcript demonstrated that transcripts are spliced at the 5' end before transcription is complete providing strong evidence for cotranscriptional splicing. The rate of transcript accumulation was reduced at the 3' end of the gene relative to the 5' end, perhaps due to premature termination of transcription complexes.

[1]*Department of Genetics and Research Institute, Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada*
[2]*Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, M5G 1X8, Canada*

Correspondence should be addressed to R.G.W.

The human dystrophin gene encodes a high molecular weight cytoskeletal protein. Mutations in this gene are the cause of the X-linked lethal degenerative muscle disease, Duchenne muscular dystrophy (DMD), in which dystrophin is usually absent and its milder allelic variant, Becker muscular dystrophy (BMD), in which dystrophin is altered or present in reduced amount. The human gene spans at least 2,300 kb[1-3]. The 79 exons of the gene[1,4] encode a 14 kilobase (kb) mRNA[5], indicating that 99.4% of this gene is comprised of introns.

The extraordinary nature of the dystrophin gene with respect to size and exon number points to several problems concerning the production and processing of dystrophin transcripts. First, tremendous transcriptional effort is implied with a predicted transcription time of 14 to 24 h based on published elongation rates for RNA polymerase II[6,7]. Second, the primary transcript is potentially the same size as the gene and stabilization of such a large RNA molecule may be difficult. Third, the gene contains at least three internal promoters[8-10] downstream of the muscle promoter indicating that RNA polymerase must traverse several regions that may be bound to regulatory proteins involved in controling other transcripts. This may result in loss of transcriptional efficiency across the gene. Finally, the cell must ensure orderly splicing of 79 exons distributed over a great physical distance.

To assess how the cell handles these challenges, we have measured the time required for transcription, assessed whether transcripts are spliced as they are synthesized (cotranscriptionally) or only after transcription is complete (posttranscriptionally), and measured transcriptional efficiency at various places along the gene. This was done using quantitative RT-PCR to monitor the accumulation of dystrophin transcripts following induction of expression in differentiating clonal human myogenic cultures. In these cultures, low levels of dystrophin transcript are detectable in myoblasts by RT-PCR[11] and these levels increase significantly during differentiation into multinucleated myotubes[12,13]. These increases are due at least in part to a muscle-specific, differentiation-dependent promoter[14], providing an excellent system for our kinetic studies.

By comparing the time at which sequences begin to accumulate in transcripts from four different sites within the gene, we have found that approximately 12 h are required to transcribe 1,770 kb, extrapolating to a transcription time of approximately 16 hours for the entire gene. Therefore the gene appears to be transcribed in a predictable manner and not by a novel mechanism such trans-splicing of multiple short transcripts as occurs for some genes in *Caenorhabditis elegans*[15]. We also show that dystrophin transcripts are spliced at the 5' end before transcription is complete, providing perhaps the first strong evidence for cotranscriptional splicing in a mammalian system.

Dystrophin expression in human myogenic cells

Clonal human myogenic cells were plated at high density, incubated for 24 h in growth medium and transfered to fusion medium. Total RNA was isolated from cell cultures over the next 40 h as the cells fused to form multinucleated myotubes. Dystrophin transcript levels, determined by RT-PCR, increase dramatically after transition of the cells to fusion medium (Fig. 1). This is similar to the profile of dystrophin transcript expression observed with differentiating myogenic cultures prepared from rat muscle[12]

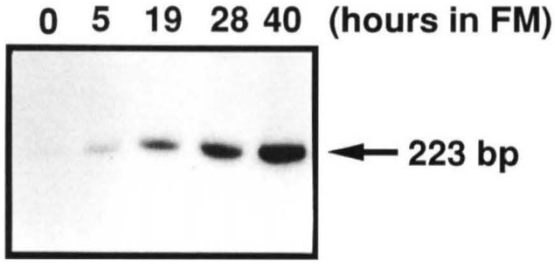

Fig. 1 Dystrophin transcript levels in differentiating clonal human myogenic cells. High density human myogenic cultures were allowed to differentiate and total RNA was prepared at the times shown following the transition to fusion medium (FH). From 1.5 υg of RNA, cDNA was made using primer 6R located in exon 6 of the dystrophin gene. One fifth of the cDNA was amplified for 21 cycles using end-labeled primer 4R and 1F. One-fifth of the PCR product was separated using an 8% non-denaturing polyacrylamide gel. The autoradiogram of the gel shows the 223 bp product. The location of primers 6R, 4R and 1F is provided in Fig. 3.

Quantitative RT-PCR of dystrophin transcript

To obtain accurate, quantifiable measurements of transcript levels by PCR, synthetic internal standards were reverse transcribed and amplified with the endogenous transcript using the same primers. Synthetic control RNA (cRNA) was identical in sequence to the endogenous transcript with the exception of a 46 nucleotide insertion (Fig. 2*a*). The small insertion did not significantly alter the efficiency with which the sequences were amplified (data not shown). A known amount of *in vitro* synthesized cRNA was mixed with a known amount of total RNA prepared from muscle cell cultures. Serial dilutions of the mixture were amplified by RT-PCR using a radiolabelled primer and the control and endogenous products were separated by non-denaturing polyacrylamide gel electrophoresis.

To determine the number of dystrophin transcripts per nanogram (ng) of total RNA, densitometry was used to quantitate the amount of control and endogenous products from an autoradiogram of the gel (Fig. 2*b*). The densitometric readings obtained for the control and endogenous amplification products were plotted against the amount of total muscle RNA (ng) and the number of molecules of cRNA (Fig. 2*c*). The parallel increase of control and endogenous products indicates that the reaction is in the linear range. In the example presented in Fig. 2*c*, 10 ng of total muscle RNA contains 3.85×10^4 molecules of dystrophin transcript or 3,850 molecules ng^{-1} total RNA. A typical yield of RNA from the myogenic cultures was 1 ng of RNA per 100 nuclei, therefore in this experiment there were approximately 38.5 molecules of dystrophin transcript per nucleus. This approach was used to determine the number of dystrophin transcripts for each set of PCR primers at each time point.

Transcription time for exons 3–69

To determine the time required to transcribe the dystrophin gene, transcript accumulation was monitored with primer pairs near the 5' and 3' ends of the gene. Cell samples were collected every 5–6 h over 40 h following transfer of human myogenic cells to fusion medium and

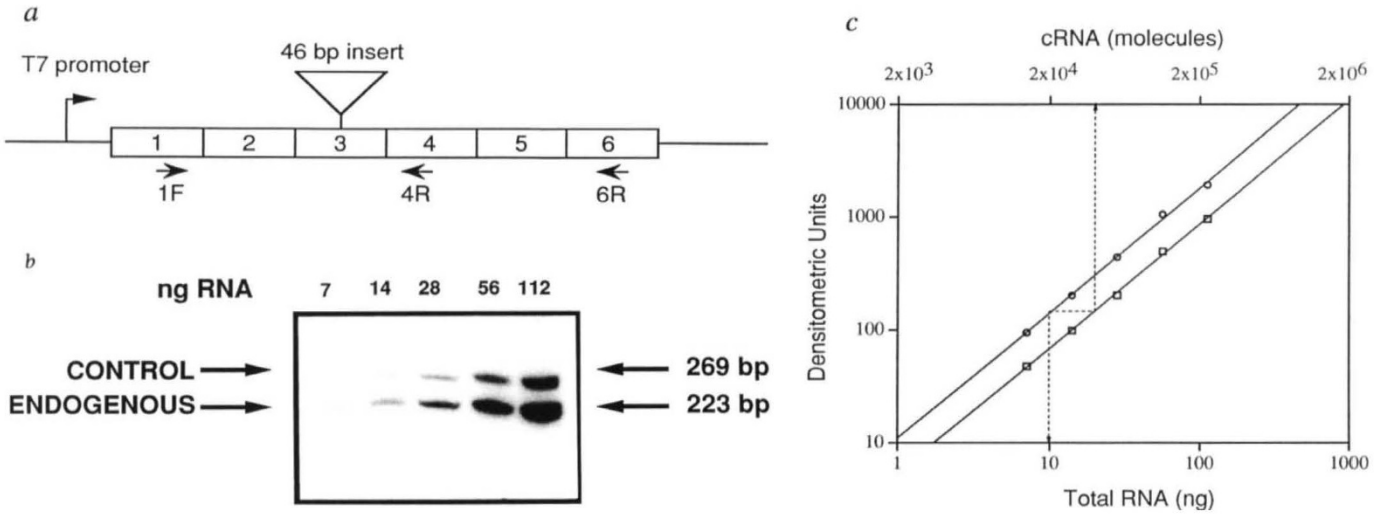

Fig. 2 Estimation of dystrophin transcript levels by quantitative RT-PCR. *a*, Schematic of a typical control template with a 46 bp insertion. To generate control templates dystrophin cDNA sequences from the region to be analysed were cloned immediately downstream of a T7 promoter. A 46 bp DNA fragment was inserted within an exon (rectangles). The position of the insertion permitted reverse transcription and amplification using primers within an exon (for detection of both spliced and unspliced transcript) or using primers in separate exons as shown (for detection of spliced transcripts only). *b*, Autoradiogram of RT-PCR amplified transcripts from control template and muscle cell RNA. Total RNA was isolated from myogenic cells 28 h after the transition to fusion medium. First strand cDNA was synthesized from a mixture containing 1.5 υg of total RNA and 3×10^6 molecules of cRNA (a ratio of 2,000 molecules cRNA per ng cellular RNA) by reverse transcription from primer 6R. cDNA was serially diluted 1:2 and an aliquot of each dilution was amplified for 23 cycles using primers 1F and 4R. One-fifth of the PCR product was separated on an 8% non-denaturing polyacrylamide gel. This autoradiogram shows the control (269 bp) and endogenous (223 bp) amplification products using cDNA from 7.0, 14, 28, 56 and 112 ng of muscle cell RNA. *c*, Densitometric units for control (open squares) and endogenous (open circles) amplification products were plotted against molecules of control RNA and the amount of endogenous RNA amplified. The dashed line indicates that 10 ng of total muscle RNA contains 3.85×10^4 molecules of spliced dystrophin transcript containing exons 1 to 6.

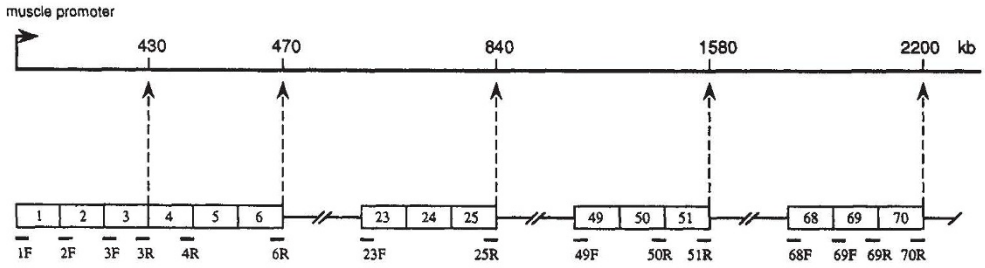

Fig. 3 Location of primers used to analyse dystrophin transcript levels by quantitative RT-PCR. This shows the 2300 kb dystrophin transcription unit in muscle as a line and shows the primers used for RT and PCR for different regions of the gene. The approximate genomic location of the primers used for RT is shown above the exons in which the primers are located[2]. The primer number indicates the exon in which the primer is located.

total RNA prepared for each time point. Total transcript levels (spliced plus unspliced transcripts) were measured using a primer pair in exon 3 (3F and 3R) and a second pair in exon 69 (69F and 69R) for quantitative RT-PCR (Fig. 3). Exon 69 was chosen as the downstream exon since previous studies have shown alternative splicing of more distal exons in muscle tissue[16,17].

The graph shown in Fig. 4*a* plots molecules of dystrophin transcript per nucleus against time in fusion medium. Transcripts containing exon 3 begin to accumulate 8.3 h following transfer of myogenic cells to fusion medium, and increase over the next 30 h. Transcripts containing exon 69 begin to accumulate later, approximately 20.4 h after transfer to fusion medium. The primer pairs in exon 3 and exon 69 are located approximately 430 and 2,200 kb downstream of the muscle promoter, respectively (Fig. 3)[2]. Therefore, approximately 12.1 h are required for RNA polymerase to transcribe 1770 kb of the dystrophin gene, indicating an average elongation rate of 2.4 kb min^{-1} (40 nt s^{-1}). By extrapolation, this implies a transcription time of approximately 16 h for the entire 2,300 kb transcription unit.

Recent studies have shown that dystrophin transcripts can be initiated from promoters located in introns 55 and 62 in tissues other than muscle[18–21]. In myogenic cultures, we did not detect expression from the intron 55 promoter using RT-PCR. However, over the period covered in our experiments, a low but constant level of transcript is expressed from the intron 62 promoter (data not shown).

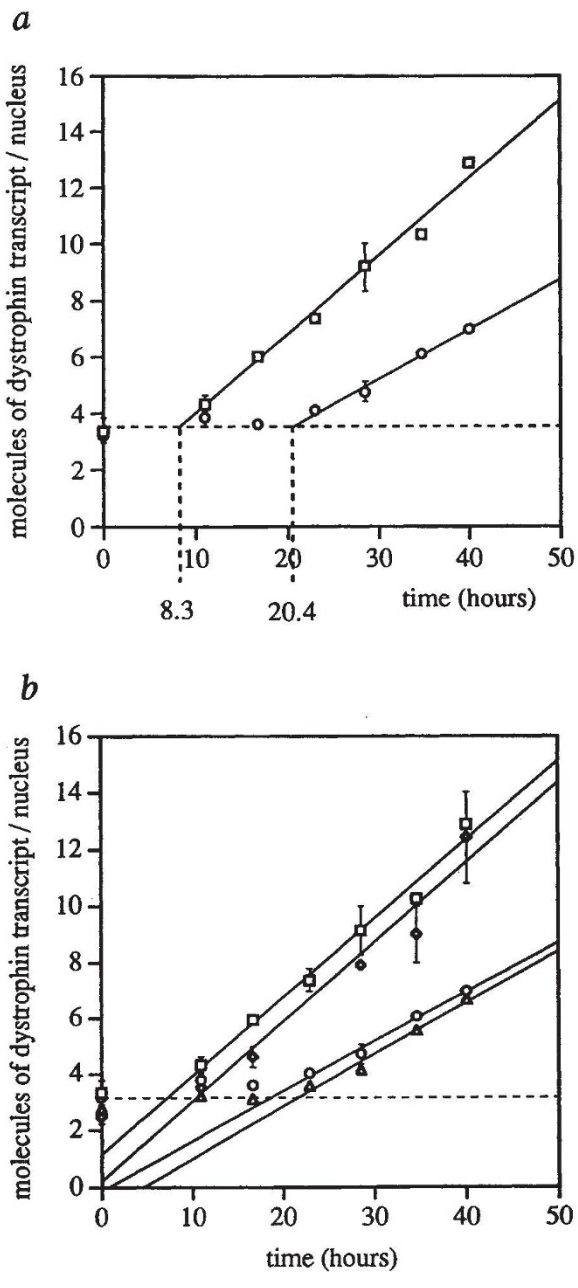

Fig. 4 *a*, Accumulation of total (spliced plus unspliced) transcript from the 5' and 3' ends of the dystrophin gene during myogenic differentiation. Total RNA was isolated from myogenic cultures at approximately 6 hour intervals during the early stages of differentiation and levels of exon 3-containing (open squares) and exon 69-containing (open circles) transcripts were measured by quantitative RT-PCR. This graph shows the average of two independent RT-PCR reactions on each RNA sample. The error bars show the difference between the measured values. A third trial using the same RNA gave a similar pattern but all absolute values were approximately 30% higher. To determine the time at which transcripts containing exon 3 begin to accumulate the last six time points were used to determine a best fit line which was extrapolated to the transcript level prior to induction (defined by the horizontal dashed line). A similar extrapolation for exon 69-containing transcripts was done using the last four time points. A linear regression line was determined given the apparent linearity of the transcript accumulation. *b*, Comparison of spliced and total (spliced plus unspliced) transcript accumulation at the 5' and 3' ends of the dystrophin gene during myogenic differentiation. The level of transcript containing exons 2 and 3 spliced together (open diamond) and exons 68 and 69 spliced together (open triangle) was measured and the results were overlaid on the graph in Fig. 4*a* which shows the level of total transcript containing exon 3 (open squares) and exon 69 (open circle). The rate of transcript accumulation was approximated by linear regression using the last six time points for exon 3-containing transcripts and the last four time points for exon 69-containing transcripts. Error bars as described in Fig. 4*a*. The dashed line represents the average basal level of spliced and total transcript.

Subtracting this transcript would lower the curve attained for the 3' end of the gene but would not alter the time at which transcripts begin to accumulate.

It is also clear from Fig. 4*a* that once transcription begins for exon 69, the rate of accumulation of exon 69-containing transcripts is less than the rate for exon 3-containing transcripts. Linear regression studies show that the rate of transcript accumulation (slope) is 0.28 and 0.18 molecules per nucleus per hour for exon 3- and exon 69-containing transcripts, respectively. Therefore, the rate of transcript accumulation toward the 3' end of the gene reaches only 64% of the rate attained near the 5' end of the gene. This implies that actively transcribing polymerases are more widely spaced at the 3' end of the gene as compared to the 5' end.

Timing of splicing relative to transcription

Comparing the accumulation of spliced transcript with that obtained for total (spliced plus unspliced) transcript from the two ends of the gene would indicate if dystrophin transcripts are generally spliced as exons become available (cotranscriptionally) or only after transcription is complete (posttranscriptionally). Spliced transcript levels near the 5' end of the gene were measured using primers 2F and 3R in adjacent exons (Fig. 3). At the 3' end of the gene, primers 68F and 69R (Fig. 3) were used to measure spliced transcript levels. Both spliced and total transcript containing exon 3 accumulate approximately 12 h before exon 69-containing transcript, indicating that intron 2 is spliced long before transcription is complete (Fig 4*b*). The time scale used in these experiments makes it difficult to determine whether intron 68 is also spliced cotranscriptionally or after transcription has been completed.

It is also possible to compare the rate of spliced transcript accumulation with the rate of total transcript accumulation. A significant difference between these rates would indicate a rate limiting splicing reaction. At the 5' end of the gene the rates (slopes) of total and spliced transcript accumulation are similar to one another (Fig. 4b). Although the rates are similar, the line representing the accumulation of spliced transcript is displaced to the right of the line representing total transcript by approximately 2–3 h. This may indicate that the initiation of splicing is delayed, although in these experiments interpretation of such time differences is difficult. Interestingly, similar statements can be made concerning the accumulation of spliced and total transcript at the 3' end of the gene. Therefore, the inclusion of the splicing event in the analysis does not significantly alter the measured rates of mature transcript accumulation but may slightly delay the time at which mature transcript begins to accumulate.

Splicing occurs 5' to 3' as predicted

To determine whether dystrophin transcripts are generally spliced in a 5' to 3' direction, the accumulation of spliced transcript from four sites within the dystrophin gene was monitored in differentiating myogenic cultures by RT-PCR. The four regions were exons 1–6, 23–25, 49–51 and 68–70, located approximately 470, 840, 1,580 and 2,200 kb downstream of the muscle promoter, respectively (Fig. 3). These distances were determined by the position of the primer used for reverse transcription as we monitored the accumulation of transcripts completed to that point. Figure 5*a* shows that spliced transcript accumulates first at the 5' end of the gene and at progressively later times as one moves further downstream from the muscle promoter. Specifically, extrapolation of the best fit lines to the basal transcript level indicates that transcript begins to accumulate from the four regions at approximately 6.2, 9.1, 14.0 and 20.2 h. This supports the proposal that splicing occurs in a 5' to 3' direction and over a time period consistent with cotranscriptional splicing. The distance between the regions and the time required for polymerase to traverse the gene segment was used to calculate the average elongation rate of RNA polymerase for each segment (Table 1). The rate of elongation ranges from 100–151 kb hr^{-1} or 1.7–2.5 kb min^{-1}, consistent with the value of 2.4 kb min^{-1} measured in the previous experiment.

Also apparent from Fig. 5a, and in agreement with the previous experiment, is that the rate of transcript accumulation (slope) is lower at the 3' end of the gene relative to the 5' end. This is clearly demonstrated in Fig. 5b where relative transcript accumulation rates are plotted as a function of position within the gene.

Table 1 Transcript elongation rate across the dystrophin gene

Region (exons)	Distance (kb)	Time (h)	Rate (kb per h)
6–25	370	2.9	128
25–51	740	4.9	151
51–70	620	6.2	100

Discussion

Time for transcription. Dystrophin gene transcription is upregulated as myogenic cultures undergo differentiation, providing a system for studying the kinetics of dystrophin transcription and splicing. By comparing the time at which transcripts begin to accumulate near the 5' and 3' ends of the gene we have shown that approximately 12 h are required to transcribe 1,770 kb (77%) of the dystrophin transcription unit. Thus, the average rate of elongation for RNA polymerase across the dystrophin gene from exon 3 to exon 69 is approximately 2.4 kb min^{-1} (40 nt s^{-1}), consistent with elongations rates for other genes[6,7]. Extrapolation using our rate of 2.4 kb min^{-1} implies that 16 h are required for polymerase to traverse the 2,300 kb dystrophin transcription unit in muscle.

The length of a transcription unit influences the time at which mature transcript becomes available for translation and the time at which transcript levels decline following transcriptional repression. In *Drosophila*, gene length is thought to contribute to the temporal regulation of the ecdysone inducible genes, *E74A* and *E74B*. These nested transcription units both respond to ecdysone but *E74B* mRNA accumulates before *E74A* as its transcription unit is much smaller[22,23]. Our results indicate that the length of the dystrophin gene contributes to a significant time delay in the appearance of dystrophin mRNA during muscle differentiation. It follows that a similar time delay would be expected following transcriptional repression.

Gene length is also thought to play a role in limiting transcript production from the developmental gene, *Ultrabithorax (Ubx)*, which is involved in the late stages of embryonic pattern formation in *Drosophila*. Studies have shown that expression of the 77 kb *Ubx* transcription unit requires 55 minutes and that expression is limited in early

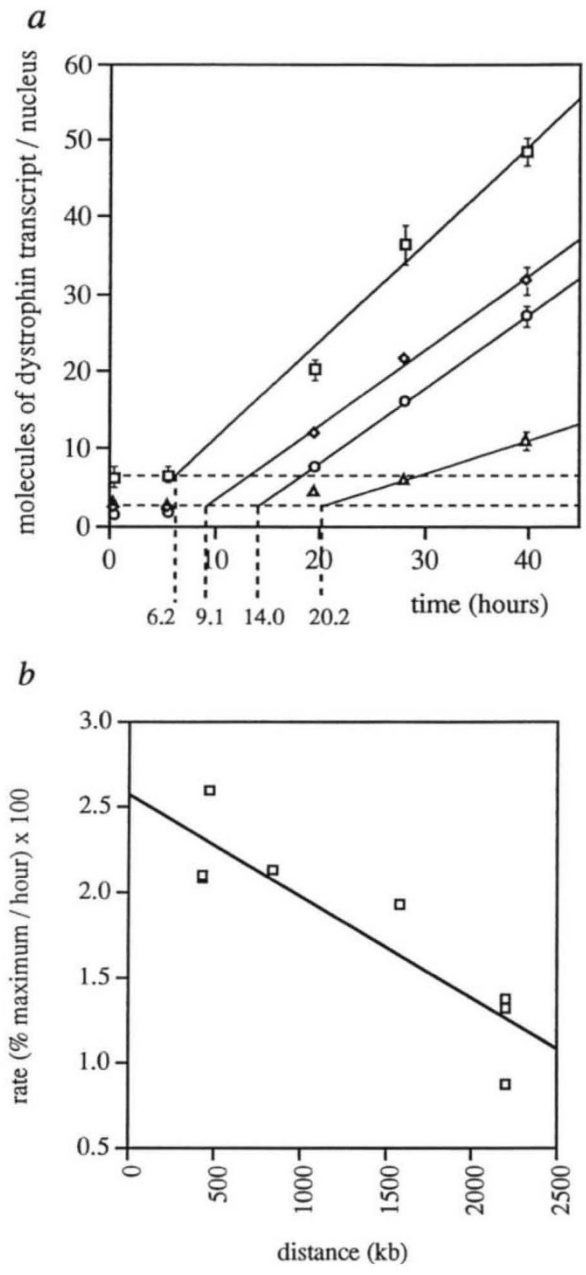

Fig. 5 *a*, Accumulation of spliced transcript from various regions of the dystrophin gene during myogenic differentiation. Total RNA was isolated from myogenic cultures over the course of differentiation and cDNA was made by reverse transcription using primers 6R, 25R, 51R and 70R (Fig. 3). cDNA was amplified using primer pairs 1F and 4R (open square), 23F and 25R (open diamond), 49F and 51R (open circle) and, 68F and 69R (open triangle). Each point represents the average of two independent quantitations by RT-PCR and the error bars show the difference between the two values. As in the previous experiment, linear regression was used to predict the approximate time at which transcript begins to accumulate at each point along the transcription unit. Note that in this experiment the basal level of transcript was higher for the exon 1 to 4 product than for the other three, necessitating the use of two different baselines for determination of the start of transcript accumulation. *b*, Rate of transcript accumulation across the dystrophin gene. This graph plots the rate of transcript accumulation (as determined from the slopes of the curves in Fig. 4*a*,*b* and 5*a*) against distance of the region from the muscle promoter. To compare the rates from different experiments the values in each experiment were converted to percent maximum and the rate determined for each data set.

Drosophila embryos due to disruption of transcription by cell division[24]. A similar mechanism could limit dystrophin transcript production in dividing myoblasts. The cell cycle time for normal human myoblasts in culture is approximately 16–17 h[25] similar to the time required to transcribe dystrophin. Previous studies have provided evidence for a low level of dystrophin transcript in myoblast cultures[12]. We have confirmed that low levels of dystrophin transcripts exist in our clonal human myoblast cultures. The low levels are determined in part by a promoter that is relatively weak in myoblasts[14]. However, our measurements of the transcription time also indicate that few initiated transcription complexes would complete transcription prior to a round of DNA synthesis and mitosis.

Cotranscriptional splicing. Our results indicate that sequences from the 5' end of the gene accumulate in both total and spliced transcript approximately 12 h before sequences from the 3' end of the gene. Therefore, it appears that the 5' end of the transcript is spliced before transcription is complete and before polyadenylation. Processing of transcripts involves several steps including capping, splicing and polyadenylation. Early studies showed that polyadenylation of viral transcripts precedes splicing[26–28]. This order of events is also seen for some cellular genes where the pattern of alternative splicing is determined by the poly(A) site that is selected, as in the case of transcripts from the IgM heavy chain gene[29]. Our results deviate significantly from the model that transcripts are synthesized, polyadenylated and then spliced, and provide perhaps the first evidence for cotranscriptional splicing in vertebrates.

Studies in other organisms, however, provide evidence for cotranscriptional splicing. Beyer and Osheim[30], for example, reported electron microscopic images of apparently spliced nascent RNAs from transcriptionally active genes of the early *Drosophila* embryo. LeMaire and Thummel[31] provided direct evidence for cotranscriptional splicing by monitoring the accumulation of spliced and unspliced transcript along the length of the *Drosophila* steroid responsive gene *E74A* using RNase protection. More recently it was shown that intron 3 of the *Balbiani ring 1* gene in the dipteran *Chironomus tentans* is cotranscriptionally spliced [32]. Finally, adenovirus RNA can be spliced in the absence of a poly(A) tail[33], indicating that polyadenylation is not required for splicing. Our finding that dystrophin transcripts are cotranscriptionally spliced is not surprising given the size of the gene and the number of exons, and demonstrates that this type of RNA processing can occur in human cells.

The mechanism that ensures orderly splicing of exons and prevents exon skipping is still unknown. Abei and Weissmann[34] proposed a 'first come, first served' mechanism of splicing in which splice site selection occurs on the nascent RNA by the formation of a committed splicing complex, which completes splicing at some later time. Our results indicate that for dystrophin transcripts, both splice site selection and splicing occur on the nascent transcript. This is an effective way to limit the number of

splice sites available for splicing and decrease the probability of incorrect splicing of 79 exons. Our analysis indicates that splicing generally occurs in a 5' to 3' direction on the dystrophin transcript. The order of splice sites is likely to be important in a global sense but it is possible that transcription of several exons at the 5' end may occur before splicing begins. In the case of the *E74A* transcripts, although they are cotranscriptionally spliced it appears that exons 2–5 are spliced before exons 1 and 2 (ref. 31). Therefore, other factors likely contribute to orderly splicing of exons including, for example, flanking sequences[35,36] and trans-acting factors[37].

Changes in polymerase density. Our results indicate a 30–40% decrease in the rate of transcript accumulation from the 5' to the 3' end of the gene. This could be due to a lower density of polymerases at the 3' end of the gene possibly caused by premature termination of transcription complexes. Although polymerase normally has a stable association with the DNA template, some DNA sequences can cause pausing and/or premature termination of some transcription complexes during elongation as displayed by the heat shock[38] and c-*myc* genes[39]. The exceptional length of the dystrophin gene may lead to an increased probability that the polymerase will encounter sequences that are difficult to transcribe and result in premature termination. Both the nature of the transcription complex and the DNA sequences being transcribed can influence the probability of stalling and/or premature termination[40].

Our data show a clear trend of decreased transcript accumulation rates toward the 3' end of the gene (Fig. 5*b*), but they are not extensive enough to determine if the loss of polymerase is stochastic or if the most significant loss occurs at one or a few sites along the gene. There does seem to be greater loss of transcriptional capacity in the interval from 1,500 to 2,200 kb. If this region is confirmed as a preferred site of polymerase release or stalling, it could be a consequence of the presence of downstream promoters in this region. These promoters, located in introns 55 and 62 are not highly expressed in muscle and are likely to contain sequences that bind transcriptional repressors or other factors that might interfere with the normal traverse of the muscle specific polymerase complex. Whatever the mechanism, it is perhaps not surprising that polymerase II has difficulty traversing a 2,300 kb gene in a continuous 16 h reaction. Indeed we consider it amazing that the transcriptional machinery is able to operate at all over this distance and time.

Methodology

Culture of human myogenic cells. Clonal populations of myogenic cells prepared from muscle of a 16 week human fetus were used for all differentiation experiments. The procedures to prepare clonal human myoblast cultures have been described[14]. A myogenic clone was expanded in growth medium (alpha minimal essential medium (GibcoBRL) containing 16 mM glucose, 10% fetal bovine serum, and 40 υg of gentamicin per ml) to approximately five million cells at low density to maintain a myoblast phenotype. The cells were then plated at high density (approximately 320 cells per mm^2) in growth medium, incubated for 24 h and then transfered to low serum fusion medium (alpha minimal essential medium (GibcoBRL) containing 16 mM glucose, 2% fetal bovine serum, and 100 U of penicillin and 0.1 mg of streptomycin per ml) to induce differentiation into multinucleated myotubes.

RNA preparation. Total RNA was isolated from differentiating myogenic cultures as described[41]. RNA concentration was determined from O.D. (260 nm) using a spectrophotometer.

Construction of control templates for *in vitro* transcription. A 46 bp DNA fragment was prepared from pUC19 by digestion with *Eco*RI and *Hin*dIII and treatment with mung bean nuclease to create blunt ends. The 46 bp fragment was inserted at the unique *Stu*I site in a plasmid encompassing dystrophin cDNA sequences 1–1538 (ATCC 57667). A *Hin*dII fragment encompassing dystrophin cDNA sequences 2839–4550 was prepared from the plasmid cDMD 4-5a (ATCC 57671) and cloned into the *Sma*I site of Bluescript (Stratagene) and the 46 bp fragment was inserted at the *Sty*I site. A *Hin*dIII fragment encompassing dystrophin cDNA sequences 7193–7800 from the plasmid cDMD 8 (ATCC 57675) was cloned into the *Hin*dIII site of Bluescript and the 46 bp insert was cloned into the *Bgl*II site. A DNA fragment encompassing dystrophin cDNA sequences 10120–10907 was obtained by digestion with *Hin*dII and ligated into the *Hin*dII site of Bluescript. The clone was digested with *Pst*I to remove a fragment with an unwanted *Nco*I site and religated. The fragment removed was not within the PCR amplified region. The 46 bp fragment was inserted at the remaining *Nco*I site. All clones were sequenced to confirm the presence of the insertion and the integrity of the surrounding sequences.

Preparation of cRNA from control template. Control template DNA was isolated by alkaline lysis and purified using equilibrium centrifugation in CsCl as described[42]. Control constructs encompassing cDNA regions 1–1538, 2839–4550, 7193–7800 and 10120–10907 were linearized using *Sty*I, *Bgl*II, *Cla*I, and *Bgl*II, respectively. Control RNA (cRNA) was generated in a 20 υl reaction using the T7 MEGAscript kit (Ambion) and treated with DNase I to remove the template DNA. To this was added 0.5 ml of guanidine solution (19 ml 7.5 M guanidine hydrochloride plus 1 ml 2 M potassium acetate, pH 5.5, and DTT added to 10 mM prior to use) and 0.25 ml ethanol and the solution stored at –20 °C overnight. The cRNA was collected by centrifugation in a microfuge, resuspended in 100 υl of 10 mM EDTA, pH 7, and extracted with chloroform-butanol (4:1). The cRNA was precipitated, resuspended in 50 υl of DEP-treated water and spun through an RNase-free G-50 sephadex Quick Spin column (Boehringer Mannheim). The concentration of cRNA was determined from O.D. (260 nm) using a spectrophotometer.

Quantitative RT-PCR. Quantitative RT-PCR was performed as described[43]. The products were separated by non-denaturing polyacrylamide gel electrophoresis. The gel was dried and autoradiograms at different exposure times were obtained. The relative amount of control and endogenous products was determined by scanning laser densitometry. When necessary, control reactions which exclude reverse transcriptase were run to ensure no detectable levels of DNA in the RNA samples. Primer sequences are as follows: (1F-GGTGGGAAGAAGTAGAGGACT; 2F-GGGTAAATGCACAA-TTTTCTAAG; 3F-TTGGGAAGCAGCATATTGAG; 3R-CAG-TTTTTGCCCTGTCAGGC; 4R-GACATTGTTCAGGGCATGAAC; 6R-TTAACCTGTGGATAATTACG; 23F-GAAGAAATTGAGGGA-CGCTG; 25R-ACACTGTTTAGACTGGGCTG; 49F-TTGACTTTA-TCGTCAAGTTC; 50R-CAATAGTGGTCAGTCCAGGAGC; 51R-TTAACCTGTGGATAATTACG; 68F-GCAAAGAGTGTCCAAT-CATT; 69F-TACAGGAGTCTAAAGCACTT; 69R-GGAGTGCAA-TATTCCACCAT; 70R-GGGATGCTTC GCAAAATACC).

Acknowledgements

We thank P.N. Ray for helpful discussions, L.O. Bosnoyan for technical assistance in establishing the myogenic clones, and C.L. Duff for sequencing the control constructs. This work was supported by grants from the Medical Research Council of Canada. C.N.T. was the recipient of a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada.

Received 5 October; accepted 29 November 1994.

1. Den Dunnen, J.Y. *et al.* Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am. J. hum. Genet.* **45**, 835–847 (1989).
2. Coffey, A.J. *et al.* Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. *Genomics* **12**, 474–484 (1992).
3. Monaco, A.P., Walker, A.P., Millwood, I., Larin, Z. & Lehrach, H. A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. *Genomics* **12**, 465–473 (1992).
4. Roberts, R.G., Coffey, A.J., Bobrow, M. & Bentley, D.R. Determination of the exon structure of the distal portion of the dystrophin gene by vectorette PCR. *Genomics* **13**, 942–950 (1992).
5. Koenig, M., Monaco, A.P. & Kunkel, L.M. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53**, 219–228 (1988).
6. Ucker, D.S.& Yamamoto, K.R. Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. *J. biol. Chem.* **259**, 7416–7420 (1984).
7. Sehgal, P.B., Derman, E., Molloy, G.R., Tamm, I. & Darnell, J.E. 5-6-Dichloro-1-B-D-ribofuranosylbenzimide inhibits initiation of nuclear heterogeneous RNA chains in Hela cells. *Science* **194**, 431–433 (1976).
8. Lederfein, D., Yaffe, D. & Nudel, U. A housekeeping type promoter, located in the 3' region of the Duchenne muscular dystrophy gene, controls the expression of Dp71, a major product of the gene. *Hum. molec. Genet.* **2**, 1883–1888 (1993).
9. Gorecki, D.C. *et al.* Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Hum. molec. Genet.* **1**, 505–510 (1992).
10. Byers, T.J., Lidov, H.G.W. & Kunkel, L.M. An alternative dystrophin transcript specific to peripheral nerve. *Nature Genet.* **4**, 77–81 (1993).
11. Chelly, J., Montarras, D., Pinset, C., Berwald-Netter, Y. & Kaplan, J.C. Quantitative estimation of minor mRNAs by cDNA-polymerase chain reaction: application to dystrophin mRNA in cultured myogenic and brain cells. *Eur. J. Biochem.* **187**, 691–698 (1990).
12. Nudel, U., Robzyk, K. & Yaffe, D. Expression of the putative Duchenne muscular dystrophy gene in differentiated myogenic cell cultures and in the brain. *Nature* **331**, 635–638 (1988).
13. Lev, A.A., Feener, C.C., Kunkel, L.M. & Brown, R.H. Expression of the Duchenne's muscular dystrophy gene in cultured muscle cells. *J. biol. Chem.* **262**, 15817–15820 (1987).
14. Klamut, H.J., Gangopadhyay, S.B., Worton, R.G. & Ray, P.N. Molecular and functional analysis of the muscle-specific promoter region of the Duchenne muscular dystrophy gene. *Molec. cell. Biol.* **10**, 193–205 (1990).
15. Krause, M. & Hirsh, D. A trans-spliced leader sequence on actin mRNA in C. elegans. *Cell* **49**, 753–761 (1987).
16. Feener, C.A., Koenig, M. & Kunkel, L.M. Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. *Nature* **338**, 509–511 (1989).
17. Bies, R.D. *et al.* Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart & brain development. *Nucl. Acids Res.* **20**, 1725–1731 (1992).
18. Bar, S. *et al.* A novel product of the Duchenne muscular dystrophy gene which greatly differs from known isoforms in its structure and tissue distribution. *Biochem. J.* **272**, 557–560 (1990).
19. Byers, T.J., Lidov, H.G.W. & Kunkel, L.M. An alternative dystrophin transcript specific to peripheral nerve. *Nature Genet.* **4**, 77–81 (1993).
20. Hugnot, J.P. *et al.* Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in non-muscle tissues. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7506–7510 (1992).
21. Lederfein, D. *et al.* A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other non-muscle tissues. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5346–5350 (1992).
22. Thummel, C.S., Burtis, K.C., Hogness, D.S. Spatial and temporal patterns of E74 transcription during drosophila development. *Cell* **61**, 101–111 (1990).
23. Karim, F.D. & Thummel, C.S. Ecdysone coordinates the timing and amounts of E74A and E74B trancription in Drosophila. *Genes Dev.* **5**, 1067–1079 (1991).
24. Shermoen, A.W. & O'Farrell, P.H. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303–310 (1991).
25. Blau, H.M., Webster, C., Pavlath, G.K. & Chiu, C. Evidence for defective myoblasts in Duchenne muscular dystrophy. *Adv Exp. Medicine Bio.* **182**, 85–110 (1985).
26. Nevins, J.R. & Darnell, J.E. Steps in the processing of Ad2 mRNA: Poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell* **15**, 1477–1493 (1978).
27. Ford, J.P. & Hus, M. Transcription pattern of in vivo-labeled late simian virus 40 RNA: equimolar transcription beyond the mRNA 3' terminus. *J. Virol.* **28**, 795–801 (1978).
28. Weber, J., Blanchard, J.M., Binsberg, H., Darnell, J.E. Order of polyadenylic acid addition and splicing events in early adenovirus mRNA formation. *J. Virol.* **33**, 286–291 (1980).
29. Galli, G., Guise, J., Tucker, P.W., Nevins, J.R. Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2439–2443 (1988).
30. Beyer, A.L. & Osheim, Y.N. Splice site selection, rate of splicing & alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754–765 (1988).
31. LeMaire, M.F. & Thummel, C.S. Splicing precedes polyadenylation during drosophila E74A transcription. *Molec. cell Biol.* **10**, 6059–6063 (1990).
32. Bauren, G. & Wieslander, L. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183–192 (1994).
33. Zeevi, M., Nevins, J.R. & Darnell, J.E. Jr. Nuclear RNA is spliced in the absence of poly (A) addition. *Cell* **26**, 39–46 (1981).
34. Abei, M. & Weissmann, C. Precision and orderliness in splicing. *Trend Genet.* **3**, 102–107 (1987).
35. Reed, R. & Maniatis, T. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **46**, 681–690 (1986).
36. Solnick, D. Alternative splicing caused by RNA secondary structure. *Cell* **43**, 667–676 (1985).
37. Hedley, M.L. & Maniatis, T. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell* **65**, 579–586 (1991).
38. Lee, H., Kraus, K.W., Wolfner, M.F. & Lis, J.T. DNA sequence requirements for generating paused polymerase at the start of hsp70. *Genes Dev.* **6**, 284–295 (1992).
39. London, L., Keene, R.G. & Landick, R. Analysis of premature termination in c-myc during transcription by RNA polymerase II in a Hela nuclear extract. *Molec. cell. Biol.* **11**, 4599–4615 (1991).
40. Spencer, C.A. & Groudine, M. Transcription elongation and eukaryotic gene regulation. *Oncogene* **5**, 777–785 (1990).
41. MacDonald, R.J., Swift, G.H., Przybyla, A.E. & Chirgwin, J.M. Isolation of RNA using guanidinium salts. *Meth. Enzymol.* **152**, 219–227 (1987).
42. Maniatis, T., Fritsch, E.F. & Sambrook, J. *Molecular cloning: a laboratory manual.* (Cold Spring Harbor Laboratory, New York, 1989).
43. Wang, A.M., Doyle, M.V. & Mark, D.F. Quantitation of mRNA by the polymerase chain reaction. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9717–9721 (1989).