Zebrafish dystrophin and utrophin genes: Dissecting transcriptional expression during embryonic development

DANIEL LAI\textsuperscript{1}, CHUAN-CHING LAN\textsuperscript{1}, IVONE UN SAN LEONG\textsuperscript{1} and DONALD R. LOVE\textsuperscript{1,2}

\textsuperscript{1}School of Biological Sciences, University of Auckland, Auckland 1142; \textsuperscript{2}Diagnostic Genetics, LabPlus, Auckland City Hospital, Auckland 1148, New Zealand

Received October 21, 2011; Accepted November 23, 2011

DOI: 10.3892/ijmm.2011.865

Abstract. Some genes can encode multiple overlapping transcripts, and this can result in challenges in identifying transcript-specific developmental expression profiles where tools such as RNA in situ hybridisations are inapplicable. Given this difficulty, we have undertaken a preliminary analysis of the developmental expression profile of selected transcripts of the dystrophin and utrophin genes of the zebrafish (\textit{Danio rerio}) by targeting unique and common regions of each of these transcripts. The dystrophin and utrophin genes of zebrafish were identified by bioinformatic analysis and the dystrophin gene predictions were confirmed by transcript sequencing. These data enabled primer pairs to be designed in order to determine the expression profiles of unique, but overlapping transcripts, throughout embryonic development using quantitative real time reverse transcription PCR (qRT-PCR). The data indicated the early expression of the short carboxyl-terminal dystrophin transcript, with expression of the full length muscle transcript occurring during myogenesis. Importantly, a composite of these two profiles appeared to comprise the major transcriptional load of the zebrafish dystrophin gene. In contrast, utrophin gene expression was dominated by the full length transcript throughout embryogenesis. The approach described here provided a means by which a gene’s transcriptional complexity can be deconvoluted to reveal transcriptional diversity during embryogenesis. This approach, however, required the identification of unique regions for transcript-specific targeting, and an appreciation of alternative splicing events that may compromise the design of primers for qRT-PCR.

Introduction

The human dystrophin gene spans 2.5 million base pairs and is the largest gene in the human genome (1). The dystrophin gene contains at least 7 known promoters that drive the expression of multiple transcripts, together with multiple splice variants (Fig. 1). The largest and most studied isoform of dystrophin, designated Dp427m, is expressed as a 427 kDa protein in human muscle. This protein comprises 3685 amino acids and plays a pivotal role in maintaining the integrity of the dystrophin:glycoprotein complex that connects cytoskeletal actin to the extracellular matrix of muscle cells. The less abundant isoforms of dystrophin are localised to other tissues such as the brain (Dp427b/c and Dp427p), the retina (Dp260), brain and kidney (Dp140), and Schwann cell/peripheral nerve (Dp116); the smallest isoform, Dp71, is ubiquitously expressed (2,3). The dystrophin gene has an equally large paralogue, termed utrophin (4,5), which expresses multiple transcripts (2,3,6-18) (Fig. 1). The full length utrophin isoform is able to fulfil the role of dystrophin when overexpressed in the absence of dystrophin (19).

Human dystrophin orthologues have been identified in many animal and insect species to date, ranging from mammals such as dog (20), cat (21) and mouse (22) to fish such as pufferfish (23) and zebrafish (24-26), together with frog (27), sea urchins (28) and fruit flies (29). Utrophin has been confirmed in many of the species in which dystrophin has been identified (4,30,31).

Importantly, Duchenne and Becker muscular dystrophies (D/BMD) are caused by mutations in the dystrophin gene [reviewed by Manzur \textit{et al} (32)] and several models of D/BMD have been constructed using mice (33) and dogs (20), among others. In addition, there is increasing interest in using the zebrafish to model muscular dystrophies such as DMD (26,34-36).

Given the transcriptional complexity of the dystrophin gene and its role in diseases, mutations generally impact not only on Dp427m but also other dystrophin isoforms. In the absence of dystrophin expression, upregulation of full-length utrophin appears to occur (19). Against this background, little is known of the complexity of the developmental expression profile of dystrophin, as well as the utrophin, gene transcripts. In order to address this deficiency, we turned our attention to the zebrafish (\textit{Danio rerio}). Zebrafish offers advantages over other model organisms in that they are optically transparent during embryonic development and ex utero development allows for easy manipulations (34,36-38). The main qualities of using zebrafish are their rapid generation cycles, low maintenance requirements and cost effectiveness at high densities.
The aim of our study was to provide for the first time a developmental expression profile of dystrophin and utrophin gene transcripts. Expression profiling posed challenges due to the transcriptional complexity of both genes: nucleotide sequences of each gene's transcripts are largely identical (albeit varying in length), and any differences are largely confined to short unique 5'-end sequences. These challenges highlighted the difficulty in applying a conventional RNA in situ hybridisation approach, which also suffers from sensitivity issues, therefore we targeted multiple exons in a quantitative real-time PCR (qRT-PCR) strategy.

Materials and methods

Primers. Primers were designed using the free online primer design software Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Advanced settings were used where appropriate to minimise the formation of secondary structures while reducing non-specific products being amplified by the primers. All primers designed were subjected to manual quality analysis using NetPrimer (http://www.premierbiosoft.com/netprimer/). qRT-PCR primers (Table I) were designed to produce product sizes between 100-160 base pairs with an approximate Tm of 60°C. All primers were designed so that the product would encompass at least 2 exons, with preference given to primers that annealed to exon/exon boundaries, providing that primer quality was not compromised.

Primers were ordered from Invitrogen Corporation (http://www.invitrogen.com) with all primers used for qRT-PCR subjected to quality checks. These checks involved the confirmation of only one PCR amplification product following electrophoresis on a 1.5% agarose 1X Tris/Borate/EDTA (TBE) gel stained with ethidium bromide. All PCR products and excised gel bands were subjected to spin-column purification using a GE Healthcare Illustra GFX PCR DNA and Gel Band Purification kit according to the manufacturer's recommended instructions (GE Healthcare Life Sciences).

Sequencing. Sequencing reactions were carried out by the Genomics Unit, School of Biological Sciences, University of Auckland. The ABI PRISM™ BigDye Terminator sequencing kit version 3.1 was used to incorporate fluorescent dyes as recommended by the manufacturer. Thermal cycling was carried out using an Applied Biosystems GeneAmp® PCR System 9700. Unincorporated fluorescent dyes were removed using Agencourt® CleanSEQ® magnetic beads. Capillary electrophoresis of the sequencing products was carried using the ABI PRISM™ 3130xl Genetic Analyzer.

Zebrafish husbandry. Wild-type zebrafish (Hollywood Fish Farms, Auckland, New Zealand) were grown and maintained at densities of 3-8 fish per 2.75 liters tank, in a water recirculation rack system (Aquatic Habitats) located inside a dedicated zebrafish facility. The zebrafish were exposed to 14 h of light per day with temperatures kept between 26-28°C. Water quality was monitored on a daily basis. Adult fish were fed a range of dry fish food supplemented with live Artemia, and juveniles were fed a mix of dry fish food and live Rotifer until at least two weeks of age. Adult male and female fish were...
kept mixed until one week prior to spawning at which point they were separated by gender. Embryos were collected after spawning and thoroughly rinsed. The embryos were visually inspected to ensure that they were at the same developmental stage and then kept immersed in 0.5X E3 media (1X E3: 5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl$_2$ and 0.33 mM MgSO$_4$) in 15 mm plastic Petri dishes at a density of 20/dish and kept at a constant 28˚C. Embryo E3 media was changed every 24 h.

Developmental time course. The zebrafish embryonic developmental stages described in this paper are reported as hours post fertilisation (hpf) at 28˚C: 3.5 hpf (high stage), 10 hpf (bud stage), 11 hpf (3-somite stage), 14 hpf (10-somite stage), 16 hpf (14-somite stage), 19.5 hpf (21-somite stage), 22 hpf (26-somite stage), 25 hpf (prim-6 stage), 31 hpf (prim-16 stage), 42 hpf (high pec stage), 60 hpf (pec fin stage) and 72 hpf (protruding mouth stage).

RNA extraction and purification. At each appropriate developmental stage, 20 embryos were inspected to ensure the correct developmental stage was reached, then they were pooled and snap frozen. RNA extraction and purification was carried out according to the protocol described earlier (39).

Reverse transcription. One microgram total-RNA from each extraction was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen Corporation) primed with random hexamers (Invitrogen Corporation) in a 20 µl reaction volume according to the manufacturer's instructions. The final SuperScript III inactivation step (70˚C for 15 min) was omitted and cDNAs were diluted 1:16 before use in qRT-PCR assays.

Quantitative real-time reverse transcription polymerase chain reaction. All qRT-PCR assays were carried out using an Applied Biosystems 7900HT platform running SDS v2.3 software (Applied Biosystems). Reactions were set up using an Eppendorf epMotion 5075 (Eppendorf) automated liquid handling platform to minimise pipetting errors. Ten microliter reaction volumes comprised the following: 5 µl Platinum SYBR-Green qPCR SuperMix-UDG with ROX (Invitrogen Corporation), 1 µl of 2 µM forward primer, 1 µl of 2 µM reverse primer and 2.6 µl of template; the template was either 16X diluted cDNA, or water in the case of negative controls. The expression of two zebrafish reference genes were used as normalisers, $Rpl13a$ and $EF1a$ (40), with all reactions performed in triplicate. Cycling programs consisted of 40 cycles of 15 sec at 94˚C and 60 sec at 60˚C. Dissociation curve analysis was performed for each qRT-PCR run.

qRT-PCR data analysis. The data obtained from the SDS v2.3 software were inspected manually to omit any well that showed abnormal amplification curves, followed by manual baseline and threshold setting adjustments as recommended by the SDS manual. Amplification efficiencies were determined using LinRegPCR v11.1 (41,42). Baseline and threshold corrected Ct values, together with amplification efficiencies, were processed as described in the geNorm v3.5 manual (version: July 8, 2008) (43). This processing provided relative expression levels (arbitrary quantities) while allowing for normalisation against the two reference genes.

Results

Dystrophin and utrophin sequences. At the start of this study, the available sequences for zebrafish dystrophin were fragmented and incomplete. These limited data were used to design primers in order to amplify overlapping fragments of the full length zebrafish dystrophin muscle transcript, which

Table I. Primers used for qRT-PCR.

<table>
<thead>
<tr>
<th>Primer designations</th>
<th>Target Gene</th>
<th>Target exons</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFdysX01F5</td>
<td>Dystrophin</td>
<td>1-2</td>
<td>GAAGCAGTGAAACCACAGGAC</td>
</tr>
<tr>
<td>ZFdysX02R5</td>
<td>Dystrophin</td>
<td>2-3</td>
<td>TGAGTTATCCACTTTTGGTGAGG</td>
</tr>
<tr>
<td>ZFdysX02F1</td>
<td>Dystrophin</td>
<td>75-76</td>
<td>CGGCACAAGAGAACAGCA</td>
</tr>
<tr>
<td>ZFdysX03R1</td>
<td>Dystrophin</td>
<td>75-76</td>
<td>GAGGCAAGGAGGAGGA</td>
</tr>
<tr>
<td>ZFdysX75F5</td>
<td>Dystrophin</td>
<td>Dp71-63</td>
<td>GCTTTTTCCACTGAGAGGAG</td>
</tr>
<tr>
<td>ZFdysX76R5</td>
<td>Dystrophin</td>
<td>Dp71-63</td>
<td>TGGTTTGGTCGTGGTATT</td>
</tr>
<tr>
<td>ZFdysDp71X01F2</td>
<td>Dystrophin</td>
<td>1-2</td>
<td>CTGACGAACTGACGTTGA</td>
</tr>
<tr>
<td>ZFdysX63R2</td>
<td>Dystrophin</td>
<td>1-2</td>
<td>AGATTCTGACCAGTGAG</td>
</tr>
<tr>
<td>ZFutrX01F1</td>
<td>Utrophin</td>
<td>69-70</td>
<td>AGGCTGACGTCAGAAAAC</td>
</tr>
<tr>
<td>ZFutrX02R3</td>
<td>Utrophin</td>
<td>69-70</td>
<td>TCTGGAGGACTTGAAAGAG</td>
</tr>
<tr>
<td>ZFutrX60F1</td>
<td>Utrophin</td>
<td>69-70</td>
<td>AGGAGGAGGAGAGGAGG</td>
</tr>
<tr>
<td>ZFutrX70R1</td>
<td>Utrophin</td>
<td>69-70</td>
<td>TCTGGAGGACTTGAAAGAG</td>
</tr>
<tr>
<td>EF1aF</td>
<td>Elongation factor 1α</td>
<td></td>
<td>ATCAAGAAGAGATGACCCGCTAGCATA</td>
</tr>
<tr>
<td>EF1aR</td>
<td>Elongation factor 1α</td>
<td></td>
<td>ATCAAGAAGAGATGACCCGCTAGCATA</td>
</tr>
<tr>
<td>Rpl13aF</td>
<td>Ribosomal protein L13α</td>
<td></td>
<td>TCTGGAGGACTTGAAAGAG</td>
</tr>
<tr>
<td>Rpl13aR</td>
<td>Ribosomal protein L13α</td>
<td></td>
<td>TCTGGAGGACTTGAAAGAG</td>
</tr>
</tbody>
</table>
Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m.
Figure 2. Continued.
Figure 2. Continued.
Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m. (A) Dp427m nucleotide sequence based on double-stranded sequencing of overlapping amplicons derived from reverse-transcribed RNA isolated from zebrafish embryos. Untranslated regions are shown in red, translated regions in blue, bases at exon-exon boundaries are underlined in turquoise and shaded in yellow, and translation start and stop codons are shaded in blue. (B) The translated protein sequence.

Figure 3. Differences between the actual and predicted sequences of zebrafish Dp427m. Five regions of the dystrophin amino acid sequence are shown that differ between the Vega predicted zebrafish dystrophin sequence (Vega Transcript ID: OTTDART00000044458; Vega Zdys) and that predicted from our cDNA sequence (Zdys).
were subsequently sequenced. The entire experimentally determined nucleotide sequence together with the predicted amino acid sequence are shown in Fig. 2. During the course of this study, updates to the zebrafish genome sequence assembly allowed for a more accurate prediction of the full-length dystrophin gene; however, while this predicted sequence is largely correct, there are a few regions in the distal half of the full-length transcript that do not agree with our sequence data (Fig. 3). This discrepancy may be due to alternative splicing or minor sequencing artefacts. Our sequence also differs from that reported by Guyon et al. (44), which is a composite of direct cDNA sequence data and EST sequences. This earlier sequence carries an additional exon (number 72, compared to annotations of the human Dp427m transcript, and so probably represents an alternatively spliced transcript), but also 45 amino acids at positions 495-539 (inclusive), which appears to be a repeat of amino acids at positions 585-629 (inclusive).

The utrophin gene proved difficult to predict using bioinformatic approaches, and the sequences appeared to be fragmentary using the currently available zebrafish genome builds. The utrophin sequence presented here is based on an incomplete utrophin gene prediction, which primarily covers the amino- and carboxy-terminal regions (Fig. 4). A comparison of the predicted exons for zebrafish utrophin and those from human utrophin via ClustalX DNA alignment showed 64.3 and 55.9% of the base pairs to be identical for the amino- and carboxy-terminal regions, respectively.

**Developmental expression profiling.** Expression profiling of the zebrafish dystrophin gene involved targeting four distinct regions of this gene in order to allow for the simultaneous analysis of its transcriptional diversity; our own sequence data was used for designing most of the primers. The muscle-specific first exon was targeted to follow the expression of the muscle isoform, the second exon of Dp427m allowed the profiling of all full-length transcripts, the expression of Dp71 was followed by targeting this transcript's unique first exon (determined bioinformatically), and the full transcriptional load of the dystrophin gene was determined by targeting exons 75-76 (Fig. 5).

The expression profile of Dp427m approximates that of total full-length dystrophin transcripts (Fig. 5A and B) with expression starting at 16 hpf, coinciding with myogenesis, and a reduced rate of expression after 30 hpf. These data suggest that for the most part, the Dp427m isoform is dominant among full-length isoforms expressed during zebrafish embryo development. The expression profile of the ubiquitous Dp71 isoform (Fig. 5C) suggests maternal inheritance as post-zygotic transcription begins at 3-3.5 hpf (45). Dp71 transcripts decrease to their lowest level at approximately 16 hpf, and are then largely maintained at a low steady state level. Interestingly, the expression profile of total dystrophin gene transcripts (Fig. 5D) appears to be a composite of Dp71 and Dp427m expression profiles.

Total utrophin gene transcript expression and full-length Utr427 expression exhibit similar profiles (Fig. 5E and F), suggesting that the full length Utr427 is the dominant isoform that is expressed by the utrophin gene during embryonic development. The data suggest maternal inheritance of Utr427 transcripts with low level de novo expression starting at approximately 16 hpf.
**Discussion**

The data presented here used an approach that allowed for the simultaneous expression profiling of the Dp427m transcript, all full-length dystrophin transcripts, the Dp71 transcript, and all transcripts expressed by the dystrophin gene. This approach involved targeting common exons that are expressed among isoforms, and targeting unique exons to allow for the profiling of specific isoforms. In addition, the profiles of all full length utrophin gene transcripts, as well as all transcripts expressed by the utrophin gene, were examined throughout zebrafish embryonic development. In terms of the former, it was considered critical to avoid exons involved in alternative splicing, which comprised exons 4 (46), 39, 41-42, 45-50 (47), 66-68 and 71-74 (24,25,48) and 78 (49). As such, the applicability of the approach described here requires knowledge of alternative splicing of target genes. The results reflect expression across the whole embryo as development progresses, and represent an average across all tissues and cell types.

The total expression profile of the dystrophin gene comprises that of Dp71 and the full length expression profiles, with Dp71 being the most abundant isoform present during the early stages of embryonic development, up until 16 hpf. From 16 hpf onwards, full length dystrophin expression appears to be present in higher abundance relative to other dystrophin isoforms. It should be noted that while other potential dystrophin transcripts were not investigated, they do not appear to be major players based on the expression profiles reported here. Experimental validation of this conclusion would be difficult in the context of qRT-PCR analysis as the sequences of the unique first exons of other dystrophin isoforms thought to exist are currently unknown or are too short to design transcript-specific primers.

In contrast to dystrophin, in which the predominantly expressed isoforms during embryonic development change from Dp71 to the full length dystrophin, the dominant isoform of utrophin that is expressed throughout development appears to be the full length Utr427. This conclusion does not rule out the presence of other as yet uncharacterised isoforms, but their expression levels would not be expected to be significant given our analysis herein.

The caveat to looking at gene expression of a whole embryo, as opposed to individual tissues, is that an absence of noticeable expression changes between two developmental stages does not imply the absence of differences between individual tissues. Conversely a small change in expression across the whole embryo may represent significant changes in a single tissue type. For example, a moderate increase in the levels of total full length dystrophins could reflect a significant increase in the expression of non-muscle isoforms.

---

![Figure 5](image-url)  
**Figure 5.** Developmental expression profiles of dystrophin and utrophin gene transcripts. The developmental stage in hpf is shown on the X-axis, and the relative expression levels (arbitrary units) are shown on the Y-axis. The latter have been normalised against the two reference genes Rpl13a and EF1a.
Localised expression, not undertaken here, might be evident by RNA in situ analysis (24,25), but as suggested earlier, the design of appropriate probes suffers from sensitivity and specificity issues that are not easily resolved in the case of the two genes studied. With these effects in mind, the total expression profile should be interpreted with caution as it does not reflect expression in different tissues, which is driven by tissue-specific promoters. Ideally for genes with tissue-specific expression such as dystrophin, tissue-specific investigations would be preferred over whole embryo studies. However, attempting to isolate individual tissue types from a developing embryo would be technically challenging. In conclusion, the expression data are the first reported attempt to follow multiple dystrophin and utrophin gene transcripts simultaneously throughout vertebrate development. The approach offers opportunities to determine the impact of targeted dystrophin gene mutagenesis in the zebrafish on the expression profile of the dystrophin gene, and the possible upregulation of utrophin gene transcripts.

Acknowledgements

We thank the Genomics Unit, DNA Sequencing Facility at the School of Biological Sciences, University of Auckland for their sequencing services. We acknowledge the assistance of Dr Andrew Dodd and Dr Rongying Tang who provided the foundation for the research presented herein. We acknowledge the financial assistance of the Maurice and Phyllis Paykel Trust (http://www.paykeltrust.co.nz/) and the School of Biological Sciences, The University of Auckland (http://www.sbs.auckland.ac.nz/), for partial funding of the study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References


