A low absolute number of expanded transcripts is involved in myotonic dystrophy type 1 manifestation in muscle

Anke E. E. G. Gudde, Anchele González-Barriga†, Walther J. A. A. van den Broek†, Bé Wieringa‡ and Derick G. Wansink*‡

Department of Cell Biology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands

*To whom correspondence should be addressed. Tel: +31 243610566; Email: rick.wansink@radboudumc.nl

Abstract

Muscular manifestation of myotonic dystrophy type 1 (DM1), a common inheritable degenerative multisystem disorder, is mainly caused by expression of RNA from a (CTG·CAG)n-expanded DM1 locus. Here, we report on comparative profiling of expression of normal and expanded endogenous or transgenic transcripts in skeletal muscle cells and biopsies from DM1 mouse models and patients in order to help us in understanding the role of this RNA-mediated toxicity. In tissue of HSA LR mice, the most intensely used 'muscle-only' model in the DM1 field, RNA from the α-actin (CTG)250 transgene was at least 1000-fold more abundant than that from the Dmpk gene, or the DMPK gene in humans. Conversely, the DMPK transgene in another line, DM500/DMSXL mice, was expressed ~10-fold lower than the endogenous gene. Temporal regulation of expanded RNA expression differed between models. Onset of expression occurred remarkably late in HSA LR myoblasts during in vitro myogenesis whereas Dmpk or DMPK (trans)genes were expressed throughout proliferation and differentiation phases. Importantly, quantification of absolute transcript numbers revealed that normal and expanded Dmpk/DMPK transcripts in mouse models and DM1 patients are low-abundance RNA species. Northern blotting, reverse transcriptase–quantitative polymerase chain reaction, RNA-sequencing and fluorescent in situ hybridization analyses showed that they occur at an absolute number between one and a few dozen molecules per cell. Our findings refine the current RNA dominance theory for DM1 pathophysiology, as anomalous factor binding to expanded transcripts and formation of soluble or insoluble ribonucleoprotein aggregates must be nucleated by only few expanded DMPK transcripts and therefore be a small numbers game.

Introduction

Myotonic dystrophy type 1 (DM1, OMIM no. 160900) is an autosomal dominant repeat expansion disorder, affecting skeletal and smooth muscle as well as the heart, the endocrine system, the eye and the central nervous system (1). The multisystemic manifestation and progression of DM1 are caused by expansion of a (CTG·CAG)n repeat, located in the 3'-untranslated region (3' UTR) of the dystrophia myotonica protein kinase (DMPK) gene (2) and in an overlapping antisense transcription unit in the DM1 locus (3). In DM1 families the expanded repeat is unstable, both...
somatically and intergenerationally, with a bias toward expansion, causing progression of disease symptoms during ageing and over successive generations (4).

Several mechanisms may contribute to the molecular pathogenesis of DM1 (5). Expanded DMPK transcripts are retained in the nucleus, where they form focal complexes in insoluble or diffuse-soluble state by abnormal association with transcription factors and RNA-binding proteins, like members of the muscleblind-like family (MBNL1–3), DEAD-box helicases and hnRNP proteins (1,6,7). In turn, abnormal phase transitions in RNP complexes lead to sequestering of factors needed for processing of other transcripts with in trans consequences for faithful alternative splicing and polyadenylation and expression of miRNAs (7,8). Production of proteins by ribosomes that decode the normally translated (CUG)ₙ-repeat tract in DMPK mRNA by a newly discovered process, coined repeat-associated non-ATG (RAN) translation, is also possible (9,10). Similar toxic events may occur with antisense transcripts originating from the complementary strand of the DM1 locus, overlapping the 3’ end of the DMPK gene. Abnormal RNAs are thus formed with an expanded (CAG)ₙ-repeat segment that entail the repeat segment are therefore critical parameters therein. Comparison of pathobiological findings between models and extrapolation to the situation in patients remain therefore difficult. Work of others has already demonstrated that the timing of DMPK expression, i.e. the onset of potential RNA toxicity, influences phenotypic severity (15). Expression of RNA with an abnormal repeat tract in satellite cells or neuronal progenitor cells may affect proper muscle and brain development (16–19) and have serious consequences for tissue regenerative capacity in adulthood. The absolute number of expanded RNAs and their structure at any given moment may also be crucial, as these ultimately will influence the extent of toxicity caused by abnormal RNP binding or abnormal properties of RAN translation products (20–22). The type of gene promoters, whether from endogenous or ectopic origin, that drive transcription during development and ageing, and the structure of the transcripts that entail the repeat segment are therefore critical parameters in animal models and patients.

Here, we report on comparison of expression and measurement of absolute numbers of (CUG)ₙ-repeat containing RNAs in muscle cells and tissues of four commonly used mouse DM1 models and in cells and biopsies from patients. DM1 mouse models express transgenes with different promoters, different structural organization and different repeat lengths: DM500, DMSXL, Tg26 and HSA LR mice, the transgene is under control of the ACTA1 promoter and the repeat is embedded in the context of the ACTA1 gene (27). DM500 and DMSXL mice are both descendants of the DM300-328 line, which was subject to intergenerational repeat expansion. These mice carry a complete human DM1 locus (23,24). The DMPK transgene in Tg26 mice carries a tandem insert of ~25 copies of the complete human DMPK gene, with a normal-sized (CTG)₁₁ repeat (25,26). In HSA LR mice, the transgene is under control of the ACTA1 promoter and the repeat is embedded in the context of the ACTA1 gene (27).

Table 1. Characteristics of DM1 mouse models used in this study.

<table>
<thead>
<tr>
<th>Mouse model</th>
<th>Transgene</th>
<th>Transgene copy number</th>
<th>Promoter</th>
<th>Expression</th>
<th>(CTG)ₙ</th>
<th>Genetic background</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM500 (DM300-328 line)</td>
<td>Human DM1 locus (43 kb transgene)</td>
<td>1</td>
<td>Human DMPK (~11.5 kb region upstream of main TSS)</td>
<td>All DM1-related tissues (e.g. skeletal muscle, heart and CNS)</td>
<td>500–600</td>
<td>&gt;90% C57BL/6</td>
<td>(23)</td>
</tr>
<tr>
<td>DMSXL (DM300-328 line)</td>
<td>Human DM1 locus (43 kb transgene)</td>
<td>1</td>
<td>Human DMPK (~11.5 kb region upstream of main TSS)</td>
<td>All DM1-related tissues (e.g. skeletal muscle, heart and CNS)</td>
<td>~1300</td>
<td>&gt;90% C57BL/6</td>
<td>(24)</td>
</tr>
<tr>
<td>Tg26</td>
<td>Human DMPK gene (14 kb transgene)</td>
<td>~25</td>
<td>Human DMPK (~1.9 kb region upstream of main TSS)</td>
<td>All DM1-related tissues (e.g. skeletal muscle, heart and CNS)</td>
<td>11</td>
<td>FVB/n</td>
<td>(25)</td>
</tr>
<tr>
<td>HSA LR (LR20b line)</td>
<td>Human α-actin gene; CTG repeat inserted in 3’ UTR (7.1 kb transgene)</td>
<td>2</td>
<td>Human α-actin (~2.1 kb region upstream of TSS)</td>
<td>Skeletal muscle only</td>
<td>220–250</td>
<td>FVB/n</td>
<td>(27)</td>
</tr>
<tr>
<td>WT</td>
<td>No transgene</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>&gt;90% C57BL/6</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

TSS, transcription start site; n.a., not applicable.
variation exists in level and developmental timing of transgene expression in DM1 cell and animal models. A remarkable low level of expression with absolute numbers of, at most, a few dozen RNA molecules per cell was observed for DMPK transcripts in human samples. Our findings highlight the hitherto unrecognized involvement of low-abundance RNA molecules in DM1 pathophysiology, altering our current view on the RNA gain-of-function theory, which explains the role of repeat RNA in DM1 manifestation. We discuss the possible implications of our findings for future interpretation of data from fundamental and translational studies in which these DM1 models and patient cells will be used.

Results

Derivation of myogenic cell lines from DM1 mouse models

Characteristics of mouse models included in this study are listed in Table 1. For profiling of transgene expression at the cellular level, we established conditionally immortalized myoblast populations from each model by pooling clones of individual cells derived from the calf muscle complex from double hemizygous mice carrying one transgenic DM1 allele and one H-2Kb-tsA58 allele (28). As there is strong evidence that satellite cells from different inbred mice behave intrinsically differently (29,30), it is important to note that crossings included different genetic backgrounds to generate the double hemizygous animals. The cell populations have therefore distinct mixed genetic backgrounds with contributions of C57BL6, FVB/n, CB6C7 and C57BL/10. We do believe, however, that these differences have no major impact on transcriptome composition and therefore should not overtly confound our comparison.

An important feature of the immortalized myoblasts is that during prolonged passaging in culture under permissive conditions cells have the tendency to undergo polyploidization, as an effect of the presence of the temperature-sensitive SV40 large T-antigen (31). This leads to a mix of 2N and 4N cells in populations of the different lineages and to variation in the absolute abundance of individual transcripts per cell (Supplementary Material, Fig. S1A). Hence, all possible care was taken to compare myoblasts from similar passage numbers.

DM500, DMSXL, HSA LR and WT myoblast populations had normal morphological appearance and showed apparently normal proliferative capacity and terminal differentiation features upon shift to myogenesis-promoting conditions (Fig. 1). Multinuclear contractile myotubes appeared after 3–5 days of differentiation. The Tg26 cell population had normal morphological

Figure 1. Morphology of myoblasts derived from DM1 and control mouse models. Conditionally immortalized myoblasts were derived from the GPS muscle complex of DM1 mouse models. Myoblasts differentiated into contractile myotubes under low-serum conditions. Representative images of cultures during proliferation (Day -2), confluency (Day 0) and differentiation (Days 1, 3 and 7) are shown. Bar 100 µm.
appearance, but showed slightly disordered cell alignment and diminished fusion capacity upon induction of differentiation. This may be due to overproduction of certain DMPK protein isoforms (26,32,33).

To quantify gene expression during proliferation and differentiation of the myoblast populations, RNA was isolated at various time points and analyzed by northern blotting and reverse transcriptase–quantitative polymerase chain reaction (RT-qPCR). Progression of myogenic differentiation in vitro was accompanied by a clear increase in skeletal muscle α-actin (Acta1) mRNA (Fig.2A), a well-known differentiation marker encoding a major constituent of the contractile apparatus (34–36). The increase was less profound in Tg26 cultures, in accordance with their diminished fusion capacity. Whereas Acta1 was induced, β-actin (Actb) expression decreased during the 7-day differentiation period (Supplementary Material, Fig. S2) (37).

For reference, primary human muscle cells obtained from a healthy individual and two DM1 patients were also included in our study. Myoblasts with healthy (CTG)5/(CTG)5 or disease-specific (CTG)21/(CTG)200 or (CTG)11/(CTG)760 repeat combinations appeared morphologically diverse, which can be explained by their different origin and culture history (Fig. 3). Multinucleated myotubes were formed during differentiation, but spontaneous contractions were never observed (Fig. 3, data not shown). RNA expression analysis corroborated this observation by showing that ACTA1 expression was minimal at all time points measured and remained low in comparison with Acta1 levels in DM500 GPS tissue (Fig. 4A). We, therefore, conclude that human cultures did not attain the same endpoint of terminal differentiation as mouse myoblasts under our in vitro conditions.

Transgene expression differs between DM1 myoblast models

Transgene expression was assayed in the myoblast populations during proliferation and differentiation (Fig. 2B and Supplementary Material, Fig. S3). DMPK transgene expression in differentiating DM500 and DMSXL cells transiently increased 3- to 10-fold and then returned to basal levels in proliferating myoblasts. A modest increase in transgene expression was also observed in differentiating Tg26 myoblasts, but this change appeared to be more permanent. In contrast, transcript levels from the Dmpk gene hardly varied with differentiation state in any of the five myoblast populations (Fig. 2C). The observed differential regulation of endogenous/transgenic Dmpk/DMPK expression may be caused by (i) species differences between the structure and function of Dmpk/DMPK promoters, (ii) differences in the length of the region upstream of the TSS in the human transgenes used (Table 1), (iii) differences between chromosomal insert sites and the endogenous Dmpk locus or (iv) different epigenetic alterations across the transgenic loci. Influence of the (CUG)m-repeat length on RNA stability is unlikely, because temporal profiles of transgenic RNA expression in DM500 and DMSXL cells during myogenic differentiation levels were highly similar.

For DMPK expression in human myoblasts, we performed separate quantitative analysis of healthy and expanded transcripts, which migrate differentially on northern gels. Steady-state levels of normal and expanded DMPK RNA molecules were approximately similar for both the 21/200 and the 11/760 patient cell cultures (Fig. 4B). Comparison of total DMPK expression levels between healthy 5/5 and patient cell cultures showed variation.
However, temporal profiles had a similar shape with a peak in expression during early differentiation, alike expression behavior of DMPK transcripts from the DMPK transgene in mouse myoblasts. This supports our idea that species-specific differences in the DMPK/Dmpk promoter explain the differential expression pattern of DMPK and Dmpk transcripts.

For better interpretation of in vitro and in vivo findings, we compared DMPK transgene expression in myogenic cultures to that in parental GPS muscle in the DM1 models. Transgene expression was ~0.6- to 6-fold higher, 6- to 18-fold higher and 3- to 11-fold lower for DM500, DMSXL and Tg26 myoblast/myotube cultures, respectively (Fig. 2B). Of particular, interest was the expression profile of the expanded ACTA1 transgene in differentiating myoblasts from the HSA$_{18}$ model, showing that transcripts were expressed only late in differentiation (Fig. 2B). Even in fully differentiated hemizygous HSA$_{18}$ myotubes in vitro, expanded ACTA1 transgene expression remained ~200-fold lower than in GPS muscle from homozygous HSA$_{18}$ mice.

Figure 3. Morphology of primary human DM1 and healthy myogenic cells. Human primary myoblasts aligned and fused under low-serum conditions. Representative images at different time points during proliferation (Day -2), confluency (Day 0), and differentiation (Days 1, 3 and 7) are shown. The number of CTG triplets for the two DMPK alleles are indicated for the one healthy (5/5) and two DM1 (21/200 and 11/760) cultures. Bar 100 µm.

Figure 4. Expression profile of proliferating and differentiating primary human DM1 myogenic cells. Assessment of ACTA1 (A) and DMPK (B) transcript levels by northern blot analysis. Signal strength in each sample was normalized to that of 18S rRNA, as outlined in the legend of Figure 3. To facilitate direct comparison with DM1 mouse model-derived myoblasts, ratios were related to Acta1 and Dmpk levels in DM500 GPS muscle. Data were obtained from two independent culture series per cell line, for which triplicate cultures per series were pooled and analyzed; bars represent mean ± SEM. Stacked bar graphs show levels of normal-sized DMPK (black part) and expanded DMPK (white part) in patient-derived cultures. On average, the ratio expanded versus normal-sized DMPK transcripts was 0.63 for 21/200 and 0.84 for 11/760. Timing of differentiation on the x-axis refers to Figure 3.
Transgene expression in distinct DM1 mouse models differs over a 1000-fold range

By extending our analysis to mouse model in vivo, using the GPS complex as RNA source and 18S rRNA for normalization, our comparison revealed that Acta1 expression was remarkably similar between models (Fig. 5 and Supplementary Material, Fig. S4A), despite differences in their genetic background (Table 1). Also Dmpk expression showed only minor fluctuations. Whereas Acta1 and Dmpk expression did not vary >2- to 4-fold between mouse models, the mean level of Acta1 RNA was ~2,500-fold higher than that of Dmpk RNA.

We next compared transgene expression in GPS muscles from DMS500, DMSXL and Tg26 mice. Quantification was straightforward, since all models expressed differentially expanded, but otherwise similar and intact DMPK transgenes. Transgenic DMPK RNA accumulated in muscle from hemizygous DMS500 and DMSXL mice to a 10-fold lower level than RNA from the endogenous Dmpk genes (Fig. 5). In contrast, hemizygous Tg26 GPS muscle expressed 20- to 25-fold more DMPK mRNA than Dmpk mRNA (assuming that Dmpk expression in Tg26 mice [FVB background] was similar to that in DMS500 mice [C57BL/6 background]). Determination of relative levels of transgenic RNA in HSA LR muscle appeared more challenging. Use of a (CAG)9 probe to detect (CUG)9 segments, the only sequence shared between transgenic products, proved unreliable to compare RNAs with different repeat lengths (data not shown). Therefore, a balanced mix of Acta1 and Dmpk probes, generated by random-primed labeling on cDNA templates of equal size was used instead (see the Materials and Methods’ section). With this approach we found ACTA1 (CUG)n RNA over 1000-fold higher expressed than Dmpk mRNA in homozygous HSA LR muscle (Fig. 5). This means that the concentration of expanded transcripts in HSA LR muscle is extraordinarily high and similar to that of Acta1 mRNA. This finding is perhaps not too surprising as both promoter and backbone of the human transgene and the Acta1 gene share strong homology.

Analysis of DMPK RNA levels in human muscle allows direct interpretation of transgene dosage

Skeletal muscle samples from healthy humans and DM1 patients were included in our study to extend our comparisons. DMPK mRNA expression in the samples varied, but was >3-fold higher than the level of Dmpk mRNA in mouse GPS tissue (Fig. 6). Expression of normal-sized and of expanded DMPK alleles were about equal in each of the patient samples. Since DMPK levels in human muscles and Dmpk levels in mouse muscle show high similarity, observations about transgene expression dosage effects in DM1 mice may be directly translated to relevance for DM1 patients. Any difference in DMPK mRNA content between human samples probably represents variation in fiber type and muscle origin or must be caused by differentiation or disease state. Note that, effects of ageing on repeat length heterogeneity were clearly visible on northern blot: in samples from adult DM1 patients a smear was observed for RNA from the expanded allele, whereas in congenital DM (CDM) patient material a defined signal was apparent (Fig. 6) (38).

Multi-pronged analysis reveals low copy number of expanded DMPK transcripts

For further stoichiometric and pathomechanistic considerations, we decided to determine the absolute number of DMPK mRNA copies per cell. Quantification on northern blot, after normalization for signal strength and correction for probe length, revealed that DMPK/Dmpk expression was 400- to 2000-fold lower than that of ACTB/Actb in proliferating human/mouse myoblasts (Fig. 7A and Supplementary Material, Fig. S4B). Based on known values for the copy number of ACTB/Actb RNA, which ranges from 350 to 8000 per cell (39–42) (Supplementary Material, Table S1), we inferred that the absolute number of DMPK transcripts must be in the range of 1–20 per cell. In patient myoblasts, this population consists of approximately equal numbers of normal-sized and expanded DMPK transcript molecules (Figs. 4 and 7A).

To validate this rough estimate, DMPK RNA copy number was quantified using two in vitro transcribed DMPK RNAs as standard references in RT-qPCR (Supplementary Material, Fig. S5). We found 20–25 DMPK transcripts per proliferating 11/760 myoblast and 45–50 DMPK molecules in myoblasts just prior to the onset of myogenic differentiation (Fig. 7B). This number includes both healthy and expanded DMPK transcripts. On average about four expanded DMPK transcripts per cell were detected in DMS500 myoblasts.

Thirdly, we used RNA-sequencing data of healthy and DM1 skeletal muscle (available via www.dmseq.org) to estimate DMPK transcript copy number. We calculated the DMPK-GAPDH ratio based on RNA-sequencing signal and found that DMPK mRNA molecules were 60- to 160-fold less abundant than GAPDH mRNA (Fig. 7C). Based on current estimates for the prevalence of GAPDH RNA, which is in the order of 250–2900 molecules per cell (39,42–45) (Supplementary Material, Table S2), DMPK transcript copy number must be between 2 and 50 per cell.

Finally, we determined the number of fluorescent in situ hybridization (FISH)-detectable RNP complexes in mouse and human myoblasts using a Cy3-labeled (CAG)7 probe. Initial
analyses demonstrated that the number of FISH signals slightly varied between myoblast pools and experiments and was influenced by culture conditions, such as cell density, feeding regime or changes in the culture medium (data not shown). This variation in foci number appeared suppressible by maintaining a very strict scheme for cell culture. On average, 5–9 foci per cell were detected in DM500 myoblasts (Fig. 8A and B). The number of foci differed between diploid and tetraploid DM500 myoblasts, with averages of ∼5 and ∼9 foci per cell, respectively (Supplementary Material, Fig. S1B). For reasons unknown, we also consistently observed signals in WT mouse myoblasts (Fig. 8A and B). These signals could not be distinguished from expanded DMPK (CUG)n mRNA signals and may also be present in other cell types from mouse. In cells from human origin, we did not observe overt background signals, so the problem may be species specific. To deal with this issue, we decided to apply a correction by defining a background interval in Figure 8. After this correction, an average of two transgenic expanded (CUG)n-specific foci per cell remained in DM500 myoblasts. Foci counts in Tg26 and HSA LR myoblasts were within the background range, as expected, based on the idea that the (CTG)11 repeat in Tg26 RNA cannot yield sufficient signal and our finding that the HSA LR transgene is barely expressed in proliferating myoblasts. In 11/760 patient myoblasts 3–4 foci per nucleus were detected, whereas no signal was observed in 21/200 and 5/5 myoblasts (Fig. 8C and D), obviously caused by lack of FISH-signal strengths by the limited length of the repeat target.

All data combined, we conclude that mouse and human myoblasts each contain up to half a dozen foci. We have to keep in mind that FISH protocols may detect transcripts with an efficiency of only 30–50% (41), for example, due to inaccessibility of the transcripts to the probe or loss of RNA molecules from the fixed cells during washing. Thus, the actual number of RNP complexes that appear as foci may be 2- to 3-fold higher. Given our estimates for the absolute number of expanded transcripts, our data suggest that every RNP complex that forms a FISH-visible aggregate is nucleated by one or only few expanded (CUG)n transcripts.

Discussion

DM1 is considered a prototypical RNA-dominant disorder, because its neurodegenerative and myopathic manifestation is thought to be based on processes wherein repeat-containing RNAs play a crucial role. Much of our current knowledge on the presumed toxic role of RNA in DM1 pathophysiology originates from direct comparison of findings in mutation carriers with findings in transgenic animals or genome edited cells. This is often done with simple bypassing of the fact that models may differ profoundly in nature, structural organization and chromatin context of the transgene and in length of the contained (CTG·CAG) repeat (13,14). A major unsolved aspect of DM1 etiology is how repeat length and dose and nature of abnormally expanded RNA transcripts affect onset and complexity of disease manifestation and its rate of progression and severity. Quantitative studies of transcript production could thus help to explain differential experimental findings with distinct DM1 models and add to conceptual progress. Here, we used gene expression profiling to compare relative abundance and absolute copy number of expanded repeat RNAs between muscle cells from patients and mouse models that are among the most commonly used in the DM1 field.

Our analyses unveiled differences in timing of production during myogenic differentiation as well as in the abundance of (CUG)n transcripts between DM1 mouse model and patient muscle cells. Expression of DMPK transgenes in DM500, DMSXL and Tg26 mice is at a basal level in proliferating myoblasts and peaks early in differentiation in a manner similar to the profile of endogenous DMPK expression in human myoblasts. Increase in DMPK expression at the start of myogenic differentiation was earlier reported for human cells (46) and also seen for C2C12

Figure 6. DMPK expression in human control and DM1 patient tissue. Northern blot (left) of human control and DM1 muscle tissue using a DMPK and a (CAG)9 repeat probe. Healthy DMPK mRNA appeared as a defined band in all tissues. Expanded DMPK mRNA appeared as a defined band in congenital DM1 tissue (CDM-e and -f), but as a smear (representing somatic mosaicism of repeat size) in adult DM1 tissue (DM1-c and -d), as size markers, we indicated the location of DMPK (CUG)700 and (CUG)1300 transcripts. DMPK transcript levels, normalized to those of 18S rRNA, were plotted relative to Dmpk transcript levels in DM500 GPS muscle to allow for comparison with DM1 mouse models (right). Stacked bar graphs show levels of normal-sized DMPK (black part) and expanded DMPK (white part) in patient-derived samples. The ratio expanded versus normal-sized DMPK transcripts was 0.33 in DM1-c; 0.92 in DM1-d; 0.84 in CDM-e and 1.79 in CDM-f.
mouse myoblasts (47). Our group (25) and others (32) have demonstrated that the actual onset of DMPK expression is already seen in somites in the developing embryo, well before the actual commitment to specific muscle cell fate and the onset of myogenesis.

DM500/DMSXL- and Tg26-derived myoblasts carry transgenes of different size that may lack regulatory elements upstream of the DMPK TSS, but the proximal promoter sequence and the enhancer sequence in Intron 1 (elements responsive to MyoD via conserved E-boxes (48)), are present. Thus, regulatory sequences from the human locus may drive ‘near-natural’ behavior of these transgenes and control the peak-shaped up-regulation with transient increase in transcriptional activity during early differentiation. The ‘human-like’ regulation of the DMPK transgene renders these mouse myoblasts useful models for study of toxic effects of normal or expanded DMPK RNA in early development. Others have demonstrated that early-onset expression of repeat-containing RNA may influence phenotypic severity by affecting proper tissue development in animals in vivo (16–19).

In HSA18 myoblasts, expression of repeat-containing RNA was hardly detectable until late in differentiation in vitro. Conspicuously, the Acta1 gene, whose human counterpart served as the recipient body for the transgene with (CTG)250 repeat insertion, showed an earlier onset of expression in differentiating mouse cells. The explanation for this differential behavior may thus be that the proximal promoter is present in the transgene, but that a cis-regulatory module in a region that activates transcription in differentiating myoblasts, >20 kb downstream of the basal promoter (49), is lacking.

We found pronounced differences in the abundance of (CUG)n-rep RNA products from transgenes and RNA products from the endogenous Dmpk gene and between expanded RNAs from cell lineages and the transgenic mouse tissues from which they originate. For proliferating cells, comparison was only meaningful for Tg26, DM500 and DMSXL myoblasts, as HSA18 myoblasts essentially lacked expression of expanded (CUG)n RNA. As explained by the presence of multiple copies of the DMPK gene in the transgenic insert in Tg26 cells, DMPK levels were relatively high. Transcripts from the
DMPK transgene in DM500 and DMSXL cells were ~10-fold lower than transcripts from the Dmpk gene in these myoblasts. Even if we take into account that these observations reflect the production rate from one DMPK transgene versus two Dmpk genes in the DM500 cell model, transgenic RNA production is still ~5-fold lower. Intrinsic differences in regulatory capacity of DMPK and Dmpk promoters may partly explain this observation, but we cannot exclude involvement of progressive epigenetic alterations, like DNA methylation or heterochromatization of the transgene or the DMPK promoter. Brouwer et al. (50) observed decreased transgene expression with age in the DM300-328 lineage, the ancestral mouse model from which DM500 and DMSXL mice and thus DM500 and DMSXL cells originate. Cis effects of repeat presence may have contributed to methylation state alterations in the DMPK transgene (50). Also transient propagation of mice in the homozygous state may have triggered partial gene silencing even before the myoblasts were derived in vitro (51–53). Finally, we cannot exclude effects of experimental handling, for example, by the...
use of dimethylsulfoxide during repeated cell freezing, which can affect the epigenetic state and is a known modulator of differentiation programming (54). Although we did not systematically study expression levels over time, we have the impression that DMPK expression in DM500 cells decreases with increasing passage number [compare data shown in (55)]. This could point to the influence of epigenetic silencing, explaining also variability in observations of different research groups in the DM1 field.

Also in muscle tissue obtained from hemizygous DM500 and DMSXL mice DMPK transcripts were 10-fold less abundant than Dmpk RNAs, confirming the observation by Huguet et al. (56). The overall low level of transgene expression may explain why disease manifestation in these models is relatively mild, even in homozygous animals. Conversely, however, one can also argue that the presence of these few expanded RNAs is already sufficient to cause mild myotonia, slow progressive muscle weakness, nuclear (CUG)n foci and splicing abnormalities (1,14,27), for which compelling evidence was provided in various studies.

Our most conspicuous observation was that ACTA1 (CUG)n expression in GPS muscle of homozygous HASL mouse was over 1000-fold higher than that of Dmpk, and remarkably similar to Acta1 expression. The high (CUG)n RNA expression may explain why the HASL mouse lineage has by far the most severe DM1-like phenotype of all models, including myotonia, extensive splice abnormalities and a conspicuous abnormal density of intranuclear foci, all limited to skeletal muscle (1,14,27).

The use of expression ratios between transgenic and endogenous transcripts (e.g. Dmpk, Acta1, Arch and 18S RNA) helped us to compare repeat RNA levels between muscle cells and mouse and human tissues, but we reasoned that the knowledge of absolute transcript copy numbers in muscle would be mechanistically more informative. Absolute numbers of molecules can only be meaningfully assessed if studied on a per cell basis. Analysis of DMPK expression in the three human myoblast lineages with no, intermediate or long repeat expansions confirmed that transcripts from the wild-type and mutant allele are about equally abundant. We found no evidence for an inverse correlation between repeat length and RNA expression level (46,58). Because endogenous DMPK/Dmpk transcripts appeared equally abundant in human and mouse myoblasts, quantitative comparison between patients and DM1 models was straightforward. Estimates for RNA expression ratios of ACTB and GAPDH, data from northern blot and RNA-sequencing analysis and absolute RT-qPCR experiments all point to the presence of at most 50 DMPK transcripts per cell. We thus propose that in patient cells <25 are expanded (CUG)n transcripts. In myoblasts from the DM500 and DMSXL mouse models around four RNA molecules per cell originate from the DMPK transgene.

Estimates for DMPK transcript numbers correspond remarkably well with the observed foci number in DM500 cell nuclei. Also in patient myoblasts the number of foci per nucleus and the number of DMPK transcripts per cell are in the same order of magnitude. Other research groups, studying different cell types from DM1 patients, have reported similar amounts of foci from their FISH experiments (59–62). Different and variable amounts of foci were observed for MyoD-transduced DM1 fibroblasts, which may be best explained by the distinct MyoD induction of DMPK expression, caused by differential vector systems or MyoD promoters (60,63). We, therefore, predict that RNP complexes that assemble in (CUG)n foci contain, on average, one to maximally six expanded transcripts. Along the same line of arguments, we propose that RNP aggregate formation is nucleated by one or very few expanded RNA molecules, thereby acting as individual entities. Phase transition of RNP complexes to insoluble aggregates (which appear as foci after FISH or MBNL antibody staining) may therefore not require further fusion with additional naked or RNP-decorated RNAs, but could be merely a protein-based event.

Many mechanistic avenues are now awaiting further exploration as our findings have important implications for the RNA-gain-of-function hypothesis that is currently in use for explaining DM1 features. How low numbers of only 1–25 normally folded RNA scaffolds can have a negative impact on ribostasis regulation—via titration of protein molecules from the total available pools of RNP proteins—is only one of the mechanistic questions that must be answered. The extent of loss of function of proteins may determine the gain of function toxicity of mutant RNA—and ultimately control the extent of cell stress caused by the DM1 mutation. Whether there is a direct relationship with the magnitude of temporal or permanent sequestration from the cellular pool of these RNP proteins is still difficult to answer. Stochastic events could be involved, as formation of an abnormal type of—microscopically visible or invisible diffusive—RNP aggregate on a repeat-containing transcript (6,7) might cause temporal or permanent perturbation of a specific nuclear pathway for mRNA processing or transport. Dominant effects of low-abundance triplet repeat RNA on RNA splicing and polyadenylation may be expected if effects on inhomogeneity and compartmentalization of protein factor pools (e.g. MBNL isoforms) in different nuclear trajectories for mRNA RNA processing play a decisive role. Finally, amplification of toxic effects by rare triplet repeat RNAs may occur otherwise: various studies have demonstrated that activation of stress-signaling cascades that involve GSK3β (64), PKC (65), double-stranded RNA-dependent protein kinase PKR (66) or sensors of foreign RNA that normally initiate immune responses (67) is involved in disease manifestation in DM1. Whatever the exact course of cellular events in these scenarios, even with low-expressed, abnormally expanded repeat RNAs the probability of exceeding the toxic threshold increases over time for any permanently expressing cell in the tissue population. Ultimately, this may result in accumulation of stress in an increasing number of cells, and progressive cell loss and loss of function of tissue during ageing.

In a parallel study, we have recently found that expression of another possible player, a DM1 antisense transcript carrying a (CAG)n repeat, is even 10-fold lower than levels observed for DMPK (Gudde et al. manuscript in preparation). Thus also here, the question remains how such an extremely rare transcript could contribute to disease manifestation. One unifying answer may be that the rare DMPK sense RNA and the even rarer antisense transcript both serve as templates for RAN translation (9,10). Homopolymeric protein products thus formed could possibly contribute to disease manifestation via initiation of aggregation of metastable proteins, initiating a prion-like cascade of events (68–70). Again, further analysis of steady-state levels of DMPK (CUG)n and antisense transcripts in cycling or resting cells and study of the frequency of use of RAN translation on these rare RNAs is necessary to better understand this possible distinct aspect of DM1 etiology.

Our findings highlight that careful choice of cellular and animal model systems that take structural properties of transgenes and transgenic products into account in combination with quantitative modeling is imperative for such studies. For this work not only knowledge of the ‘per cell’ presence of individual endogenous and mutant RNA molecules is important, but also stoichiometric considerations on the binding of MBNL1–3, Staufen, DDX, HnRNP or other RNP proteins (7) by one (CUG)n repeat
tract are needed to better understand nucleation and phase transition events in abnormal aggregates that could form around expanded DMPK mRNAs. Whatever the molecular mechanism involved, based on our findings, we propose that DM1 is caused by anomalous behavior of only very few mutant RNA molecules per muscle cell.

**Materials and Methods**

**Human material**

Human skeletal muscle samples were obtained from pre-consented post-mortem donors for research purposes in accordance with local guidelines in The Netherlands more than 18 years ago. No additional approval by an ethics committee was required at that time. Muscle autopsies were obtained from patients with a confirmed clinical and DNA diagnosis of adult-onset DM1 (male DM1-c, 55 years, psoas muscle (CTG)5/(CTG)exp) and congenital DM1-c, 55 years, psoas muscle (CTG)5/(CTG)exp; female DM1-d, detailed information could be traced.

The current guidelines for written consent were enforced and no additional approval by an ethics committee was required at that time. Muscle autopsies were obtained from patients with a confirmed clinical and DNA diagnosis of adult-onset DM1 (male DM1-c, 55 years, psoas muscle (CTG)5/(CTG)exp) and congenital DM1-c, 55 years, psoas muscle (CTG)5/(CTG)exp; female DM1-d, detailed information could be traced.

**Cell culture**

Conditionally immortalized myoblasts from the gastrocnemius-plantaris-solaeus (GPS) muscle of DMSXL, Tg26 and HSA1 mice after crossing with H-2Kb-tsA58 mice were derived as described (23, 55, 71). Individual myoblast lineages were obtained by ring cloning and selected for myotube formation ability. Myoblast populations used in this study were established by forming equal mixtures of five cell clones for DMSXL, seven cell clones for Tg26 and four cell clones for HSA1. DM500 and WT myoblasts were derived previously (55).

Myoblasts were grown on 0.1% (w/v) gelatin-coated culture dishes in the proliferation medium containing Dulbecco’s modified Eagle’s medium containing DMEM supplemented with 5% (v/v) horse serum and 50 µg/ml gentamicin at 7.5% CO2 and 37°C. Differentiation conditions were maintained for a maximum of 7 days. Spontaneous contractions started to appear around Day 3.

Human myoblast cultures isolated from skeletal muscle of fetuses, one healthy control (5/5) and two DM1 affected lines [21/200 with (CGG·CGG)n interruptions in the 3rd of the expanded (CTG·CA)n repeat; data not shown] and 11/760 were a gift of Dr Furling et al. (46). Myoblasts were maintained on 0.1% (v/v) gelatin-coated culture dishes in proliferation medium containing Ham’s F10 medium (Gibco) supplemented with GlutaMAX, 20% bovine growth serum (Thermo Scientific) and 25 µg/ml gentamicin at 7.5% CO2 and 37°C. Differentiation to myotubes was induced by placing confluent myoblast cultures in the differentiation medium containing DMEM supplemented with 4 mM l-glutamine, 1 mM pyruvate, 10 µg/ml insulin (Sigma), 100 µg/ml apo-transferrin (Sigma) and 25 µg/ml gentamicin.

Phase contrast images of cell cultures were taken with a Zeiss Axiovert 35 M light microscope, 10×/0.30 objective.

**Fluorescence-activated cell sorting**

Immortalized WT mouse myoblasts were detached from the culture surface using trypsin, washed in phosphate-buffered saline (PBS) and fixed in 70% (v/v) ethanol in PBS overnight at −20°C. After two times PBS wash, cells were stained in 20 µg/ml propidium iodide, 0.2 mg/ml RNase A, 0.1% Triton X-100 in PBS for 15 min at 37°C. Fluorescence-activated cell sorting analysis was performed on the Beckman Coulter Epics Altra cell sorter. Immortalized DM500 myoblasts were cultured in proliferation medium at 37°C for 24 h, detached and incubated with 3 µg/ml Hoechst 33342 at 37°C for 30 min and then sorted. Diploid and tetraploid DM500 cells were seeded on gelatin-coated glass cover slips, cultured for another 24 h at 37°C and analyzed by FISH.

**RNA isolation**

RNA from cultured cells was isolated using the Aurum Total RNA Mini Kit (Bio-Rad), according to the manufacturer’s protocol. RNA from muscle tissue was isolated using TRIzol reagent (Invitrogen), according to the manufacturer’s protocol.

**Northern blotting**

Northern blotting was performed according to standard procedures. RNA was subjected to electrophoresis in a 1.2% agarose gel under denaturing conditions. Depending on sample availability, 5–12 µg RNA was loaded per lane. RNA was transferred to Hybond-XL nylon membrane (Amersham Pharmacia Biotech) by capillary transfer in 10× SSC and hybridized with random-primer 32P-labeled cDNA-based probes or 32P-end-labeled oligonucleotides: a 2.6 kb DMPK cDNA covering the entire ORF and 3′ UTR (55) and the complete 1.9 kb human 18S rRNA cDNA were used as templates in random priming reactions. A (CAG)9 oligo (55), an Acta1 oligo (5′-ACCTGCGAACACAGCAAGCAATGCTGATTT-3′) (72) and a mixture of three oligos complementary to both Acta1 and Acta1(5′-GGGCTGTGGTCTCTGCTGATGTGCAT-3′, 5′-TGCCCTACAGGTCCTCTGCTGATGTGCAT-3′, 5′-GCTGTGCTGACATCTCACTGCACCTGATTC-3′) were 5′-end labeled. The Acta1/Acta1 oligo mix and the DMPK cDNA probe were used to quantify Acta1 and DMPK mRNA levels, respectively, in human samples. In murine samples, the single Acta1 oligo, DMPK cDNA and (CAG)9 oligo probes were used to quantify Acta1, Dmpk and transgenic DMPK (DM500 and DMSXL) and transgene (Tg26 and HSA1) mRNA.
levels, respectively. Blots were washed and exposed to X-ray film (Kodak, X-OMAT AR). Quantification of signals was performed by phosphor-imager analysis (Molecular Imager FX, Bio-Rad) and analyzed with Quantity One (Bio-Rad) and FIJI (73) software. 18S rRNA levels were used for normalization.

To be able to compare transcript expression in the DM1 mouse models, carrying different transgenes (i.e. DMPK- or ACTA1-type), a mix of ACTA1 cDNA probe (1.3 kb, Exon 2–7, 27% G-content) and DMPK cDNA probe (1.3 kb, Exon 3–12, 34% G-content) was used. Both genes are ~90% identical between mouse and human, so simultaneous detection of endogenous and transgenic transcripts was possible. (α-32P)-dCTP label was diluted with non-radioactive dCTP for labeling of the ACTA1 probe to obtain comparable signal strengths for simultaneous phosphor-imager detection of Acta1/ACTA1 and Dmpk/DMPK on the same blot. Signals were corrected for label dilution and G-content in the two probes. Since ACTA1 and Actb are 84% identical, the ACTA1 probe was also used for detection of Actb. A similar method was used to compare DMPK/Dmpk and ACTB/Actb expression. A 1.3 kb ACTB cDNA probe was used (Exon 1–6, 27% G-content), which has a 90% identical sequence with the mouse variant.

**In vitro transcription**

DNA templates corresponding to DMPK regions Exon 1–6 (0.6 kb) and Exon 11–15 (1.0 kb) were generated by PCR using primers 5′-GAATTATAGCGTACCTATAGGGAGAGGCAGTCCAGCCAGC-3′ and 5′-CCCCAGGGTTAGGCAAGAGG-3′, and 5′-GAATTATATAGCGTACCTATAGGGAGAGGCAGTCCAGCCAGC-G-3′ and 5′-GCCAGGGTTAGGCAAGAGG-3′, and 5′-GCCAGGGTTAGGCAAGAGG-3′ and 5′-GTCCTGTAGCTGAGA-3′, respectively (T7 promoter underlined). DNA purity and sequence identity were confirmed by sequencing. For in vitro transcription, 200 ng of the templates was used in the MEGAscript T7 Transcription Kit (Ambion) following manufacturer’s instructions. RNA products were purified using the Aurum Total RNA Mini Kit (Bio-Rad). Transcript length, purity and concentration were verified by gel electrophoresis and absorbance at 260/280 nm (NanoVUE spectrophotometer, GE Healthcare Life Sciences). Known amounts of RNA fragments DMPK Exon 1–6 and Exon 11–15 were mixed with WT mouse total RNA (as carrier RNA) and used as standards in RT-qPCR.

**Reverse transcriptase-quantitative polymerase chain reaction**

Using 500 ng RNA template per reaction and random hexamers as primers, cDNA synthesis was performed with SuperScript™ II Reverse Transcriptase (Invitrogen) or iScript™ cDNA Synthesis Kit (Bio-Rad). For RT-qPCR, 3 µl of 10–100-fold diluted cDNA preparation was mixed in a final volume of 10 µl containing 5 µl 2× Sybr Green mix (Roche Applied Science) or 5 µl iQ™ SYBR® Green Supermix (Bio-Rad) and 4 pmol of each primer: DMPK Exon 1–2: 5′-ACTGCGCGAGAAGATACG-3′ and 5′-CCTCTGAATCCACACCCG-3′; DMPK Exon 15(5′): 5′-AGAATGTCTTCTCAGTCCCGG-3′ and 5′-TGCAGCGGGATGGGAACTG-3′; ACTA1: 5′-CACCTCCAGCGGACTT-3′ and 5′-CGATGGCAGCAACGGAAGAATACG-3′; DMPK: 5′-TTTGTGACGTAACGACCCCTG-3′ and 5′-CCTTCTGACGATCCCGCCT-3′; 18S rRNA: 5′-GTAACCCCGGATACCCAGATAC-3′ and 5′-CATTCAATCGGTAGATGGACG-3′. Samples were analyzed using the CFX96 Real-time System (Bio-Rad). A melting curve was obtained for each sample in order to confirm single product amplification. cDNA samples from no template control and no reverse transcriptase control (RT−) were included as negative controls. Transgene (DMPK and ACTA1) and endogenous Dmpk mRNA levels were normalized to 18S rRNA expression.

**Fluorescent in situ hybridization**

Human myoblasts (5/5, 21/200 and 11/760) and mouse myoblasts (DM500, Tg26, HSA18 and WT) grown on gelatin-coated glass cover slips were washed once with PBS and fixed in 4% (v/v) formaldehyde, 2 mM MgCl2 in PBS for 10 min at room temperature. Cells were stored under 70% ethanol at 4°C. After two times PBS wash, cells were pre-hybridized in 40% (v/v) deionized formamide in 2× SSC for 20 min at room temperature, followed by high hybridization with a 0.1 ng/µl Cy3-(CAG)7 probe (2′-O-methyl phosphorothioate-modified) in 40% deionized formamide, 10% (v/v) dextran sulfate, 0.1% Triton X-100, 1 mg/ml herring sperm DNA, 100 µg/ml yeast tRNA, 0.2% (v/v) BSA, 2 µM VRC, 2× SSC in a humidified chamber at 37°C. After two times PBS wash, cells were counterstained with 0.33 µg/ml DAPI in PBS for 10 min at room temperature, followed by a PBS wash, dehydration in methanol and mounting in Mowiol. Images were acquired using the Zeiss Axioshot2 Fluorescence microscope or Olympus IX-71 wide field fluorescence microscope.

**RNA-sequencing**

RNA-sequencing data from healthy and DM1-affected human tissue samples [tibialis anterior (TA), quadriceps and heart] were obtained from www.dmseq.org. RNA-sequencing signals of full-length DMPK and GAPDH were used as an independent measure for transcript abundance.

**Statistical analysis**

Endogenous and transgenic mRNA levels were compared between time points and models, using a one-way analysis of variance (ANOVA) test, followed by a post-test for linear trend or Tukey’s multiple comparison test. All values in graphs are presented as means±SEM. Two tailed unpaired Student’s t-tests were used to compare between different myoblast lines with a one-way ANOVA test and Tukey’s multiple comparisons post-test. Data are visualized as scatter plot, each dot representing one observation, and the mean is shown. Statistical analyses were performed with GraphPad Prism version 5.01 for Windows.

**Supplementary Material**

Supplementary Material is available at HMG online.

**Acknowledgements**

We thank Geneviève Gourdon (Paris, France) for the DM500 and DMSXL mouse lines, Charles Thornton (Rochester, NY, USA) for the HSA18 line and Denis Furling (Paris, France) for human myoblasts. We thank the founders and contributors of the Myotonic Dystrophy Deep Sequencing Data Repository available via www.dmseq.org. We thank the members of the Department of Cell Biology for discussions.

**Conflict of Interest statement. None declared.**

**Funding**

This work was supported by the Prinses Beatrix Sfierfonds (grant number W.OR10-04; with contribution from the Stichting Sfieren voor Sfieren) and by ZonMW (TOP grant NL91212009). Funding to pay Open Access charges was provided by both organizations and Radboudumc.
References


