Friedreich ataxia, myotonic dystrophy type 1, and 3 forms of intellectual disability, Fragile X syndrome, FRAXE mental retardation and FRA12A mental retardation are Repeat Expansion Diseases caused by expansion of CTG•CAG, GAA•TTC, or CGG•CCG-repeat tracts. These repeats are transcribed, but not translated. They are located in different parts of different genes and cause symptoms that range from ataxia and hypertrophic cardiomyopathy to muscle wasting, male infertility and mental retardation. Yet recent reports suggest that, despite these differences, the repeats may share a common property, namely the ability to initiate repeat-mediated epigenetic changes that result in heterochromatin formation.

Background
The Repeat Expansion Diseases arise from the expansion or increase in the number of repeats in a specific tandem repeat array. Pathology only becomes apparent when the repeat number exceeds a certain critical threshold (1). Many diseases in this group arise from expansion of a repeat located in an open reading frame. Expansion in these instances results in a protein with a long polyglutamine (PolyQ) tract that is toxic to neurons (1). Other diseases like Fragile X syndrome (FXS), FRAXE mental retardation (MR), FRA12A MR, Friedreich ataxia (FRDA) and Myotonic dystrophy type 1 (DM1) are caused by expansion of repeats that are transcribed but not translated (Table 1). A more detailed description of the genetics and clinical presentation of these disorders can be found in 2 recent books on the subject (2,3), and in this minireview series. In this minireview, we will briefly describe the genetic basis of the non-coding repeat expansion disorders listed above, and focus on recent findings suggesting that the consequences of the repeat expansion in these disorders may be more similar than originally thought.

FXS, FRAXE MR and FRA12A MR. FXS is an X-linked disorder that is the most common heritable cause of intellectual disability, and the most common known cause of autism. FRAXE MR and FRA12A MR are less common forms of MR that produce much milder neurocognitive deficits. The repeat unit responsible for all 3 disorders is CGG•CCG. The repeat tract is located in the 5’ UTR of the affected gene, FMR1 in the case of FXS, FMR2 in the case of FRAXE MR (4,5), and the DIP2B gene in the case of FRA12A-associated MR (6). FMR1 alleles with 55-200 repeats produce elevated levels of FMR1 mRNA, and carriers of such alleles are at risk for Fragile X associated tremor and ataxia syndrome and premature ovarian insufficiency. FXS only becomes apparent when the repeat number exceeds 200 (7). The same threshold is seen for FRAXE MR, but the threshold for FRA12A has not yet been well defined. All 3 MR disorders are associated with gene silencing or reduced gene expression. Relatively little is known about events occurring at the FMR2 and DIP2B loci. Thus, we will focus here on what is known for FXS, bearing in mind that similar mechanisms may be responsible for the other 2 disorders.
MINIREVIEW: Repeat-mediated epigenetic changes

In the case of FXS, it was appreciated early on that the reduced gene expression was associated with DNA methylation (8-10). The CCG•CGG-repeats responsible for FRAXE MR (5) and FRA12A MR are also methylated (6), as are all the other long CCG•CCG-repeat tracts studied to date (11). The CCG-strand of the repeat is a good substrate for DNA methyltransferases in vitro (12). This led to the idea that the propensity of the repeats to undergo DNA methylation was responsible for the initiation of gene silencing in FXS (12). This methylation was thought to facilitate the recruitment of various chromatin-modifying enzymes, including histone deacetylases that resulted in the formation of transcriptionally silent chromatin or heterochromatin. This heterochromatin was then envisioned to spread into the adjacent promoter, resulting in gene silencing.

FRDA. FRDA is a recessively inherited, early onset, ataxia, with an associated, and frequently fatal, hypertrophic cardiomyopathy. In the case of FRDA, the responsible repeat, which has a GAA•TTC-repeat unit, is located in intron 1 of the frataxin (FXN) gene (13). Repeat expansion causes a decrease in FXN mRNA, resulting in a deficiency of frataxin, a protein important for normal mitochondrial function. Pathology is seen when both FXN alleles have more than 65 repeats.

The absence of methylatable CpG residues in the GAA•TTC-repeat led to an initial focus on mechanisms that did not involve epigenetic changes. A number of models based on the ability of polypurine:polypyrimidine sequences to form triplexes, or other forms of triple-stranded structures, have been proposed (14). Work in vitro and in bacteria supports the idea that such structures can form during transcription. Once formed, they are thought to trap RNA polymerase on the template, thus preventing transcription elongation and reducing the accumulation of mature FXN mRNA (15,16).

DM1. DM1 is a dominantly inherited multisystemic disorder whose symptoms include myotonia, muscle weakness, cataracts, cardiac conduction defects, as well as endocrine and gonadal abnormalities. DM1 has a congenital form, as well as juvenile and adult onset forms. Congenital DM (CDM) is associated initially with severe neonatal hypotonia, respiratory problems, delayed motor development and mental retardation. Symptoms more typical of the postnatal onset forms develop later. Alleles with more than 50 CTG•CAG-repeats in the 3’ UTR of the DMPK gene result in DM1, while CDM only arises when the repeat number exceeds 1000 (17).

While juvenile and adult onset forms are thought to be related to CUG-RNA-mediated sequestering of MBNL-family members (17), it was noticed early on that this chromatin in the region containing the repeat was more compact in CDM cells (18). The relationship, if any, of this compaction to disease pathology is still unresolved. It may be that the unique aspects of CDM pathology are related to this phenomenon, since attempts to model these symptoms in transgenic and knockout mice have failed (17).

Since the DM1 repeat, like the FRDA repeat, lacks CpG residues, explanations quite different from those proposed for FXS were initially advanced to explain the chromatin compaction. One early suggestion was that this more compact chromatin structure might have something to do with the fact that CTG•CAG-repeats are very strong nucleosome positioning signals (19). Since CCG•CCG-repeats exclude nucleosomes and GAA•TTC-repeats have no effect on nucleosome positioning (20), this reinforced the idea that the effect of each of these repeats was quite distinct.

These repeats do share a common property

However, recently, a rare unsilenced and unmethylated FXS allele was shown to be associated with a mark of silent chromatin, histone H3 dimethylated at lysine 9 (H3K9Me2) (21). Furthermore, this same modification was seen on FXS alleles in differentiating embryonic stem cells prior to the detection of DNA methylation (22). Thus, DNA methylation is probably not the first step in FXS gene silencing. Therefore, by extension, the lack of
CpG-residues per se does not preclude an effect of the repeat on chromatin remodeling.

Consistent with this idea, recent studies have shown that the chromatin in the region flanking the GAA•TTC-repeat in the FXN gene is also associated with repressive histone marks (23-25). This includes hypoacetylation of many lysines on H3 and H4, and di- and trimethylation of H3K9 (23,25). This region also displays aberrant DNA methylation (23,24). This methylation, which is seen in both affected and unaffected individuals, extends further upstream of the repeat in patient cells. In addition, a methylation footprint seen on normal alleles is absent in patients. This finding suggests a model in which DNA methylation occurs secondarily to chromatin compaction. The more compact chromatin could affect the accessibility of the flanking DNA to proteins that might bind and protect individual bases from methylation. Part of the methylation footprint seen in normal cells is due to binding of a positive regulator of FXN transcription (24). Thus, a loss of binding of this regulator may be one way in which heterochromatin formation in the intron negatively affects transcription. Intragenic methylation can also reduce the efficiency of transcription elongation (26).

The fact that histone deacetylase inhibitors correct the frataxin deficit, in both patient lymphoblasts and a mouse model of FRDA, supports the idea that repeat-mediated epigenetic effects underlie disease pathology (25,27).

The repeat region, on both unaffected and patient DMPK alleles, is enriched for methylated H3K9. CpG residues flanking the repeat tract are also methylated in CDM (28). Repeat-mediated epigenetic effects in both CDM and FRDA is in agreement with the finding that insertion of the responsible repeats into arbitrary locations in the mouse genome results in a high frequency of silencing of a linked transgene (29). Thus, there is reason to think that the ability to assemble heterochromatin is a general feature of the disease-associated repeats.

RNA-based models for heterochromatin formation

Clues to what the initial trigger for heterochromatin formation may be have emerged from a characterization of the DMPK locus (28,30). Small RNAs, 21 nucleotides (nt) in length derived from the 3’ UTR of the DMPK gene, have been found in cells from both unaffected individuals and those with CDM (30). A variety of similarly sized RNAs are produced in cells. Many of these RNAs associate with members of the Argonaute/PIWI family of proteins, that are involved in transcriptional and posttranscriptional gene silencing (31). The best characterized of these small RNAs are those derived from longer double-stranded RNAs (dsRNAs) or imperfect RNA hairpins, by the action of ribonucleases like Dicer. After Dicer digestion, these RNAs can enter the RNA interference (RNAi) pathway. In organisms like fission yeast, these RNAs can be incorporated into RNA-induced transcriptional gene silencing (RITS) complexes that mediate heterochromatin formation (32). There is evidence from model systems to suggest that a similar mechanism may operate in mammalian cells (33).

Sense-antisense transcripts may provide a source of dsRNA. Double-stranded RNA may arise from the production of an antisense transcript from the affected gene. Such a transcript could interact with the sense transcript to produce a sense-antisense pair as illustrated in Fig. 1 and 2. An antisense transcript has been identified for the DMPK gene that originates downstream of the repeat (30). This results in a potential region of overlap between the 3’ end of the DMPK sense transcript and the antisense transcript, as shown in Fig. 1A. Only the antisense strand is represented in the pool of 21 nt DMPK RNAs (30). This would be consistent with the antisense strand being the “guide” strand in the RITS complex. The sense “passenger” strand would then be rapidly degraded. The small RNAs from the DMPK gene are limited to the region between 2 CTCF-binding sites that contains the repeat, despite the larger region of potential transcript overlap (Fig. 1A). It is possible that some property of the
dsRNA from this region makes it a particularly good target for Dicer. In normal and non-congenital forms of DM1 heterochromatin is also limited to this region. In CDM cells, the region of heterochromatin spreads beyond the CTCF sites. This may be because more extensive epigenetic modification, resulting from the presence of a much longer repeat tract, blocks CTCF binding (28).

The situation in FXS is quite complex since, as shown in Fig. 1B, multiple sense and antisense transcripts with different start sites and alternative splice sites are produced (34). Furthermore, on active alleles not only does the transcription start site usage and splicing profile change with increasing repeat number, but the levels of both sense and antisense transcripts also increases (34,35). However, despite the presence of a potential source of dsRNA, the FMRI gene does not get silenced until the repeat number exceeds ~200. Perhaps, prior to silencing, FXS alleles produce a unique sense-antisense variant that triggers heterochromatin formation. Alternatively, it may be that the critical mass of the sense-antisense pair necessary for silencing is only reached when the repeat number exceeds 200.

Since genes with antisense transcripts are relatively common in the human genome, a sense-antisense mechanism that accounts for heterochromatinization of the other long CGG-repeat tracts and the FRDA repeat is certainly possible. In the case of FRDA, the FXN locus contains an Alu element 3’ of the repeat that is oriented so as to be transcribed in the opposite direction (24) as illustrated in Fig. 1C. Alu elements contain regulatory sequences that can act as Pol II promoters and thus could potentially initiate the production of an antisense transcript.

**RNA hairpins formed by the repeats are another potential source of dsRNA.** However, since not all genes with antisense transcripts are silenced, it is possible that some other source of dsRNA is involved. In the case of both the CGG•CCG-repeat disorders and DM1, another source of dsRNA is known. Both strands of the responsible repeats form stable RNA hairpins that are substrates for Dicer (36,37). The repeat itself may thus play a role in the silencing process as illustrated in Fig. 2A. Long FRDA repeats have also been shown to form DNA hairpins (38) but it is not known whether they form RNA hairpins, or whether the RNA could be a Dicer substrate.

However, if the repeats themselves are responsible for gene silencing, it is unclear why heterochromatinization begins on DMPK alleles with as few as 5 repeats while >200 repeats are necessary for heterochromatin formation in FXS (30). It may be that the threshold for silencing is determined not only by the number and perhaps the properties of these repeats, but the expression level, stability and availability of the repeat-containing transcript to enter the gene-silencing pathway. The chromatin context of the repeat-containing gene may also play a role. However, the very low repeat number associated with heterochromatinization of the DMPK allele makes it seem unlikely that this chromatin remodeling is due to the repeats alone. If that were the case, repeat-mediated gene silencing would be much more commonly observed. It may be that sequences flanking the repeat also form part of the Dicer substrate. In this regard, it is interesting to note that the region between the CTCF sites on the DMPK gene is GC-rich, and is predicted to form very stable hairpins using standard structural prediction algorithms (unpublished observations).

**DNA-based models for heterochromatin formation**

It is also possible that an RNA-independent mechanism is responsible for the epigenetic changes in one or more of these diseases, as illustrated in Fig. 2B. A number of specific DNA-binding proteins are known that cause epigenetic reprogramming by recruiting histone deacetylases and proteins like HP1 (e.g., the large KRAB-ZNF family (39)). DNA-binding proteins that bind to CGG•CCG, CTG•CAG and GAA•TTC-repeatss have been described (40-42). Whether any repeat-binding proteins are able to act in an analogous way to effect chromatin remodeling remains to be seen.
MINIREVIEW: Repeat-mediated epigenetic changes

Perspective

More work is needed to understand the mechanism(s) responsible for repeat-mediated chromatin remodeling, and in the case of the DMPK gene, what significance it has, if any, for disease pathology. However, despite these uncertainties, it is apparent that these repeats all have the ability to affect chromatin structure. This raises the possibility that this may also be true of other transcribed repeats, including those responsible for some of the other non-coding repeat expansion diseases.

Furthermore, the similarities in the chromatin modifications present on silenced alleles of different disease genes suggest not only that a similar mechanism of gene silencing occurs at these loci, but also that similar approaches to gene reactivation may be possible.

Recent work also suggests that PolyQ-containing proteins sequester histone modifying proteins (43). Thus, while the mechanisms involved are quite different, both PolyQ diseases and the diseases that involve long transcribed but untranslated repeats may be “chromatinopathies”. Histone deacetylase inhibitors have also shown promise in alleviating the PolyQ-mediated toxicity (43) and therefore it may be that this class of drugs has therapeutic potential in treating both subsets of the Repeat Expansion Diseases.

REFERENCES

### Table

#### Table 1. Summary of key disease features

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene</th>
<th>Repeat</th>
<th>Location</th>
<th>Symptoms</th>
<th>Mutation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>FXS</td>
<td>FMR1</td>
<td>CGG•CCG</td>
<td>5’ UTR</td>
<td>Intellectual disability, depression, autistic symptoms, behavior problems, insomnia</td>
<td>LOF</td>
</tr>
<tr>
<td>FRAXE MR</td>
<td>FMR2</td>
<td>CGG•CCG</td>
<td>5’ UTR</td>
<td>Mild intellectual impairment, behavior problems, obsessive compulsive disorder (OCD)</td>
<td>LOF</td>
</tr>
<tr>
<td>FRA12A MR</td>
<td>DIP2B</td>
<td>CGG•CCG</td>
<td>5’ UTR</td>
<td>Intellectual impairment</td>
<td>LOF</td>
</tr>
<tr>
<td>FRDA</td>
<td>FXN</td>
<td>GAA•TTC</td>
<td>Intron 1</td>
<td>Early onset progressive gait and limb ataxia, dysarthria, areflexia in all 4 limbs, axonal sensory neuropathy, hypertrophic cardiomyopathy</td>
<td>LOF</td>
</tr>
<tr>
<td>DM1</td>
<td>DMPK</td>
<td>CTG•CAG</td>
<td>3’ UTR</td>
<td><strong>Juvenile and adult onset forms:</strong> Myotonia, muscle weakness, cataracts, cardiac conduction defects, insulin insensitivity, endocrine and gonadal abnormalities, hypersonnia and dysphagia.</td>
<td>GOF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>CDM:</strong> severe neonatal hypotonia, respiratory problems, delayed motor development and mental retardation.</td>
<td>GOF +LOF</td>
</tr>
</tbody>
</table>

Abbreviations: FXS, Fragile X syndrome; MR, mental retardation; FRDA, Friedreich ataxia; DM1, myotonic dystrophy type 1; CDM, congenital myotonic dystrophy; LOF, loss-of-function mutation; GOF, gain-of-function mutation.
FIGURE LEGENDS

Fig. 1. A) The organization of the human DMPK locus. i) The region shown encompasses the sequence of chromosome 19 from ~50,960,300 to ~50,988,000, using the numbering from the March 2006 build of human genome database (UCSC Genome Bioinformatics site; http://genome.ucsc.edu/). ii) Expansion of the region between the end of the DMPK ORF and the start of SIX5 transcription. The location of the CTG•CAG repeat is indicated by the green arrowhead. The 2 CTCF sites are shown as stippled ovals (28). The black box indicates the DNase I hypersensitive site that contains the promoter for the antisense transcript (28). B) The organization of the human FMRI locus. (i) The numbering used in this diagram is relative to the major FMRI sense transcription start site in cells from normal individuals. The green arrowhead indicates the location of the CGG•CCG-repeat. CTCF sites are shown as stippled ovals (ii) Sense and antisense transcripts produced from the FMRI gene. The sense transcripts initiate from 1 of 3 major transcription start sites located close to exon 1. Start site usage changes with increasing repeat number with start sites II and III being used more heavily as repeat number increases (44). Two major promoters are used for the production of antisense transcripts, one in the promoter of the sense transcript and one >10 kb downstream. Transcripts initiating at this second promoter predominate in carriers of long, but still active alleles. C) The organization of the human FXN locus. (i) The numbering used in this diagram is relative to the major FXN sense transcription start site in cells from normal individuals. The green arrowhead indicates the location of the GAA•TTC-repeats. The arrows indicate the orientation of the 2 Alu elements in intron 1. (ii) The sense and hypothetical antisense transcripts from the FXN gene. The grey dashed line illustrates the potential antisense transcript originating from the upstream AluSp element.

Fig. 2. Schematic representation of potential mechanisms of heterochromatin formation at loci containing long CGG•CCG, CTG•CAG and GAA•TTC-repeats. The green segments of the DNA and RNA represent the repeat-containing region. The nucleosome octamer is shown in pale blue. The orange and blue lines emerging from each nucleosome represent a histone H3 and H4 tail respectively. For simplicity only 1 tail is shown per nucleosome. The acetyl groups (purple circles) represent those attached to H3K9 and H4K16 as seen on the active FMRI and FXN alleles. Many other potentially acetylatable residues exist. H3K9 methylation seen on all 3 disease alleles is indicated by the blue circles. A) RNA-based models. In the mechanism illustrated on the left hand side of panel A, sense and antisense transcripts from the affected gene generate a region of dsRNA that is a substrate for Dicer. The region of overlap may or may not include the repeat. On the right hand side of panel A, the repeats in the transcript form an RNA hairpin that is a Dicer substrate. Irrespective of the source of the dsRNA, the small Dicer products become loaded into the RITS complex and target the complex to the affected gene. The RITS complex then facilitates hypoacetylation of histone H3 and H4, methylation of H3K9 and DNA methylation. These, and presumably other epigenetic changes, are then propagated to the surrounding regions perhaps because they abolish the activity of an insulator element like CTCF (not shown). B) DNA-based model. Repeat-binding proteins may bind to the expanded repeat and recruit proteins like HP1 that lead to heterochromatin formation. RITS: RNA-induced initiation of transcriptional gene silencing complex; HDACs: histone deacetylases; HMTs: histone methyltransferases that establish repressive chromatin marks; DMNT: de novo DNA methyltransferase.
Fig. 1
Fig. 2

A

HDACs
HMTs
DMNT
HP1
H3
H4

Dicer

RITS

Repeat-binding protein

B

Ac
Me
CpG
C^mepG

Fig. 2
Chromatin remodeling in the non-coding repeat expansion diseases
Daman Kumari and Karen Usdin

J. Biol. Chem. published online October 28, 2008

Access the most updated version of this article at doi: 10.1074/jbc.R80026200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts