Mechanisms of RNA-mediated Disease*

Jason R. O'Rourke and Maurice S. Swanson

Recent mapping of functional sequence elements in the human genome has led to the realization that transcription is pervasive and that noncoding RNAs compose a significant portion of the transcriptome. Some dominantly inherited neurological disorders are associated with the expansion of microsatellite repeats in noncoding regions that result in the synthesis of pathogenic RNAs. Here, we review RNA gain-of-function mechanisms underlying three of these microsatellite expansion disorders to illustrate how some mutant RNAs cause disease.

Dynamic Mutations and Hereditary Disease

The human genome contains thousands of microsatellites, which are short (generally 2–6 bp) polymorphic repetitive sequences that can expand or contract because of mistakes in DNA replication, repair, and recombination (1). Microsatellite expansions are particularly pathogenic and associated with a number of hereditary disorders. It is not surprising that repeat expansions in protein-coding regions result in many of these diseases, including Huntington, spinal bulbar muscular atrophy (Kennedy disease), and at least six types of spinocerebellar diseases, including Huntington, spinal bulbar muscular atrophy (Kennedy disease), and at least six types of spinocerebellar ataxias. Perhaps more unexpected is that a number of dominantly inherited disorders result from microsatellite expansions in noncoding regions. These diseases include DM,2 FXTAS, SCA8, SCA10, SCA12, and Huntington disease-like 2 (2). Interestingly, some of these expansions may also be located in a coding region. In this review, we focus on mechanisms that have been proposed to explain RNA-mediated pathogenesis in DM, FXTAS, and SCA8 and discuss how mutations in other types of ncRNAs might result in deleterious RNA gain-of-function effects.

DM, FXTAS, and SCA8

The idea that transcription of noncoding repeat expansions leads to the production of pathological, or toxic, RNAs origin-ated from studies on DM (2, 3). DM is a dominantly inherited and multisystemic disease characterized by a distinctive combination of clinical features, including skeletal muscle myotonia and weakness/wasting, heart conduction defects, and cerebral atrophy. The genetic basis of this disorder is atypical because it is caused by the expansion of structurally related microsatellites in two unlinked genes. DM1 is caused by CTG repeat expansions (50 to >3500 repeats) in the 3′-UTR of the DMPK gene, whereas DM2 is associated with (CCTG)75–11,000 expansions in intron 1 of ZNF9 (2). Although DM1 and DM2 are adult-onset degenerative diseases, some DM1 patients develop a more severe congenital disease with neonatal hypotonia and mental retardation when (CTG)n expansions exceed ~1000 repeats.

FXTAS was identified as a late adult-onset disorder in families of children with fragile X syndrome (4), the most common congenital mental retardation syndrome caused by (CGG)200–4500 expansions in the 5′-UTR of FMR1 (5). This gene encodes the fragile X mental retardation protein, thought to influence synaptic plasticity through its roles in regulating mRNA transport and translation (6). In contrast to the large expansions characteristic of fragile X syndrome, FXTAS-associated microsatellites are more restricted in length (55–200 repeats). Indeed, mutant FMR1 genes in this range were originally considered premutation alleles (7). Affected carriers of FXTAS alleles are predominantly older males (>50 years old) and present with progressive intention tremor, gait ataxia, and parkinsonism accompanied by loss of cognitive function and cerebral/cerebellar atrophy. In females, premature ovarian failure, or cessation of menses prior to age 40, occurs in premutation carriers (4).

SCA8, an autosomal dominant disorder with reduced penetrance, is characterized by progressive gait ataxia, nystagmus (involuntary jerky eye movements), and dysarthria (slurred speech) (2). Whereas both DM and FXTAS are caused by microsatellite expansions in the untranslated regions of protein-coding genes, the SCA8 (CTG)n–>1300 expansion mutation was initially described as a CTG expansion located at the 3′-end of a gene transcribed only in the CTG orientation (8). However, recent evidence demonstrates that this locus is bidirectionally transcribed and produces both CUG and CAG expansion RNAs. The CAG expansion transcripts produce a nearly pure polyglutamine protein, so the gene encoding this novel protein is now called ATXN8, whereas the gene in the CTG direction is ATXN8OS (9).

Disease Mechanisms

C(C)UG and rCGG Expansion RNAs Are Pathogenic—Sev-eral observations argue that the DM, SCA8, and FXTAS microsatellite expansions are pathogenic at the RNA level. First, there are no examples of DMPK, ZNF9, or FMR1 coding region mutations associated with DM or FXTAS, so it is unlikely that these diseases result from haploinsufficiency or protein gain-of-function effects (2–4). Second, C(C)UGexp and rCGGexp RNAs accumulate within the nucleus in either nuclear RNA
MINIREVIEW: RNA-mediated Disease

(ribonuclear) foci (DM1 and DM2) or large intranuclear inclusions (FXTAS) (2–4). Although full-length DMPK mRNAs are trapped in ribonuclear foci and DMPK protein levels may therefore be compromised, ZNF9 pre-mRNA splicing is not affected by CCUG expansions (2, 10). Indeed, only the intronic CUG repeats localize in ribonuclear foci, and ZNF9 protein levels are normal in DM2 (10). The fate of SCA8 CUG\textsuperscript{exp} RNAs has not been determined. However, bidirectional transcription at the SCA8 locus suggests that this disease is caused by gain-of-function effects at both the RNA and protein levels (9). In FXTAS cells, FMR1 mRNA levels are elevated 2–10-fold with a concurrent accumulation of FMR1 mRNAs in large (2–5 μm) ubiquitin-positive intranuclear inclusions in neurons and astrocytes (4, 11). Third, animal models for these diseases show that C(C)TG and CGG expansion mutations are pathogenic independent of gene context or flanking sequence and must be transcribed to cause disease. For example, a transgenic mouse model for DM expresses a (CTG)\textsubscript{250} repeat within the 3’-UTR of the human skeletal actin gene (HSA\textsuperscript{ATR}) (12). These mice develop DM-associated myotonia, skeletal muscle pathology, and ribonuclear foci, whereas mice expressing a wild-type length (CTG)\textsubscript{3} repeat (HSA\textsuperscript{ATR}) are normal. Notably, the severity of these muscle phenotypes correlates with the level of HSA\textsuperscript{ATR} transgene expression, suggesting that the gain of function is at the RNA level.

More recently, tissue-specific bitransgenic mouse models that express CUG\textsuperscript{exp} within ncRNAs have been generated. Inducible lines were created by constructing transgenes that express either an interrupted (CTG)\textsubscript{960} repeat (EpA960) or no repeats (EpA0), positioned in the DMPK 3’-UTR (13, 14). Expression of EpA960 and EpA0 is dependent on Cre-mediated excision of an upstream SV40 polyadenylation cassette, so bitransgenic lines were created by crossing EpA0 and EpA960 mice with tissue-specific and inducible Cre lines. Expression of the EpA960 transgene in heart causes DM-associated cardiac defects, including cardiomyopathy and arrhythmias, whereas expression in skeletal muscle leads to myotonia and muscle wasting. For SCA8, mouse bacterial artificial chromosome transgenic lines expressing human SCA8 with (CTG-CAG)\textsubscript{116} (but not (CTG-CAG)\textsubscript{11}) repeats show a progressive movement disorder (9). Because (CTG-CAG)\textsubscript{116} animals express both CUG\textsuperscript{exp} transcripts and a polyglutamine protein, it is not yet clear which potentially pathogenic molecule plays the more significant role in disease.

Drosophila models have also been developed to examine CUG\textsuperscript{exp} and CGG\textsuperscript{exp} toxicity. Transgenic flies expressing a noncoding interrupted (CTG)\textsubscript{140} expansion show extensive muscle and eye degeneration concomitant with the development of ribonuclear foci (15, 16). Interestingly, Drosophila lines that express a GFP-DMPK 3’-UTR (CTG)\textsubscript{162} transgene develop ribonuclear foci, but they do not display any abnormal pathology (17). In contrast, overexpression of human SCA8/ATXN8OS cDNAs containing either 9 or 112 CTG repeats causes retinal neurodegeneration in Drosophila (18). However, genetic modifier screens have revealed significant differences between expression of nonpathogenic versus SCA8 pathogenic repeats (18). These and other observations suggest that whereas the pathogenic threshold for CUG repeats is dependent on the animal model, transgene, and expression level, repeats exceeding this threshold are pathogenic independent of flanking sequence context and whether these repeats are expressed as mRNAs or ncRNAs.

Elevated Fmr1 mRNA levels and intranuclear inclusions are also present in mouse (CGG)\textsubscript{98} (19) and (CGG)\textsubscript{120} (20) knock-in models for FXTAS, although only the larger repeat line recapitulates disease-associated loss of Purkinje cells. Evidence for dose- and repeat length-dependent toxicity of rCGG\textsuperscript{exp} repeats independent of FMR1 has emerged from Drosophila studies, where expression of (CGG)\textsubscript{160} or (CGG)\textsubscript{90}–EGFP transgenes leads to neurodegeneration accompanied by the presence of ubiquitin-positive intranuclear inclusions in neurons (21).

Expansions Create Toxic RNA Structures—What is the gain of function at the RNA level? For DM and SCA8, RNA structure prediction tools suggest that C(C)UG repeat expansions form extended hairpins with G-C Watson-Crick base pairs and U-U mismatches (Fig. 1A). This prediction has been confirmed experimentally by chemical and enzymatic structure probing, thermal denaturation studies, and visualization of these dsRNAs in the electron microscope (22–24). Additionally, the x-ray crystal structure of an 18-bp (CUG)\textsubscript{6} repeat has been solved to 1.58 Å resolution (25). This short repeat stacks upon itself to form pseudo-continuous A-form helices in which the U-U mismatches fail to base pair, resulting in an undistorted backbone. Although a CCUG repeat structure has not been determined, chemical/enzymatic probing indicates that these RNAs also form hairpins with consecutive C-U and U-C mismatches (22).

The RNA secondary and higher order structures formed by FXTAS-associated (CGG)\textsubscript{55–200} RNAs have been more controversial. Although structure probing and NMR spectroscopy indicate that rCGG repeats form stem-loop structures containing G-G mismatches (22, 26, 27), these repeats might also promote the formation of tetraplexes formed by a planar arrangement of four guanines stabilized by Hoogsteen-type hydrogen bonds (28).

Several cautionary notes should be mentioned for the interpretation of these structures and their pathological relevance. In general, these RNA structures have been determined using repeats in the normal range, and alternative conformational states may exist for longer repeats. Additionally, RNA folding pathways in cells are modulated by numerous interactions with a variety of factors, including RNA-binding proteins and small ncRNAs, so it is not yet clear if the RNA secondary and tertiary structures deduced from in vitro analyses are essential pathological features in vivo (29). Nevertheless, structural studies on DM, FXTAS, and SCA8 repeats indicate that expansion mutations result in the creation of novel dsRNA structures that gain a deleterious function.

Trans-effects of Pathogenic RNAs—ncRNAs perform a variety of cellular functions by forming specific complexes with proteins. To explain the mechanistic basis of RNA toxicity, the protein sequestration hypothesis predicts that C(C)UG\textsuperscript{exp} and rCGG\textsuperscript{exp} RNAs bind specific proteins and inhibit their normal cellular functions (2, 3). Alternatively, RNA repeat expansions could exert other effects on gene expression. Both of these
hypotheses appear to be true for DM, where C(C)UGexp RNAs have opposing effects on the activities of two RNA splicing factors, the MBNL1 (muscleblind-like 1) and CUGBP1 proteins. Indeed, current evidence suggests that DM is caused by a fundamental defect in alternative splicing regulation during development, resulting in the persistence of fetal protein isoforms in adult tissues (Fig. 1B).

The MBNL proteins were originally proposed to be the major sequestered factors in DM based on their identification as repeat length-dependent double-stranded CUG-binding proteins (30). Supporting evidence for this MBNL loss-of-function model comes from several observations. MBNL proteins colocalize with ribonuclear foci in DM1 and DM2 tissues, and Mbnl1 knock-out mice recapitulate disease-associated phenotypes, including myotonia, dystrophy, and mis-splicing of specific exons during postnatal development (2, 3). Moreover, overexpression of a single Mbnl1 isoform in the skeletal muscles of a poly(CUG) model for DM, HSA1L5 mice, reverses myotonia (31). Other proteins have been proposed to be sequestered in DM, including hnRNP H as well as several transcription factors, but the corresponding animal models have not been generated to test whether loss of function of these factors recapitulates the DM phenotype (2, 3). Finally, CUGBP1, a member of the CELF (CUG binding protein- and ETR3-like factor) family, was initially suggested as a sequestered factor, but this protein fails to colocalize with C(C)UGexp RNAs in ribonuclear foci, and its level and activity actually increase in DM (2, 3).

These and additional studies on the RNA splicing activities of the MBNL and CELF proteins have led to the following pathogenesis model for DM. During the embryonic and early neonatal developmental periods, CUGBP1 activates the splicing of specific fetal exons that are aberrantly retained in adult DM tissues (Fig. 1B). For example, CUGBP1 promotes the inclusion of the skeletal muscle troponin T (TNNT3) exon as well as skipping of the SERCA1 adult-specific exon 22 in fetal and neonatal tissues. During the postnatal fetal-to-adult transition, CUGBP1 levels decline, and MBNL1 relocalizes to the nucleus, where it promotes the opposite splicing pattern or SERCA1 exon 22 inclusion and TNNT3 fetal exon skipping. This normal transition in splicing pattern is blocked in DM because MBNL1 proteins are sequestered by C(C)UGexp RNAs, which also activate protein kinase C, leading to hyperphosphorylation and stabilization of CUGBP1 (32). The possibility that phosphorylation and elevated CUGBP1 levels are primary events in DM pathogenesis is supported by a recent study using EpA960 mice, where CUGexp RNAs were detected by RNA fluorescence in situ hybridization as early as 6 h following transgene induction with tamoxifen (13). Concurrent with the appearance of these RNAs, MBNL1 proteins colocalized with these RNAs, and CUGBP1 levels increased but only in those cells expressing CUGexp RNA.

An argument against the MBNL protein sequestration hypothesis is that nonpathogenic CTG repeats appear to be toxic in certain sequence contexts. Although HSA1L5 (CTG)15 lines are normal, mice that overexpress an inducible GFP-DMPK 3’-UTR (CTG)15 transgene develop myotonia and muscle wasting (33). Both of these (CTG)15 transgenes are expressed at relatively high levels in skeletal muscle. As noted previously, a potential complication with the GFP-DMPK 3’-UTR (CTG)15 transgenic model is that GFP expression is deleterious to muscle function and that regulation of GFP mRNA levels and/or translation by the DMPK 3’-UTR may contribute to the observed toxicity (3). Nevertheless, this result suggests that sequence context (DMPK versus HSA 3’-UTR) of the CUG repeat influences toxicity of short repeats. As described below, this conclusion led to further studies designed to clarify the binding properties of MBNL1 for both pre-mRNA splicing target and pathogenic RNAs.

The MBNL proteins contain either one or two pairs of N-terminal CCCH zinc finger motifs that are required for C(C)UGexp binding (30, 34, 35). The MBNL1 C-terminal region is more
variable and unstructured than the CCCH region but has been proposed to mediate MBNL homotypic and heterotypic interactions (36). Two recent reports indicate that MBNL1 CCCH motifs also recognize GC-rich RNA hairpins containing pyrimidine mismatches in normal cellular target RNAs, including TNNT2/cardiomyosin T and TNNT3 (36, 37). Indeed, the higher affinity of MBNL1 for CUG<sup>exp</sup> versus CUG<sup>esp</sup> RNAs suggests a prominent role for pyrimidine mismatches for MBNL recognition of pathogenic RNAs (34, 37). A fundamental mystery that remains to be solved is how MBNL1 proteins are sequestered by C(C)UG<sup>exp</sup> RNAs in DM cells particularly because they recognize similar structures on splicing precursor and pathogenic RNAs.

Protein sequestration has also been implicated as a pathogenesis trigger in FXTAS. Protein composition analysis of nuclear inclusions obtained from FXTAS patient autopsied brain tissues revealed >20 proteins, including MBNL1 and hnRNPs A2/B1 (11). In agreement with the hypothesis that sequestration of rCCGG-interacting proteins is an important event in FXTAS pathogenesis, overexpression of hnRNPs A2/B1, CUGBP1, or Purα suppresses the neurodegeneration phenotype observed in the Drosophila (CGG)<sub>90</sub>-EGFP transgenic model (38, 39). Purα is also detectable in FXTAS intranuclear inclusions, and Purα knock-out mice show neurological abnormalities reminiscent of FXTAS (40). It is not yet clear how loss of hnRNPs A2/B1, CUGBP1, or Purα might result in FXTAS, although these proteins have been implicated in various pathways, including DNA replication and transcription and mRNA transport and translation (39, 40). Conversely, other proteins are present in FXTAS intranuclear inclusions, including lamin A/C, which may be relevant because LMNA mutations cause a variety of diseases or lamiopathies (11, 41).

Pathogenic RNAs May Trigger dsRNA Pathways—The observation that expanded microsatellites are transcribed into RNA hairpins raises the possibility that a dsRNA-induced pathway is activated and contributes to pathogenesis. RNAi uses miRNAs and siRNAs, which are 20–22-nucleotide noncoding RNAs, to inhibit gene expression by binding to complementary (CNG) motifs and to promote mRNA degradation by transcriptional base pairing to mRNA 3′-UTRs (42). Strikingly, (CNG)<sub>n</sub> RNAs are incorporated into the RNAi pathway (43). These repeat RNAs form dsRNA structures similar to precursor miRNA and siRNA hairpins, and they are cleaved by Dicer to generate small interfering CNG sequences and to inhibit gene expression by binding to complementary (CNG)<sub>n</sub> sequences in target mRNA 3′-UTRs (Fig. 1C). Therefore, expanded repeat RNAs can function in trans to negatively regulate gene expression.

Although the RNAi pathway may contribute to disease pathogenesis, it is uncertain if this mechanism is causative or plays a secondary role. One argument against a primary effect is that the dsRNA substrates and catalytic enzymes are not localized in the same subcellular compartment because C(C)UG<sup>esp</sup> and CGG<sup>esp</sup> RNAs localize to discrete nuclear foci or inclusions, whereas Dicer is predominantly cytoplasmic (42).

Contribution of Antisense Transcription to Pathogenesis—As noted previously, transcription at the SCA8 locus produces two potentially pathogenic molecules, the polyglutamine protein ATXN8 and CUG<sup>esp</sup> RNA from ATXN8OS (Fig. 2A) (9). Two groups have also identified antisense transcripts at the FMR1 locus. ASFMR1 encodes ncRNAs that are transcribed from two alternative promoters (Fig. 2B) (44, 45). One of these promoters is bidirectional and produces a transcript, also named FMR4, which may possess an anti-apoptotic function (44). The other promoter, which drives the major transcription initiation site in premutation cells, is located in FMR1 intron 2 and produces a transcript that overlaps the CGG repeat region (45). Following splicing and polyadenylation, this transcript is exported from the nucleus and potentially encodes a polyproline protein. Although the function of antisense transcripts from these loci remains elusive, these results suggest that both protein and RNA gain-of-function effects contribute to SCA8 and possibly FXTAS pathogenesis.

A different mechanism has been proposed for DM1 in which bidirectional transcription and the RNAi pathway collaborate to induce regional heterochromatin formation and gene silencing (Fig. 2C) (46). The DMPK CTG repeat, which is flanked by CTCF-binding sites to form an insulator element, lies just upstream of the SIX5 transcriptional regulator, the HSE. The extension of an antisense transcript, which initiates in the HSE, through the CTG repeat is normally inhibited by CTCF binding. Unlike SCA8, this antisense transcript does not contain a polyglutamine open reading frame. However, when this region is perturbed by (CTG)<sub>1-1000</sub> expansions in congenital DM, enhanced Cpg methylation and impaired CTCF binding result in regional heterochromatization and possibly down-regulation of SIX5 mRNA levels (46). In contrast, DMPK levels may be

FIGURE 2. Antisense transcription in unstable microsatellite diseases. A, the SCA8 locus produces several overlapping transcripts (arrows, transcription initiation sites): ATXN8OS ncRNA, which is composed of exons D–A (open boxes with (CTG)<sub>n</sub>, expansion in exon A); ATXN8, which encodes a polyglutamine protein (unlabeled black box); and KLHL1. B, antisense transcription at the FMR1 locus (red oval; CTCF-binding sites; open box, FMR1 exon 1 5′-UTR with CGG repeat expansion; black box, open reading frame) produces ASFMR1/FMR4 RNAs (dotted line). ASFMR1 transcripts that initiate at the upstream site (+10243) contain (CCG), expansions. C, DMPK exon 15, which contains the 3′-UTR (CTG) flanked by CTCF-binding sites, expands in DM1 and lies immediately upstream of the SIX5 HSE (blue oval) that is also the initiation site for a (CAG)<sub>10</sub>-containing antisense transcript (ASDMPK).
up-regulated in congenital DM because of loss of insulator function.

Perspective

Noncoding mutations underlie disease pathogenesis in several hereditary disorders. Although additional microsatellite expansion diseases have been proposed to be caused by RNA gain-of-function effects, a broader question is whether other types of ncRNA mutations are pathogenic. For example, miRNA/mRNA specificity is largely determined by the miRNA seed sequence (42). Although point mutations in a miRNA seed could inhibit the normal regulation of its RNA targets, these mutations might also result in the recognition of a new set of miRNAs and gain-of-function effects.

REFERENCES

1. Mirkin, S. M. (2007) Nature 447, 932–940
2. Ranum, L. P., and Cooper, T. A. (2006) Annu. Rev. Neurosci. 29, 259–277
3. Wheeler, T. M., and Thornton, C. A. (2007) Curr. Opin. Neurol. 20, 572–576
4. Hagerman, P. J., and Hagerman, R. J. (2004) Am. J. Hum. Genet. 74, 805–816
5. O’Donnell, W. T., and Warren, S. T. (2002) Annu. Rev. Neurosci. 25, 315–338
6. Jin, P., and Warren, S. T. (2003) Trends Biochem. Sci. 28, 152–158
7. Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., and Hagerman, P. J. (2000) Am. J. Hum. Genet. 66, 6–15
8. Koob, M. D., Moseley, M. L., Day, J. W., and Ranum, L. P. (1999) Nat. Genet. 21, 379–384
9. Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., Chen, G., Weatherspoon, M. R., Clark, H. B., Ebner, T. J., Day, J. W., and Ranum, L. P. (2006) Nat. Genet. 38, 758–769
10. Margolis, J. M., Schoser, B. G., Moseley, M. L., Day, J. W., and Ranum, L. P. (2006) Hum. Mol. Genet. 15, 1808–1815
11. Iwashashi, C. K., Yasui, D. H., An, H. J., Greco, C. M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C. B., Hagerman, R. J., and Hagerman, P. J. (2006) Brain 129, 256–271
12. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Hender, T., and Thornton, C. A. (2007) J. Clin. Investig. 117, 2802–2811
13. Oren, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Darbinian, N., Otte, J., Meier, E., Johnson, E. M., Daniel, D. C., Kinoshita, Y., Amini, S., and Gordon, J. (2003) Mol. Cell. Biol. 23, 6857–6875
14. Stewart, C. L., Roux, K. J., and Burke, B. (2007) Science 318, 1408–1412
15. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008) Nat. Rev. Genet. 9, 102–114
16. Krol, J., Fischer, A., Mykowska, A., Sobczak, K., de Mezer, E., and Krzyzosiak, W. J. (2007) Mol. Cell 25, 575–586
17. Khalil, A. M., Faghihi, M. A., Modarresi, F., Brothers, S. P., and Wahlstedt, C. (2008) PLoS ONE 3, e1846
18. Ladd, P. D., Smith, L. E., Rabaia, N. A., Moore, J. M., Georges, S. A., Hansen, R. S., Hagerman, R. J., Tassone, F., Tapscott, S. J., and Filippova, G. N. (2007) Hum. Mol. Genet. 16, 3174–3187
19. Cho, D. H., Thienes, C. P., Mahoney, S. E., Analau, E., Filippova, G. N., and Tapscott, S. J. (2005) Mol. Cell 20, 483–489