it was initially thought to consist of two distinct electronic states, E and F, with internuclear distances of 1.01 and 2.31 Å, respectively. Theoretical analysis (*6*) showed the existence of a single adiabatic curve. The idea of multiple isomers on a single electronic potential surface is accepted as standard for polyatomic molecules but was thought to be a rarity for diatomic molecules, which have a single nuclear geometry coordinate. Several electronic states of $^1\Sigma_g^+$ symmetry with multiple minima are now known (*4*). Theory and experiment have also shown this phenomenon for a second electronic symmetry (*7*). The potential energy curve in the latter system has minima at 1.05 and 5.7 Å.

The complexity of the electronic energy levels of H_2 is also seen in their electric and magnetic susceptibilities. The surprising paramagnetism observed in one state (*4*) is explained by the mixing in of another nearby excited state. A molecular Stark effect has also been reported for excited states of H_2 (*8*). The apparent electric dipole moment of these states results from the frequent close proximity of gerade and ungerade states, that is, states whose wave functions are respectively symmetric and antisymmetric with respect to inversion.

The reactivity of H_2 depends on which electronic states are accessible. Among the simplest processes are electron and proton transfer. The rich chemistry of organic species that has been observed by radioastronomers (*9*) occurs primarily in massive dense dark molecular clouds, which cannot be penetrated by starlight. Here, molecular processing is initiated by cosmic ray ionization of H_2 and He. The latter is 1000 times more abundant than CO.

The secondary reactions of the ion products, H_2^+ and He^+, with the prevalent neutral H_2 exhibit electronic complexity. Collisions result in facile conversion of H_2^+ to H_3^+ (*10*). The highly exothermic reactions $He^+ + H_2 \rightarrow HeH^+ + H$, $He + H_2^+$, or $He + H + H^+$ essentially do not occur, whereas the much less exothermic reaction $H_2^+ + He \rightarrow HeH^+ + H$ occurs at collision frequency (*11*). It is the lack of H_2 reactivity with He^+ that permits the rich interstellar organic chemistry observed by radioastronomers. The latter has its origin in the copious production of carbon ions by the reaction $He^+ + CO \rightarrow C^+ + O + He$. C^+ adds efficiently to existing organic species, thereby increasing their carbon chain length and producing an interesting organic-rich molecular universe.

References and Notes

1. T. R. Geballe, *ASP Conf. Ser.* **41**, 147 (1993).
2. L. Spitzer, W. D. Cochran, A Hirshfeld, *Astrophys. J. Suppl. Ser.* **28**, 373 (1974).
3. D. C. Morton, *Astrophys. J.* **197**, 85 (1975).
4. O. Biham, I. Furman, V. Pirronello, G. Vidale, *Astrophys. J.* **553**, 595 (2001).
5. J. Rychlewski, *Adv. Quantum. Chem.* **31**, 173 (1999).
6. E. R. Davidson, *J. Chem. Phys.* **33**, 1577 (1960).
7. A. de Lange, W. Hogervorst, W. Ubachs, L. Wolneiwicz, *Phys. Rev. Lett.* **86**, 2988 (2001).
8. J. Stark, *Wien and Harms Handbuch d. Experimentalphysik*, vol. XXI (Akademische Verlagsgesellschaft, Leipzig, 1927), p. 442.
9. E. Herbst, *Annu. Rev. Phys. Chem.* **46**, 27 (1995).
10. B. J. McCall, T. R. Geballe, K. H. Hinkle, T. Oka, *Astrophys. J.* **522**, 338 (1999).
11. B. H. Mahan, *Acc. Chem. Res.* **8**, 55 (1970).

PERSPECTIVES: BIOMEDICINE

Reconstructing Myotonic Dystrophy

Stephen J. Tapscott and Charles A. Thornton

Myotonic dystrophy (DM) is an inherited disease characterized by progressive muscle weakness, hyperexcitability of the muscle membrane (myotonia), ocular cataracts, and cardiac arrhythmias. Some families with DM have a mutation in a gene on chromosome 19q13.3 (the *DM1* locus), whereas others have a mutation in a gene on chromosome 3q21 (the *DM2* locus). In 1992, the DM1 mutation was identified as an unusually large CTG trinucleotide repeat in the *DMPK* gene. However, as the mutation is in the noncoding region of *DMPK*, it is not clear how this expanded repeat actually causes DM1. Many investigators have been hoping that identification of the DM2 mutation would reveal a common pathogenic mechanism underlying both forms of the disease. Now, these hopes have been realized with the report by Liquori *et al.* (*1*) on page 864 of this issue. These investigators reveal that the DM2 mutation is a huge expansion of a tetranucleotide repeat (CCTG) in a noncoding region of the *ZNF9* gene. This finding adds strong support to the growing suspicion that the mutant RNA contributes to the constellation of features that characterize DM.

In many diseases caused by expansions of a trinucleotide repeat, such as Huntington's disease, a CAG repeat is located in a region of the gene that is translated into protein. Consequently, the protein contains an expanded stretch of polyglutamines that confers a toxic property on the protein, leading to the pathology of the disease. In contrast, the CTG repeat at the *DM1* locus is located in the 3′-untranslated region of the *DMPK* gene, a region that is transcribed into RNA but never translated into protein. Because the CTG expansion at the *DM1* locus does not alter the protein sequence encoded by *DMPK*, the mechanism of pathogenesis in DM1 must be different from that in the polyglutamine encoding CAG-repeat diseases.

Three different theories have been proposed to explain how repeat expansions in the *DMPK* gene result in DM1. The expanded CUG repeat in RNA transcribed from the mutant gene has important effects on the metabolism of this and other transcripts (*2*). Most dramatically, the mutant RNA accumulates in discrete foci in the cell nucleus rather than being transported to the cytoplasm, where translation of mRNA into protein normally takes place (see the figure) (*3, 4*). Therefore, DM could be caused by a decrease in the amount of DMPK protein (because the RNA from the mutant allele is in the nucleus and is not available for translation into protein) or by the abnormal RNA itself, which may interact with and disrupt the activity of nuclear RNA binding proteins. A third possibility is that DM is caused by a decrease in the amount of the SIX5 transcription factor, because the CTG expansion at the *DM1* locus suppresses the expression of the adjacent *SIX5* gene.

There is experimental support for each of these theories. Disruption of the *Dmpk* gene in mice causes a cardiac arrhythmia similar to that seen in DM (*5–7*), but does not cause other features of the disease such as myotonia or cataracts. Myotonia and muscle disease can be induced in mice with a transgene that is transcribed into an RNA with a large CUG repeat in its untranslated region (*8*). It is the mutant RNA that promotes muscle disease in these mice, because animals do not develop disease if the transgene containing the expanded repeat is not transcribed into RNA. Because expression of the transgene in this experiment was restricted to skeletal muscle, it remains unclear whether expression of the mutant RNA in other cell types could produce additional features of the disease. Finally, disruption of *Six5* in mice causes ocular cataracts (*9–12*). From these

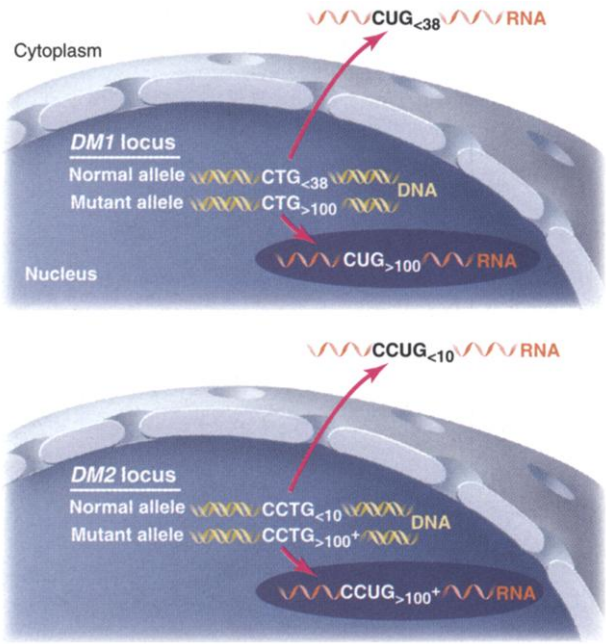
S. J. Tapscott is in the Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. E-mail: stapscot@fhcrc.org. C. A. Thornton is in the Department of Neurology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA. E-mail: charles_thornton@urmc.rochester.edu

findings, it is tempting to speculate that the cardinal features of DM are attributable to the combined effects of a CUG repeat in the RNA (causing myotonic myopathy) and a deficiency in DMPK and SIX5 proteins (causing cardiac conduction defects and cataracts, respectively) (*13*).

Liquori *et al.* (*1*) present good evidence that the principal features of DM2 are caused by a huge expanded repeat in the *ZNF9* gene. The repeat is in the first intron of *ZNF9*, a region that is transcribed into RNA but not translated into protein. Similar to the CUG-containing RNA in DM1, the CCUG-containing RNA in DM2 accumulates in a small number of foci in the cell nucleus. Notably absent from the *DM2* locus are genes homologous to *DMPK* or *SIX5*. Therefore, the simplest explanation is that expression of an RNA containing either a CUG or CCUG expanded repeat, rather than a deficiency in the DMPK, SIX5, or ZNF9 proteins, accounts for the disease features common to both DM1 and DM2.

Now that a repeat containing RNA has been implicated as the cause of both DM1 and DM2, effort can be devoted to identifying the common features of the two mutant RNAs. In the case of DM1, the expanded CUG repeats can form a hairpin structure with an extended region of double-stranded RNA (dsRNA) (*14*). The periodic U·U mismatch in the dsRNA might affect interactions with binding proteins and RNA-modifying enzymes. Although CUG repeats above a threshold length of about 20 repeats are recognized by dsRNA binding proteins, they are not cleaved by ribonuclease III (an enzyme that cleaves dsRNA) (*15*). Also, as they lack adenosines, expanded CUG repeats are not substrates for adenosine deaminases that act on RNA (ADARs), a family of proteins that modify dsRNAs, targeting some of them for rapid degradation. All of these characteristics may be shared by the mutant DM2 RNA, because RNA folding algorithms predict that CCUG repeats can also form stable hairpins with a periodic mismatch in the stem.

Perhaps the most striking characteristic shared by the mutant RNAs, however, is the sheer size of the expanded repeat: up to 15 kilobases in *DMPK* and 44 kilobases in *ZNF9*. The focal accumulation of enormously expanded mutant RNA transcripts in the nucleus might lead to the sequestration or activation of specific RNA binding proteins. CUG repeats interact with several proteins in vitro, including CUG-binding protein 1, an elav-type ribonucleoprotein, a dsRNA binding protein kinase, and a muscleblind-like protein (*15–18*). Comparative analysis of proteins that become localized to the nuclear foci of CUG- and CCUG-repeat RNAs will help to identify candidate proteins that could contribute to the pathology of DM1 and DM2. The neurodegenerative disease spinocerebellar ataxia type 8 (SCA8) is also caused by a CTG-repeat expansion in an untranslated region of a gene (*19*, *20*). It is possible that SCA8, DM1, and DM2 may all share a common disease pathway.



Repetitive injuries on a molecular scale. Abnormally large expanded repeats in RNAs cause DM1 and DM2. (**Top**) A large expanded CTG repeat in the *DMPK* gene at the *DM1* locus on chromosome 19 causes DM1. (**Bottom**) A large expanded CCTG repeat in the *ZNF9* gene at the *DM2* locus on chromosome 3 causes DM2. In both cases, the mutant RNAs containing the huge repeat expansions accumulate in a small number of foci in the nucleus. In contrast, RNA from the normal allele is spliced to form mRNA, which is exported to the cytoplasm and then translated into protein.

Identifying the DM2 mutation will focus attention on how mutant RNAs containing expanded repeats are involved in disease. We should not, however, ignore the possible contributions of the DMPK, SIX5, and ZNF9 proteins. For example, the cardiac arrhythmia in *Dmpk*-deficient mice is very similar to the arrhythmia in DM1, and it remains plausible that both a deficiency in the DMPK protein and the expanded CUG repeat in the RNA cause specific aspects of cardiac disease in DM1. Similarly, SIX5 may also contribute to cataract formation in DM1. The next steps will be to carefully compare shared characteristics and differences between DM1 and DM2 patients, and to develop animal models that explore the effects of expanded CUG or CCUG RNAs in appropriate tissues.

It is curious that some expanded CTG repeats in genes are not associated with disease. For example, some individuals with large CTG expansions at the *SCA8* locus do not have cerebellar disease, and moderately large expansions of a CTG repeat in an intron of the *E2-2* gene are also not associated with disease (*19–21*). Understanding these exceptions to the emerging rule that CUG and CCUG expansions lead to disease may reveal ways in which the toxicity of a repeat-containing RNA can be modulated, opening the door to new therapeutic strategies.

Once again, human genetics has challenged us to expand our understanding of cell biology. Identifying mutations in several different genes associated with familial Alzheimer's disease has led to a realization of the normal proteolytic events that occur in the endoplasmic reticulum and at the plasma membrane. In similar fashion, identifying noncoding DM1 and DM2 mutations might lead to a better understanding of how nuclear RNAs regulate cellular processes. Of more immediate importance, particularly for families with DM, accurate genetic diagnosis for both DM1 and DM2 is now possible. In addition, therapies designed to limit the production or accumulation of the repeat-containing RNAs might prove to be beneficial in treating these debilitating diseases.

References

1. C. L. Liquori *et al.*, *Science* **293**, 864 (2001).
2. A. V. Philips *et al.*, *Science* **280**, 737 (1998).
3. K. L. Taneja *et al.*, *J. Cell Biol.* **128**, 995 (1995).
4. B. M. Davis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7388 (1997).
5. C. I. Berul *et al.*, *J. Clin. Invest.* **103**, R1 (1999).
6. S. Saba *et al.*, *J. Cardiovasc. Electrophysiol.* **10**, 1214 (1999).
7. W. J. Groh, *J. Cardiovasc. Electrophysiol.* **10**, 1221 (1999).
8. A. Mankodi *et al.*, *Science* **289**, 1769 (2000).
9. T. R. Klesert *et al.*, *Nature Genet.* **16**, 402 (1997).
10. C. A. Thornton *et al.*, *Nature Genet.* **16**, 407 (1997).
11. T. R. Klesert *et al.*, *Nature Genet.* **25**, 105 (2000).
12. P. S. Sarkar *et al.*, *Nature Genet.* **25**, 110 (2000).
13. S. J. Tapscott, *Science* **289**, 1702 (2000).
14. M. Napierala, W. J. Krzyzosiak, *J. Biol. Chem.* **272**, 31079 (1997).
15. B. Tian *et al.*, *RNA* **6**, 79 (2000).
16. L. T. Timchenko *et al.*, *Nucleic Acids Res.* **24**, 4407 (1996).
17. X. Lu *et al.*, *Hum. Mol. Genet.* **8**, 53 (1999).
18. J. W. Miller *et al.*, *EMBO J.* **19**, 4439 (2000).
19. M. D. Koob *et al.*, *Nature Genet.* **21**, 379 (1999).
20. M. L. Moseley *et al.*, *Hum. Mol. Genet.* **9**, 2125 (2000).
21. T. S. Breschel *et al.*, *Hum. Mol. Genet.* **6**, 1855 (1997).