

number are very loose: many of the data are based upon reassociation techniques which have technical limitations, and also upon biological assumptions that are often untested (for example, that there is no significant proportion of genes that are expressed in very tight developmental windows). The question is explicit. Given

that we do not know the gene density of any but tiny regions of vertebrate genomes, what exactly is the experimental basis of contentions that cDNA sequencing alone is the way forward? □

Peter Little is in the Department of Biochemistry, Imperial College, Prince Consort Road, London SW7 2AZ, UK.

HUMAN GENETICS

The costs of instability

Kay E. Davies

THE identification over the past year of the mutations involved in three neurological disorders — fragile X syndrome¹⁻³, spinal bulbar muscular atrophy⁴ and, now, myotonic dystrophy⁵⁻¹⁰ — has revealed that a common genetic mechanism underlies them all. This creates a very exciting prospect: the rapid identification of other disease genes having a similar defect, the possession of sequences of unstable DNA.

The myotonic dystrophy (DM) story has emerged very quickly (as described in News and Views last week), the elements coming from several laboratories and appearing in *Nature*⁵⁻⁷ last month, and in the latest issues of *Cell*⁸ and *Science*^{9,10}. The disease is an autosomal dominant multisystem disorder; mildly affected individuals have cataracts but little or no myopathy, whereas severe cases are characterized by muscular hypoplasia and mental retardation¹¹. An increase in severity of the disease in successive generations is observed in susceptible families (genetic anticipation). Ascertainment bias has been suggested to explain this pattern of inheritance¹², but the new work shows that anticipation is due to the progressive amplification of a DNA sequence.

The amplification of sequence was first observed using genomic DNA from DM individuals⁵⁻⁷, and has now been shown to be due to the expansion of a CTG repeat at the 3' end of a transcript designated DM-1 (refs 8-10). Normal individuals also show the variable numbers of the CTG repeat with a heterozygosity frequency of 75 per cent; the modal number of repeats is 5 with the largest being 77 (ref. 8). In contrast, minimally affected DM individuals have at least 52 copies of the repeat. Assuming that larger amplifications are expansions of CTG, then severely affected cases must often possess more than 1,000 copies of the CTG trinucleotide. This type of mutation readily explains the molecular basis of genetic anticipation. Asymptomatic carriers in early generations pass on the mutant chromosome to their offspring with an increase in the number of repeats. Consequently, in successive generations, the repeat length

increases and a greater severity of the disease and earlier age of onset are observed. Amplification of the repeats occurs whether the mutation is passed through the male or female line. The present data do not, however, bear upon the question of why severe congenital cases are almost exclusively maternally transmitted.

The expansion of the CTG repeat seems to have occurred only in a specific population subgroup, because there is linkage disequilibrium between DM and polymorphisms closely linked to the disorder on chromosome 19 (refs 13, 5). The initial amplification event (perhaps a doubling of the number of repeats) may have existed in populations for some time before finally expanding to give a clinical phenotype. Studies of populations as yet unaffected by the disorder will be necessary to investigate this further.

An unusual characteristic of CTG amplification is that it is also associated with somatic heterogeneity. Individuals with large amplifications show a smear of fragments rather than a discrete allele. It will be of great interest to determine whether the same multiple pattern of mutant fragments is observed in all tissues, as the phenomenon could also contribute to clinical variability.

Sequencing studies demonstrate that the CTG repeat occurs within the 3' untranslated region of a gene. The transcript is approximately 3 kilobases in length and is expressed at high levels in heart and to a lesser extent in skeletal muscle and brain^{8,9}. Database sequence searches reveal similarities with genes encoding cyclic-AMP-dependent protein kinase, the closest similarity being to protein kinase TKR-YKR from *Saccharomyces cerevisiae*. Abnormalities in function or regulation of such a molecule would fit well with the pleiotropic effects exhibited by DM patients; indeed, a defect in such a protein was predicted by Roses and Appel almost two decades ago^{14,15}.

It is not immediately apparent why a mutation at the 3' end of a gene should result in a dominant phenotype. It seems unlikely that the expression of such a

gene would be sensitive to dosage. There may be critical regulatory factors in the 3' region of the gene which are disrupted by the mutation in a way that would have a dominant effect (there are examples of such regulatory elements in the nematode^{16,17}). Detailed knowledge of the regulation of DM-1 will be required to elucidate this further.

Is the genetic mechanism underlying DM instructive for determination of the molecular basis of other genetic diseases? The fragile X syndrome is already well documented; in this case, genetic anticipation is also explained by the amplification of a trinucleotide repeat (in this instance CGG)¹⁻³. In a similar manner to DM, an increase in repeat size is seen in successive generations and somatic heterogeneity is evident. In fact, it was the nature of the fragile X mutation that prompted Fu *et al.*⁹ to scan cosmids covering the DM candidate region for trinucleotide repeats.

Trinucleotide expansion also underlies X-linked spinal bulbar muscular atrophy⁴. Here the number of CAG repeats in the androgen-receptor gene increases with either the maternal or paternal transmission 30-35 per cent of the time overall (K. H. Fischbeck, personal communication). But there is no genetic anticipation, and dramatic expansions have not been observed (patients, 39-60 repeats; controls, 13-30 repeats). This lack of large numbers of repeats may explain why somatic instability is not evident in these mutant alleles.

The three examples discussed here show that it may well be worthwhile to scan complementary DNA libraries for cDNAs carrying trinucleotide repeats. It seems likely that several other genetic disorders are caused by the instability of GC-rich trinucleotide repeat elements, a prospect that will keep human molecular geneticists busy for some time. □

Kay E. Davies is in the Molecular Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK.

- Oberle, I. *et al.* *Science* **252**, 1097 (1991).
- Kremer, E. J. *et al.* *Science* **252**, 1711-1714 (1991).
- Fu, Y-H. *et al.* *Cell* **67**, 1047-1058 (1991).
- La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. *Nature* **352**, 77-79 (1991).
- Harley, H. G. *et al.* *Nature* **355**, 545-546 (1992).
- Buxton, J. *et al.* *Nature* **355**, 547-548 (1992).
- Aslanidis, C. *et al.* *Nature* **355**, 548-551 (1992).
- Brook, J. D. *et al.* *Cell* **53**, 799-808 (1992).
- Fu, Y-H. *et al.* *Science* **255**, 1256-1258 (1992).
- Mahadevan, M. *et al.* *Science* **255**, 1253-1255 (1992).
- Harper, P. S. *Myotonic Dystrophy*, 2nd edn (W. B. Saunders, London, 1989).
- Howeler, C. J., Busch, H. F. M., Geraedts, J. P. M., Niermijer, H. F. & Staal, A. *Brain* **112**, 779-797 (1989).
- Harley, H. G. *et al.* *Hum. Genet.* **87**, 73-80 (1991).
- Roses, A. D. & Appel, S. H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1855-1859 (1973).
- Roses, A. D. & Appel, S. H. *Nature* **250**, 245-247 (1974).
- Ahringer, J. & Kimble, J. *Nature* **349**, 346-348 (1991).
- Wrightman, B., Bürglin, T., Gatto, J., Arosu, P. & Ruvbun, G. *Genes Dev.* **5**, 1813-1824 (1991).