Triplet repeat genes raise questions

The novel mechanism for genetic mutation in Huntington's disease offers no comfort for neo-Lamarckians, but raises many other questions likely to prove difficult as well as intriguing.

Is THE Human Genome Project about to rehabilitate Lamarckian inheritance, or at least the inheritance of some acquired characteristics? That seems to be the unspoken question attending the recognition in the past few years of the role of repetitive and simple DNA sequences in the inheritance of several diseases, all of them (so far) conditions with neurological consequences.

Seven conditions have so far been well characterized, including fragile X syndrome, Huntington's disease (HD) and myotonic dystrophy, in which the underlying genetic mutation is an increase of the number of nucleotide triplets outside its normal range. In HD, for example, the gene for a protein now called huntingtin (for want of a clear idea of its function) on the shorter arm of chromosome 4 normally contains between 6 and 37 repeats of the nucleotide sequence CAG (which codes for the amino acid glutamine), but people who inherit from either parent a chromosome 4 in which the number of CAG repeats lies above that range are at risk of developing HD.

Indeed, in HD and the other diseases, there is a close inverse correlation between the number of CAG repeats and the age of onset. With 40 repeating units, the onset of HD is likely to be delayed until middle age, but with a chromosome containing more than 50 repeats, HD is likely to appear in a person's teens or even infancy. HD is dominantly inherited; either a maternal or a paternal chromosome 4 with an excess of repetitive CAG units is likely to be damaging.

Worse than that, the phenomenon of what the geneticists call anticipation applies. In other words, the age of onset tends to be reduced and the severity of the disease increased in successive generations in families susceptible to the disease. That in turn implies that the longer the repetitive units in one generation, the longer still they are likely to be by the next.

Is that not simply Lamarckian? The old chestnut about the length of the giraffe's neck almost literally applies. Lamarckians would have it that earlier generations of animals stretched their necks to reach into trees and were then able to pass on this advantage to their offspring; Darwinians, while readily confessing that there is no evidence to show what would have been the selective advantage of longer necks at intermediate stages, would insist that the eventual inheritance of this characteristic would have arisen from the natural selection of naturally arising variations.

Sadly (at least for Lamarckians) the analogy is false. Both Darwinians and Lamarckians are concerned with the evolution of physical characteristics, aspects of the phenotype. Precisely what features of the genome are responsible for them is strictly speaking irrelevant to last century's argument, although it may have a bearing on the more recent question whether the concept of the 'selfish gene' is a faithful representation of the darwinian process. For do not CAG repeats that tend to lengthen *in situ* have an intrinsic selfishness? Again, the analogy is false; for a selfish gene to succeed, it must populate as many organisms as possible, not debilitate a few.

In reality, of course, HD is a familial disease and the susceptibility is inherited by the mendelian rules that apply to genetically dominant conditions. Moreover, it must be plain that neither the particular CAG repeat involved in HD nor triplet repeats in general can be inherently unstable, for then we should have to face the chilling prospect that all families would eventually carry HD, as well as, no doubt, fragile X syndrome and the other conditions. From there it would be a short step to the conclusion that *Homo sapiens* will soon come to a sticky end by the prevalence of neurodegenerative disorders whose potential is embodied in its genome.

Two questions then arise. Given that the length of the repetitive CAG sequence in the HD gene is an unambiguous determinant of disease, what is the physiological connection between the length of the repeat and the damage done in the brain (chiefly the death of neurons)? And what feature of the genome is responsible for the instability of the CAG repeating unit in HD families, or is the true genetic cause of familial HD?

What follows is a necessarily random gleaning from the recent literature that seems to throw some light on that question. Remarkably, as it happens, the nucleotide sequence of huntingtin as originally published (*Cell* **72**, 971–983; 1993) showed that immediately downstream of the CAG repeat is a different trinucleotide repeat, formed of the nucleotides CCG (coding for the amino acid proline). And now it turns out that the length of this repetitive unit is variable between individuals ('polymorphic' as they say in the trade); it may even be part of the explanation why the CAG repeating unit is unstable in HD families.

Two separate reports in the same issue of *Human Molecular Genetics* seem to show, when taken together, that the polymorphism of the CCG repeat helps to determine the stability of the CAG repeat. Thus Susan A. Andrew *et al.* (*Hum. molec. Genet.* **3**, 65–67; 1994) from the University of British Columbia show that in 205 normal people, the CCG repeating unit can contain 7, 9, 10, 11 or 12 repeating units, of which 7 and 10 respectively account for 67 per cent and 30 per cent of the population, but that HD patients almost invariably have 7 CCG units.

At the same time, Lilias H. Barron *et al.* from the Human Genetics Unit of the University of Edinburgh describe a study of the CCG polymorphism in HD patients and normal people from Scotland (*Hum. molec. Genet.* **3**, 173–175; 1994) in which there appear to be five different versions of the CCG repeating unit. The numbers are not exactly comparable with those in the Vancouver study, because the Edinburgh group has measured both the first stretch of CCG repeats and a second separated from it by 18 nucleotides, but again they find that HD patients invariably carry just one of these, that which happens to be most common in the Scottish population.

One should not jump too quickly to conclusions, and it cannot be the case that the most common CCG repeating unit is the sole decisive determinant of the instability of the CAG repeat that lies upstream of it, for 60 per cent of us would then be at risk of HD. But the 7-CCG repeating unit (in the Vancouver nomenclature) looks as if it is a necessary, but not a sufficient, condition for instability. That should be an entertaining conundrum for structural biologists, but the search for the other determinants of instability remains a problem for geneticists.

On the first question, the mechanism by which the CAG repeat causes HD if there are more than 40 units in it, there is less to say. Because HD is dominantly inherited, there must be 'gain of function', but that says nothing of what the function may be. So it is as well that people have made a start on telling where huntingtin is to be found in the body. A group from Leiden and Rotterdam (A. T. Hoogeven, *et al. Hum. molec. Genet.* **2**, 2069–2073; 1993) has, for example, used antibodies against a huntingtin peptide to show the presence of huntingtin in brain, testis and other tissues. The distribution appears the same in normal people and those with HD. The protein appears in the nuclei of neurons but not of other somatic cells.

That, sadly, is only a beginning. And no amount of elegant histochemistry will tell what is the natural function of huntingtin. What gain of function there can be that enables the aberrant protein to kill neurons is something else again. **John Maddox**