Human Molecular Genetics, 1999, Vol. 8, No. 2 157–164

ARTICLE

Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis

Ammar Al-Chalabi[1,*], Peter M. Andersen[2], Peter Nilsson[3], Barry Chioza[1], Jörgen L. Andersson[3], Carsten Russ[1], Christopher E. Shaw[1], John F. Powell[1] and P. Nigel Leigh[1]

[1]Departments of Neuroscience and Clinical Neurosciences, Institute of Psychiatry and King's College School of Medicine and Dentistry, London SE5 8AF, UK and [2]Department of Neurology and [3]Department of Chemistry, Umeå University, S-901 85 Umeå, Sweden

Received July 20, 1998; Revised and Accepted November 19, 1998

Amyotrophic lateral sclerosis (ALS) is a progressive motor neuron degeneration resulting in paralysis and death, usually within 3 years of onset. Pathological and animal studies implicate neurofilament involvement in ALS, but whether this is primary or secondary is not clear. The heavy neurofilament subunit (NFH) tail is composed of a repeating amino acid motif, usually X-lysine-serine-proline-Y-lysine (XKSPYK), where X is a single amino acid and Y is one to three amino acids. There are two common polymorphic variants of 44 or 45 repeats. The tail probably regulates axonal calibre, with interfilament spacing determined by phosphorylation of the KSP motifs. A previous study suggested an association between sporadic cases of ALS and NFH tail deletions, but two subsequent studies have found none. We have analysed samples from two different populations (UK 207, Scandinavia 323) with age-matched controls for each group (UK 219, Scandinavia 228) and have found four novel NFH tail deletions, each involving a whole motif. These were found in three patients with sporadic ALS and a family with autosomal dominant ALS, although another was also found in two young controls. In all cases motif deletions were only associated with disease when paired with the long NFH allele. The deletions all occurred within a small region of the NFH tail. This has allowed us to propose a structural organization of the tail as well as allowing observed deletions both from this study and previous reports to be organized into logical groups. These results strongly suggest that NFH motif deletions can be a primary event in ALS but that they are not common.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a progressive motor neuron degeneration resulting in paralysis and death, usually from respiratory failure, in 3–5 years (1,2). In 5–10% of cases there is a family history suggesting autosomal dominant inheritance. Mutations in the gene for Cu/Zn superoxide dismutase have been found in ~20% of familial cases (3,4) and, more recently, in ~2% of apparently sporadic cases (4–6).

Pathologically, neurofilament accumulation is a hallmark of ALS and occurs early in the disease, in particular in the proximal axons and cell bodies of lower motor neurons (7). Neurofilaments are 10 nm intermediate filaments specific to neuronal cells (8). They have a globular head, α-helical rod region and a globular tail. There are three neurofilament subunits of light, medium and heavy molecular weight: NFL, NFM and NFH. The neurofilaments are often found in association with ubiquitin immunoreactive inclusions, suggesting a protein abnormality or a defect of the normal degradation pathway (9–11). Evidence of a primary role for abnormal neurofilaments in motor neurodegeneration comes from animals. The Japanese quivering quail is a naturally occurring recessive variant which lacks NFL and quivers as a result of damage to the motor system, with accompanying pathological changes (12,13). Transgenic mice with point mutations of the NFL rod, overexpression of human NFH or overexpression of mouse NFL all develop motor neuron pathology and mice transgenic for mutant *SOD1* develop neurofilament accumulations (14–17), as do humans with D48Q, A4V and I113T SOD1 mutations (18,19). Neurofilaments may therefore have a primary role in ALS pathogenesis or accumulate secondary to other cell death mechanisms, and neurofilament genes are therefore candidates for mutations in ALS. There are two key areas in neurofilament subunits which are of interest in this regard: the ends of the rod domain, which are highly conserved, and the tail domains of NFM and NFH, which contain a repeated motif, X-lysine-serine-proline-Y-lysine (XKSPYK),

*To whom correspondence should be addressed. Tel: +44 171 346 5172; Fax: +44 171 346 5186; Email: ammar@iop.bpmf.ac.uk

where X is a single amino acid and Y is one to three amino acids. In humans, there are two common variants of the NFH tail, one with 45 motif repeats and one with 44 repeats (although four repeats vary slightly from this theme) (20). These have been mislabelled in the published literature as 44 and 43 repeat variants, respectively, and will therefore be referred to simply as long (L) and short (S) alleles.

Previous studies have addressed the issue of neurofilaments as candidate genes in ALS but without a clear resolution. In a study of the NFH tail in 356 cases of sporadic ALS with 306 controls, deletions of two types were found in five of the cases and in none of the controls, although they were also found in healthy relatives (20,21). Four apparently unrelated individuals had a deletion of a codon for valine (referred to in this text as allele Δ2) after KSP repeat 40 by the numbering of Lees *et al.* (22), destroying a recognition sequence for cdk5, a neurofilament kinase. The other individual had deletion of a section coding for 33 amino acids, encompassing repeats 5–9 (deletion of 1582–1683, allele Δ1). In contrast, a study of 117 cases of familial ALS found no deletions (23) and a study of all neurofilament subunits in 100 familial and 75 sporadic cases found point mutations in cases and controls but no deletions (24). It is possible that the NFH deletions which the original study found in the sporadic cases were unrelated to disease. An alternative explanation is that deletions have a low penetrance and familial studies will therefore select against them in the sampling process. They were found at a frequency of ~2% in the original study and the study of 75 sporadic cases therefore has insufficient power to reliably detect deletions.

In view of the strong circumstantial evidence implicating neurofilaments and with the issue of NFH tail deletions unresolved, we performed a study of UK and Scandinavian cases with matched controls. We have found four novel deletions occurring in four cases of ALS, three unaffected relatives, one individual with suspected ALS and two unrelated controls. Three of the deletions occurred in the same region and the fourth was in a homologous region downstream. All followed a pattern, allowing a recoding of the NFH tail and previous mutants and the known polymorphism by the pattern. In addition, a point mutation associated with the S allele of NFH was found, acting as a marker for the origin of the deletion mutants. These results support the role of NFH tail deletions in ALS and shed light on the functional organization and evolution of the NFH tail.

RESULTS

In order to maximize the power of our study, we identified samples in the UK and Scandinavia, with age-matched controls. In all, 977 samples were analysed. From the UK there were 207 cases, consisting of 19 FALS and 188 SALS patients. Four of those with FALS had SOD1 mutations, as did four of those with apparently sporadic ALS (including one patient with the D90A mutation which is prevalent in Scandinavia). In addition, 17 relatives of patients were screened and 202 unrelated controls. The mean age of onset for sporadic cases was 55.2 years (58% male) and mean age of the unrelated controls 55.8 years (*t*-test, $P = 0.657$) (44% male). From Scandinavia, there were 323 cases (Denmark eight, Norway 33, Sweden 222 and the ethnically distinct Finland 60), consisting of 59 FALS and 264 SALS patients. Fourteen of those with FALS had mutations of SOD1 and 11 of these had the D90A mutation. In addition, 67 relatives of patients were screened and 161 unrelated Scandinavian controls. Mean age at sample donation was 61.5 for cases (60% male) and 56.8 for controls (44% male).

For patients and unrelated controls from both populations, NFH alleles were in Hardy–Weinberg equilibrium. From the UK, 78 patients were LL genotype, 90 LS, 37 SS and two L/deletion ($n = 207$), compared with 127 LL, 147 LS, 47 SS and two L/deletion in the Scandinavian patients ($n = 323$); unrelated controls from the UK, 73 LL, 95 LS, 34 SS ($n = 202$); unrelated controls from Scandinavia, 55 LL, 77 LS, 27 SS, one L/deletion and one S/deletion ($n = 161$). This gives 37% LL, 46% LS and 16% SS, which is in broad agreement with previous reports, although there is a slightly higher than expected frequency of the SS genotype. There was no association between L or S allele genotype and ALS (χ^2 test, $P = 0.82$ UK, 0.54 Scandinavia), family history (χ^2 test, $P = 0.99$ UK, 0.48 Scandinavia), mode of presentation (limb or bulbar onset) (χ^2 test, $P = 0.67$ UK, 0.12 Scandinavia), age of onset (linear regression, $r^2 = 0.006$ UK) or survival (Kaplan–Meier log rank 0.19 UK).

To prevent confusion, the numbering of the published sequence by Lees *et al.* (22) (S NFH allele) will be used to describe deletions. For the L allele, the 18 bp insertion has been taken as occurring after base 1964, but is unnumbered. In addition, the original paper has misnumbered the repeats, with two of them labelled as repeat 34. This too will be maintained through this paper. It should also be borne in mind that as the sequence is so repetitive, deletions can be regarded as starting from different positions but with the same resulting sequence (as an analogy, 'DILATATION' to 'DILATION' may be produced by removal of the first AT, second AT or middle TA) and base numbering in the text may therefore be different from that cited elsewhere.

In the UK cases, there were two deletions: a 24 bp deletion of 1965–1988 (Δ5) and an 18 bp deletion of 1989–2006 (Δ6), both on the background of the S allele (Figs 1 and 2 and Table 1). The clinical features of these cases were entirely typical of ALS. The first individual (D141, genotype LΔ5) was a man with El Escorial probable ALS and lower limb onset at age 66 who survived 19 months. The second individual (C72, genotype LΔ6) was a woman, had El Escorial definite ALS with upper limb onset at age 73 and survived 33 months. Neither patient had a family history of ALS, atypical clinical features or SOD1 mutations and in both cases there was electromyographic (EMG) evidence to support the diagnosis and no evidence of conduction block. DNA was not available from relatives of these cases.

In the Scandinavian samples there were two deletions (Fig. 3 and Table 1). A 42 bp deletion of 1989–2030 (Δ3) on the background of the L allele and an 18 bp deletion of 2230–2247 (Δ4) on the background of the L allele. The Δ3 deletion was found in a FALS pedigree (Fig. 4), but it is not known whether the deletion segregates with the disease. The index case (29, genotype LΔ3) had entirely typical ALS, with lower limb onset at age 66, bulbar symptoms within 1 year and surviving 23 months. EMG supported the diagnosis and there was no electromyographic evidence of multifocal conduction block, but there were elevated anti-LK-1 ganglioside antibodies (1:400). Levels of anti-GM-1, GA-1, 3′LM1, GM-2, GD-1a and GD-1b antibodies were normal. Her mother (genotype unknown) developed ALS at age 71 with a similar phenotype and survived 5 years. One of her children (627), who is younger than 50, carries the Δ3 mutation but remains well. Of interest is the phenotype of her brother (622), who since the age of 71 has developed a gait disturbance with symptoms in the right leg and signs of weakness and wasting in

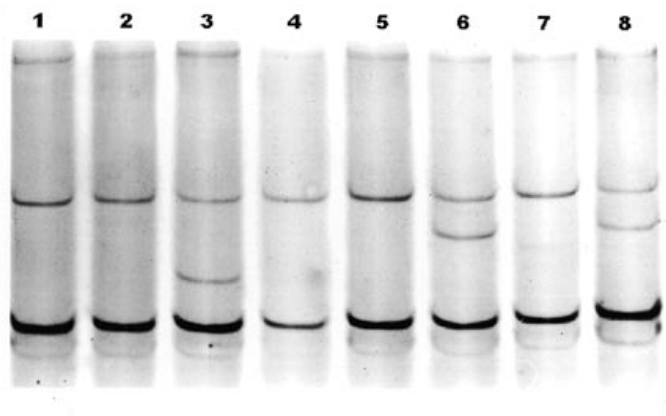

Figure 1. SSCP gel of UK cases. Lanes 1, 2, 4, 5 and 7, individuals with LL genotype; lanes 6 and 8, individuals with LS genotype; lane 3, patient C72 with LΔ6 genotype.

the right foot and lower leg. This has been only very slowly progressive over a follow-up period of 7 years. EMG revealed neurogenic changes distally in both legs and in the right hand, but there were no proximal changes. He can therefore be classified as having suspected ALS by the El Escorial criteria, but he also carries the Δ3 mutation. Three other members of this pedigree with ALS have died, but no samples were available for testing.

The patient with the Δ4 mutation had El Escorial definite ALS, with onset in the upper limb at 50 years old and survival for 38 months. EMG supported the clinical diagnosis. His mother remains well aged >70, his father died of a dementing illness aged 79. His brother (838, genotype SΔ4) is 48 and well. Two Scandinavian unrelated controls had the Δ6 deletion (P107 and P150), but are aged 38 and 24, respectively, making it unlikely that they would have yet developed ALS.

Table 1. Genotype and phenotype correlations for the known motif deletion mutants

Sample	Genotype	Affected?	Deletion	Age
C72	L Δ6	ALS	1989–2006	66
D141	L Δ5	ALS	1965–1989	70
622	L Δ3	Suspected ALS	1989–2030	71
29	L Δ3	ALS	1989–2030	66
627	L Δ3	Unaffected	1989–2030	46
401	L Δ4	ALS	2230–2247	50
809	S Δ4	Unaffected	2230–2247	66
838	S Δ4	Unaffected	2230–2247	48
Figlewicz AC	L Δ1	ALS	1582–1683	72
P150	S Δ6	Unaffected	1989–2006	24
P107	L Δ6	Unaffected	1989–2006	38

Age is the age at onset for those affected. It is striking that in those affected the normal allele is always L and those with the S allele are unaffected, suggesting interactions which affect the penetrance of the deletion mutant. Figlewicz AC is discussed in Figlewicz *et al*. (21).

The C2414A mutation (A805E) reported previously (21,24) was found associated with a silent point mutation C2232T distributed equally in cases and controls and found in three of six sequenced instances of the S allele, but never in 16 sequences of the L allele. These mutations were also found in the Δ5 and Δ6 deletions but not in the Δ3 or Δ4 deletions, providing an insight into the evolution of the mutations. The Δ1 and Δ2 deletions previously reported were not found in this study.

DISCUSSION

Clinical features

We have found four novel NFH deletions, occurring in four of 530 cases of ALS and two of 379 neurologically normal unrelated controls. There was nothing remarkable about the disease phenotype of the cases. The age of onset, gender distribution, sites of onset and duration of disease were all typical of ALS or varied as expected. The Scandinavian FALS pedigree (Fig. 4) with the Δ3 deletion has an individual (LΔ3) with a non-progressive gait disturbance and focal amyotrophy of one limb. He has declined further contact and there is no more information regarding his disease. The other affected individual (LΔ3) for whom data were available had entirely typical ALS. Within this family, an unaffected offspring also has the deletion; however, it is paired with the S allele. This is the first NFH deletion described in a pedigree with FALS, but it cannot be determined whether the mutation segregates with the disease or not. No deletions were found in the eight specimens extracted from brain tissue and the pathological changes associated with NFH deletions remain unknown.

Structural organization of the tail and mechanism of mutation

Three of the four novel deletions are in the same region as the 18 bp variation between the L and S alleles of NFH. There are thus five versions of NFH consisting of insertions or deletions within the region 1965–2030 inclusive, making this a hotspot for deletion events (Fig. 5). This region is particularly repetitive at both the DNA and amino acid levels and it is possible that the mutations arise through unequal crossing over events. All four deletions, the known polymorphism and the longest of the deletions reported in the original paper are of units of 18 or 24 bp, which is the length required to encode one repeat motif. Inspection of the regions deleted reveals a pattern with the format 13+5 bp and 13+11 bp and further analysis shows the whole KSP repeat region of the NFH tail to be composed of this structure. A small region incorporating two deletions and the long/short polymorphism is shown colour coded to reveal this (Fig. 6).

It can be seen that allele Δ5 could easily be derived from the L or S allele, whereas Δ6 could only easily be derived from the S allele. Both these sequences contain the C2232T and C2414A mutations which are only otherwise found in the S NFH allele, suggesting that this was the ancestor for these deletions. These point mutations are not found in the Δ3 or Δ4 deletions.

At the amino acid level the repeat motif consists of X-Lys-Ser-Pro-Y-Lys (where X is unspecified and Y is one to three amino acids) at all except four repeats. (At these four repeats there is a single amino acid variation, although the S/TP motif required for phosphorylation is maintained. Repeat 2 is A<u>A</u>SPEK, the first repeat labelled as 34 is AK<u>T</u>PEK, repeat 38 is AK<u>T</u>PAK and

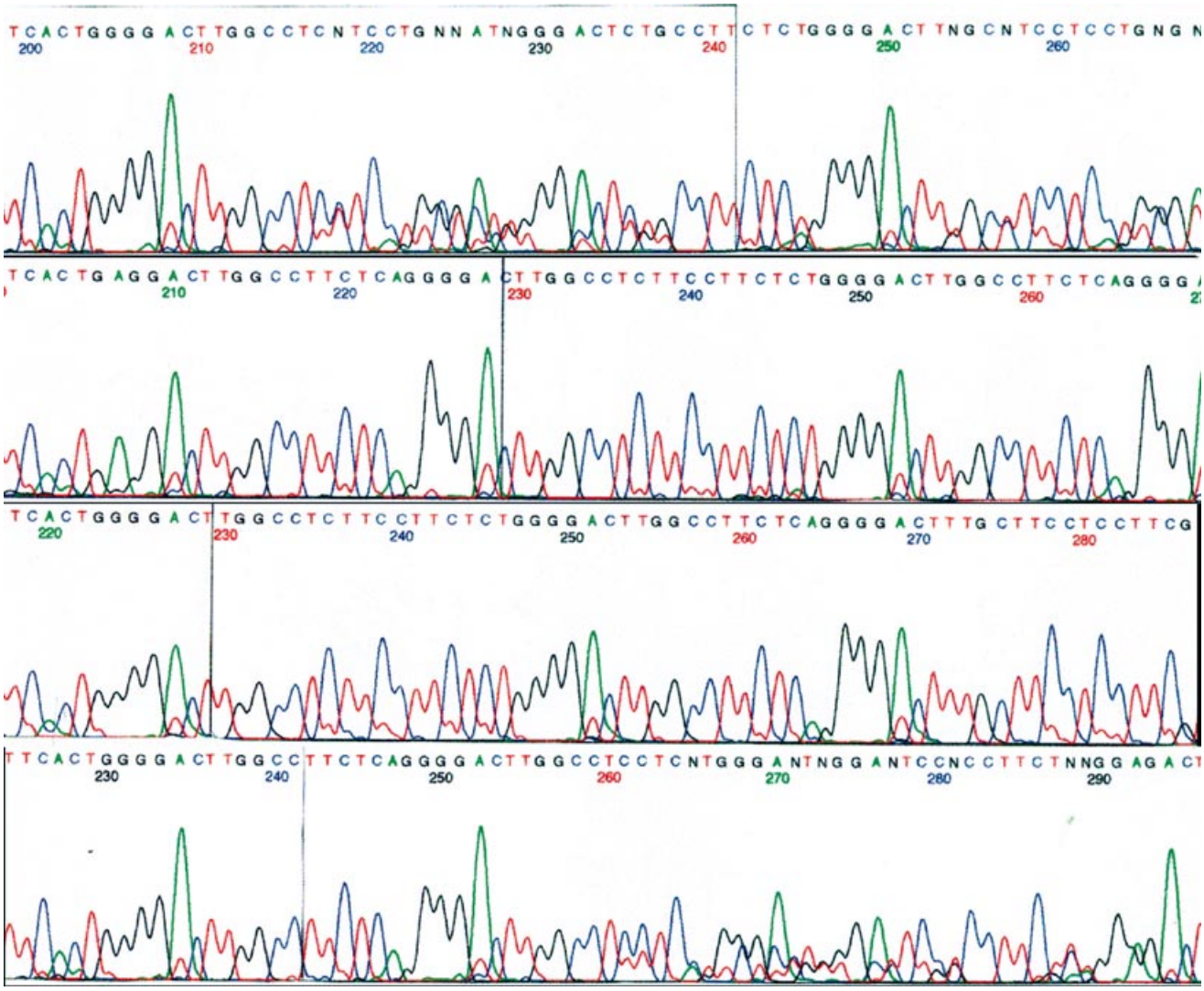

Figure 2. Sequence from patients C72 and D141. Row 1, patient C72 genomic DNA; row 2, subcloned L allele; row 3, subcloned Δ6 allele; row 4, patient D141 genomic DNA. The four sequences have been aligned. Superimposing rows 2 and 3 produces sequence row 1.

repeat 39 is ARSPADK.) Using a motif-based analysis, the NFH tail can be described by a string of letters each representing a single repeat motif and grouped into domains separated by spaces thus: D EVC ECEAC ECEAC EVC VCET EEEE VE VE VE VE VE VE EBAB VELV, where the motif X-Lys-Ser-Pro-X-Lys (XKSPXK) is represented by E, V, L, A or T corresponding to the second X in the sequence in single letter amino acid code; XKSPXXK is designated B, XKSPXXXK is designated C and any other sequence is designated D (Fig. 7). The sequence separating the proline residue from the final lysine in each repeat is critical for recognition by some kinases (25–27) and this analysis therefore has functional relevance. Other than the Δ2 allele, all variants of the NFH tail so far described differ by whole XKSPYK units and therefore by designated letters or domains. The third E in domain 7 represents the 18 bp insertion to form the long polymorphism. Of the two previously reported mutations, the Δ1 allele, which seems to be an otherwise random and large deletion, corresponds exactly to loss of domain 3 (ECEAC). The UK mutations Δ5 and Δ6 correspond to loss of the third and fourth E and second and third E in domain 7, respectively. The Scandinavian mutations are Δ3, loss of domain 8, and Δ4, loss of E from domain 13. The Δ2 deletion does not fit neatly into this classification but is loss of a part of the first V in domain 15.

The tail appears to be in two distinct halves separated by the region which is the deletion hotspot. The first half contains some repeats not seen in the second half and is more complex. The second half shows a 'VE' alternating repeat. The rabbit NFH gene is also polymorphic in length, with either 55 or 49 KSP repeats, and is also in two halves, with a more complex first half, repeats from the first half not seen in the second and a simple 'VE' repeat in the second half (28). This is in contrast to rat NFH which does not have this underlying structure and is not length polymorphic (29,30). This implies a functional organization to the repeats and may be related to cross-linking of NFH tails of different lengths. In rabbits, it is thought that the tail repeats have evolved by unequal crossover events (28) and the same may be true in humans. Indeed, the NFH tail may be predisposed to these events, due to the similarity of part of the DNA sequence to transposon-like

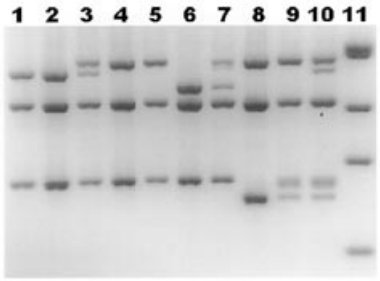

Figure 3. RFLP analysis of Scandinavian cases and subcloned samples (see Materials and Methods). Lane 1, control sample SS genotype; lane 2, plasmid containing subcloned S allele; lane 3, control individual LS genotype; lane 4, plasmid subcloned L allele; lane 5, control individual LL genotype; lane 6, plasmid subcloned Δ3 allele; lane 7, patient 29, genotype LΔ3; lane 8, plasmid subcloned Δ4 allele; lane 9, patient 401, genotype LΔ4; lane 10, individual 809, genotype SΔ4; lane 11, pUC19/*Msp*I standard, molecular sizes 501/489, 404, 331, 242.

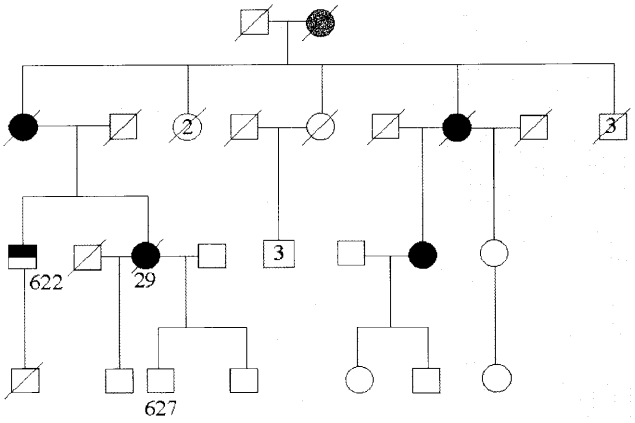

Figure 4. Scandinavian pedigree with the Δ3 allele. Numbering as in Table 1. Diagonal lines indicate those who have died. Squares, males; circles, females; black, ALS; grey, dementing illness with swallowing difficulties; half-filled, monomelic ALS.

```
     1951                                                      1982
 L   AAGGCCAAGT CCCCTGAGAA GGCCAAGTCC CCAGAGAAGG AAGAGGCCAA
 S   AAGGCCAAGT CCCC...... .......... ..AGAGAAGG AAGAGGCCAA
Δ3   AAGGCCAAGT CCCCTGAGAA GGCCAAGTCC CCAGAGAAGG AAGAGGCCAA
Δ5   AAGGCCAAGT CCCC...... .......... .......... ..........
Δ6   AAGGCCAAGT CCCC...... .......... ..AGAGAAGG AAGAGGCCAA

     1983                                                      2032
 L   GTCCCCTGAG AAGGCCAAGT CCCCAGTGAA GGCAGAAGCA AAGTCCCCTG
 S   GTCCCCTGAG AAGGCCAAGT CCCCAGTGAA GGCAGAAGCA AAGTCCCCTG
Δ3   GTCCCC.... .......... .......... .......... ........TG
Δ5   ......TGAG AAGGCCAAGT CCCCAGTGAA GGCAGAAGCA AAGTCCCCTG
Δ6   GTCCCC.... .......... ....AGTGAA GGCAGAAGCA AAGTCCCCTG
```

Figure 5. Five NFH sequences have been aligned. This shows the mutation hotspot region from 1965 to 2030 (numbering of published sequence) where deletions occur most frequently. This region is particularly repetitive and separates the tail domain into two halves (Fig. 6).

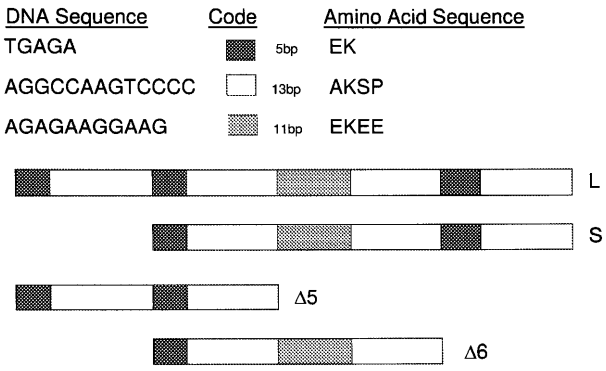

Figure 6. Diagram to show the region where most deletions are found, colour coded to reveal the underlying repeat structure. The diagram can be interpreted at the DNA or amino acid level. Whereas the Δ6 allele could only easily be derived from the S allele, the Δ5 allele could be produced by a single deletion event from either the L or S allele. Both these deletions contain the C2232T polymorphism which is only found in the S allele and this is therefore their ancestor.

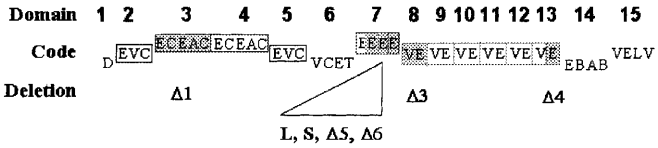

Figure 7. The NFH phosphorylation repeat domain has been encoded according to the amino acid(s) separating the proline from the subsequent lysine at each XKSPYK motif. A, XKSPAK; E, XKSPEK; V, XKSPVK; T, XKSPTK; L, XKSPLK; B, XKSPXXK; C, XKSPXXXK; D, other KSP sequence. Seven of the eight variants of the NFH tail repeats consist of loss of whole motifs compared with the L allele (shaded), the exception being the Δ2 allele. Five of these deletions cluster in a mutation hotspot. The tail consists of two different sections separated by the mutation hotspot. The large previously described deletion of 102 bp (Δ1) can be seen to knock out a region which is repeated.

elements and its minisatellite-like structure. Such sequences are prone to deletions or insertions and there are several neurological diseases in which this is the case, in particular the trinucleotide repeat diseases (31). More than a third of the human genome is composed of interspersed repeat sequences, most of which are copies of transposable elements (32). Interestingly, 36 bp of a long interspersed element (LINE-2) sequence (TCCCCAGTGAAG-GAAGAAGCAAAGTCACCGGCTGAG) is found in the NFH tail from base 1816 and is partially repeated throughout, 443 bp of the rod have a 65% identity with the telomere-associated repeat, TAR1, and 66 bp of human endogenous retrovirus repeat are in the linker region joining rod and tail. The NFH tail may therefore have evolved from amplification of transposable elements and its underlying structure may make it prone to loss or addition of repeat units.

Predisposition to ALS and relevance to previous studies

The failure of two previous studies to replicate the findings of the first may be due in part to selection bias and small sample size. If NFH tail deletions are causative in ALS, they may have a low

penetrance, as shown by unaffected relatives of cases with deletions. Looking at familial ALS will therefore select against NFH deletions because, generally, familial ALS is taken to mean autosomal dominant ALS, records and diagnostic labels may be obscure and many people are unable to give a clear account of their family history beyond their immediate relatives. FALS samples will therefore be unlikely to contain low penetrance mutations, except in inbred populations or those with a founder effect where homozygosity may be an issue. It is striking that affected individuals always carry an L allele, while all S allele carriers are unaffected (Table 1), although there is an age difference between the two groups which might account for this. The alternative is that the normal allele may modify the penetrance of the deletion mutant so that relatives may be unaffected despite carrying mutations if they also carry the S allele.

Pathological mechanisms

The functions of the NFH tail include interaction with cytoskeletal transporters, MAP2 and microtubules, regulation of interfilament spacing and regulation of axonal calibre. All deletions have resulted in loss of a whole XKSPYK motif or multiples of this except for Δ2, which destroys a consensus cdk5 site. XKSPXK is the site of binding to microtubules when not phosphorylated (33) and when phosphorylated the tail forms cross-bridges with microtubules through MAP2A (34). At the electron microscope level, NFH tail domains appear to cross-link neurofilaments (35) and thus mediate interaction with adjacent neurofilaments. It is likely therefore that disruption of tail interaction with other cytoskeletal elements occurs, although given the late onset of ALS, this is probably a mild effect. For all those individuals for whom data is available from this and previous studies, the deleted allele is paired with the L allele in those affected, whereas for the majority of controls, this is not the case. This may be because of a simple steric effect of different length alleles or because there is stronger or weaker cross-linking between neurofilaments. Accumulations of neurofilaments in cell bodies and proximal axons are highly phosphorylated in antibody studies (36) and appear as filamentous whorls, implying that assembly is not disrupted. Phosphorylation and tail extension are intimately related to transport of the filaments down the axon, both increasing as the filament leaves the cell body (37,38). Neurofilaments form part of slow axonal transport (SCa) with tubulins and an unknown motor (39,40). At the nerve terminal they are degraded by calcium-dependent kinases. Slowing of axonal transport due to stronger cross-links or faster axonal transport due to disrupted phosphorylation without an increase in degradation at the synapse could both lead to accumulation of neurofilament. It is also plausible that with the filament assembled and phosphorylated, it is less susceptible to normal lysosomal degradation and the ubiquitin pathway is invoked.

Summary

Motif deletions in the NFH KSP repeat domain are found in patients with typical ALS and penetrance may be determined by the normal allele. We report four novel NFH deletions, including the first instance of an NFH deletion in a familial ALS pedigree. The pattern of the deletions sheds light on the structural organization of the NFH KSP repeat domain and reveals a mutation hotspot. The findings support a primary role for NFH deletions in ALS but confirm that they are not common.

MATERIALS AND METHODS

Clinical and patient resources

There was ethical approval at each centre. Patients with sporadic and familial ALS attending the Umeå University ALS Clinic and the King's MND Clinic were included after fully informed written consent. The diagnosis of ALS was made by a consultant neurologist after full investigation, according to the ICD-10 classification of disease, and patients were categorized according to the El Escorial criteria for ALS (41). Controls were unrelated, neurologically normal and matched as far as possible by age, gender and ethnicity. For those with familial ALS, attempts were made to include unaffected relatives. Whole blood was taken and DNA extracted from blood or brain by standard methods. For samples from the King's College brain bank, DNA was extracted from brain which had been frozen at –80°C.

SOD1 genotyping

Samples were genotyped for SOD1 mutations as described previously (4,5).

NFH PCR amplification

Oligonucleotide primers for PCR amplification were based on those of Figlewicz *et al.* (21) and designed to amplify the same region of the NFH tail containing the KSP repeats, from 1336 to 2509. For sequencing, the same oligonucleotide primers were used. Internal primers were designed, but because of the GC-rich highly repetitive nature of the sequence internally, only one could be used for sequencing, from 2291 on the non-coding strand. Oligonucleotides used were: forward, 5′-GTGGTGGAGAAGTCTGAGAA-3′; reverse, 5′-CTTTGACTTTCACCTCCTGGG-3′; internal, 5′-GGAGACTTCACATCAAGAG-3′. Reactions contained 1 ng/µl DNA, 0.7 µM each oligonucleotide, 40 mM dNTPs and reaction buffer of 50 mM KCl, 10 mM Tris–HCl, pH 8.3 at 25°C, 1.5 M $MgCl_2$, 0.001% (w/v) gelatin and 5 U/µl Taq polymerase. Samples were denatured at 95°C for 5 min, followed by 40 cycles of denaturing at 95°C for 30 s, annealing at 62°C for 30 s and extension at 72°C for 90 s. There was a final extension of 10 min.

Genotyping

All samples were digested as described by Figlewicz *et al.* (21). Single strand conformation polymorphism analysis (SSCP) of products was used for genotyping and mutation detection of all except the Scandinavian cases and relatives. A 50:50 mix of sample and 95% deionized formamide, 10 mM NaOH loading buffer was denatured at 95°C for 3 min and snap cooled on ice. Electrophoresis was in 0.5× MDE gel, 0.6× TBE buffer at 8 W and 25°C for 6 h. Products were visualized after SSCP by silver staining. Of the 955 samples screened, four (all controls from Scandinavia) persistently had non-specific banding, making the gel difficult to interpret. Direct sequencing and attempts at subcloning of these samples were unsuccessful. Southern hybridization by standard methods was used to determine the specific banding of these samples. Probe was prepared by PCR amplification of a previously sequenced sample from a control individual, known to be homozygous for the S NFH allele, and labelled by the random

priming method. Filters were hybridized at high stringency. Two of these samples had band shifts consistent with deletions. Because of the difficulty in sequencing or subcloning these samples, they were electrophoresed in 10% acrylamide, visualized by ethidium bromide staining and directly sequenced from the gel. The Scandinavian cases and relatives were screened by PCR–digestion as above, followed by restriction fragment length polymorphism analysis (RFLP) by electrophoresis in 2% Metaphor agarose (FMC), 1× TBE buffer for 7 h at 130 V. Visualization was by ethidium bromide staining.

Sequencing

For samples with band shifts on SSCP or RFLP, genomic DNA PCR-amplified products were sequenced using fluorescent labelled primers and analysed by ABI377 sequencer (PE Biosystems, Warrington, UK) and associated Genescan software. In addition, samples (except for the two difficult controls described above) were subcloned into JM109 *Escherichia coli* cells with the pGEM-T vector (Promega, Southampton, UK) according to the manufacturer's instructions. Subclones were then sequenced in the same way as genomic samples. Genomic electrophoretograms were analysed by subtracting the known sequence from the two superimposed sequences, leaving the mutant sequence. This was then confirmed on the subcloned sequence.

Searching for repetitive DNA

The deletions found could be interpreted as removal of sequence from several alternative positions as explained above. The oligonucleotide and oligopeptide words each deletion represented were used to encode the whole NFH KSP repeat domain, revealing the underlying repeat structure. In addition, known human repeat elements were searched for using the CENSOR (42) and RepeatMasker programs (A.F.A. Smit and P. Green, unpublished data).

ACKNOWLEDGEMENTS

We are grateful to all patients and relatives who contributed samples to this study. We thank all clinicians who contributed samples and, in particular, Dr Hugh Markus and Dr George Kirov. This work was supported by grants from the Medical Research Council (UK) and the Motor Neurone Disease Association of Great Britain to the King's MND Care and Research Centre; the Council of Västerbotten, Sweden, and the Maj Fants ALS Research Foundation. At the time of this work, A.A.C. was a Medical Research Council Clinical Training Fellow.

NOTE ADDED IN PROOF

An 84 bp insertion has now been reported in a study of the NFH tail in 164 ALS cases and 209 age matched controls [Tomkins, J., Usher, P., Slade, J.Y., Ince, P.G., Curtis, A., Bushby, K. and Shaw, P.J. (1998) Novel insertion in the KSP region of the neurofilament heavy gene in amyotrophic lateral sclerosis (ALS). *Neuroreport*, **9**, 3967–3970]. This insertion occurs at 2124 and results in an extra 4 KSP repeat motifs. By the classification system used here, it is duplication of domains 10 and 11 (2080–2163) with insertion at 2080 (start of domain 10). The normal allele in this case was the S allele, consistent with the hypothesis that length differences or steric effects between NFH tails may be important. As with the deletions discussed, it is whole domains which are involved.

REFERENCES

1. Charcot, J.M. and Joffroy, A. (1869) Deux cas d'atrophie musculaire progressive avec lesions de la substance grise et des faisceaux antero-lateraux de la moelle epiniere. *Arch. Physiol. Neurol. Pathol*., **2**, 744.
2. Nelson, L.M. (1995) Epidemiology of ALS. *Clin. Neurosci*., **3**, 327–331.
3. Cudkowicz, M.E., McKenna-Yasek, D., Sapp, P.E., Chin, W., Geller, B., Hayden, D.L., Schoenfeld, D.A., Hosler, B.A., Horvitz, H.R. and Brown, R.H. (1997) Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis [see comments]. *Ann. Neurol*., **41**, 210–221.
4. Andersen, P.M., Nilsson, P., Keranen, M.L., Forsgren, L., Hagglund, J., Karlsborg, M., Ronnevi, L.O., Gredal, O. and Marklund, S.L. (1997) Phenotypic heterogeneity in motor neuron disease patients with CuZn-superoxide dismutase mutations in Scandinavia. *Brain*, **10**, 1723–1737.
5. Jackson, M., Al-Chalabi, A., Enayat, Z.E., Chioza, B., Leigh, P.N. and Morrison, K.E. (1997) SOD-1 and sporadic ALS: analysis of 155 cases and identification of a novel insertion mutation. *Ann. Neurol*., **42**, 803–806.
6. Jones, C.T., Swingler, R.J., Simpson, S.A. and Brock, D.J. (1995) Superoxide dismutase mutations in an unselected cohort of Scottish amyotrophic lateral sclerosis patients. *J. Med. Genet*., **32**, 290–292.
7. Delisle, M.B. and Carpenter, S. (1984) Neurofibrillary axonal swellings and amyotrophic lateral sclerosis. *J. Neurol. Sci*., **63**, 241–250.
8. Steinert, P.M. (1996) Intermediate filaments in health and disease. *Exp. Mol. Med*., **28**, 55–63.
9. Leigh, P.N., Anderton, B.H., Dodson, A., Gallo, J.M., Swash, M. and Power, D.M. (1988) Ubiquitin deposits in anterior horn cells in motor neurone disease. *Neurosci. Lett*., **93**, 197–203.
10. Lowe, J., Aldridge, F., Lennox, G., Doherty, F., Jefferson, D., Landon, M. and Mayer, R.J. (1989) Inclusion bodies in motor cortex and brainstem of patients with motor neurone disease are detected by immunocytochemical localisation of ubiquitin. *Neurosci. Lett*., **105**, 7–13.
11. Lowe, J., Lennox, G., Jefferson, D., Morrell, K., McQuire, D., Gray, T., Landon, M., Doherty, F.J. and Mayer, R.J. (1988) A filamentous inclusion body within anterior horn neurones in motor neurone disease defined by immunocytochemical localisation of ubiquitin. *Neurosci. Lett*., **94**, 203–210.
12. Mizutani, M., Nunoya, T., Yamasaki, H. and Itakura, C. (1992) The hypotrophic axonopathy mutant in Japanese quail. *J. Hered*., **83**, 234–235.
13. Ohara, O., Gahara, Y., Miyake, T., Teraoka, H. and Kitamura, T. (1993) Neurofilament deficiency in quail caused by nonsense mutation in neurofilament-L gene. *J. Cell Biol*., **121**, 387–395.
14. Julien, J.P., Cote, F. and Collard, J.F. (1995) Mice overexpressing the human neurofilament heavy gene as a model of ALS. *Neurobiol. Aging*, **16**, 487–490.
15. Gurney, M.E., Cutting, F.B., Zhai, P., Doble, A., Taylor, C.P., Andrus, P.K. and Hall, E.D. (1996) Benefit of vitamin E, riluzole and gabapentin in a transgenic model of familial amyotrophic lateral sclerosis. *Ann. Neurol*., **39**, 147–157.
16. Xu, Z., Cork, L.C., Griffin, J.W. and Cleveland, D.W. (1993) Increased expression of neurofilament subunit NF-L produces morphological alterations that resemble the pathology of human motor neuron disease. *Cell*, **73**, 23–33.
17. Tu, P.H., Gurney, M.E., Julien, J.P., Lee, V.M. and Trojanowski, J.Q. (1997) Oxidative stress, mutant SOD1 and neurofilament pathology in transgenic mouse models of human motor neuron disease [review]. *Lab. Invest*., **76**, 441–456.
18. Shaw, C.E., Enayat, Z.E., Powell, J.F., Anderson, V.E., Radunovic, A., al-Sarraj, S. and Leigh, P.N. (1997) Familial amyotrophic lateral sclerosis. Molecular pathology of a patient with a SOD1 mutation. *Neurology*, **49**, 1612–1616.
19. Rouleau, G.A., Clark, A.W., Rooke, K., Pramatarova, A., Krizus, A., Suchowersky, O., Julien, J.P. and Figlewicz, D. (1996) SOD1 mutation is associated with accumulation of neurofilaments in amyotrophic lateral sclerosis. *Ann. Neurol*., **39**, 128–131.
20. Figlewicz, D.A., Rouleau, G.A., Krizus, A. and Julien, J.P. (1993) Polymorphism in the multi-phosphorylation domain of the human neurofilament heavy-subunit-encoding gene. *Gene*, **132**, 297–300.
21. Figlewicz, D.A., Krizus, A., Martinoli, M.G., Meininger, V., Dib, M., Rouleau, G.A. and Julien, J.P. (1994) Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. *Hum. Mol. Genet*., **3**, 1757–1761.

22. Lees, J.F., Shneidman, P.S., Skuntz, S.F., Carden, M.J. and Lazzarini, R.A. (1988) The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it. *EMBO J.*, **7**, 1947–1955.
23. Rooke, K., Figlewicz, D.A., Han, F.Y. and Rouleau, G.A. (1996) Analysis of the KSP repeat of the neurofilament heavy subunit in familial amyotrophic lateral sclerosis. *Neurology*, **46**, 789–790.
24. Vechio, J.D., Bruijn, L.I., Xu, Z., Brown, R.H. Jr and Cleveland, D.W. (1996) Sequence variants in human neurofilament proteins: absence of linkage to familial amyotrophic lateral sclerosis. *Ann. Neurol.*, **40**, 603–610.
25. Sun, D., Leung, C.L. and Liem, R.K.H. (1996) Phosphorylation of the high molecular weight neurofilament protein (NF-H) by Cdk5 and p35. *J. Biol. Chem.*, **271**, 14245–14251.
26. Starr, R., Hall, F.L. and Monteiro, M.J. (1996) A cdc2-like kinase distinct from cdk5 is associated with neurofilaments. *J. Cell Sci.*, **109**, 1565–1573.
27. Jaffe, H., Veeranna, Shetty, K.T. and Pant, H.C. (1998) Characterization of the phosphorylation sites of human high molecular weight neurofilament protein by electrospray ionization tandem mass spectrometry and database searching. *Biochemistry*, **37**, 3931–3940.
28. Soppet, D.R., Beasley, L.L. and Willard, M.B. (1992) Evidence for unequal crossing over in the evolution of the neurofilament polypeptide H. *J. Biol. Chem.*, **267**, 17354–17361.
29. Lieberburg, I., Spinner, N., Snyder, S., Anderson, J., Goldgaber, D., Smulowitz, M., Carroll, Z., Emanuel, B., Breitner, J. and Rubin, L. (1989) Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide (NF-H): developmental and tissue expression in the rat and mapping of its human homologue to chromosomes 1 and 22. *Proc. Natl Acad. Sci. USA*, **86**, 2463–2467.
30. Robinson, P.A., Wion, D. and Anderton, B.H. (1986) Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H). *FEBS Lett.*, **209**, 203–205.
31. Muller, U. and Graeber, M.B. (1996) Neurogenetic diseases: molecular diagnosis and therapeutic approaches. *J. Mol. Med.*, **74**, 71–84.
32. Smit, A.F. (1996) The origin of interspersed repeats in the human genome [review]. *Curr. Opin. Genet. Dev.*, **6**, 743–748.
33. Pant, H.C. and Veeranna (1995) Neurofilament phosphorylation [review]. *Biochem. Cell Biol.*, **73**, 575–592.
34. Hirokawa, N., Hisanaga, S.I. and Shiomura, Y. (1988) MAP2 is a component of crossbridges between microtubules and neurofilaments in the neuronal cytoskeleton: quick-freeze, deep-etch immunoelectron microscopy and reconstitution studies. *J. Neurosci.*, **8**, 2769–2779.
35. Gotow, T., Takeda, M., Tanaka, T. and Hashimoto, P.H. (1992) Macromolecular structure of reassembled neurofilaments as revealed by the quick-freeze deep-etch mica method: difference between NF-M and NF-H subunits in their ability to form cross-bridges. *Eur. J. Cell Biol.*, **58**, 331–345.
36. Leigh, P.N., Dodson, A., Swash, M., Brion, J.P. and Anderton, B.H. (1989) Cytoskeletal abnormalities in motor neuron disease. An immunocytochemical study. *Brain*, **112**, 521–535.
37. Nixon, R.A. and Sihag, R.K. (1991) Neurofilament phosphorylation: a new look at regulation and function. *Trends Neurosci.*, **14**, 501–506.
38. Nixon, R.A. (1993) The regulation of neurofilament protein dynamics by phosphorylation: clues to neurofibrillary pathobiology. *Brain Pathol.*, **3**, 29–38.
39. Nixon, R.A. (1992) Slow axonal transport. *Curr. Opin. Cell Biol.*, **4**, 8–14.
40. Bray, J.J. and Mills, R.G. (1991) Transport complexes associated with slow axonal flow. *Neurochem. Res.*, **16**, 645–649.
41. Brooks, B.R. (1994) El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial 'Clinical limits of amyotrophic lateral sclerosis' workshop contributors. *J. Neurol. Sci.*, **124** (suppl.), 96–107.
42. Jurka, J., Klonowski, P., Dagman, V. and Pelton, P. (1996) CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.*, **20**, 119–121.