Volume 241, number 1,2, 213–218 FEB 06569 December 1988

Partial sequence of the rat heavy neurofilament polypeptide (NF-H)

Identification of putative phosphorylation sites

Kieran C. Breen, Philip A. Robinson*, Didier Wion+ and Brian H. Anderton

Dept of Immunology, St. Georges Hospital Medical School, Cranmer Terrace, London SW17 0RE, England

Received 29 September 1988

A 3 kb cDNA clone has previously been isolated in this laboratory corresponding to the rat heavy neurofilament polypeptide (NF-H). This clone, equivalent to approximately 70% of the total mRNA of the protein has been sequenced and shown to contain the carboxy-terminal region of the message. This contains 51 of the Lys-Ser-Pro repeat triplets which are reported to be the site of neurofilament phosphorylation. The sequence obtained was subsequently compared to those of mouse and human NF-H, showing a homology of approximately 85%. There is, however, one region which is variable between the species, this being the highly phosphorylated region of the protein containing the Lys-Ser-Pro triplet repeat.

Neurofilament; cDNA sequence; Phosphorylation site; Interspecies comparison; (Rat brain)

1. INTRODUCTION

Neurofilaments (NF), the neuronal intermediate filaments, are composed of three polypeptides of apparent molecular mass, determined by SDS-PAGE, approx. 70 kDa (NF-L), 160 kDa (NF-M) and 200 kDa (NF-H) and are involved in the maintenance of neuronal caliber [1]. Phosphorylation probably plays a major role in the functioning of the two larger neurofilament polypeptides (NF-M and NF-H), the levels of phosphorylation being altered developmentally and coincident with a change in neurofilament function [2,3]. Both NF-M and NF-H contain a heavily phosphorylated carboxy region which may influence the formation of filament-filament axonal cross bridges [4,5].

The NF-L and NF-M genes and cDNAs have been cloned and sequenced [6–11] and also those of the mouse and human NF-H [12,13]. More recently, a partial cDNA clone for rat NF-H has been sequenced [14]. We have previously reported the isolation of a partial cDNA clone for rat NF-H [15] which extends over 200 bases further towards the N-terminal and here we describe the sequence of the clone which corresponds to approx. 70% of the total message. The clone contains the triplet repeat Lys-Ser-Pro sequence which is present in the carboxy-terminal region and is repeated 51 times, this being the putative phosphorylation region of the protein. The sequence of the rat triplet repeat region is then compared to those of human and mouse NF-H.

Correspondence address: K.C. Breen, Dept of Immunology, St. Georges Hospital Medical School, Cranmer Terrace, London SW17 0RE, England

* *Present address:* Dept of Dental Surgery, School of Dentistry, University of Leeds, Leeds LS2 9LU, England

\+ *Present address:* INSERM U298, Faculté de Médecine, Rue Haute de Reculée, 49045 Angers, Cedex, France

2. MATERIALS AND METHODS

2.1. *Preparation and screening of a rat brain cDNA library*

The cDNA library was prepared and screened as previously described [15] yielding a clone of 3 kb in length.

2.2. *DNA sequencing*

The clone obtained was inserted into the Bluescript expression vector. Deletion mutants were generated using the Exo III/mung bean nuclease system (Stratagene) and these were sequenced by the dideoxy chain termination method [16].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

```
   1 CAC CAG GAG GAG GTG GGC GAG CTG CTC GGT CAG ATT CAG GGC TGC GGT GCC GCG CAG GCG CAG GCT CAG GCC GAG GCT CGG GAC GCC CTC AAG TGC   96
     His Gln Glu Glu Val Gly Glu Leu Leu Gly Gln Ile Gln Gly Cys Gly Ala Ala Gln Ala Gln Ala Gln Ala Glu Ala Arg Asp Ala Leu Lys Cys

  97 GAC GTG ACG TCG GCG CTG CGG GAG ATC CGC GCG CAG CTC GAA GGA CAC ACG GTG CAG AGT ACG CTG CAG TCA GAG GAG TGG TTC CGA GTG AGA TTG  192
     Asp Val Thr Ser Ala Leu Arg Glu Ile Arg Ala Gln Leu Glu Gly His Thr Val Gln Ser Thr Leu Gln Ser Glu Glu Trp Phe Arg Val Arg Leu

 193 GAC CGA CTC TCA GAG GCA GCC AAA GTG AAC ACG GAT GCT ATG CGC TCT GCC CAA GAG GAG ATA ACT GAG TAC CGG CGG CAG CTG CAG GCC AGG ACC  288
     Asp Arg Leu Ser Glu Ala Ala Lys Val Asn Thr Asp Ala Met Arg Ser Ala Gln Glu Glu Ile Thr Glu Tyr Arg Arg Gln Leu Gln Ala Arg Thr

 289 ACA GAG TTG GAG GCA CTG AAA AGC ACC AAG GAG TCA CTG GAG AGG CAG CGC TCT GAG CTG GAG GAC CGT CAT CAG GTA GAC ATG GCC TCC TAC CAG  384
     Thr Glu Leu Glu Ala Leu Lys Ser Thr Lys Glu Ser Leu Glu Arg Gln Arg Ser Glu Leu Glu Asp Arg His Gln Val Asp Met Ala Ser Tyr Gln

 385 GAT GCA ATT CAG CAG CTG GAC AAT GAG CTG AGA AAC ACC AAA TGG GAG ATG GCC GCG CAG CTC CGA GAG TAC CAG GAC CTG CTC AAC GTC AAG ATG  480
     Asp Ala Ile Gln Gln Leu Asp Asn Glu Leu Arg Asn Thr Lys Trp Glu Met Ala Ala Gln Leu Arg Glu Tyr Gln Asp Leu Leu Asn Val Lys Met

 481 GCC CTG GAT CTT GAG ATC GCT GCT TAC AGA AAA CTC CTG GAA GGC GAA GAG TGT CGG ATT GGC TTT GGA CCC ATT CCC TTC TCT CTT ACT GAG GGA  576
     Ala Leu Asp Leu Glu Ile Ala Ala Tyr Arg Lys Leu Leu Glu Gly Glu Glu Cys Arg Ile Gly Phe Gly Pro Ile Pro Phe Ser Leu Thr Glu Gly

 577 CTC CCA AAA ATT CCC TCC ATG TCC ACT CAC ATA AAA GTC AAA AGC GAA GAG AAG ATA AAA GTA GTA GAA AAA TCG GAG AAG GAA ACC GTC ATT GTA  672
     Leu Pro Lys Ile Pro Ser Met Ser Thr His Ile Lys Val Lys Ser Glu Glu Lys Ile Lys Val Val Glu Lys Ser Glu Lys Glu Thr Val Ile Val

 673 GAG GAA CAG ACA GAA GAG ATC CAG GTG ACA GAA GAA GTG ACA GAA GAG GAG GAC AAA GAG GCC CAA GGG GAG GAA GAA GAA GAG GCA GAA GAG GGA  768
     Glu Glu Gln Thr Glu Glu Ile Gln Val Thr Glu Glu Val Thr Glu Glu Glu Glu Asp Lys Glu Ala Gln Gly Glu Glu Glu Glu Ala Glu Glu Gly

 769 GGA GAA GAA GCA GCA ACT ACG TCT CCC CCT GCA GAA GAG GCT GCA TCT CCA GAA AAG GAA ACC AAG TCT CCT GTG AAA GAA GAG GCC AAG TCC CCA  864
     Gly Glu Glu Ala Ala Thr Thr Ser Pro Pro Ala Glu Glu Ala Ala Ser Pro Glu Lys Glu Thr Lys Ser Pro Val Lys Glu Glu Ala Lys Ser Pro

 865 GCT GAG GCC AAG TCC CCA GCT GAG GCC AAG TCA CCA GCT GAG GCC AAG TCC CCA GCT GAG GTC AAA TCT CCA GCT GTG GCC AAG TCC CCA GCT GAA  960
     Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Val Lys Ser Pro Ala Val Ala Lys Ser Pro Ala Glu

 961 GTC AAA TCT CCA GCT GAG GTC AAA TCT CCA GCT GAG GCC AAG TCA CCA GCT GAG GCC AAG TCA CCA GCT GAG GTC AAC TCT CCA GCT ACA GTG AAG 1056
     Val Lys Ser Pro Ala Glu Val Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Val Asn Ser Pro Ala Thr Val Lys

1057 TCT CCA GGT GAG GCC AAG TCC CCA GCT GAG GCC AAG TCA CCA GCT GAG GTC AAA TCT CCA GTG GAG GCC AAG TCA CCA GCT GAG GCC AAG TCT CCA 1152
     Ser Pro Gly Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Val Lys Ser Pro Val Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro

1153 GCT TCA GTG AAG TCC CCA GGT GAG GCC AAG TCA CCA GCT GAG GCC AAG TCA CCA GCT GAG GTC AAG TCT CCA GCT ACA GTG AAG TCC CCA GTT GAG 1248
     Ala Ser Val Lys Ser Pro Gly Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Val Lys Ser Pro Ala Thr Val Lys Ser Pro Val Glu

1249 GCC AAG TCA CCA GCT GAG GTC AAA TCT CCA GTT ACA GTG AAG TCC CCA GCT GAG GCC AAG TCA CCA GTT GAG GTC AAA TCT CCA GCT TCG GTG AAG 1344
     Ala Lys Ser Pro Ala Glu Val Lys Ser Pro Val Thr Val Lys Ser Pro Ala Glu Ala Lys Ser Pro Val Glu Val Lys Ser Pro Ala Ser Val Lys

1345 TCC CCA AGT GAA GCC AAG TCA CCT GCT GGA GCC AAG TCA CCA GCT GAG GCC AAG TCA CCA GTT GTG GCC AAA TCA CCA GCT GAG GCC AAG TCA CCA 1440
     Ser Pro Ser Glu Ala Lys Ser Pro Ala Gly Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Val Val Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro

1441 GCT GGA GCC AAG CCT CCA GCT GAG GCC AAG TCA CCA GCT GAG GCC AAG TCT CCA GCT GAG GCC AAG TCT CCA GCT GAG GCC AAG TCA CCA GCT GAG 1536
     Ala Gly Ala Lys Pro Pro Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu Ala Lys Ser Pro Ala Glu

1537 GCC AAG TCA CCT GTT GAG GTA AAA TCT CCA GAG AAG GCC AAG AGC CCC GTG AAG GAA GGT GCA AAA TCC CTA GCT GAG GCC AAG TCC CCT GAG AAG 1632
     Ala Lys Ser Pro Val Glu Val Lys Ser Pro Glu Lys Ala Lys Ser Pro Val Lys Glu Gly Ala Lys Ser Leu Ala Glu Ala Lys Ser Pro Glu Lys

1633 GCC AAG TCC CCT GTG AAG GAA GAG ATC AAG CCT CCA GCT GAG GTG AAA TCC CCC GAG AAG GCC AAG AGC CCC ATG AGG AAG GAG GCC AAG TCT CCT 1728
     Ala Lys Ser Pro Val Lys Glu Glu Ile Lys Pro Pro Ala Glu Val Lys Ser Pro Glu Lys Ala Lys Ser Pro Met Arg Lys Glu Ala Lys Ser Pro

1729 GAG AAG GCC AAG ACT CTG GAT GTG AAG TCT CCA GAA GCC AAG CCT CCA GCG AAG GAG GAA GCA AAG CGC CCC GCA GAC ATC AGA TCC CCT GAG CAG 1824
     Glu Lys Ala Lys Thr Leu Asp Val Lys Ser Pro Glu Ala Lys Pro Pro Ala Lys Glu Glu Ala Lys Arg Pro Ala Asp Ile Arg Ser Pro Glu Gln

1825 GTC AAA AGT CCT GCC AAG GAG GAG GCC AAG TCC CCC GAG AAG GAA GAG ACC AGG ACT GAA AAG GTG GCT CCC AAG AAG GAA GAG GTG AAG TCC CCT 1920
     Val Lys Ser Pro Ala Lys Glu Glu Ala Lys Ser Pro Glu Lys Glu Glu Thr Arg Thr Glu Lys Val Ala Pro Lys Lys Glu Glu Val Lys Ser Pro

1921 GTG GAG GAG GTA AAA GCC AAA GAA CCC CCA AAG AAG GTG GAG GAG GAG AAG ACA CCA GCC ACA CCA AAG ACA GAG GTG AAG GAG AGC AAG AAA GAT 2016
     Val Glu Glu Val Lys Ala Lys Glu Pro Pro Lys Lys Val Glu Glu Glu Lys Thr Pro Ala Thr Pro Lys Thr Glu Val Lys Glu Ser Lys Lys Asp

2017 GAA GCT CCC AAG GAG GCC CAG AAG CCC AAG GCG GAG GAG AAG GAG CCT CTC ACA GAA AAG CCC AAG GAC TCT CCG GGG GAA GCC AAG AAG GAA GAG 2112
     Glu Ala Pro Lys Glu Ala Gln Lys Pro Lys Ala Glu Glu Lys Glu Pro Leu Thr Glu Lys Pro Lys Asp Ser Pro Gly Glu Ala Lys Lys Glu Glu

2113 GCT AAA GAG AAG AAG GCG GCG GCC CCA GAG GAG GAG ACG CCC GCC AAG TTG GGC GTG AAG GAA GAG GCT AAA CCC AAA GAG AAG GCA GAA GAC GCC 2208
     Ala Lys Glu Lys Lys Ala Ala Ala Pro Glu Glu Glu Thr Pro Ala Lys Leu Gly Val Lys Glu Glu Ala Lys Pro Lys Glu Lys Ala Glu Asp Ala

2209 AAG GCC AAA GAA CCT AGC AAA CCC TCA GAG AAG GAG AAA CCG AAG AAG GAG GAG GTG CCG GCA GCA CCA GAG AAG AAA GAC ACC AAG GAG GAG AAG 2304
     Lys Ala Lys Glu Pro Ser Lys Pro Ser Glu Lys Glu Lys Pro Lys Lys Glu Glu Val Pro Ala Ala Pro Glu Lys Lys Asp Thr Lys Glu Glu Lys

2305 ACT ACG GAG TCC AAG AAG CGT GAG GAG AAA CCC AAA ATG GAG CCA AGG CCA AGG AGG AGG ACA AGG GCC TTC CCC AAG AGC CTA GCA AAC CCA AGA 2400
     Thr Thr Glu Ser Lys Lys Arg Glu Glu Lys Pro Lys Met Glu Pro Arg Pro Arg Arg Arg Thr Arg Ala Phe Pro Lys Ser Leu Ala Asn Pro Arg

2401 CAG AAA AGG CTG AAA AGT CCT CTA GCA CAG ACC AAA AAG ACA GCC AGC CCT CAG AGA AGG CCC CAG AGG ACA AGG CTG CCA AGG GAG ACA AGT AAG 2496
     Gln Lys Arg Leu Lys Ser Pro Leu Ala Gln Thr Lys Lys Thr Ala Ser Pro Gln Arg Arg Pro Gln Arg Thr Arg Leu Pro Arg Glu Thr Ser Lys

2497 AGG ACG AGA GGG ACA CCC AGA ATA GCC AAA GAA ACT CAG GAC GGC CCC GGT ACT CAA GGG TTG GTG TAA TAA AGT TTA TTT CTT CCT TTC CCT CCG 2592
     Arg Thr Arg Gly Thr Pro Arg Ile Ala Lys Glu Thr Gln Asp Gly Pro Gly Thr Gln Gly Leu Val  *

2593 TAA GAA GAA ACA CCA CTT AGA TGG CGG GCC CGC CCT CAC CAA ACA GGA TTT CTA TTA GGA TTA AGT TAG CAA GAG AAG ATC AAC CCT GAG CCC TGC 2688

2689 CTC CCC AAC ACC AAA GCC CTC CCC AAG GTG ATG GAC AAT TAT GAT AGC TTA TCG TAG CCG AAC GGA GAT GTA TTG CTG AAC GCT CCA CGT AAA ACG 2784

2785 CGT GAC TAA AAA CTG CCC CCC CTC CTT TCC AAG TAA GTG CAT TCA CTT CCC GTA TGT CCT ACC GAC AGG TGA CCG CAG TAA TGA ATG AGC AGT TAG 2880

2881 AAA TGC ATT ATG CTT GAA ATG TTG TAA CCT ATT CCC GAA TGC CTT CTT GTT TTC CAA AGG AGC GGT CAG GCC CTT GCC CGG TAC ACG CTC CTG GAA 2976

2977 GAG CTG CAG CAG GTG AGG CAG GGC GCT GGC CGC TGA ACC AGG CCA GGG TGT GCT GTC CAC TGA AGG CCA CTT TCG ATT GCT TCC GTG CAA TAA AAC 3072

3073 CCA AGT GCT TCT GA(41) 3086
```

Fig.1. Nucleotide sequence of the 3 kb clone of rat NF-H, and the corresponding deduced amino acid sequence. The untranslated region is 529 bases long and includes the poly(A)$^+$ tail.

3. RESULTS AND DISCUSSION

The 3 kb clone obtained begins at the end of α-helix coil 1B and continues through the carboxy-terminal (stop codon) to the poly(A) tail (fig.1). The amino acid sequence of coil 2 was found to be very highly conserved with respect to both rat NF-L and NF-M and also mouse and human NF-H, demonstrating a homology of approx. 80% and 90%, respectively [6,11–13]. This is in good agreement with previous results which have shown a high degree of conservation between the coiled regions of intermediate filaments from a number of different species [4–6]. There is also a good homology between the three species in the last 200 amino acids or so at the carboxy-terminal of the NF-H protein.

In between these two regions, there is a section containing a series of 51 repeated amino acid triplets: Lys-Ser-Pro. These occur largely in a sequence of 6 amino acids, in which the Lys-Ser-Pro is highly conserved in positions 3–5 of the sextet (fig.2). The amino acid in position 2, preceding the triplet, is also highly conserved being either Ala or Val. There is, however, a greater divergence in the first and sixth amino acids of the sextet, but this difference is unlikely to influence greatly the secondary structure of the region, this probably being determined by the triplet repeat sequence. The Lys-Ser-Pro sequence has previously been identified as the phosphorylation site within NF-H [18] and has also been recognized in other neurofilament proteins as well as in neurofilament and microtubule-associated proteins including

	RAT	MOUSE	HUMAN
Glu-Ala-Lys-Ser-Pro-Ala	21	15	6
Glu-Val-Lys-Ser-Pro-Ala	4	1	-
Glu-Ala-Lys-Ser-Pro-Val	3	2	2
Glu-Val-Lys-Ser-Pro-Val	3	1	1
Val-Ala-Lys-Ser-Pro-Ala	2	-	-
Glu-Ala-Lys-Ser-Pro-Ala	2	-	-
Glu-Val-Lys-Ser-Pro-Ser	2	-	-
Glu-Val-Lys-Ser-Pro-Glu	2	-	4
Glu-Ala-Lys-Ser-Pro-Glu	2	-	10
Glu-Thr-Lys-Ser-Pro-Val	1	1	-
Thr-Val-Lys-Ser-Pro-Gly	1	1	-
Ser-Val-Lys-Ser-Pro-Val	1	-	-
Thr-Val-Lys-Ser-Pro-Val	1	-	-
Thr-Val-Lys-Ser-Pro-Ala	1	-	-
Gly-Ala-Lys-Ser-Pro-Ala	1	1	-
Gly-Ala-Lys-Ser-Pro-Glu	1	1	-
Lys-Ala-Lys-Ser-Pro-Met	1	-	-
Asp-Val-Lys-Ser-Pro-Glu	1	1	-
Gln-Val-Lys-Ser-Pro-Ala	1	-	-
Ala-Val-Lys-Ser-Pro-Gly	-	1	-
Glu-Ala-Lys-Ser-Pro-Ile	-	1	-
Gln-Val-Lys-Ser-Pro-Glu	-	1	-
Glu-Ala-Lys-Ser-Prp-Gly	-	7	-
Glu-Ala-Lys-Ser-Pro-Ser	-	1	-
Glu-Pro-Lys-Ser-Pro-Ala	-	3	-
Glu-Thr-Lys-Ser-Pro-Pro	-	-	1
Lys-Ala-Lys-Ser-Pro-Ala	-	-	2
Glu-Ala-Lys-Ser-Pro-Pro	-	-	1
Lys-Ala-Lys-Ser-Pro-Val	-	-	8
Lys-Ala-Lys-Ser-Pro-Thr	-	-	1
Lys-Ala-Lys-Ser-Pro-Glu	-	1	3
Lys-Ala-Lys-Ser-Pro-Leu	-	-	1
Glu-Val-Lys-Ser-Pro-Gly	-	3	-
Ala-Val-Lys-Ser-Pro-Ala	-	1	-
TOTAL	51	43	41

Fig.2. Sequence of the amino acid sextets containing the Lys-Ser-Pro triplet repeat of NF-H of rat, mouse [12] and human [13].

MAP-2 and Tau, indicating that it too, is highly conserved [17].

A small number of triplet repeats are also present in NF-M [9–11,18], but are absent from NF-L [6–8]. Previous studies have shown that the number of alkaline phosphatase-sensitive phosphate groups on NF-H and NF-M are approximately proportional to the number of triplet repeats, these phosphate groups being absent from NF-L [19]. This conforms with the proposal that the triplet is indeed the site of neurofilament phosphorylation.

There is however, a difference between both the total number of triplet repeats present and the intervening amino acids in the heavy neurofilament protein from rat, mouse and human. The mouse and human NF-H messages are fairly similar with 43 and 40 repeat sequences, respectively, whereas the rat NF-H has 51 repeats (fig.3). A similar interspecies difference has also previously been

```
R   - - - - - - T S P P A E E A A S P E K E T K S P V K E E A K S P A E A K S P
H   G G E E E T K S P P A E E A A S P E K E A K S P V K E E A K S P A E A K S P
M   E L A A A - T S P P A E E A A S P E K E T K S P V K E E A K S P G E A K S P

R   A - - E A K S P A E A K S P A E V K S P A - - V A K S P A E V K S P A - - E
H   E K E E A K S P A E V K S P E K A K S P A K E E A K S P P E A K S P E K E E
M   G - - E A K S P A E A K S P G E A K S P G - - E A K S P G E A K S P A - - E

R   V K S P A E A K S P A E A K S P A - - E A K S P A T V K S P G E A K S P A -
H   A K S P A E V K S P E K A K S P A K E E A K S P A E A K S P E K A K S P V K
M   P K S P A E P K S P A E A K S P A - - E P K S P A T V K S P G E A K S P S -

R   - E A K S P A E V K S P V - - E A K S P A E A K S P A S V K S P V - - E A K
H   E E A K S P A E A K S P V K E E A K S P A E V K S P E K A K S P T K E E A K
M   - E A K S P A E A K S P A - - E A K S P A E A K S P A E V K S P G - - E A K

R   S P A E A K S P A - - E V K S P S T V K S P V - - E A K S P A E V K S P V -
H   S P E K A K S P E K E E A K S P E K A K S P V K A E A K S P E K A K S P V K
M   S P A E P K S P A - - E A K S P A E V K S P A - - E A K S P A E V K S P G -

R   - T V K S P A E A K S P V - - E V K S P A S V K S P S - - E A K S P A G A K
H   A E A K S P E K A K S P V K E E A K S P E K A K S P V K E E A K S P E K A K
M   - E A K S P A A V K S P A - - E A K S P A A V K S P G - - E A K S P G E A K

R   S P A - - E A K S P V V A K S P A - - E A K S P A E A K S P A E A K S P A E
H   S P V K E E A K T P E K A K S P V K E E A K S P E K A K S P E K A K T L D V
M   S P A - - E A K S P A E A K S P I - - E V K S P G G A K T P V - - - - - - -

R   A K S P A E A K S P A E A K S P A E A K S P V E V K S P E K A K S P V K E G
H   K - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
M   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - E E

R   - A K S P E K A K S P V K E E I K P P A E V K S P E K A - - - - - - - K S P
H   - - K S P E - A K T P A K E E A - - - - - - - R S P A D K F P E K A - - K S P
M   G A K S P A G A K S P - - E E A - - - - - - - K S P V E E D I P P A E A K S P

R   M R K E A K S P E K A K T L D V - - - - - - - - K S P E A K P P A K E E A K
H   V K E E V K S P E K A - - - - - - - - - - - - - K S P L K A D A K A P E K E
M   - G E A - K S P V K E G A K P P E K A K P L D V K S P E A Q T P V Q E E A T

R   A P A D I R S P E Q V K S P A K E E A K S P E K E E T R - T E K V A P K K E
H   I P K - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - K E
M   Y P T D I R P P E Q V K S P A K E K A K S P E K E E A K T S E K V A P K K E

R   E V K S P V
H   E V K S P V
M   E V K S P V
```

Fig.3. Comparison of the amino acid sequences of the repeat region of rat NF-H, human NF-H [13], and mouse NF-H [12]. Deletions (dashes) allow for better alignment of the sequences.

noted in the corresponding region of NF-M [4,9–11].

In the rat NF-H cDNA sequence, 41 of the triplet Lys-Ser-Pro repeats occur in a regular pattern of six amino acids (fig.3). In the mouse, also, this regular pattern was seen with 33 of the 43 triplet repeats present. In the human, however, it can be noted that the spacing of the triplets is slightly altered with them being separated alternately by three and five amino acids. There is also a considerable difference between the amino acid composition of the repeat region between the three species. While Ala and Val are conserved in position 2 of the sextet in all three species, there is a considerable difference in the amino acid composition of positions 1 and 6 (fig.2). The rat NF-H favours Glu in position 1 and a nonpolar amino acid (Ala or Val) in position 6. In the mouse NF-H, however, there is a much greater variation in the amino acids present at positions 1 and 6, with no clear pattern emerging. In the human NF-H sequence, while there is a broad range of amino acids in position 6, two main amino acids, Glu and Lys, dominate position 1. As previously mentioned, the triplet repeats in the human NF-H are separated alternatively by 5 and 3 amino acids. One can notice that when one of the extra two amino acids present in a 5 amino acid bridge is a lysine residue and the subsequent amino acid at position 1 of the sextet is usually a Glu residue. Alternatively, when there are only three residues between the repeats, position 1 is always occupied by a Lys residue. From this, one can conclude that the presence of a basic amino acid between the triplet repeats may be important in human NF-H. The reason for the differences seen between species, considering the high degree of homology seen in the rest of the proteins, is unknown. However, given that the region in question is that of phosphorylation, it is possible that the interspecies differences between the amino acid sequences separating the triplet repeats may reflect different binding requirements for the associated protein kinase enzymes, which may exhibit a high degree of species specificity.

Neurofilament phosphorylation has been shown to be related to the neuronal calibre [19] and is very specifically developmentally regulated [2,20]. In the perikarya and developing axons, NF-H is in a relatively unphosphorylated form. In mature axons, however, NF-H is in a highly phosphorylated form, and it is thought that phosphorylation of the polypeptide results in the formation of interfilament cross bridges that are important in the maintenance of axonal calibre [1,4,5]. Therefore the triplet repeats in the carboxy region are likely to play an important role in the functioning of neurofilaments. The acidic carboxy tailpiece of neurofilaments has been implicated in the formation of interneurofilament cross bridges [4,5] and the proximity of the potential phosphorylation site to this region would have a major influence on protein structure and function.

In conclusion, therefore, while there are some minor differences in the repeat region of the protein in the three species examined, these are unlikely to greatly affect the functioning of this region of the protein. Overall, therefore, the amino acid sequence of the NF-H protein is highly conserved between the mouse, rat and human species.

Acknowledgements: This work was supported by the Medical Research Council. We acknowledge the technical assistance of Ms Suzanne Payne and Mr Stephen Field.

REFERENCES

[1] Hoffman, P.N., Cleveland, D.W., Griffin, J.W., Landen, P.W., Cowan, N.J. and Price, D.L. (1987) Proc. Natl. Acad. Sci. USA 84, 3472–3476.
[2] Foster, G.A., Dahl, D. and Lee, V.M.Y. (1987) J. Neurosci. 7, 2651–2663.
[3] Oblinger, N.M. (1987) J. Neurosci. 7, 2510–2521.
[4] Geisler, N., Fischer, S., Vanderkerchove, J., Plessman, U. and Weber, K. (1984) EMBO J. 3, 2701–2706.
[5] Geisler, N., Fischer, S., Vanderkerchove, J., Van Damme, J., Plessmann, U. and Weber, K. (1985) EMBO J. 4, 57–63.
[6] Julien, J.P., Ramachandran, K. and Grosveld, F. (1985) Biochim. Biophys. Acta 825, 398–404.
[7] Lewis, S.A. and Cowan, N.J. (1986) Mol. Cell Biol. 6, 1529–1534.
[8] Julien, J.P., Grosveld, F., Yazdanbaksh, K., Flavell, D., Meijer, D. and Mushynski, W. (1987) Biochim. Biophys. Acta 909, 10–20.
[9] Levy, E., Liem, R.K.H., D'Eustachio, P. and Cowan, N.J. (1987) Eur. J. Biochem. 166, 71–77.
[10] Myers, M.W., Lazzarini, R.A., Lee, V.M.Y., Schlaepfer, W.W. and Nelson, D.L. (1987) EMBO J. 6, 1617–1626.
[11] Napolitano, E.W., Chin, S.S.M., Colman, D.R. and Liem, R.K.H. (1987) J. Neurosci. 7, 2590–2599.
[12] Julien, J.P., Cote, F., Beaudet, L., Sidky, M., Flavell, D., Grosveld, F. and Mushynski, W. (1988) submitted.
[13] Lee, J.F., Shneidman, P.S., Skuntz, S.F., Carden, M.J. and Lazzarini, R.A. (1988) EMBO J. 7, 1947–1955.

[14] Dautigny, A., Pham-Edinh, D., Roussel, C., Felix, J.M., Nussbaum, J.L. and Jolles, P. (1988) Biochem. Biophys. Res. Commun. 154, 1099–1106.
[15] Robinson, P.A., Wion, D. and Anderton, B.H. (1986) FEBS Lett. 209, 203–205.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Lee, V.M.Y., Otvos, L., Carden, M.J., Hollosi, M., Dietzschold, B. and Lazzarini, R.A. (1988) Proc. Natl. Acad. Sci. USA 85, 1998–2002.
[18] Geisler, N., Vanderkerchove, J. and Weber, K. (1987) FEBS Lett. 221, 403–407.
[19] Carden, M.J., Schlaepfer, W.W. and Lee, V.M.Y. (1985) J. Biol. Chem. 260, 9805–9817.
[20] Carden, M.J., Trojanowski, J.Q., Schlaepfer, W.W. and Lee, V.M.Y. (1987) J. Neurosci. 7, 3489–3504.