Volume 221, number 2, 403–407 FEB 05101 September 1987

Location and sequence characterization of the major phosphorylation sites of the high molecular mass neurofilament proteins M and H

Norbert Geisler, Joel Vandekerckhove* and Klaus Weber

Max Planck Institute for Biophysical Chemistry, 3400 Göttingen, FRG and **Rijksuniversiteit Gent, Laboratory of Genetics, B-9000 Gent, Belgium*

Received 20 July 1987

Diagonal fingerprinting allows the specific purification of those tryptic peptides which change electrophoretic mobility due to a dephosphorylation step introduced after the first dimension. Nine tryptic peptides from the tail domain of porcine neurofilament M protein identify a minimum of 6 phosphorylated serines. Unexpectedly, four of the nine peptides characterize a region of degenerate repetitive sequences. Results on neurofilament H tail, although less complete, yield longer sequences of degenerate repetitive character. Here, all serines present appear to be contained in a lysine-serine-proline unit. This motif also occurs in some but not all M peptides. We suggest that degenerate repetitive sequences in neurofilament M and H tails have a high species-specific drift.

Axon; Cytoskeleton; Intermediate filament; Neurofilament; Protein kinase; Serine phosphate

1. INTRODUCTION

The three neurofilament (NF) proteins, L, M and H, are major constituents of the neuronal cytoskeleton. The two higher molecular mass proteins are strongly phosphorylated [1,2]. Recently phosphate values around 8 and 13 were obtained for porcine M and H preparations [3]. The serine phosphates have been located to the carboxy-terminal tails [4–6]. These protrude as autonomous domains from the filament wall which arises from the interaction of the α-helical rod domains [4–8]. Given their unique amino acid compositions dictated by a wealth of lysine and particularly glutamic acid [4], the tail domains were not covered in previous protein sequence studies of porcine M and H [4,5,7,8]. NF phosphorylation poses several interesting problems. For instance, monoclonal antibodies distinguishing phosphorylated and unphosphorylated M and H are known [9,11,12] and the pattern of phosphorylation seems to differ in cell bodies and axons (e.g. see [9,10]). As a first step towards understanding NF phosphorylation we have concentrated on the phosphorylated regions of M and H present in standard neurofilament preparations.

Correspondence address: N. Geisler, Max Planck Institute for Biophysical Chemistry, 3400 Göttingen, FRG

2. MATERIALS AND METHODS

NF polypeptides L, M and H were isolated from porcine spinal cord and the tail domains of H and M were excised by limited chymotryptic cleavage [7]. NF tails were extensively digested with trypsin (1–2%, w/v) in 50 mM NH_4HCO_3 at 37°C for 12 h and lyophilized. The diagonal fingerprinting was modelled on an earlier procedure [13] to which several modifications were added. Tryptic peptides dissolved in a small volume of pH 6.5 buffer (10% pyridine, 0.5% acetic acid) were subjected to high-voltage electrophoresis on a strip of paper. After

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

air drying the strip was sprayed with a solution of *E. coli* alkaline phosphatase (0.2 mg/ml, Sigma, type III) in 0.2 M NH_4HCO_3 adjusted to pH 10 with dilute ammonia. $MgCl_2$ and $ZnSO_4$ were present at 1 mM. After 1 h at 20°C in a damp dessicator the strip was air dried and sewn onto a paper sheet. Electrophoresis was repeated at right angles to the first run using as before 1000 V per 50 cm. Aspartic acid and lysine added as markers separated by about 45 cm. The diagonal fingerprints were lightly sprayed with fluorescamine (0.0005% in acetone). Peptides located off the diagonal (fig.1) were eluted and characterized by composition and sequence. In the case of spot M4 peptide material was subjected to HPLC on a C18 column. Molar yields after diagonal fingerprinting are expressed in % based on the amount of M and H tail domains processed.

3. RESULTS AND DISCUSSION

The tryptic digests of M and H tails were subjected to electrophoresis at pH 6.5. The paper was sprayed with alkaline phosphatase and electrophoresis was repeated at right angles. The few spots separated from the diagonal on which the bulk of the peptides are found showed a relative decrease in negative charge after dephosphorylation (fig.1a,c). Control experiments in which alkaline phosphatase was omitted from the solvent used in the dephosphorylation step yielded perfect diagonal patterns without any extra spots (fig.1b,d). As a positive control a thermolytic and a chymotryptic digest of ovalbumin were subjected to the diagonal fingerprinting procedure. After alkaline phosphatase treatment two peptides off the diagonal were detected. Their sequences agreed with the two phosphorylation sites of ovalbumin [13] (not shown).

Fig.2 summarizes the sequences of the peptides purified by diagonal electrophoresis and therefore originally carrying at least one phosphate. It also gives their molar yield. Material from spot M4 provided 4 peptides after HPLC. As indicated in fig.2 peptides M2a, M2b and M2 are related as M2a and M2b cover the amino-terminal 7 residues of the longer M2 peptide. Similarly, peptides M4a–M4d describe a 39-residue-long sequence given in fig.3. Whether the repeat pattern extends over a longer sequence is not yet known. Since M1 and M4a have

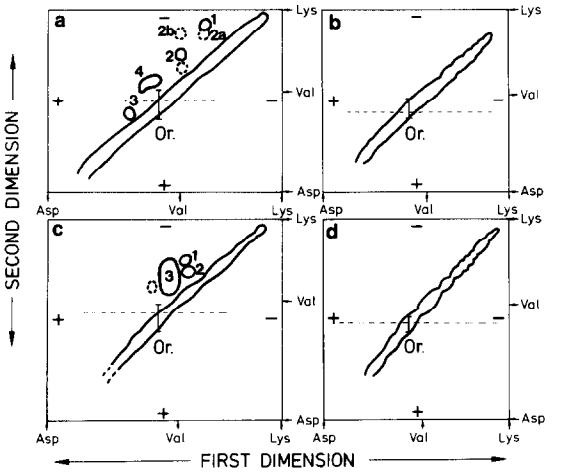

Fig.1. Diagonal fingerprinting of tryptic phosphopeptides from tail domains of M (a,b) and H (c,d) with (a,c) or without (b,d) alkaline phosphatase present in the incubation step after the first electrophoresis. Note a limited number of spots located off the diagonal in a,c and their absence in b,d. Dotted spots indicate low-yield peptides or as yet unidentified mixtures obtained in low yield. For the sequence of the major peptides located off the diagonal see fig.2 and text. (Bar Or.) Origin where digests were applied. Asp, Lys and Val are free amino acids used as markers in both dimensions.

only a single serine residue the position of the phosphates in the NF-M tail in this region can be immediately assigned. As M2b was more acidic in the first dimension than M2a it originally carried two phosphates while M2a had only one. The precise number and location of the phosphates in peptides M2a, M2, M3, M4b, M4c and M4d are not yet known. Nevertheless, the combined results identify a minimum of six phosphorylation sites (M1, M2b, M3, M4a, M4b) but the actual number could be higher (see fig.2). This result compares favourably with a total of 8 serine phosphates reported for porcine NF-M [3].

HPLC provided two fractions of M tail. The major peak (80%) had the amino-terminal sequence VEEIIEETKVEDE. The same sequence preceded by a phenylalanine (position 452 in human NF-M) is also present in the NF-M sequences predicted from DNA studies in man [14], rat [15] and chicken [16]. Thus, the major M tail domain starts about 40 residues past the previously documented end of the α-helical rod domain [4,7].

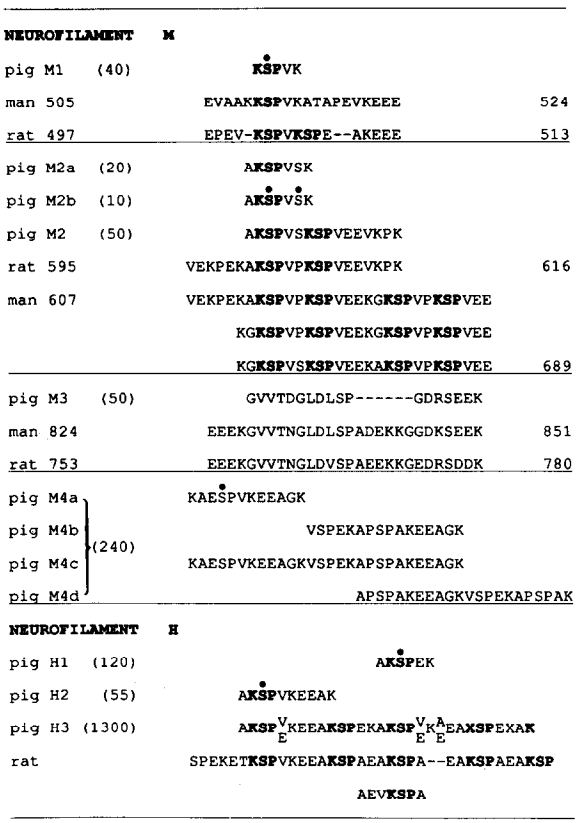

Fig.2. Amino acid sequences of tryptic peptides purified by diagonal electrophoresis. Numbers correspond to fig.1. Peptides M4a–d are from HPLC of spot M4. Although all peptides must have contained phosphoserine residues after the first dimension, a filled circle above a serine is only used for those peptides where the particular phosphorylated serine is clearly identified (see text). Sequences M1–M3 are tentatively arranged along the corresponding regions of the human and rat proteins [14,15]. Numbers on the right reflect position number of the last residue from these regions. For M4 peptides see text. H peptides are arranged along a short sequence predicted for rat H [17]. X, as yet unestablished residue. Dashes allow for better alignment. H3 material shows microheterogeneity at a few positions. Residues set lower were present at 10%. The KSP motif when present is given in bold-face letters. Molar yields obtained from the fingerprint are given on the left by numbers in parentheses. Values above 100% point to repetitive sequences. The yields of M4 and H3 peptides were calculated based on the 16- and 14-residue repeats outlined in fig.3.

The minor peak (20%) contained two sequences. The more abundant one (15%) coincided with the amino-terminal sequence of CNBr7 described earlier [4]. It is due to cleavage at the methionine situated 32 residues past the major phenylalanine cleavage site. The less abundant species (5%) started 4 residues prior to the methionine. From the isolation of a tryptic peptide corresponding to residues 861–872 in the human sequence [14] we assume that the M tail isolated after chymotryptic cleavage lacks the carboxy-terminal 43 residues of porcine NF-M.

When we compared the human, rat and chicken sequences [14–16] we noticed a striking feature. While the amino and carboxyl regions of the different tails are highly homologous the middle regions display a strong species-specific drift (alignment not shown). Here, for instance, the human protein has 6 nearly perfect repeats of 13 residues [14] while the rat protein has only 1 segment [15] (see fig.2). The chicken protein [16] is even more remote. It has at least two distinct repeat patterns. One involves 8 degenerate repeats of 12 residues which are quite distinct from those found in man. This species-specific drift places strong restrictions on the alignment of the porcine M peptides with the complete tail sequences from other species. As shown in fig.2 peptides M1 and M3 are readily placed but already in the case of the related peptides M2a, M2b and M2 it is not clear whether they arise from a repeat pattern as in human, or only from a single segment as in rat. Although peptide M4 fits in its amino-terminal 10 residues positions 733–742 of the human protein it differs distinctly in the remaining part. As M4 shows a degenerate repeat pattern (described in fig.3) it could also arise from the middle region of the M tail where at least chicken and man show multiple repeats. In spite of these ambiguities the combined results show that the phosphorylated regions are spread over a large part of the M tail domain. While some of them display KSP-type sequences (M1, M2) others clearly do not (M3, M4).

The results on phosphorylation sites in the porcine H tail are less complete (fig.2). Peptides H1 and H2 contain a single serine residue which was originally phosphorylated. The high yield of H1 is indicative of a repeat pattern, which is very clearly seen in H3. It was sequenced over 30 residues. The microheterogeneity observed at a few positions

```
M 4:          * *
              KAESPVK          7

              EEAGKVSPEKAPSPAK 23

              EEAGKVSPEKAPSPAK 39

H 3:          *
              AKSPVKEEAKSPEK   14

              *
              AKSPVKAEAKSPEX   28

              AK               30
```

Fig.3. Repeat patterns in phosphopeptides M4 and H3. The 39 residues predicted from the M4 peptides show two completely identical 16-residue repeats between positions 8 and 23, and 24 and 39, respectively. Each longer repeat reveals two short degenerate repeats (underlined) and a similar degenerate repeat occurs in residues 1–7 (underlined). The main sequence of H3 shows two nearly identical 14-residue repeats. Again each repeat shows two shorter degenerate repeats (underlined). For the total number of repeats see text. Asterisks indicate amino acid exchanges which perturb the repeat pattern.

(fig.2) indicates the presence of two long peptides which share a degenerate repetitive sequence theme. This seems to be based on 14 residues containing two repeats of AKSP followed either by E or V (see also fig.3). H3 was recovered at about 13-fold molar yield relative to this 14-residue repeat. Thus, there should be either one very long sequence of degenerate repeats or several shorter ones. Although H1 and H2 sequences can be excised from H3 (fig.2) we do not know whether these shorter peptides arose from H3-type regions or from other regions of the H tail. The only extended H tail sequence currently known is that of 72 amino acids predicted from a partial characterization of a rat cDNA clone [17]. This sequence reveals in its carboxy-terminal part 6 degenerate repeats (fig.2) which are clearly related but not identical to porcine H3 (fig.2). At least 13 serine phosphates have been reported for porcine NF-H [3]. Our results indicate that they may all involve KSP-type sequences as all peptides isolated by the diagonal procedure show their serines in this motif.

A search of the protein sequence bank shows that repetitive degenerate repeats containing KSP sequences are rare. They are found, however, in the very basic regions of several histone H1 molecules including rabbit histone H1.3 [18]. This probably explains why certain monoclonal antibodies to phosphorylated M and H detect epitopes shared with chromatin proteins including histone H1 [19]. Thus, the histone sequence AKSPKK, for instance, is very similar to the AKSPEK sequence found in the H peptides and to the KSPVK sequence present in M1 (see fig.2).

In conclusion, we have identified sequences of NF proteins M and H which are strongly phosphorylated. Chemically synthesized peptides, as well as the dephosphorylated tail domains, can now be used to search for those protein kinases which are involved in NF phosphorylation. Future work is required in order to decide whether there are NF-specific kinases and whether these could give rise to progressive phosphorylation during axonal transport. Our results increase the evidence for degenerate repetitive sequence regions in some NF proteins. Although their functional importance is not understood, they seem to show a high species-specific drift and account for at least some of the major phosphorylation sites.

ACKNOWLEDGEMENTS

We thank Drs R. Lazzarini, R. Liem and H. Betz for providing sequence data prior to publication and A. Gruber and J. Van Damme for expert technical assistance. This work was supported in part by the Deutsche Forschungsgemeinschaft (We 338/4-3).

REFERENCES

[1] Julien, J.-P. and Mushynski, W.E. (1982) J. Biol. Chem. 257, 10467–10470.
[2] Jones, S.M. and Williams, R.C. jr (1982) J. Biol. Chem. 257, 9902–9905.
[3] Georges, E., Lefebvre, S. and Mushinsky, W.E. (1986) J. Neurochem. 47, 477–483.
[4] Geisler, N., Fischer, S., Vandekerckhove, J., Plessmann, U. and Weber, K. (1984) EMBO J. 3, 2701–2706.
[5] Geisler, N., Fischer, S., Vandekerckhove, J., Van Damme, J., Plessmann, U. and Weber, K. (1985) EMBO J. 4, 57–63.

[6] Julien, J.-P. and Mushynski, W.E. (1983) J. Biol. Chem. 258, 4019–4025.
[7] Geisler, N., Kaufmann, E., Fischer, S., Plessmann, U. and Weber, K. (1983) EMBO J. 2, 1295–1302.
[8] Geisler, N., Plessmann, U. and Weber, K. (1985) FEBS Lett. 182, 475–478.
[9] Sternberger, L.A. and Sternberger, N.H. (1983) Proc. Natl. Acad. Sci. USA 80, 6126–6130.
[10] Bennett, G.S. and DiLullo, C. (1985) J. Cell Biol. 100, 1799–1804.
[11] Carden, M.J., Schlaepfer, W.W. and Lee, V.M.-Y. (1985) J. Biol. Chem. 260, 9805–9817.
[12] Dahl, D. (1983) Exp. Cell Res. 149, 397–408.
[13] Milstein, C.P. (1968) Biochem. J. 110, 127–134.
[14] Myers, M.W., Lazzarini, R.A., Lee, V.M.Y., Schlaepfer, W.W. and Nelson, D.L. (1987) EMBO J. 6, 1617–1626.
[15] Napolitano, E.W., Chin, S.S.M., Colman, D.R. and Liem, R.K.H. (1987) J. Neurosci., in press.
[16] Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E.D. and Betz, H. (1987) Genes Dev., in press.
[17] Robinson, P.A., Wion, D. and Anderton, B.H. (1986) FEBS Lett. 209, 203–205.
[18] Strickland, W.N., Strickland, M., Brandt, W.F., Von Holt, C., Lehman, A. and Wittmann-Liebold, B. (1980) Eur. J. Biochem. 104, 567–578.
[19] Wood, J.N., Lathangue, N.B., McLachlan, D.R., Smith, B.J., Anderton, B.H. and Dowding, A.J. (1985) J. Neurochem. 44, 149–154.