Identification of Phosphorylation Sites on Neurofilament Proteins by Nanoelectrospray Mass Spectrometry*

(Received for publication, November 7, 1996, and in revised form, February 27, 1997)

Joanna C. Betts‡§, Walter P. Blackstock‡, Malcolm A. Ward‡, and Brian H. Anderton‡

From the ‡Department of Neuroscience, The Institute of Psychiatry, De Crespigny Park, London, SE5 8AF and §Glaxo Wellcome Research and Development, Stevenage SG1 2NY, United Kingdom

Neurofilament (NF) proteins are intermediate filaments found in the neuronal cytoskeleton. Phosphorylation of these proteins is considered an important factor in the assembly of filaments and determination of filament caliber and stability. Mammalian neurofilaments are composed of three polypeptide subunits, NF-L, NF-M, and NF-H, all of which are phosphorylated. Here we used techniques for the mass spectrometric sequencing of proteins from polyacrylamide gels to analyze in vivo phosphorylation sites on NF-M and NF-L. Neurofilaments were isolated from rat brain and enzymatically digested in gel. The resulting peptides were analyzed and sequence data obtained by nanoelectrospray mass spectrometry. Four phosphorylation sites have been found in the C-terminal domain of NF-M: serines 603, 608, 666, and 766. Two of these are found in lysine-serine-proline (KSP) motifs and two in the variant motifs, glutamic acid-serine-proline (ESP) and valine-serine-proline (VSP). Serine 55 in NF-L was not found to be phosphorylated, which confirms the possible role of phosphorylation and dephosphorylation of this site in early neurofilament assembly. The techniques used enable sequence data and characterization of posttranslational modifications to be obtained for each individual subunit directly from polyacrylamide gels.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Biomolecular Structure Units, 2F 116, Glaxo Wellcome Research and Development, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK. Tel.: 44 (0) 1438 763331; Fax: 44 (0) 1438 763330; E-mail: wxxs5951@grg.co.uk.

§ The abbreviations used are: NFP, neurofilament protein; HPLC, high performance liquid chromatography; PAGE, polyacrylamide gel electrophoresis; IMAC, immobilized metal affinity chromatography; nanoES, nanoelectrospray; MS, mass spectrometry; MS/MS, tandem mass spectrometry; CID, collision-induced dissociation; pS, Ser(P).
peptides from in vitro 32P-labeled cytoskeletal preparations with those from metabolically labeled dorsal root ganglia. Peptides from unlabeled NF-M eluting at the same retention time as the labeled phosphopeptides were then sequenced by Edman degradation, and the phosphorylation sites were determined. Elhanany et al. (38) used a combination of mass spectrometry and mass spectrometry to identify nine endogenous phospho-rylated KSP sites in the C-terminal region of rat NF-H.

Recently, a technique for the sequencing of proteins directly from polyacrylamide gels using electrospray (39) in combination with tandem mass spectrometry (40) has been developed (41). This allows sequence data and characterization of post-translational modifications, using parent ion scans (42, 43), to be obtained from total digest mixtures using, exclusively, mass spectrometric techniques. Proteins are able to be uniquely identified with the use of peptide tags and data base searching (44).

We have used this technology in conjunction with small scale immobilized metal affinity chromatography (IMAC) (45, 46) to identify phosphopeptides in digest mixtures of NFPs and then to characterize the sites of phosphorylation. Having covered more than 80% of the sequence of NF-L, no endogenous phosphorylation sites were found, suggesting dephosphorylation of the head terminal region during filament assembly. Each endogenous phosphorylation sites have been found within the C-terminal domain of NF-M, including one that has not been previously reported.

**EXPERIMENTAL PROCEDURES**

**Materials—**Except where otherwise noted, all chemicals used were purchased from Sigma and were of the highest quality available. Hydroxylapatite (HAP) was from Bio-Rad Laboratories Ltd. (Hertfordshire, England). Acetic acid and formic acid (AnalaR grade) were from BDH Laboratory Supplies (Merck Ltd., Leicester, UK). The proteolytic enzymes trypsin (bovine, sequencing grade), endoproteinase Asp-N (from Pseudomonas fragi, sequencing grade), and endoproteinase Glu-C (from Staphylococcus aureus strain V8, sequencing grade) were obtained from Boehringer Mannheim UK Ltd. (East Sussex, UK). For mass spectrometric analysis and gel spot preparation, HPLC grade methanol and acetonitrile (Rathburn Chemicals, Scotland) were used.

**Purification of NF Polypeptides—**A Triton-X-insoluble pellet was prepared from about 15 g of rat brain by homogenization in phosphate-buffered saline, pH 7.2, containing KCl (600 mM), EDTA (2 mM), EGTA (2 mM), phenylmethanesulfonyl fluoride (1 mM), tetrasodium pyrophosphate (10 mM), sodium fluoride (20 mM), sodium orthovanadate (100 mM), and 0.1% (v/v) Triton X-100, at a ratio of 1:20 (w/v). After centrifugation (Sorval SS34 rotor, 15,000 rpm, 20 min, 4°C), the pellet was washed by resuspension in the same buffer and then in the absence of Triton X-100. The final pellet was dissolved in buffer A (10 mM sodium phosphate, pH 7.4, made up in 8% urea containing 1 mM dithiothreitol), and batchwise HTP fractionation was performed as described (47). The eluate was dialyzed against buffer B (4 mM urea, 20 mM Tris, pH 7.5, 1 mM dithiothreitol), centrifuged (Beckman type 65 rotor, 50,000 g, 1 h, 4°C), and loaded on to a Mono Q HR 5/5 column (Pharmacia Biotech Europe, Sweden). The sample was concentrated by eluting with a gradient of 0–100% 1 M NaCl in buffer B developed over 10 ml at a flow rate of 0.5 ml/min.

**In-gel Digestion—**Approximately 50 μg of NFPs were separated by one-dimensional SDS-polyacrylamide gel electrophoresis using precast 8% acrylamide Novex (San Diego, CA) mini-gels (48). Gels were stained with 0.2% (w/v) Coomassie Brilliant Blue R250 in 50% (v/v) methanol in water containing 2% (w/v) acetic acid and destained in 50% (v/v) methanol, 2% (v/v) acetic acid. Protein bands were cut from the gel, reduced, acetylamidated, and tryptically digested as described (41). Peptides were extracted twice with 100 mM NH4HCO3 and acetonitrile followed by two extractions with 5% (v/v) acetic acid and acetonitrile and then dried. Additional peptides in which phosphopeptides were also subjected to enzymatic digestion using endoproteinase Glu-C (50 mM NH4HCO3, pH 7.8, 25°C) and endoproteinase Asp-N (50 mM sodium phosphate buffer, pH 8.0, 37°C) by the same method.

**Immobilized Metal Affinity Chromatography—**Empty miniature Protein Chemistry Systems (PCS) desalting columns (Hewlett-Packard, Cheshire, UK) were packed with chelating Sepharose high performance slurry (70 μl in 20% (v/v) ethanol, Pharmacia) and washed with water (2 ml) followed by 0.1 M acetic acid, pH 3.1 (solution A, 500 μl). 0.1 M FeCl3 solution (50 μl in solution A) was added followed by washing with solution A (500 μl) to remove excess iron. The dried peptide mixture in the gel digest was then loaded (dissolved in 50 μl of solution A), and the column was washed with buffer A (500 μl). The phosphopeptides were eluted with 0.1 M Tris-HCl (300 μl), pH 8.5, and the eluate dried before being analyzed by nanoelectrospray mass spectrometry.

**Nanoelectrospray (nanoES) Mass Spectrometry (MS)—**Needles for nanoelectrospray mass spectrometry were made with a micropipette puller (Sutter Instrument Co., Novato, CA) from borosilicate glass capillaries (Clark Electromedical Instruments, Pangbourne, Reading, UK) as described by Wilm and Mann (49). They were gold-coated in a vapor desorption instrument. Dried protein digests were dissolved in 5% (v/v) formic acid and desalted on a miniature PCS column self-packed with (∼20 μl of POROS R2 sorbent (PerSeptive Biosystems, Framingham, MA) as described (41). The sample was not eluted directly into the spraying needle but was dried and then taken up in 10 μl of spraying solution (50% (v/v) methanol, 1% (v/v) formic acid in water for positive ion or 50% (v/v) methanol, 5% (v/v) ammonia in water for negative ion), and 1 μl was inserted into the needle. Electrospray mass spectra were acquired on an API III triple quadrupole machine (Perkin-Elmer Sciex, Ontario, Canada) equipped with a nanoES ion source developed by Wilm and Mann (49, 50). Q1 scans were performed with 0.1% (v/v) acetic acid step. For operation in the MS/MS mode, Q1 was set to transmit a mass window of 2 Da for both parent and product ion scans, and spectra were accumulated with 0.2-Da mass steps. Dwell time was 1 ms for all scans except for parent ion scans, where it was 3 ms. Resolution was set so that fragment masses could be assigned to more than 1 Da. Collision energy was individually tuned for each peptide for optimum MS/MS spectra. A new needle was used for each experiment. Spectra interpretation was performed using BioMultiView (Sciex) software.

**RESULTS**

**Analysis of Tryptic Digests—**NFps purified from rat brain were run on SDS-polyacrylamide gel electrophoresis, and NF-M and NF-L were tryptically digested in gel. After desalting on POROS, analysis of the total digest mixtures of NF-M and NF-L was performed by nanoES mass spectrometry. Fragmentation of peptide ions by collision-induced dissociation (CID) tandem MS resulted in partial sequences being obtained that, in combination with the mass, were sufficient to characterize unambiguously the peptides from the known protein sequences (1, 2, 44).

Phosphopeptides were isolated from the digest mixture using a small scale IMAC technique, which takes advantage of the affinity of phosphopeptides for immobilized Fe3+ ions (45). Following small scale desalting, the samples were analyzed by nanoES MS/MS. Phosphopeptides were identified within the digest mixtures using scans for the parents of m/z 79 in negative ion mode (42, 43) and, for neutral loss of 49, [M–H–PO4]2H2O, in positive ion mode (53). The parent ion scan shows ions that fragment to produce an ion at m/z 79 (the phospho group, PO32−). The neutral loss scan shows the masses of precursor ions that lose the phosphate group as a neutral fragment.

The Q1 spectra of IMAC-purified samples mainly showed ions corresponding to phosphopeptides in comparison with the total digest spectra (Fig. 1, A and B). A few non-phosphopeptides containing histidine residues were also seen, since histidine is thought to have some affinity to the packing used. The use of the IMAC column reduced the problem of electrospray ion suppression by reduction of the complexity of the peptide mixture. This reduction increased the phosphorylated peptide ion signal strength and thus improved the quality of the microsequencing spectra.

**Identification of Serines 603, 608, 666, and 766 as Phosphorylated Sites on NF-M**—Four phosphorylation sites, which must have been generated in vitro, must have been generated in vivo, were found within the C-terminal region of NF-M. Three of these, two in KSP motifs and one in the variant ESP motif, have been reported previously by...
This shows the masses of ions that fragment to produce an ion at m/z 79 (PO₃⁻). This is the expected site for phosphorylation on this peptide since trypsin has cleaved at lysine 602, which it may not have done had serine 603 been phosphorylated.

Fragmentation of the 2+ ion of P4 (m/z 798.2) gave a full series of Y⁺ ions (Fig. 2B), thereby characterizing the peptide as residues 757–771 and confirming that serine 766 is phosphorylated (Mᵣ 2026.2). This is the expected site for phosphorylation on this peptide because trypsin has cleaved at lysine 602, which it may not have done had serine 603 been phosphorylated.

Fragmentation of the 2+ ion of P6 gave a series of b ions (sequence VK) that enabled characterization of the peptide as residues 601–620 with both serines phosphorylated (AK⁶⁰³pSPVPK⁶⁰⁸pSPVEEVKPKPEAK, Mᵣ 2305.3). Fragments corresponding to the loss of ⁶⁰⁸pS were also seen as singly charged b ions.

MS/MS of the 3+ ion (m/z 742.1) of P2 resulted in doubly charged Y⁺ ions, giving the partial sequence SPVP from doubly charged Y ions (Fig. 2A), which, from the known sequence, is sufficient to characterize the peptide and sites of phosphorylation as residues 601–620 with both serines phosphorylated (AK⁶⁰³pSPVPK⁶⁰⁸pSPVEEVKPKPEAK, Mᵣ 2305.3). Fragments corresponding to the loss of ⁶⁰⁸pS were also seen as singly charged b ions.

Fragmentation of the 3+ ion of P1 (m/z 769.6) in positive ion mode gave the partial sequence ⁶⁰⁸pSPVP from doubly charged Y ions (Fig. 2A), which, from the known sequence, is sufficient to characterize the peptide and sites of phosphorylation as residues 601–620 with both serines phosphorylated (AK⁶⁰³pSPVPK⁶⁰⁸pSPVEEVKPKPEAK, Mᵣ 2305.3). Fragments corresponding to the loss of ⁶⁰⁸pS were also seen as singly charged b ions.

The product ion spectrum of the 3+ ion of P3 (m/z 676.0) displayed doubly charged Y⁺ ions corresponding to the sequence VPK and confirming the identity of the peptide as residues 603–620 with serine 608 phosphorylated (Mᵣ 2225.4). This indicates that serine 603 is heterogeneously phosphorylated within the NF-M molecule.

The signal-to-noise ratio of the CID spectrum of P5 was not sufficient to produce any useful sequence information.

CID MS/MS of the ion at m/z 546.4 corresponding to the 2+ of P6 gave a series of mass fragments characteristic of phosphorylated peptides (residues 663–671, Mᵣ 1095.1) with serine 666 carrying a phosphate. Analysis of Endoproteinase Glu-C and Asp-N Digests—NF-M and NF-L were digested in gel with endoproteinases Glu-C and Asp-N. Desalted samples were analyzed by nanoES MS/MS, and the peptides were identified, thus giving greater sequence coverage for each protein. Total sequence coverage gained for both proteins was 81% and for NF-M, 64% (Fig. 3).

Endoproteinase Glu-C-digested samples were subjected to IMAC and then analyzed by nanoES MS/MS. Scans for fragment masses characteristic of phosphorylated peptides were performed. The negative ion parents of m/z 79 scan (not shown) revealed two phosphopeptides in the NF-M sample corresponding to phosphorylation at serines 603 and 608 (P8 and P9, mode. This shows the masses of precursor ions that lose the phosphate group as a neutral fragment, [M–HPO₄⁻ + 2H]²⁺.
The signal-to-noise ratio of the CID spectra of these peptides was not sufficient to produce any useful sequence information. Fragmentation in positive ion mode of the ion at \( m/z \) 738.4 seen in the neutral loss scan (not shown) gave sequence data corresponding to residues 600–612, with serine 608 phosphorylated (P7, Table I).

Analysis of NF-L—81% of the sequence of NF-L was covered by mass spectrometric sequencing; however, no phosphopeptides were found within the NF-L molecule. IMAC of the digest mixtures also failed to reveal any phosphopeptides. Upon phosphorylation with cAMP-dependent protein kinase, phosphorylated residues were identified within the NF-L molecule, suggesting that the absence of endogenous sites was not due to a technical problem. These findings seem consistent with the notion that NF-L is mainly phosphorylated at its head domain and that these phosphates are removed before assembly into filaments. The glutamic acid-rich tail region of NF-L containing 74 residues, including serine 473, which has been found to be phosphorylated in vivo (36), could not be located despite the use of several different proteases. This may perhaps be due to incomplete digestion of this region or suppression of ionization of the peptides in the total digest mixture by these residues.

**DISCUSSION**

Phosphorylation of neurofilament proteins is considered an important factor in the assembly of filaments, determination of filament caliber, and stability and plays a potential role in the pathology of several neurodegenerative diseases. Until now, conventional approaches have been employed in the analysis of phosphorylation of NFPs. These methods have involved metabolic labeling, either in vivo or of cultured neurons, and two-dimensional phosphopeptide mapping or lengthy chromatographic separations after various enzymatic digestions. Phosphopeptides have then been sequenced and sites located

---

2 J. C. Betts, K. E. Cleverley, W. P. Blackstock, and B. H. Anderton, manuscript in preparation.
Phosphopeptides identified in the enzymatic digests of NF-M

Phosphopeptides found within NF-M digests are shown. Peptides fragmented have residues found as a continuous Y or b ion series underlined. The identified phosphorylated residues are boldface.

| Phosphopeptide | Residue no. | Peptide sequence |
|----------------|-------------|-----------------|
| P1             | 601–620     | (K) AKPSVFKpSPVEVVKPEAK (A) |
| P2             | 601–620     | (K) AKPSVFKpSPVEEVKPEAK (A) |
| P3             | 603–620     | (K) SPVFPSVVEEVKPEAK (A) |
| P4             | 707–771     | (K) GVSVEKAKSPVEEVVKPEAK (K) |
| P5             | 749–771     | (K) GSGGEVKVNGVDVpPAEKEK (K) |
| P6             | 663–671     | (K) KAPePpVEAK (K) |
| P7             | 600–612     | (E) KAPSVFKpSPVEEVK (E) |
| P8             | 603–612     | (E) IKVERKApKSPVFpVEEV (E) or (E) IKVERSEKpKSPVFpVEEV (E) |
| P9             | 597–618     | (E) KAPKSVFKpSPVEEVKPE (E) or (E) KAPKSVFKpSPVEEVK (E) |

by the use of conventional protein sequencing. These methods nearly always require the purification to homogeneity of often limited quantities of the neurofilament subunits and separation of the digest mixtures and are limited by the sensitivity of Edman degradation.

Here we use techniques developed by Mann and co-workers (41) for the mass spectrometric sequencing of proteins directly from polyacrylamide gels, hence elevating the necessity for subunit separation and enabling sequencing of peptides from a total digest mixture at lower sensitivity levels. The nature of the fragments generated by tandem MS of peptides allows unique characterization of the peptide, both from an unknown protein with the use of peptide tags and database searching (44) and, as in this case, from a known protein sequence.

Phosphopeptides were identified within the digest mixtures by the use of parent ion (42, 43) and neutral loss (53) scans. Once distinguished from the peptide mixture, the phosphopeptides could be fragmented during the same experiment, yielding sequence data and hence uniquely characterizing the peptide and site of phosphorylation. The extremely low flow rate of the nanospray technique (49) requires the use of just 1 μl of sample/30 min of analysis and therefore enables the acquisition of a large quantity of sequence data with low sample consumption.

A small scale IMAC technique was developed using mini-desalting columns, supplied empty (Hewlett Packard), which were self-packed with a small amount of chelating Sepharose high performance slurry (Pharmacia). This enabled small scale separation of phosphopeptides from the digest mixture due to their affinity for Fe⁺³ ions (45), resulting in improved sensitivity for the phosphopeptide ions and augmenting the quality of the sequence data obtained. This can be explained due to the fact that in a mixture of peptides, the more strongly ionizing components can often suppress the signal from those that are more weakly ionizing. Therefore, by reducing the complexity of the mixture by IMAC, the likelihood of suppression is also reduced (46).

Using these methods, we have characterized four phosphorylation sites within the C-terminal domain of NF-M and covered 64% of the sequence by mass spectrometric sequencing (Fig. 3). Three of the sites have been reported previously; two in KSP motifs and one in the variant motif, ESP. In addition, we have characterized a novel site found within a VSP motif that may also be phosphorylated by separate kinase-directed kinase.

We have also been able to gain some insight in the heterogeneity of phosphorylation of some of these residues. Serine 603 has been found in both its phosphorylated and non-phosphorylated forms, as has serine 766. There is evidence to suggest that the phosphorylation of NFPs is heterogeneous (54, 55) and that the state of phosphorylation changes as the proteins are transported down the axon (56). Our results are in agreement with these findings and identify particular sites displaying this heterogeneity, although we cannot unambiguously rule out loss of phosphate during neurofilament manipulation.

We have covered 81% of the NF-L sequence (Fig. 3) by mass spectrometric sequencing but found no sites of phosphorylation. The fact that no sites were found in the head region of NF-L seems to confirm the hypothesis that these sites undergo phosphorylation and dephosphorylation during filament assembly (12, 30, 31), since we started with assembled neurofilaments for the preparation. When we phosphorylated NF-L with cAMP-dependent protein kinase, which is known to phosphorylate the head domain of NF-L (30, 57), we were able to identify sites of phosphorylation by mass spectrometric sequencing,² which suggests that the lack of endogenous sites found is a true result rather than a practical limitation.

Some regions of the sequence of both proteins, namely those that are rich in glutamic acid residues, remain uncovered. This may be caused either by incomplete digestion or by poor ionization of these peptides due to the presence of these residues.

In addition to phosphorylation, neurofilaments have been shown to be posttranslationally modified by O-linked N-acetylglucosamine (51, 52). The parent ion scan approach may also be used in the identification of glycopeptides in a digest mixture (42) as N-hexosamines give a characteristic oxonium ion at m/z 204. We have used this scan for parents of m/z 204 in the analysis of NF-M and NF-L but found no evidence of glycopeptides. This may be due to the relatively low amounts of these species within neurofilaments (51, 52) compared with those modified by phosphorylation (11).

The sensitivity of this technique, coupled with the ease and speed of sample preparation and analysis, should enable NFPs and other proteins to be analyzed for posttranslational modifications from a variety of sources. These may include cells stimulated to activate signal transduction cascades and tissue from diseases such as motor neuron disease.

REFERENCES
1. Chin, S. S. M., and Liem, R. K. H. (1989) Eur. J. Cell Biol. 50, 475–490
2. Napolitano, E. W., Chin, S. S. M., Colman, D. R., and Liem, R. K. H. (1987)
