Molecular Weight Determination and Identification of Endogenous Phosphorylation Sites of Bovine Neurofilament Protein Subunits

Sarah Trimpin[1], April E. Mixon[1], Moo-Young Kim[1], Martha Stapels[1], Peter S. Spencer[2], Max L. Deinzer[1]

[1] Department of Chemistry, Oregon State University, Corvallis, OR 97331, USA and [2] CROET, Oregon Health & Science University, Portland, OR 97201, USA

Overview

Purpose:
- Molecular weight determination
- Identification of phosphorylation sites

Method:
- HPLC-MALDI-TOF-MS *off line*
- Peptide mapping

Result:
- Molecular weight determination of all three bovine NFP subunits
- Identification of phosphorylation sites in bovine NFL and bovine NFM

Introduction

Neurofilament proteins (**NFP**s) are major cytoskeletal components of neurons and are composed of three different polypeptide subunits: **NFL** (low molecular weight), **NFM** (intermediate molecular weight), and **NFH** (high molecular weight). Each polypeptide subunit is organized into three domains: a central relatively conserved segmented α-helical rod domain of approximate 310 amino acids separates a variable N-terminal head domain from a hypervariable C-terminal domain. NFPs are reported to be **highly phosphorylated** *in vivo* [1]. Since it is likely that modulation of the level of phosphorylation at these specific sites will affect the structural and functional interaction of filament assembly and interfilament spacing, it is important to identify endogenous phosphorylation sites unambiguously. Various approaches have been used to evaluate the extent and sites of NFP phosphorylation *in vivo* as for example ^{32}P labeling.

Little is known about **bovine NFP** except that it is a protein mixture with relatively **high molecular weight subunits**. Only the full sequence information is accessible for bovine NFL. In case of bovine NFM only a partial sequence containing the C-terminus of the protein is available. Currently, there is no information available on the *in vivo* phosphorylation sites present in bovine NFPs. The lack of this important information made bovine NFP the perfect model protein mixture for evaluating a **mass spectrometric approach** towards **molecular weight determination** of individual subunits, ***de novo* sequencing** of the N-terminus of NFM and **identification of their posttranslational phosphorylation sites** of NFL and NFM *in vivo*. Peptide mapping investigations, including MS and MS/MS experiments, were obtained utilizing Q-TOF and MALDI-TOF/TOF mass spectrometers. These fundamental investigations give the opportunity to establish a foundation for further studies.

Methods

- **Molecular weight:**
  - **Separation**: HPLC-fractionation *off line*
  - **MS**: costumer build MALDI-TOF mass spectrometer
- **Peptide mapping:**
  - **Proteases**: trypsin, Lys C, pepsin
  - **Environment**: in-gel, in-solution
  - **Incubation conditions**: time-dependent
  - **Extraction conditions**: methylation/IMAC-separation
  - **MS**: LC-MS/MS (Quadrupole Time-of-Flight mass spectrometer, Q-TOF, Global MS from Micromass, Manchester, UK; MALDI-MS/MS (Applied Biosystems ABI4700 TOF/TOF mass spectrometer, Applied Biosystems, Inc., Framingham, MA)
  - data interpretation: Mascot

Results

- **HPLC-MALDI-TOF-MS *off line*** determined the molecular weights of NFP: **NFL** (63 kDa), **NFM** (105 kDa), **NFH** (125 kDa)
- **Peptide Mapping** determined the following aspects:
  - Coverage: **NFL** (99%), **NFM** (94%)
  - *De novo sequencing* of the N-terminus of bovine NFM: 115 amino acids (98% confidence; side chain fragmentation suggest two Ile residues at positions 26 and 104)
  - phosphorylation: **NFL** (three heterogeneous sites), **NFM** (sixteen homogenous, six heterogeneous, and two unconfirmed sites)
  - Further modifications in both NFL and NFM: Deamidation, oxidation, pyro-glu, N-terminal acetylation
  - Methylation/IMAC-approach: Improved quality in the mass spectra but not in sequence coverage because it was difficult to yield quantitative esterfication and thereby increased complexity of data interpretation
  - Protease: Trypsin is the most valuable with respect to highest sequence and phosphopeptide coverage and showed the best reproducibility

Overview of bovine NFL

	trypsin	Lys C	pepsin	overall
total number of identified aa [out of 554 possible]	468	329	376	546
total coverage [in %]	**84.5**	**59.4**	**67.9**	**98.6**
unspecific cleavage #/number of peptides	n.o.	n.o.	35/48	35/48
total number of modification sites*2	26	7	2	26
#/554 [in %]	4.7	1.3	0.4	4.7
#/individual coverage [in %]	5.6	2.1	0.5	4.8
total number of phosphorylation sites	**3**	**1**	n.o.	**3**
Ser [number of aa]	**3**	**1**	n.o.	**3**
confident	--	--	--	3
tentative	--	--	--	n.o.
homogenous	--	--	--	n.o.
heterogenous	--	--	--	3
Thr [number of aa]	**n.o.**	**n.o.**	**n.o.**	**n.o.**

Overview of bovine NFM

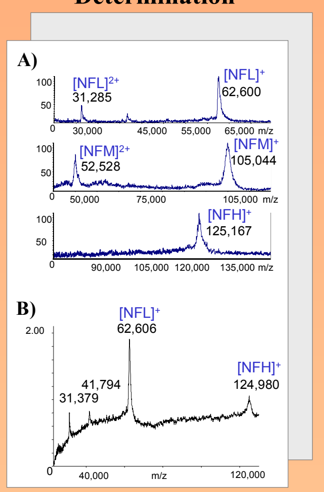
	trypsin	Lys C	pepsin	overall
total number of identified aa [out of 810 possible]	646	492	410	753
total coverage for partial NFM sequence [in %]	79.3	60.7	50.6	93
total number of identified aa [out of 925 possible]	720	523	491	868
total coverage for total NFM sequence [in %]	**77.8**	**56.5**	**53.1**	**93.8**
unspecific cleavage #/number of peptides	--	--	62/71	62/71
total number of modification sites in total NFM sequence*2	45	16	4	51
#/925 [in %]	4.9	1.7	0.4	5.5
#/individual coverage incl N-term [in %]	6.3	3.1	0.8	5.9
total number of phosphorylation sites	**21**	**7**	**1**	**24**
confident [+ tentative]	--	--	--	**22 [+2]**
Ser [number of aa]	**16**	**7**	**1**	**20**
confident	--	--	--	19
tentative	--	--	--	"0.5"
homogenous [+ tentative]	--	--	--	13 [+"0.5"]
heterogenous [+ tentative]	--	--	--	6
Thr [number of aa]	**5**	**n.o.**	**n.o.**	**5**
confident	--	--	--	2
tentative	--	--	--	"2.5"
homogenous	--	--	--	3 [+"0.5"]
heterogenous	--	--	--	0 [+1]

Conclusion

The determined molecular weights indicate one phosphate group in bovine NFL, in "24.5" NFM, and in 108 NFH (extrapolating from mouse NFH)

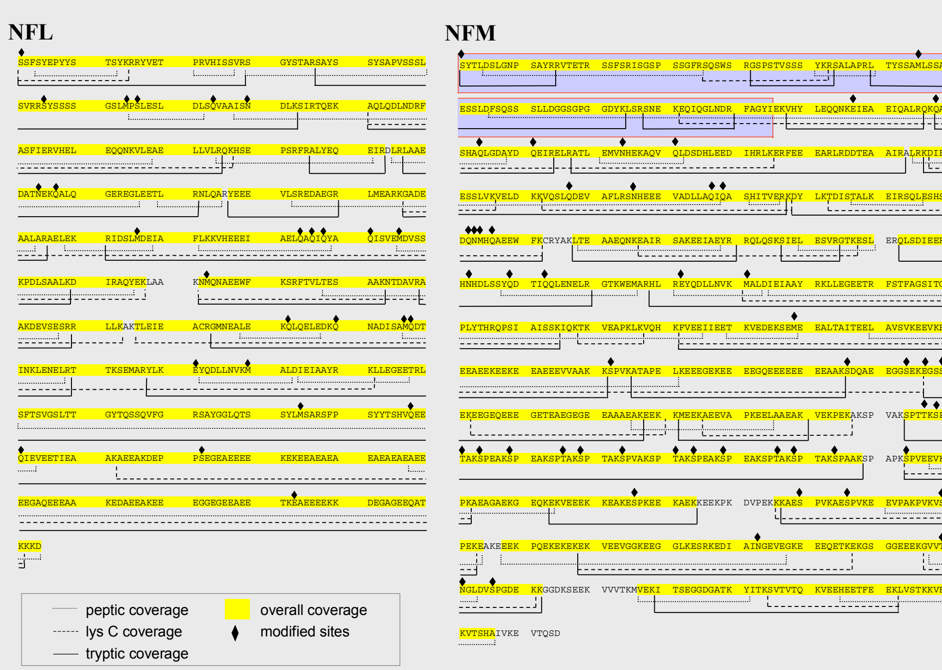
The site at Ser[66] in bovine NFL was previously uncharacterized, and the two others conform to other mammalian NFPs. Phosphorylation of bovine NFM mainly occurred in the Lys-Ser-Pro region but were also found in Val-Ser-Pro and Ser-Glu-Lys motifs. Most of them are in accordance with other mammalian phosphorylation sites. The phosphorylation sites in NFL were all heterogeneous. In NFM homogenous as well as heterogeneous phosphorylation sites were determined unambiguously.

Molecular Weight Determination



Peptide Mapping Investigations

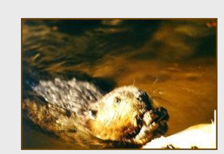
I. Sequence Coverage and *De Novo Sequencing*



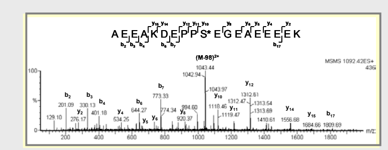
N-terminus of NFM

#	sequence	ions score	NFM species
T_{1-15}	**SYTLDSLGNPSAYRR**	90	human, mouse, rat
P_{5-12}	DSLGNPSA	56	human, mouse, rat
P_{13-23}	**YRRVTETRSSF**	49	human
P_{24-34}	**SRISGSPSSGF**	42	not observed[x]
P_{26-35}	**ISGSPSSGFR**	31	not observed
L_{36-52}	**SQSWSRGSPSTVSSSYK**	74	human, pig, rat
T_{42-52}	GSPSTVSSSYK	82	human, rat, pig
T_{42-53}	GSPSTVSSSYKR	74	human, rat, pig
P_{51-60}	**YKRSALAPRL**	31	rat, pig
T_{54-59}	SALAPR	29	rat, pig
T_{60-94}	**LTYSSAMLSSAESSLDFSQSSSLLDGGSGPGGDYK**	130	pig
P_{76-92}	FSQSSSLLDGGSGPGGD	55	pig
P_{83-92}	LDGGSGPGGD	60	pig
P_{93-114}	**YKLSRSNEKEQIQGLNDRFAGY**	38	pig
T_{98-110}	SNEKEQIQGLNDR	82	pig
{T_{98-110}	SNEKEQLQGLNDR	82	human, rat}[y]
$P_{101-109}$	KEQIQGLND	71	pig
$T_{102-110}$	EQIQGLNDR	53	pig
{$T_{102-110}$	EQLQGLNDR	53	human, rat}[y]
$L_{102-117}$	EQIQGLNDRFAGYIEK	86	pig
$T_{111-117}$	FAGYIEK	60	pig, human, rat
$P_{115-129}$	**IEKVHYLEQQNKEIE**	55	human, mouse, rat, pig

[x] error tolerant search
[y] side chain fragmentation utilizing MALDI-TOF/TOF suggests Iso instead of Leu. These peptides are therefore excluded.

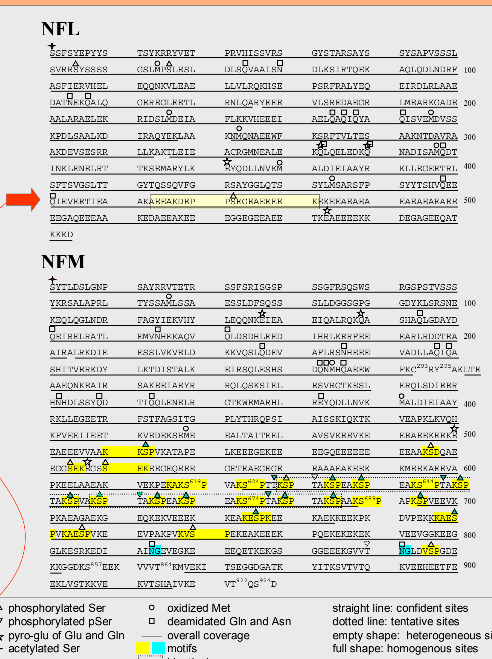


II. Determination of the Modification and Identification of the Modified Amino Acid Residues



#	peptide sequence in bovine NFL	modifications [aa: ions score]
T_{1-14}	S^{He}SFSYEPYYSTSYK	N-acetyl [S^{1}: 77]
L_{1-14}	S^{He}SFSYEPYYSTSYK	N-acetyl [S^{1}: 68]
T_{1-15}	S^{He}SFSYEPYYSTSYKR	N-acetyl [$S^{1, conf}$: **79**]
T_{54-83}	RS$^{(He)}$YSSSSGSLM^{He}PS$^{(He)}$LESLDLSQVAAISNDLK	unmodif [90]; M-oxid [M^{64}: 102]
T_{55-83}	S^{He}YSSSSGSLM^{He}PS^{He}LESLDLSQ^{He}VAAISN^{He}DLK	unmodif [235]; phospho [$S^{55, conf}$: 78]; M-oxid [$M^{64, conf}$: **232**];
		phospho [$S^{66, conf}$: **54**]; 2 deamid [$Q^{74, conf}$, $N^{80, conf}$: **78**]; deamid [N^{80}: 35]
$T_{147-156}$	LAAEDATN$^{(He)}$EK	unmodif [78]
$T_{147-163}$	LAAEDATN^{He}EKQ^{He}ALQGER	deamid [$N^{154, conf}$: **93**]; deamid [$Q^{157, conf}$: **125**]
$P_{145-166}$	LRLAAEDATNEKQ$^{(He)}$ALQGEREGL	unmodif [**32**]
$T_{211-223}$	RIDSLM^{He}DEIAFLK	unmodif [83]; M-oxid [M^{216}: 38]
$L_{211-223}$	RIDSLM^{He}DEIAFLK	M-oxid [M^{216}: 37]
$T_{211-224}$	RIDSLM^{He}DEIAFLKK	unmodif [75]; M-oxid [M^{216}: 52]
$T_{212-223}$	IDSLM^{He}DEIAFLK	unmodif [100]; M-oxid [$M^{216, conf}$: **79**]
$T_{212-224}$	IDSLM^{He}DEIAFLKK	unmodif [101]; M-oxid [M^{216}: 77]
$T_{225-259}$	VHEEEIAELQ^{He}AQ^{He}IQ^{He}YAQ^{He}ISVEM^{He}DVSSKPDLSAALK	unmodif [247]; deamid [$Q^{234, conf}$: **89**]; deamid [$Q^{236, conf}$: **65**]; deamid [$Q^{238, conf}$: **55**];
		deamid [$Q^{241, conf}$: **84**]; M-oxid [$M^{246, conf}$: **56**]
$P_{246-265}$	M^{He}DVSSKPDLSAALKDIRAQY	unmodif [56]; M-oxid [M^{246}: 45]
$T_{272-281}$	NM^{He}QNAEEWFK	unmodif [53]; M-oxid [$M^{273, conf}$: **58**]
$L_{272-281}$	NM^{He}QNAEEWFK	unmodif [65]; M-oxid [M^{273}: 48]
$P_{329-337}$	LEKQ^{He}LQELE	deamid [$Q^{332, conf}$: 58]
$L_{332-339}$	Q^{He}LQELEDK	unmodif [59]; pyro-glu [Q^{332}: 40]
$T_{332-353}$	Q$^{(He)}$LQELEDKQ^{He}NADISAM^{He}QDTINK	unmod [119]; deamid [Q^{340}: 83]; M-oxid [M^{347}: 40]
$T_{332-359}$	Q^{He}LQELEDKQ^{He}NADISA$M^{He}$$Q^{He}$DTINKLENELR	unmodif [91]; pyro-glu [$Q^{332, conf}$: **96**]; pyro-glu, M-oxid [Q^{332}, M^{347}: 57]
		2 deamid [$Q^{340, conf}$, $Q^{348, conf}$: 106]; M-oxid [M^{347}: 55]
$T_{340-353}$	QNADISAM^{He}QDTINK	unmodif [101]; M-oxid [$M^{347, conf}$: **75**]
$T_{340-359}$	QNADISAM^{He}QDTINKLENELR	unmodif [178]; M-oxid [M^{347}: 45]
$L_{340-362}$	Q^{He}NADISAMQ^{He}DTINKLENELRTTK	unmodif [76]; pyro-glu [$Q^{340, conf}$: 61]; deamid [Q^{348}: 66]
$T_{368-379}$	YLKE$^{(He)}$YQDLLNVK	unmodif [93]
$T_{371-379}$	E^{He}YQDLLNVK	unmodif [56]; pyro-glu [$E^{371, conf}$: **78**]
$T_{380-390}$	M^{He}ALDIEIAAYR	unmodif [93]; M-oxid [$M^{380, conf}$: **85**]
$T_{422-437}$	SAYGGLQTSSYLM^{He}SAR	unmodif [120]; M-oxid [$M^{434, conf}$: **56**]
$T_{438-462}$	SFPSYYTSHVQ^{He}EEQ^{He}IEVEETIEAAK	unmodif [115]; deamid [$Q^{448, conf}$: **72**]; 2 deamid [Q^{448}, Q^{451}: 51]; deamid [$Q^{451, conf}$: 61]
$T_{463-481}$	AEEAKDEPPS^{He}EGEAEEEEK	unmodif [43]; phospho [S^{472}: **83**]
$L_{463-481}$	AEEAKDEPPS^{He}EGEAEEEEK	unmodif [108]; phospho [S^{472}: **69**]
$T_{463-483}$	AEEAKDEPPS^{He}EGEAEEEEKEK	phospho [$S^{472, conf}$: **88**]
$L_{468-481}$	DEPPS^{He}EGEAEEEEK	unmodif [52]; phospho [S^{472}: 64]
$T_{468-481}$	DEPPS^{He}EGEAEEEEK	phospho [S^{472}: 55]
$T_{468-483}$	DEPPS^{He}EGEAEEEEKEK	phospho [S^{472}: 32]
$T_{533-551}$	E^{He}AEEEEKKDEGAGEEQATK	unmodif [151]; pyro-glu [$E^{533, conf}$: **91**]

bold numbers indicate highest ions score
conf confident identification
tent tentative identification
Ho homogenous modified site
He heterogeneous modified site



Acknowledgements

We thank Brain Arbogast and Lilo Barofsky for assistance. This work was supported by a subproject of grant NIH/NIEHS P42 ES10338-02 to MLD.

Literature

[1] Pant, HC, Veeranna Biochemistry & Cell Biology (**1995**) 73: 575-592.