JOURNAL OF MASS SPECTROMETRY, VOL. 32, 483–493 (1997)

Matrix-assisted Laser Desorption/Ionization Mass Spectrometric Peptide Mapping of the Neural Cell Adhesion Protein Neurolin Purified by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis or Acidic Precipitation

Martin Kussmann,[1]* Ute Lässing,[2] Claudia A. O. Stürmer,[2] Michael Przybylski[3] and Peter Roepstorff[1]

[1] Department of Molecular Biology, Protein Research Group, Odense University, Campusvej 55, DK-5230 Odense M, Denmark

[2] Faculty of Biology, University of Konstanz, Postfach 5560 M 625, D-78434 Konstanz, Germany

[3] Faculty of Chemistry, University of Konstanz, Postfach 5560 M 731, D-78434 Konstanz, Germany

Neurolin is a cell surface protein involved in the neural regeneration and neogenesis of the central nervous system of goldfish. Its theoretical molecular mass, based on the amino acid sequence translated from the cDNA, is 58 kDa, but in SDS-PAGE it shows an apparent MW of 86 kDa. Neurolin is stated to be a glycoprotein and it contains five potential *N*- and 96 potential *O*-glycosylation sites. The complete characterization of the primary structure and initial investigations on the postulated glycosylation of neurolin, immunopurified from goldfish brains, are described. The protein was either digested *in situ* in the sodium dodecyl sulfate polyacrylamide gel matrix or digested after trichloroacetic acid precipitation. Trypsin and endoprotease Glu-C were used as proteases and matrix-assisted laser desorption/ionization mass spectrometry was applied for direct peptide mapping analysis of the proteolytic mixtures. Various sample preparation techniques were performed and the mass spectra were recorded in both positive- and negative-ion modes. © 1997 by John Wiley & Sons, Ltd.

J. Mass Spectrom. **32**, 483–493 (1997)
No. of Figures: 6 No. of Tables: 3 No. of Refs: 22

KEYWORDS: MALDI mass spectrometric peptide mapping; in-gel proteolytic digestion; neurolin; neural regeneration; protein glycosylation

INTRODUCTION

Axons of the central nervous system (CNS) in adult fish are able to regenerate upon injury, unlike the mammalian CNS.[1] There are many proteins known to be involved in axonal growth and guidance in the developing nervous system, and to be re-expressed in the adult fish CNS during regeneration.[2] Recently, several growth-associated cell surface proteins in the goldfish visual system have been identified on growing embryonal axons and found re-expressed on axons in the retina and optic nerve upon optic nerve transection.[3–5] Neurolin is stated to be a glycoprotein of this group, has an apparent molecular mass (M_r) of 86 kDa and was originally identified by the monoclonal antibody E21.[4] Lectin-binding assays suggested α-2,6- and α-2,3-bound sialic acid and galactose-β(1,3)-*N*-acetylgalactosamine as carbohydrate epitopes.[4] In embryonic goldfish, neurolin is present on all retinal ganglion cells (RGCs) and their axons. In the adult organism, it is absent from mature axons along most of their length, but appears on new and growing axons derived from the RGCs at the retinal margin.[4] These peripheral RGCs of the adult goldfish visual system are constantly added to the annular retinal growth zone throughout the species' lifetime.[6] The RGC-derived axons cluster into age-related bundles within the optic nerve.[2] Furthermore, neurolin turns out to remain on adult RGCs only at intercellular contact sites, and is continuously found in the retinal axon terminal arbor layers of the adult optic tectum in the brain.[4]

The reappearance of neurolin and other growth-associated cell surface proteins after optic nerve lesions in the goldfish visual system indicates that goldfish RGCs are capable of re-expressing molecules required for successful regeneration.[2] In particular, the spatially and temporarily patterned expression of neurolin assigns this protein to be involved in RGC differentiation.[7]

In addition to functional and histological investigations of the protein, the cDNA coding for neurolin has recently been isolated and sequenced.[8] Sequence homology considerations revealed close relations to the chick cell adhesion molecule (CAM) family DM-GRASP/BEN/SC-1, and furthermore assign neurolin to be a member of the Ig superfamily.[9–11]

In this paper, we describe the primary structure characterization of native, immunopurified neurolin. Tryptic

* *Correspondence to*: M. Kussmann.

CCC 1076–5174/97/050483–11 $17.50

Received 28 November 1996
Accepted 17 January 1997

and endoprotease Gluc-C digestion were carried out *in situ* in the neurolin gel band and after trichloroacetic acid (TCA) precipitation of neurolin. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was applied to direct peptide mixture analysis. MALDI-MS, among other soft ionization/desorption techniques such as electrospray ionization (ESI), allows the mass determination and molecular characterization of large biomolecules. MALDI-MS predominantly yields singly charged molecular ions of biopolymers, is relatively tolerant towards impurities often associated with biological samples (e.g. salts, detergents) and provides mass accuracies of 0.1% or better. In particular, the recent development of the acceleration mode 'delayed extraction' (DE) in MALDI time-of-flight (TOF) MS has lead to a significantly increased resolution and sensitivity.[12] Moreover, MALDI-MS has especially proved superior for the direct analysis of complex mixtures such as obtained in proteolytic digests.[13]

Mass spectrometric peptide mapping involves the digestion of the protein by specific endoproteases and the analysis of the resulting peptide mixture by, e.g., MALDI-MS.[14] The comparison of the set of peptide molecular ions obtained with the expected, computer-based digest typically yields a sequence coverage of 70% and frequently even more than 90%. In addition to the characterization of the amino acid sequence, peptide mapping allows the determination of the location and, partially, the structure of post-translational modifications, e.g. disulfide bonds, phosphorylations and glycosylations. The characterization of protein modifications can be carried out by monitoring the 'non-matching' peptides in a proteolytic digest (i.e., those which do not correspond to expected proteolytic fragments) before and after chemical or enzymatic treatment. For instance, specific endo- and exoglycosidases can be applied to a glycoprotein followed by mass spectrometric monitoring of mass shifts of the presumably glycosylated peptides in order to determine location and constitution of a glycan structure. Glycan analysis is of particular interest in the case of neural cell adhesion glycoproteins as neurolin, because glycans play a major role in intercellular contacts and, in particular, in axonal guidance and recognition.[15] Furthermore, the comparison of peptide maps before and after DTT reduction can elucidate disulfide bonds in a protein: the disulfide-bonded non-matching peptides are found as reduced cysteinyl peptides after reduction. This has recently been demonstrated by confirmation of the immunoglobulin G (IgG) loops predicted for the neural cell adhesion protein axonin.[16]

Figure 1 shows the primary structure of neurolin. It contains 533 amino acids (AAs) and has an M_r of 58 140 Da, according to the AA sequence. In sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), it shows an apparent M_r of ~ 86 kDa under reducing and 80 kDa under non-reducing conditions, indicating the presence of a considerable amount of secondary modification.[8] Hydrophobicity plots predicted for neurolin assign it as a membrane protein with a large extracellular domain from AA 1–476, a short hydrophobic transmembrane region of 24 AA and a short hydrophilic cytoplasmic domain from AA 501 to 530. Four Ig

VGTVIGLYGETIVVPCNDGTKKPDGLIFTKWKYVKDDGSPGDLLVKQAQKDEATVSATD

GYKSRVSIAANSSLLIARGSLADQRVFTCMVVSFTNLEEYSVEVKVHKKPSAPVIKNNA

KELENGKLTQLGECVVENANPPADLIWKKNNQTLVDDGKTIIITSTITKDKITGLSSTS

SRLQYTARKEDVESQFTCTAKHVMGPDQVSEPESFPIHYPTEKVSLQVVSQSPIREGED

VTLKCQADGNPPPTSFNFNIKGKKVTVTDKDVYTLTGVTRADSGIYKCSLLDNDVMEST

QFVTVSFLDVSLTPTGKVLKNVGENLIVSLDKNASSEAKVTWTKDNRKLDKLPDFSKLT

YSDAGLYVCDVSIEGIKRSLSFELTVEGIPKITSLTKHRSSDGKHKVLTCEAEGSPKPD

VQWSVNGTNDEVSYNNGKATYKLTVVPSKNLTVSCLVTNKLGEDTKEISVFSQKNEDGT

EQAKVIVGIVVGLLVAAALVGLIYWIYIKKTRQGSWKTGEKEAGTSEESKKLEENNHKP

DV[533]

Figure 1. Primary structure of neurolin and sequence coverage achieved by tryptic and endoprotease Glu-C mass spectrometric peptide mapping. The entire sequence could be covered. Double-underlined stretches were identified by both tryptic and Glu-C peptide mapping; non-underlined partial sequences were only found as tryptic peptides. The five potential Asn–Xxx–Ser/Thr glycosylation sites are highlighted in italic bold face. Tryptic and Glu-C cleavage sites identified by corresponding partial peptides are marked in bold face.

superfamily type loops are predicted between the cysteinyl residues 16 + 88, 132 + 195, 241 + 284 and 404 + 448.

Neurolin further contains five potential *N*-glycosylation sites, namely Asn-70, -148, -328, -419 and -443 (highlighted in Fig. 1, all within the extracellular domain and consensus sequences Asn–Xxx–Ser/Thr) and many potential *O*-glycosylation positions, i.e. 48 threonine and 48 serine residues. Potential glycosylation could at least partially account for the difference of ~22–28 kDa between the calculated, sequence-based M_r and the observed M_r in SDS-PAGE.

EXPERIMENTAL

Isolation and purification of neurolin

Neurolin was isolated and purified as described previously.[8] Briefly, neurolin was immunopurified with the monoclonal antibody E21 from a pool of proteins, that was solubilized by *n*-octyl glucopyranoside and stemmed from a fraction enriched in cell surface membranes of adult goldfish brains. Approximately 100 μg of protein, dissolved in 50% glycerol–50% 100 mM Tris–HCl (pH 7.5), 1 mM EDTA, 0.1% (v/v) NP-40 (Nonidet P-40, detergent), was used as the starting material for structural characterization.

Protein reduction and alkylation

First, aliquots of 50 μg of the starting material were subjected to size-exclusion chromatography on a Pharmacia Smart System using a Pharmacia Fast Desalting PC 3.2/10 column. The protein was eluted with 2% HOAc at a flow rate of 200 μl min^{-1} and the protein fraction volume was reduced to 50 μl. Aliquots of 5 μg/5 μl were frozen and stored at −20 °C and used for

reduction and alkylation. To one of those aliquots, 85 μl of alkylation buffer, containing 500 mM Tris (pH 7.6), 6 M guanidinium hydrochloride and 5 mM EDTA, were added. Then 5 μl of a 1.4 M dithiothreitol (DTT) solution were added and reduction was carried out for 1 h at 37 °C. Subsequently, 1 μl of 4-vinylpyridine (VP) was added and the reaction mixture was kept for 10 min at room temperature. Finally, in order to quench the alkylation, another 5 μl of DTT solution (1.4 M) were added. The reaction mixture was then subjected to the same gel filtration procedure as described above, the protein fraction was dried and used for subsequent SDS-PAGE or the fraction was frozen prior to TCA precipitation.

Gel electrophoresis

Standard 12.5% SDS-PAGE was carried out using the Laemmli buffer system. A 5 μg amount of reduced, alkylated and gel-filtered neurolin were dissolved in 10 μl of sample denaturation buffer and heated at 95 °C for 3 min prior to loading the gel. A prestained molecular mass marker (14, 21, 31, 43, 66, 94, 200 kDa; Bio-Rad) was used to estimate the position of the expected 86 kDa neurolin band. After 1.5 h of electrophoresis at 20 mA, the gel was stained with Coomassie Blue for 20 min and subsequently destained for the same time. Staining and destaining were kept to a minimum, because both cause partial fixing of the protein to the gel, a process that should be avoided in the case of subsequent digestion in the gel, as carried out in this study. The 85 kDa band was excised, covered with ultra-high quality (UHQ) water and stored at −20 °C. To ensure that the correct band was excised and to check for adjacent protein bands, the gel was further destained for 1 h.

In situ digestion in the gel matrix

Proteolytic digestions in the gel matrix were performed according to the procedure of Rosenfeld *et al.*[17] with certain modifications. First, the excised gel plug was destained for 1 h in 40% acetonitrile (ACN)–60% 50 mM hydrogencarbonate (pH 7.8) in order to remove Coomassie Blue, gel buffers, SDS and salts. The plug was subsequently dried in a vacuum centrifuge for 15 min. A 0.5 μg amount of modified trypsin or endoprotease Glu-C (both sequencing grade, Boehringer-Mannheim) were dissolved in 5 μl of digestion buffer and deposited on the dried gel plug (enzyme to substrate ratio (E:S) = (1:10)). Within 10 min, the protease solution was sucked up by the plug. An additional 10 μl of digestion buffer were added and digestion was carried out for 12 h at 37 °C.

The proteolytic peptide mixture was extracted into 100 μl of 60% ACN–40% UHQ water overnight at room temperature. Finally, the gel plug was removed and the peptide solution was dried for subsequent MALDI-MS analysis.

Trichloroacetic acid precipitation

To 5 μg of reduced and alkylated neurolin, dissolved in 10 μl 2% HOAc (directly taken from size-exclusion chromatography), 2 μl of 50% TCA solution were added and the protein was allowed to precipitate for 20 min at 4 °C. The solution was then centrifuged at 14 000 rpm for 10 min at 4 °C in order to pellet the protein. TCA was pipetted off and 500 μl of acetone were added. After vortex mixing and a second centrifugation (14 000 rpm, 5 min, 4 °C), the acetone was pipetted off and the protein pellet was dried in a vacuum centrifuge. This pellet was dissolved in 50 mM hydrogencarbonate (pH 7.8) for proteolytic degradation.

Proteolytic digestion in solution

A 5 μg amount of reduced and alkylated, TCA-precipitated neurolin was dissolved in 5 μl of digestion buffer (50 mM hydrogencarbonate, pH 7.8). An aliquot of 100 ng of modified trypsin (sequencing grade, Boehringer-Mannheim), dissolved in 1 μl of digestion buffer was added (E:S = 1:50) and digestion was carried out for 12 h at 37 °C. Finally, the solution was dried for subsequent MALDI-MS analysis.

Sample preparation for MALDI-MS

Aliquots of one tenth of the extracted (after *in situ* digest in the neurolin gel band) or dried (after in-solution digest of TCA-precipitated neurolin) peptide mixture (~5–10 pmol), dissolved in 0.5 μl of 30% ACN–70% TFA (2%), were used for MALDI-MS analysis. The following matrices and sample preparation techniques were applied.[18,19]

4-Hydroxy-α-cyanocinnamic acid (HCCA), sandwich method. A first HCCA solution was prepared by dissolving the matrix in acetone (20 μg μl^{-1}) and by adding ~1% (v/v) of 0.1% trifluoroacetic acid (TFA), and a second by dissolving it in 70% ACN–30% TFA (0.1%) (20 μg μl^{-1}).

A 0.5 μl volume of HCCA in acetone was deposited on the target and dried, then 1 μl of 2% TFA, 0.5 μl of sample solution and 0.5 μl of HCCA in 70% ACN–30% UHQ water were subsequently added on top of the first matrix layer and allowed to dry. Optionally, the dried sample could be washed 1–3 times with 10 μl of 0.1% TFA.

HCCA plus nitrocellulose as additive. Equal amounts of nitrocellulose (NC) and HCCA were first dissolved in acetone (40 μg μl^{-1}) and this solution was then diluted with propan-2-ol to 20 μg μl^{-1}. Again, the final solution contained ~1% (v/v) of 0.1% TFA. Twice, 10 μl of this matrix preparation were applied to the target and spin-dried. On this layer, 0.5 μl of 2% TFA and 0.5 μl of sample solution were added and allowed to dry without spinning. Finally, 0.5 μl of each of 2% TFA and HCCA in 70% ACN–30% UHQ water were deposited and allowed to dry without spinning. Alternatively, when using a multi-sample target, the NC–HCCA layer was formed on a small piece of Scotch tape (adhesive on both sides) and subsequently washed with 100 μl of 0.1% TFA. The prepared tape was transferred to the target and the following steps were identical with those described above.

2,5-Dihydroxybenzoic acid (DHB) and sinapic acid (SA), dried-droplet method. A 0.5 μl volume of matrix in 70% ACN–30% TFA (0.1%) (20 μg μl^{-1}) and 0.5 μl of sample solution were mixed on the target and allowed to air dry or to dry under a gentle stream of argon.

Mass spectrometry

MALDI/TOF-MS was performed on two different instruments, namely a Bruker Reflex and a PerSeptive Voyager Elite mass spectrometer, the latter being equipped with DE. The Bruker instrument was equipped with the data acquisition and processing software Laser One (by M. Mann and P. Mortensen, EMBL, Heidelberg, Germany). Mass spectra were recorded in linear and reflector and in positive- and negative-ion modes, with low mass gates set at 500, 1000 and 2000 Da. Instrumental calibration was routinely used on both instruments, and internal calibration with molecular ions of regularly occurring neurolin peptides was additionally performed to consolidate peptide assignments further.

Computational analysis of proteolytic digests

The computer program GPMAW (General Peptide Mass Analysis for Windows, Peter Højrup, Department of Molecular Biology, Odense University) was used for computer-assisted comparison of the tryptic and V8-proteolytic peptide mapping data with the expected set of peptides based on the neurolin amino acid sequence according to Lässing *et al.*[8] (see Fig. 1).

RESULTS

Protein and peptide sample preparation

The neurolin sample was twice subjected to gel filtration, namely before and after reduction and alkylation. This was particularly necessary in order to purify the protein from glycerol and detergent in the starting material and from DTT and VP in the reduction/alkylation mixture. Unfortunately, the reduced and alkylated protein turned out to be resistant to any kind of proteolytic digestion in solution. Neither trypsin, chymotrypsin nor the endoproteases Asp-N and Glu-C yielded a specific digest in acceptable yields. Further purification prior to digestion by means of C_4 reversed-phase microbore high-performance liquid chromatography was not successful because neurolin could not be recovered after chromatography. This could be due to the poor solubility of neurolin in the loading buffer (0.1% TFA) or to a too large hydrophobicity of the protein.

Referring to good results in our laboratory with regard to overcoming protease resistance of proteins, SDS-PAGE followed by *in situ* tryptic or endoprotease Glu-C digestion in the gel matrix were performed. Glycerol and the detergent NP-40, contained in the original neurolin sample and commonly used in the isolation and immunoaffinity purification of membrane-bound proteins, might still have stuck to the protein and thus shielded it from protease attacks. In SDS-PAGE, SDS is known to replace other detergents on the surface of the applied protein. Furthermore, it can be sufficiently removed afterwards and, thus, SDS-PAGE and *in situ* digestion are appropriate tools to make a protein amenable to proteolysis. Especially the recently developed combination of *in situ* digestion of proteins separated by SDS-PAGE with subsequent MALDI-MS peptide mapping has proved superior for fast and sensitive protein identification and characterization, particularly of complex protein mixtures and of proteolytically inaccessible proteins. Typical sequence coverages of direct MALDI-MS analysis of *in situ* digests range from 30 to 80% of the protein sequence and thereby allow, in addition to protein identification, considerable sequence verification and the partial characterization of post-translational modifications.[20]

In the reducing gel, the neurolin sample showed a single, well defined band at ~86 kDa and no protein contaminants (data not shown).

As an alternative and complementary approach, in order to make neurolin amenable to proteolysis, the reduced and alkylated protein was precipitated with TCA prior to conventional digestion with trypsin in solution.

Because of the limited amount of sample (see Experimental), the tryptic and V8 proteolytic mixtures were directly subjected to MALDI peptide mapping. With regard to previous studies,[18,19] in which the protein sequence coverage in peptide mixture analyses is dependent on the choice of matrix and the sample preparation procedure, three different matrices were used in different sample preparation techniques (see Experimental). Especially the combination of the matrix DHB with the delayed extraction acceleration mode, as chosen in this study, has proved efficient for the detection of glycopeptides.

Tryptic peptide mapping

Figure 2 shows a MALDI mass spectrum of an *in situ* tryptic digest of reduced and alkylated neurolin, recorded in the linear positive-ion mode (matrix: HCCA). Almost every signal can be assigned to tryptic peptides predicted from the translated cDNA sequence. The peptides found range from *N*-terminal (e.g. 31–50), through core (e.g. 199–220) to *C*-terminal partial sequences (e.g. 502–522). In the case of closely coinciding masses of the tryptic peptides, so that the latter could not be assigned unambiguously based on the mass accuracy of the mass spectrometer, both possible partial sequences are given. In this spectrum, ~ 60% of the amino acid sequence of neurolin could be covered (see Table 1).

Figure 3(a) shows the lower mass range of a spectrum of the same digest, recorded in the reflector positive-ion mode with delayed extraction (matrix: HCCA). As mentioned in the Introduction, DE is a recently introduced acceleration mode in MALDI/TOF-MS resulting in increased sensitivity and resolution.[12] The combined reflector–DE running mode allows the monoisotopic

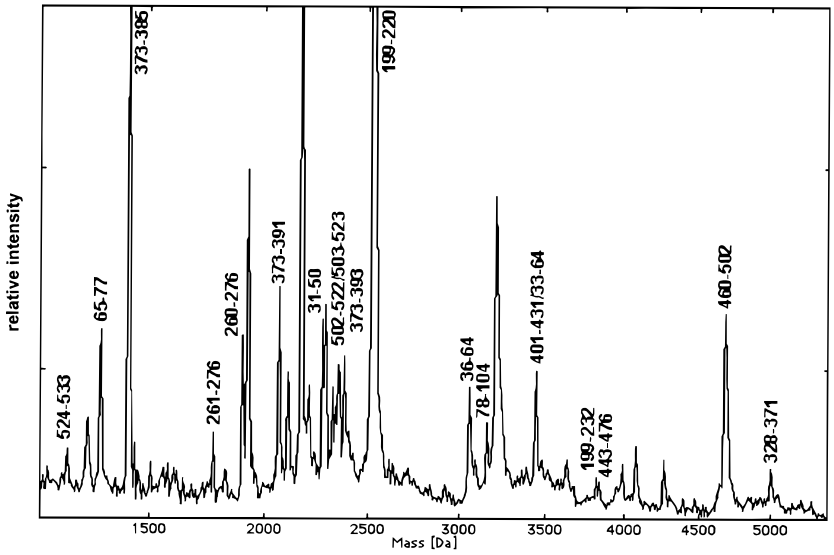

Figure 2. MALDI mass spectrum of an in-gel tryptic digest of reduced and alkylated neurolin. About 5 pmol of peptide mixture were used in a sandwich sample preparation with HCCA as matrix. The spectrum was recorded in linear positive-ion mode without delayed extraction. Tryptic peptides found are assigned.

resolution of mass signals above m/z 2500 and still provides good sensitivity. In general, monoisotopic mass determinations yield higher mass accuracies (0.01% or better) and, therefore, more definite peptide assignments than average mass determinations as obtained in the linear mode (0.1–0.01% mass accuracy).

The peptides found in the reflector mode spectrum (Fig. 3(a)) cover partial sequences from the *N*- and *C*-terminal and also the central part of neurolin, and almost every significant mass signal could be assigned (see Table 1). In order to illustrate the accuracy of the mass determination, the found and calculated peptide masses are also given. In the enlargement of the spectrum of Fig. 3(b), the monoisotopic resolution of two signals in the 2.5 kDa mass range is shown. The molecular ion at m/z 2524.22 (monoisotopic mass) corresponds to the tryptic peptide 199–220 ($m/z_{\text{calc.}}$ 2524.18). The mass difference of 16 Da and the methionine at position 201 suggest the molecular ion at m/z 2540.22 to be the oxidized form of the peptide 199–220 ($m/z_{\text{calc.}}$ 2540.79). Methionine oxidation can occur in all sample processing steps, including the gel experiment and

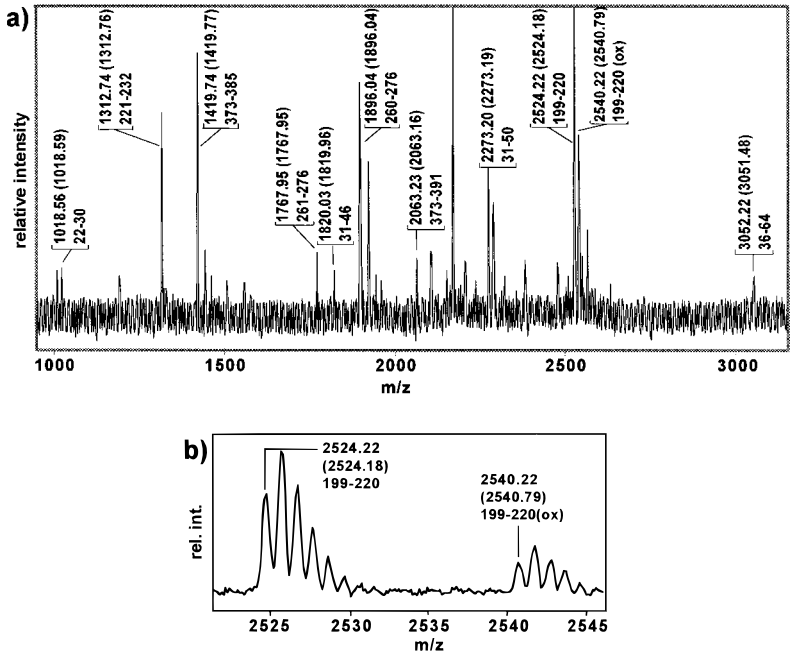

Figure 3. MALDI mass spectrum of an in-gel tryptic digest of reduced and alkylated neurolin. About 5 pmol of peptide mixture were used in a sandwich sample preparation with HCCA as matrix. The spectrum was recorded with delayed extraction in reflector positive-ion mode. The peak assignment is as follows: found m/z first, calculated m/z second in parentheses; sequence position third. (a) Lower mass range of the spectrum; (b) enlargement of the 2500 Da range: molecular ions of the native and the oxidized tryptic peptide 199–220, containing Met^{201}.

Table 1. Extract of tryptic peptide mapping data of reduced and alkylated neurolin[a]

m/z (av.) found	*m/z* (av.) calculated	*m/z* (mi), found	*m/z* (mi) calculated	Sequence position	Peptide	$[M+H]^+$	$[M-H]^-$
605.86	605.67			505–509	T63	x	
632.72	632.73	631.97	632.34	335–339	T40	(x)	x
751.90	751.86			180–185	T24	x	
837.90	838.03			108–115	T14–15		x
1019.19	1019.23	1018.56	1018.59	22–30	T2–3	(x)	x
1066.06	1066.11			514–523	T65–66	x	
1123.57	1123.25			267–276	T34		x^1
1192.90	1193.26			524–533	T67–68		x
1278.51	1278.41	1277.77	1277.66	503–513	T62–64	x	
1299.98	1299.46			316–327	T38		x^1
1311.73	1311.52			221–232	T28	(x)	x^2
1313.54	1313.54	1312.74	1312.76	221–232	T28	x	
1418.59	1418.63	1418.02	1417.75	373–385	T47	(x)	x
1463.56	1463.61			187–198	T26	x	
1506.30	1506.70	1505.76	1505.78	33–46	T5–6	x	
1557.03	1556.80			63–77	T9–10	(x)	x^2
1769.00	1769.01	1767.95	1767.95	261–276	T33–34	x	
1821.51	1821.09	1820.03	1819.96	31–46	T4–6	x	
1897.32	1897.18	1896.04	1896.04	260–276	T32–34	$x^{1,2}$	
1929.25	1929.00			514–530	T65–67		x
1941.73	1941.26			443–459	T56–57	x	
1962.06	1962.21	1960.81	1961.03	33–50	T5–7	x	
2062.75	2062.41			373–391	T47–48	(x)	x
2152.53	2152.48			353–371	T45	x	
2241.86	2241.62			1–21	T1	x	
2270.67	2270.53			392–411	T49–52	(x)	x
2274.44	2274.48			31–50	T4–7	(x)	x
2325.38	2325.49			502–522	T61–65	x	
				503–523	T62–66		
2357.32	2357.75			373–393	T47–49	x	
2483.60	2483.98			432–453	T54–56	(x)	x
2524.21	2523.78	2522.48	2522.16	199–220	T27	(x)	x
2547.90	2547.91	2546.22	2546.40	147–169	T19–22	x	
2554.53	2554.73			454–476	T57–59	x	
2786.24	2785.57			477–502	T60–61	x	
2933.77	2933.12			505–530	T63–67	x^1	
3051.73	3051.25			36–64	T6–9		x
3139.10	3139.52			233–260	T29–32		x
3156.26	3156.61			78–104	T11–12	(x)	$x^{1,2}$
3183.11	3183.67			158–186	T21–25	(x)	x^2
3022.40	3200.49	3198.35	3198.58	33–62	T5–8	x	
3314.32	3314.71	3312.96	3312.65	170–198	T23–26	x^2	
3459.23	3458.95			313–343	T37–42	x	
3474.35	3474.90			316–346	T38–43	x	
3820.96	3820.32			199–232	T27–28	x	
3833.86	3833.27			443–476	T56–59	x	
3988.40	3988.53			277–312	T35–36	x	
4464.24	4464.20			108–147	T15–19	x	
4472.84	4474.15			372–411	T47–52	x^1	
4479.05	4477.84			392–431	T49–53	x	
4505.29	4503.16			313–352	T37–44	x	
4677.58	4677.58			460–502	T68–61	x	
4696.20	4698.43			63–104	T9–12	x	
4752.26	4752.37			241–283	T30–35		x^1
	4752.50			344–385	T43–47		
4889.92	4891.52			233–276	T29–34	x	
5012.70	5013.69			328–371	T39–45	x	
5497.71	5499.40			1–50	T1–7	x	
7498.87	7499.57			454–522	T57–65	x^1	
7651.28	7651.82			344–411	T43–52	x^1	
8587.29	8587.83			108–185	T14–24	x^1	
	8587.83			109–186	T15–25		

[a] The data stem from in-gel tryptic digestions and from tryptic digestions in solution after TCA precipitation of neurolin. Average (av.) masses are obtained in linear, monoisotopic (mi) masses are obtained in reflector mode. If a peptide was detected in both positive- and negative-ion modes, only the negative-ion masses are given and (x) indicates the complementary detection in the positive mode; x indicates the detection with (NC–)HCCA, x^1 with SA and x^2 with DHB as the matrix; the table does not contain all information in the spectra of Figs 2–5 and vice versa.

MALDI sample preparation. Using the sandwich method, this can be kept to a minimum during MALDI sample preparation but it cannot be completely avoided.[18] Often, as in this case, methionine oxidation can be of further assistance in the identification of peptides.

Figure 4 shows the negative-ion MALDI mass spectrum of the same tryptic digest as described above (linear mode, DE, HCCA as matrix). All peptides identified in this spectrum could also be detected as positive ions (see Table 1). Again, in the case of closely coinciding masses of the tryptic peptides and ambiguities in terms of peptide assignments, both possible partial sequences are given. However, when using DHB, SA or HCCA plus NC as alternative matrices and different sample preparation techniques than the sandwich method, a few peptides turned out to be exclusively detectable in the negative-ion mode: T7–8 (47–62), T14–15 (108–115), T33–35 (261–283), T34 (267–276), T38 (316–327), T65–67 (514–530) and T67–68 (524–533). T38, T65–67 and T67–68 represent acidic partial peptides with an excess of glutamic or aspartic acid *vs.* the basic lysine or arginine residues, a fact that can explain the preferred formation of negative ions. By contrast, the other peptides mentioned have a balanced amount of acidic and basic residues so that the reasons for their exclusive yield of negative ions seem to be more complex.

Compared with the previous positive-ion spectra, there further appear more 'mismatches,' i.e. ions, that do not correspond to tryptic peptides. They are due to modified peptides (not further analyzed) rather than to unspecific cleavages by trypsin, because this protease is generally very specific.

Figure 5 shows the mass spectrum of a tryptic digest of neurolin in solution after TCA precipitation of the reduced and alkylated protein (linear positive-ion mode, DE, HCCA as matrix). As in previous figures, in the

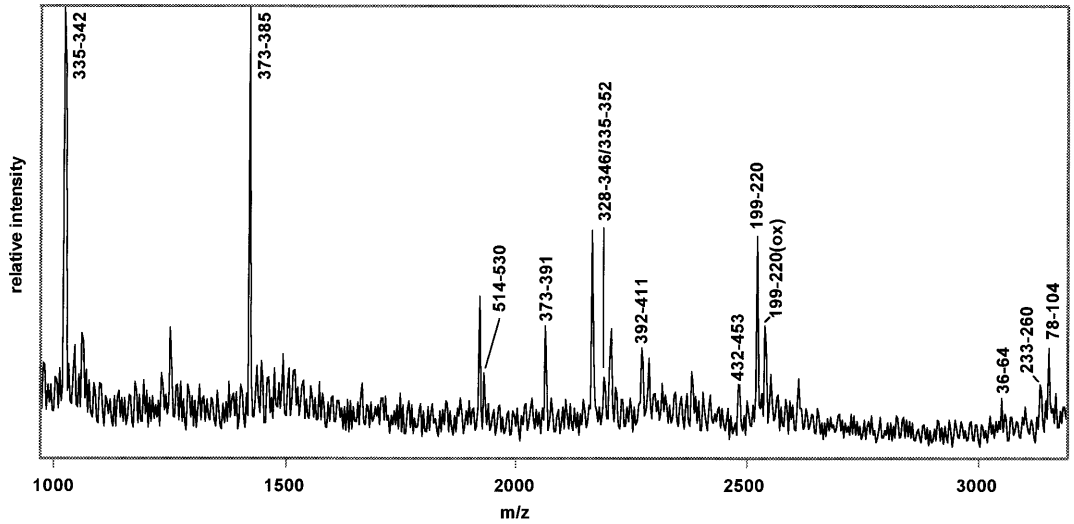

Figure 4. Negative-ion MALDI mass spectrum of an in-gel tryptic digest of reduced and alkylated neurolin. About 10 pmol of peptide mixture were used in a sandwich sample preparation with HCCA as matrix. The spectrum was recorded with delayed extraction in the linear mode. Tryptic peptides found are assigned.

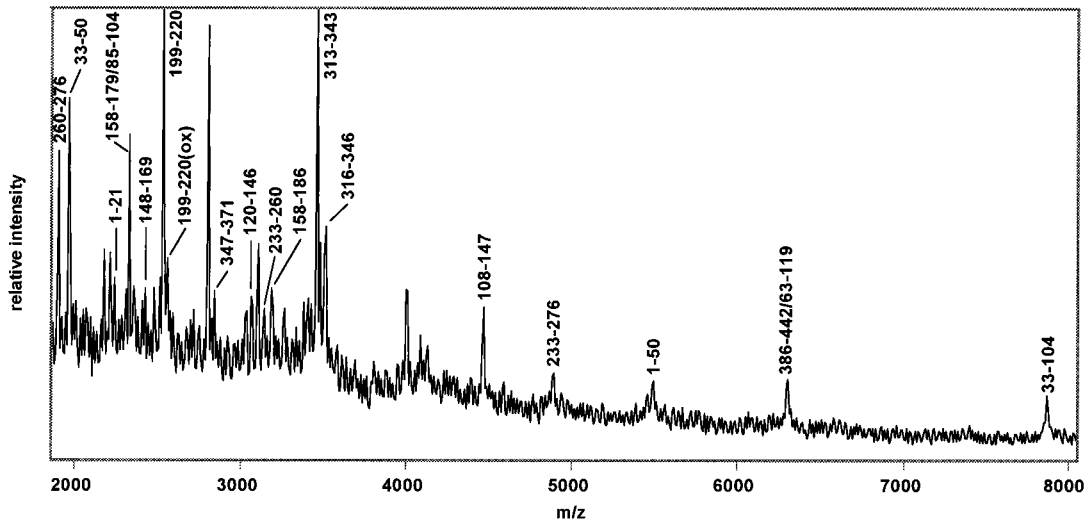

Figure 5. MALDI mass spectrum of a tryptic digest of reduced and alkylated neurolin, carried out in solution after TCA precipitation of the protein. About 5 pmol of peptide mixture were used in a sandwich sample preparation with HCCA as matrix. The spectrum was recorded with delayed extraction in the linear positive-ion mode. Tryptic peptides found are assigned.

case of closely coinciding peptide masses and ambiguous peptide assignments, both possible partial sequences are given. Complementary to tryptic peptide maps derived from in-gel digests of neurolin, this spectrum also shows peptides in the high mass range ($m/z > 5000$) that are due to non-quantitative cleavages by trypsin and contain two or more uncleaved Arg–Xxx and Lys–Xxx positions. However, the large proteolytic fragments could also be assigned, and a large part of the neurolin primary structure could be verified ($\sim 74\%$, see Table 1). In general, peptide maps of tryptic digests of TCA-precipitated neurolin showed fewer specific cleavages in the lower peptide mass range ($m/z < 3000$) and more non-quantitatively cleaved peptides compared with digestions in the gel matrix. The latter observation could on the one hand be attributed to the extraction process subsequent to the in-gel digest, which might not yield the very large peptides, whereas these can be recovered after an in-solution digest of TCA-precipitated neurolin. Another reason might be the different 'protein preparations' for the subsequent digest, i.e. that neurolin is less denaturated after TCA precipitation that it is after a gel electrophoretic procedure, so that, in the first case, fewer tryptic cleavage sites are accessible, leading to a less quantitative digestion and larger proteolytic fragments. The last conclusion turned out to be more likely because, in contrast to the tryptic degradation, the in-gel V8-proteolytic digest yielded fragments in the very high mass range (see below), meaning that the lack of high-mass peptides after tryptic in gel digestion is not likely to be due to losses during the peptide extraction process.

Table 1 shows an extract of the MALDI peptide mapping data of reduced and alkylated neurolin, obtained both from *in situ* tryptic digestion in the gel matrix and from in-solution digestion after TCA precipitation. The table encompasses the results obtained by (i) different sample preparations (i.e. the sandwich method with HCCA as matrix, the nitrocellulose variant of the sandwich–HCCA technique and the dried-droplet method with SA and DHB as matrices) and (ii) by different mass spectrometric modes (i.e. linear mode, reflector mode, positive- and negative-ion modes, each of them with and without delayed extraction).

The mass accuracy is better than 0.05% and for peptides below 4000 Da the mass deviation is between 0.7 and 0.01 Da. There are two partial peptides, namely T37–44 (313–352) and T49–53 (392–431), whose found MH^+ ions show a positive mass deviation and which further contain potential *N*-glycosylation sites, i.e. N^{328} and N^{419}, respectively. This is of particular interest, since a potential cleavage of an *N*-glycan during the neurolin purification or sample preparation process, causing a conversion of the *N*-carboxamido- into an *N*-carboxyl group, accounts for a +1 Da mass increment. The observed increased masses of these two peptides may therefore indicate a possible glycosylation site (see Discussion).

Tryptic peptide mapping yielded 100% sequence coverage of neurolin, a result that is illustrated in Fig. 1, where double-underlined partial sequences were covered by both trypsin and endoprotease Glu-C and non-underlined stretches were identified only by tryptic peptide mapping. Every part of the neurolin primary structure was covered at least twice by overlaps of quantitatively cleaved smaller tryptic peptides and larger fragments harboring several adjacent tryptic peptides.

Peptide mapping with endoprotease Glu-C

In congruence with previous studies, endoprotease Glu-C (V8-protease) turned out to cleave less specifically in the gel matrix than trypsin does. However, V8-proteolytic peptide mapping could be used for confirmation of the partial sequences already found as tryptic peptides.

Figure 6 shows the high mass range of a MALDI mass spectrum of an in-gel V8-proteolytic digest of neurolin (linear positive-ion mode, DE, HCCA as matrix). Most of the molecular ions could be assigned to endoprotease Glu-C peptides, covering $\sim 74\%$ of the neurolin primary structure. In the case of closely coin-

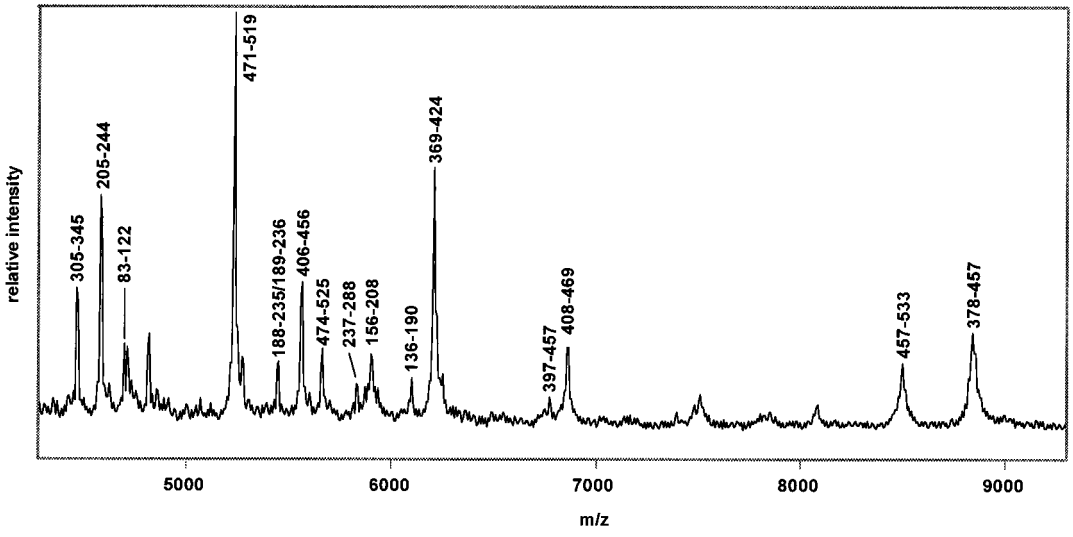

Figure 6. MALDI mass spectrum of an in-gel V8-proteolytic (endoprotease Glu-C) digest of reduced and alkylated neurolin. About 5 pmol of peptide mixture were used in a sandwich sample preparation with HCCA as matrix. The spectrum was recorded with delayed extraction in the linear positive-ion mode. V8-proteolytic peptides found are assigned.

ciding peptide masses and ambiguous peptide assignments, both possible partial sequences are given. As mentioned above, the in-gel endoprotease Glu-C digestion yielded more of the larger fragments (corresponding to non-quantitative cleavage) compared with the analogous degradation with trypsin. In particular, C-terminal and core-derived peptides of neurolin are predominant, including the long C-terminal sequence 457–533.

Table 2 gives an extract of the MALDI peptide mapping data of reduced and alkylated neurolin, obtained from *in situ* endoprotease Glu-C digestion in the gel matrix. It includes the results obtained by different sample preparations, i.e. the sandwich method with HCCA as matrix, the nitrocellulose variant of the sandwich HCCA technique and the dried-droplet method with DHB as matrix. The mass accuracy is better than 0.06%.

The V8-proteolytic sequence verification covers 93% of the entire primary structure and is shown in Fig. 1. As it is valid for tryptic peptide mapping, every neurolin partial sequence covered by V8-protease could be identified at least twice by this enzyme because of overlaps of quantitatively cleaved smaller Glu-C peptides and larger fragments encompassing several adjacent V8 peptides.

Table 2. Extract of *in situ* endoprotease Glu-C (V8-protease) peptide mapping data of reduced and alkylated neurolin[a]

MH^+ (av.), found	MH^+ (av.), calc.	Sequence position	Peptide
863.19	862.96	520–526	V66–68
1465.36	1465.69	365–377	V48–49
1524.95	1525.66	457–469	V58–60
		458–470	V59–61
1538.15	1538.70	521–533	V67–70
1701.80	1701.79	408–423	V54–55
1787.10	1788.03	350–364	V46–47
2292.57	2292.67	358–377	V47–49
2346.12	2346.51	37–59	V5–9[b]
2350.85	2350.68	60–82	V10
2465.51	2466.89	99–120	V13–14
2640.30	2640.92	188–210	V23–27
2743.29	2742.22	382–405	V51–52
2945.72	2944.27	211–236	V28–31
2946.57	2947.32	320–345	V41–44
3168.64	3170.51	209–236	V27–31
3213.60	3213.66	266–293	V34–38
3358.77	3356.82	289–319	V37–40
3651.12	3650.22	99–131	V13–16
3796.95	3798.36	333–364	V43–47
4471.17	4471.12	305–345	V40–44
4590.35	4590.26	1–42	V1–6
5233.44	5234.16	471–519	V62–65
5560.21	5560.22	406–456	V53–57
5661.68	5661.70	474–525	V63–67
5829.38	5832.73	237–288	V32–36
6094.99	6091.86	136–190	V18–24
6214.76	6215.00	369–424	V49–55
6773.57	6776.66	397–457	V52–58
6867.90	6866.67	408–469	V54–60
6869.80	6869.90	458–520	V59–66
8504.24	8504.67	457–533	V59–70

[a] The data stem from in-gel V8-proteolytic digestions and from digestions in solution after TCA precipitation of neurolin. Average (av.) masses are obtained in linear mode. The table does not contain all information in the spectrum of Fig. 6 and vice versa.
[b] Detected with DHB as the matrix; all other peptides were detected with the matrix (NC–)HCCA.

Experiments on the characterization of post-translational modifications

First, we attempted to determine the positions of the disulfide bonds in neurolin. As mentioned in the Introduction, four Ig superfamily type domains are predicted for neurolin, namely between the cysteinyl residues 16 + 88, 132 + 195, 241 + 284 and 404 + 448. *In situ* tryptic digests of unreduced neurolin were performed in the gel matrix. These digests revealed significantly less specific cleavages than those of the reduced and alkylated protein and no molecular ions could be assigned to disulfide-linked tryptic peptides. The DTT reduction of this tryptic peptide mixture did not yield reduced cysteinyl peptides, possibly originating from cysteinyl peptides, and it did not reproduce the data of the tryptic digest of reduced and alkylated neurolin. These negative results can be partially explained by the predicted Ig superfamily loop structure, which may significantly reduce the accessibility of the native protein to endoproteases compared with the reduced, alkylated and unfolded protein.

In order to obtain further information on the postulated neurolin glycosylation directly from peptide mixture analyses, PNGaseF was used to specifically remove possible *N*-glycans. Those tryptic peptides that show molecular ions exclusively after PNGaseF treatment can be assumed to be *N*-glycosylated. Neurolin was treated with PNGaseF in the gel plug[20] and subsequently digested with trypsin (E:S $\approx$ 1:10) and, alternatively, we attempted to apply PNGaseF to the extracted peptide mixture of a tryptic in-gel digestion.[21] Owing to the limited amount of sample, these two experiments could only be carried out with a few micrograms of neurolin and, therefore, MALDI-MS analyses of the PNGaseF-treated peptide mixtures revealed a poor signal-to-noise ratio and did not permit the identification of *N*-glycosylation sites.

To elucidate further the divergence between in-gel observed and real protein mass, we attempted to determine the M_r of neurolin directly by MALDI-MS. Despite the application of different matrices and sample preparation techniques, there were no molecular ions between 50 and 100 kDa or in the low molecular mass range detectable. This can be partially explained by the often observed difficulty of obtaining molecular ions of glycoproteins because of their pronounced heterogeneity.

DISCUSSION

The first aim of this study was to characterize the primary structure of neurolin isolated from goldfish brains and to establish the protein sequence identity with the translated cDNA sequence published previously.[8] This was achieved. The second task was the determination of the location and, partially, of the structure of postulated post-translational modifications,

in particular glycosylations. The preliminary results in the latter context revealed none of the 101 putative glycosylation sites as being fully glycosylated.

As can be seen in Tables 1 and 2, by far the most peptides could be detected with HCCA as the matrix, either used in the sandwich sample preparation or in the spin-dry variant with nitrocellulose as additive. This is in congruence with the experience that HCCA often yields the broadest protein sequence coverage in MALDI-MS peptide mixture analyses.[19] Nevertheless, especially the tryptic peptide mapping results (Table 1) exhibit complementary data obtained with the matrices SA and DHB. In particular, several negatively charged and some of the very large peptide molecular ions were exclusively detected in dried-droplet sample preparations with DHB and/or SA as the matrices.

According to previous immunohistological data[4] and to SDS-PAGE analysis, neurolin is stated to be a glycoprotein. Lectin-binding assays suggested α-2,6- and α-2,3-bound sialic acid, and galactose-β(1,3)-*N*-acetylgalactosamine as carbohydrate epitopes.[4] By contrast, the MALDI-MS finding of a 100% sequence coverage by tryptic and V8-proteolytic peptide mapping proves that every partial peptide exists in a non-glycosylated form. As a consequence, none of the 101 potential glycosylation sites, i.e. five Asn–Xxx–Ser/Thr consensus sequences (Asn-70, -149, -328, -419 and -443), 48 serine and 48 threonine residues, is quantitatively glycosylated. Instead, neurolin is likely to be partially and heterogeneously glycosylated. Heterogeneous glycosylation is contradicted by the finding of one well defined protein band in the SDS-PAGE analysis of neurolin showing an M_r of ~ 86 kDa in the reducing gel (AA sequence-based $M_r = 58$ kDa), rather suggesting extensive and homogeneous modification of the protein.

However, the detection of a molecular ion of an unmodified peptide in a mixture analysis does not exclude partial modification of the same species, i.e. the existence of the same peptide in the modified form, which possibly does not show a molecular ion owing to suppression effects. This is especially valid for glycosylated peptides that in general reveal a wide distribution of molecular ions due to their high degree of heterogeneity, which finally results in low ion abundances for one homogeneous species.[21] On the other hand, it has recently been shown that *N*- and *O*-glycosidic bonds are hardly susceptible to fragmentation in MALDI, meaning that a complex molecular ion pattern of a glycopeptide reflects its actual heterogeneity.[21] Furthermore, according to our experience, glycan losses during protein sample preparation prior to MALDI-MS analysis are unlikely.

Nevertheless, there have been hitherto a few MS results accounting for neurolin glycosylation. First, as already mentioned in the discussion of the data in Table 1, there are two partial peptides, namely T37–44 (313–352) and T49–53 (392–431), whose found MH^+ ions show a positive mass deviation and which further contain potential *N*-glycosylation sites, i.e. N^{328} and N^{419}, respectively. This can indicate a potential loss of a glycan, since a cleavage of an *N*-glycan during the neurolin purification or sample preparation process (although unlikely), encompassing a conversion of the *N*-carboxamido- into an *N*-carboxyl group, would account for a + 1 Da mass increment. The found increased masses of these two peptides can therefore suggest glycosylation sites.

Second, with regard to the sequence coverage of both endoproteases, there is one of those partial sequences (i.e. 142–154) only identified by tryptic, but not by V8-proteolytic peptide mapping, that contains a potential *N*-glycosylation site, namely N^{148}. According to the MALDI-MS data N^{148}, N^{328} and N^{419} therefore remain the more likely candidates for *N*-glycosylation.

Moreover, the other partial sequence that could not be covered by V8-proteolytic peptide mapping, i.e. 245–265, harbours one potential *O*-glycosylation site, i.e. S^{251}, suggesting this serine residue to be a more probable candidate for *O*-glycosylation than the remaining 47 serine and 48 threonine residues.

Appropriate tools to pursue further the characterization of the neurolin glycosylation are the carbohydrate analysis and the HPLC separation of the proteolytic digests prior to MS investigation. The carbohydrate analysis would yield the composition of the glycans, whereas the chromatographic separation would allow the isolation and separate characterization of the putative glycopeptides. Furthermore, the application of neuraminidase (sialidase) first to remove the suggested terminal sialic acids from the glycans is an alternative enzymatic approach to tackle the glycan structure elucidation. These experiments will be the subject of future studies but could not be performed hitherto, owing to the limited amount of neurolin sample.

CONCLUSIONS

The complete primary structure of neurolin, isolated from adult goldfish brains, has been characterized and the identity with the cDNA sequence published previously[8] has been established. According to the peptide mapping data presented in this study, none of the 101 potential glycosylation sites is quantitatively glycosylated. The considerable difference between the apparent M_r of 80–86 kDa in SDS-PAGE and 58 kDa based on the neurolin primary structure could not be explained in terms of extensive secondary modifications. However, the divergence between in-gel observed and real protein mass has been observed with other proteins.[22]

REFERENCES

1. J. H. P. Skene, *Annu. Rev. Neurosci.* **12**, 127 (1989).
2. C. A. O. Stürmer, M. Bastmeyer, M. Baehr, G. Strobel and K. Paschke, *J. Neurobiol.* **23**, 537 (1992).
3. M. Bastmeyer, B. Schlosshauer and C. A. O. Stürmer, *Development* **108**, 299 (1990).
4. K. A. Paschke, F. Lottspeich and C. A. O. Stürmer, *J. Cell. Biol.* **117**, 863 (1992).
5. J. Vielmetter, F. Lottspeich and C. A. O. Stürmer, *J. Neurosci.* **11**, 3581 (1991).
6. P. A. Johns, *J. Comp. Neurol.* **176**, 343 (1977).

7. U. Lässing and C. A. O. Stürmer, *J. Neurobiol*. **29**, 65 (1995).
8. U. Lässing, S. Giordano, B. Stecher, F. Lottspeich and C. A. O. Stürmer, *Differentiation* **56**, 21 (1994).
9. F. R. Burns, S. von Kannen, L. Guy, J. A. Raper, J. Kamholz and S. Chang, *Neuron* **7**, 209 (1991).
10. O. Pourquié, C. Corbel, J.-P. LeCaer, J. Rossier and N. M. LeDourain, *Proc. Natl. Acad. Sci. USA* **89**, 5261 (1992).
11. H. Tanaka, T. Matsui, A. Agata, M. Tomura, I. Kubota, K. C. McFarland, B. Kohr, A. Lee, H. S. Phillips and D. L. L. Shelton, *Neuron* **7**, 535 (1991).
12. M. L. Vestal, P. Juhasz and S. A. Martin, *Rapid Commun. Mass Spectrom*. **9**, 1044 (1995).
13. F. Hillenkamp and M. Karas, *Methods Enzymol*. **193**, 280 (1991).
14. H. R. Morris, M. Panico, M. Barber, R. S. Bordoli, R. D. Sedgwich and A. N. Taylor, *Biochem. Biophys. Res. Commun*. **101**, 623 (1981).
15. A. Varki, *Glycobiology* **3**, 97 (1993).
16. T. Denzinger, Diplomarbeit, Universität Konstanz (1995).
17. J. Rosenfeld, F. Capdevieille, J. C. Guillemont and P. Ferrara, *Anal. Biochem*. **203**, 173 (1992).
18. P. Roepstorff, H. R. Nielsen, M. R. Larsen, S. Haebel, C. Jensen, L. Palm, T. N. Krogh, E. Mirgorodskaya, E. Nordhoff and M. Kussmann, in *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, 1996, p. 1357.
19. M. Kussmann, E. Nordhoff, H. R. Nielsen, S. Haebel, M. R. Larsen, L. Jakobsen, J. Gobom, E. Mirgorodskaya, A. K. Kristensen, L. Palm and P. Roepstorff, *J. Mass Spectrom*., 1997, in press.
20. E. Mørtz, T. Sarenava, S. Haebel, I. Julkunen and P. Roepstorff, *Electrophoresis* **17**, 925 (1996).
21. E. Mørtz, T. Sarenava, I. Julkunen and P. Roepstorff, *J. Mass Spectrom*. **31**, 1109 (1996).
22. C. Jensen, S. Haebel, S. O. Andersen and P. Roepstorff, *Int. J. Mass Spectrom. Ion Processes* in press.

