

Determination of primary structure and microheterogeneity of a β -amyloid plaque-specific antibody using high-performance LC–tandem mass spectrometry

Irina Perdivara · Leesa Deterding · Adrian Moise ·
Kenneth B. Tomer · Michael Przybylski

Abstract Using the bottom-up approach and liquid chromatography (LC) in combination with mass spectrometry, the primary structure and sequence microheterogeneity of a plaque-specific anti- β -amyloid (1–17) monoclonal antibody (clone 6E10) was characterized. This study describes the extent of structural information directly attainable by a high-performance LC–tandem mass spectrometric method in combination with both protein database searching and de novo sequence determination. Using trypsin and chymotrypsin for enzymatic digestion, 95% sequence coverage of the light chain and 82% sequence coverage of the heavy chain of the 6E10 antibody were obtained. The primary structure determination of a large number of peptides from the antibody variable regions was obtained through de novo interpretation of the data. In addition, N-terminal truncations of the heavy chain were identified as well as low levels of pyroglutamic acid formation. Surprisingly, pronounced sequence microheterogeneities were determined for the CDR 2 region of the light chain, indicating that changes at the protein level derived from somatic hypermutation of the Ig V_L genes in mature B-cells might contribute to unexpected structural diversity. Furthermore, the major glycoforms at the

conserved heavy chain N-glycosylation site, Asn-292, were determined to be core-fucosylated, biantennary, complex-type structures containing zero to two galactose residues.

Keywords β -Amyloid antibody · Plaque specific · De novo sequence determination · Microheterogeneity · CDR variations · LC-MS/MS

Introduction

The amyloid β -peptide (A β) has a central role in initiating neurodegeneration in Alzheimer's disease (AD) and AD-related diseases [1, 2], and has been an active target of a number of therapeutic approaches, including antibody immunotherapy. Therapeutically active antibodies produced by active immunization with protofibrillar A β (1–42) and fibrillar preparations have been found first to reduce the amyloid burden and to improve cognitive functions in TgCRND8 transgenic mouse models of AD [3]. The antibodies against A β in the immunized TgCRND8 mice recognize a short epitope sequence located at the N-terminus of A β (FRHDSGY), which has been identified by proteolytic epitope excision and high resolution mass spectrometry (FTICR-MS) [3]. Unfortunately, an initial therapeutic trial of active immunization of AD patients with A β (1–42) was discontinued because a few patients developed significant meningo-encephalitic cellular inflammatory reactions [4]. In spite of this initial drawback, A β -specific active as well as passive immunotherapy is presently developing as a promising therapeutic approach.

Recent advances in recombinant DNA and hybridoma technologies have enabled the production of a large number of recombinant proteins for experimental and clinical use. An increasing number of therapeutic approaches have relied

I. Perdivara · A. Moise · M. Przybylski (✉)
Laboratory of Analytical Chemistry and Biopolymer Structure
Analysis, Department of Chemistry, University of Konstanz,
78457 Konstanz, Germany
e-mail: Michael.Przybylski@uni-konstanz.de

I. Perdivara · L. Deterding · K. B. Tomer
Mass Spectrometry Group, Laboratory of Structural Biology,
National Institute of Health and Environmental Health
Sciences (NIEHS/NIH),
111 TW Alexander Drive,
Research Triangle Park, NC 27709, USA

on monoclonal antibodies for treatment of both malignant [5, 6] and autoimmune diseases [7, 8]. A mouse monoclonal antibody produced in hybridoma (clone 6E10) has been widely used in AD research. Terai and coworkers have used the 6E10 antibody to characterize the major β -amyloid species in senile plaques by affinity MS and immunochemistry [9], while Maddalena et al. used protein chip technology to capture the A β -peptides in cerebrospinal fluid (CSF) with 6E10 followed by mass spectrometric characterization of the captured peptides [10]. These results and recent clinical studies confirming the epitope specificity (ADPD Congress, 14–17 March 2007, Salzburg, Austria) emphasize the therapeutic potential of plaque-specific antibodies. Using epitope excision–mass spectrometry and alanine-scanning mutagenesis, we have previously shown that the mouse monoclonal 6E10 antibody, directed against β -amyloid (1–17), recognizes the same short epitope (FRHDSGY) at the N-terminus of A β , as did the antibodies resulting from active immunization of transgenic mice with A β (1–42) [11, 12]. Therefore, it is our hypothesis that primary structure details of this plaque-specific antibody (6E10) will provide a better understanding of the antigen recognition process at the molecular level and contribute to the development of more effective vaccines.

Structural studies of immunologically important molecules, especially of antibody–antigen recognition structures, have greatly expanded with the development of analytical methods, including NMR, x-ray crystallography, and mass spectrometry (MS) [13–15]. Among these, affinity-based approaches, in conjunction with high resolution and high sensitivity mass spectrometric methods, are becoming increasingly effective for identification of antigens directly from biological material [16]. The developments in soft ionization techniques [17, 18] and mass analyzers greatly expanded the potential of mass spectrometry to analyze large biomolecules, such as antibodies, and this has become a major tool for the characterization of recombinant protein pharmaceuticals [19]. Most recently, recombinant monoclonal antibodies have been analyzed by both top-down and bottom-up approaches in the determination of: (i) protein primary structure heterogeneities, (ii) verification of amino acid sequences predicted from the cDNA, and (iii) possible post-translational modifications such as glycosylation, oxidation, deamidation, C-terminal truncation, and formation of pyro-Glu at the N-terminus.

In the absence of crystal structure data, the complete characterization of the primary structure of monoclonal antibodies produced in hybridomas, for which no cDNA information is available, represents a considerable challenge, resulting primarily from the complexity of the immune system. IgG antibodies are tetrameric molecules (ca. 150 kDa), composed of two identical heavy (gamma) and two light (kappa or lambda) chains. The tertiary structure is stabilized by intrachain and interchain disulfide

bridges. The complementary determining regions (CDRs), three on each chain, are located in the hypervariable region and are critical for the strength and specificity of antigen recognition [20, 21]. The DNA encoding the antigen-combining portion of an antibody is assembled by gene rearrangement during the early phase of B lymphocyte development and includes the fusion of gene segments encoding for variable (V), diversity (D), and joining (J) domains [22]. The combinatorial diversity resulting from the joining of V–D–J segments is expanded by imprecisions at the recombination sites, which may arise from insertions or deletions of nontemplated nucleotides. During B-cell affinity maturation, somatic hypermutations (SHMs) and class switch recombination (CSR), both occurring in the rearranged DNA of the Ig genes and targeted by an enzyme called activation-induced deaminase (AID), account for additional diversity to provide antibodies with the highest affinity for a specific antigen [23]. As a result, it is difficult to characterize the protein sequence of a full-length antibody based only on peptide mapping and information available in protein databases.

Recently, Pham et al. reported the first de novo sequence analysis of a full length monoclonal antibody (for which the cDNA was not available) with neutralizing activity against the OX40L ligand, a member of the tumor necrosis factor superfamily involved in T-cell activation [24, 25]. In these studies, they used a combination of proteolytic and chemical digestions, Edman degradation, and mass spectrometry [26]. However, a substantial portion of the data was derived from Edman sequence analysis, while mass spectrometry was rather inefficient in providing unambiguous structural information about the antigen-binding regions.

In the present study, we characterize the primary structure of the 6E10 mouse antibody using predominantly liquid chromatography (LC) in combination with tandem mass spectrometry (MS/MS). Within the amino acid sequences characterized, N-terminal truncation of the heavy chain and a low extent of pyroglutamic acid formation were determined. Unexpected microheterogeneities in the CDR2 region of the light chain have been observed by MS/MS, suggesting that they may originate from somatic hypermutations in the V gene of the light chain upon fusion of the B-cell with the myeloma cell. A large part of the variable regions was determined by de novo MS/MS analysis.

Materials and methods

Materials

Mouse anti- β -amyloid monoclonal antibody (mAb1560, clone 6E10) was purchased from Millipore (Billerica, MA). Dithiothreitol, iodoacetamide, ammonium bicarbonate, and

96% formic acid were purchased from Sigma-Aldrich (St. Louis, MO). Sequencing-grade modified porcine trypsin was obtained from Promega (Madison, WI), and sequencing-grade bovine α -chymotrypsin was obtained from Roche (Penzberg, Germany). NuPage 4–12% Bis-Tris precast gels, sample and running buffers, and Coomassie SimplyBlue were from Invitrogen (Carlsbad, CA). Acetonitrile was purchased from Caledon Laboratories, Ltd. (Georgetown, Ontario). Purified water (17.8 M Ω) was obtained from an in-house Hydro Picopure 2 system. All sequencing-grade reagents used for Edman sequencing were purchased from Applied Biosystems (Darmstadt, Germany). All chemicals were used without further purification unless otherwise specified.

Methods

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Reduced and alkylated 6E10 antibody was separated on 4–12% Bis-Tris precast gels as follows: 40 μ L of the antibody stock solution (1 μ g μ L⁻¹ in PBS buffer) was incubated for 1 h at 90 °C with 20 μ L sample buffer containing 100 mM dithiothreitol (DTT). A solution of iodoacetamide (IAA) in water was then added to the mixture at a molar ratio of DTT/IAA 1:3 and the reaction was continued for an additional hour at room temperature. The reduced and alkylated antibody was loaded onto the gel (approximately 10 μ g lane⁻¹) and the heavy and light chains were separated. Electrophoresis was performed at 200 V and a maximum of 60 mA, for 1 h. The bands were stained overnight with Coomassie Simply Blue solution.

In-gel digestion

The protein bands corresponding to antibody heavy and light chains were manually cut and digested with trypsin for 8 h at 37 °C in an automated fashion with a Progest robotic digester (Genomic Solutions). In-gel digestion with α -chymotrypsin was performed manually using an enzyme to substrate ration of 1:30, based on the initial amount of reduced and alkylated antibody. Samples were lyophilized to dryness and resuspended in 0.1% formic acid.

Mass spectrometry

LC/MS analyses were performed using a Waters Q-ToF Premier mass spectrometer equipped with a nanoAcquity UPLC system (Waters, Milford, MA). Separations were performed on a 100 μ m \times 100 mm, Atlantis 3- μ m dC18 column (Waters, nanoAcquity), using a flow rate of 300 nL min⁻¹. A C18 trapping column (180 μ m \times 20 mm) with a 5- μ m particle size (Waters, nanoAcquity) was

positioned in-line with the analytical column. Briefly, a 2- μ L aliquot of the digest sample was injected, trapped onto the column, and peptides were eluted using a linear gradient of 98% solvent A (0.1% formic acid in water (v/v)) and 2% solvent B (0.1% formic acid in acetonitrile (v/v)) to 40% solvent B over 60 min. Mass spectrometer settings for the MS analyses were capillary voltage 3.2 kV, cone voltage 33 V, collision energy 8.0 V, and source temperature 80 °C. The mass spectra were acquired over the mass range 200–2,000 Da. MS/MS data were acquired in the data-dependent mode, using collision energies based on mass and charge state of the candidate ions. An external lock mass using a separate reference sprayer (LockSpray) and a solution of Glu-Fibrinogen peptide (300 fmol μ L⁻¹) in water/acetonitrile 80:20 (v/v) and 0.1% formic acid, with a mass of 785.8496 (2+), was used for calibration. Data analyses were performed using MassLynx 4.0 software (Waters, Milford, MA).

Database search

Prior to database searching, data were processed (including ions with S/N ratio greater than 3) using Mascot Distiller software (Matrix Science, UK) and saved as an *.mgf file. These data were searched against the NCBI protein database by means of the Mascot MS/MS Ion Search engine, using a precursor tolerance of 0.2 Da and an MS/MS tolerance of 0.1 Da. The sequences determined from the MS/MS data obtained for the identified peptides were validated manually.

N-terminal Edman sequencing

Following SDS-PAGE separation of 20 μ g of the reduced and alkylated antibody, the heavy and light chains were electroblotted (45 min, 250 mA) onto a Problott poly (vinylidene difluoride) (PVDF) membrane (Applied Biosystems) in CAPS buffer (10 mM, pH 11) containing 20% methanol with 10 mM thioglycolic acid. The membrane was washed with water for 15 min followed by washing with methanol for 5–10 s and then it was incubated with the staining solution (0.1% Coomassie Brilliant Blue R-250 in 40% methanol in water) until the bands became visible. The membrane was then destained in destaining solution (50% methanol in water, freshly prepared before use) until the background stain had decreased and the bands were clearly visible. The membrane was allowed to air-dry after which the bands of interest were cut, fully destained, and kept at 4 °C in Eppendorf tubes until use. For Edman sequencing, the PVDF membranes were wetted with 100% methanol and loaded into the sequencing cartridge. Automated amino acid sequence analysis was performed using an Applied Biosystems Model 494 Procise Sequencer attached to a Model 140C Micro-

gradient System and a 785A Programmable Absorbance Detector. All solvents and reagents used were of sequencing-grade purity (Applied Biosystems). Sequencing was performed using the standard pulsed liquid method, and data were analyzed with Model 610A Data Analysis software.

Results

Primary structure determination of the 6E10 heavy chain

In order to ascertain the amino acid sequence of the 6E10 monoclonal antibody, the heavy and the light chains were separated by SDS-PAGE following in-solution reduction and alkylation. In addition to the heavy and light chain bands, with an apparent molecular weight of 50 kDa and 25 kDa, respectively, a less intense band was observed at 75 kDa which was identified as albumin. Tryptic and chymotryptic digests of the heavy and light chains were analyzed separately by LC-MS/MS on a QToF Premier mass spectrometer.

The amino acid sequence determined by MS and MS/MS for the heavy chain of the 6E10 antibody (Fig. 1) resulted in an estimated sequence coverage of 82%. The identified tryptic and chymotryptic peptides were fit to a known homologous IgG heavy chain sequence frame from the NCBI database (accession number ABD73933). Typically, γ -heavy chains contain approximately 440 residues, divided into one variable region (V_H) and three constant regions (C_{H1} , C_{H2} , and C_{H3}) with each containing about 110 amino acids. The heavy chain of 6E10 belongs to the IgG1 isotype, based on the amino acid composition of

the constant region. The locations of the three complementary determining regions (CDRs) in the V_H domain, involved in the antigen recognition, were approximated using the Kabat rules [20, 21], and are highlighted with boxes in Fig. 1. A higher sequence coverage was obtained by comparison of the MS/MS data in the NCBI database entries for the heavy chain constant region than for the V_H region. This is not unexpected as the constant region is conserved among immunoglobulins, whereas the V_H region is the result of the site-specific recombination of the V-D-J genes and affinity maturation of the antibody [22, 27], leading to a greater extent of variability. The amino acid sequence for the constant region $C_H(219-434)$ was, for the most part, ascertained by comparison of the MS/MS data with that calculated for the database entries with the exception of those amino acids indicated in italics in Fig. 1. The resulting proteolytic peptides containing these residues were possibly too small to be retained on the C18 column or yielded singly charged ions, which would not have been selected for MS/MS analysis based on the predetermined criteria for data-dependent acquisition. Lower sequence coverage for the constant region $C_H(149-218)$ was obtained; however, as the constant region is conserved among immunoglobulins, it is reasonable to assume that the undetermined amino acids are identical to those reported in the database for the IgG₁ isotype. Collectively, LC/MS analyses of the trypsin and chymotrypsin digestions provided complementary information; first, by disclosing antibody regions observed with only one of the two enzymes; and secondly, by generating overlapping peptides, which confirmed the correct succession of the amino acids in the sequence frame.

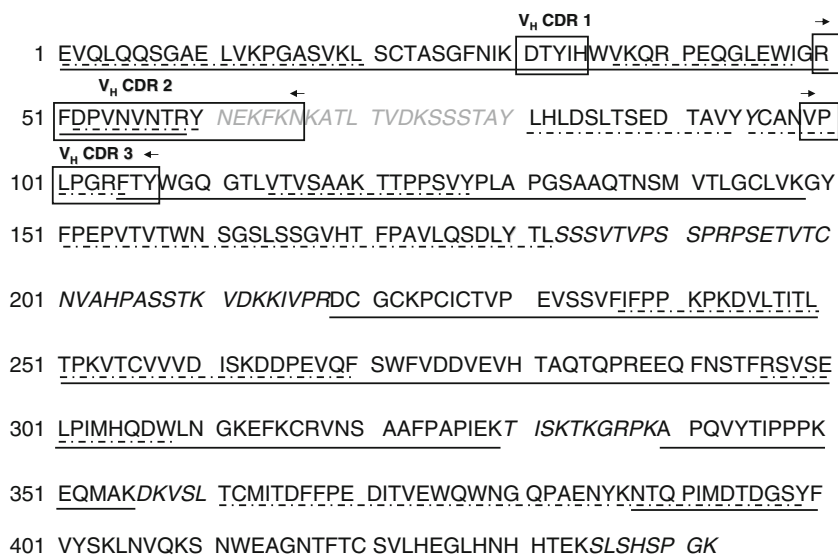


Fig. 1 Amino acid sequence determined for the 6E10 antibody heavy chain. The CDRs are highlighted with *black boxes* and *arrows*; *solid lines* regions identified using trypsin; *dotted-dashed lines* regions

were identified using chymotrypsin. The amino acids in *italics* were fitted in the heavy chain sequence based on homology with other antibody sequences (database accession number ABD73933)

The specificity of an antibody for an antigen is dictated by the amino acids in the complementary determining regions (CDRs). The V_H CDR3 represents a signature of each immunoglobulin, as it is generated somatically, in the process of B-cell maturation. Unless a specific antibody was previously sequenced and the information deposited into the database, it is not possible to characterize the variable region using the database searching approach. In the case of the 6E10 heavy chain, only the variable region $V_H(1-50)$ spanning the CDR1 (residues 31–35) could be assigned by searching the MS/MS data against the NCBI nr database. The fragment ion spectra (Electronic supplementary material) for peptides containing the regions $V_H(1-19)$, $V_H(20-38)$, and $V_H(39-50)$ are consistent with the sequence indicated in Fig. 1. According to the Kabat definition, the V_H CDR1 is located nine residues after the first cysteine of the variable region ($^{22}CXXXXXXXX$) and is followed always by a tryptophane [20, 21]. In the case of the 6E10 heavy chain, the V_H CDR1 spans the residues $V_H(31-35)$ and contains five amino acids (Fig. 1).

From the LC/MS analysis of the tryptic digest, other ions for which abundant fragmentation was obtained were de novo sequenced assuming trypsin specificity. As an example, the MS/MS spectrum of the ion of m/z 531.278 is shown in Fig. 2. From these data, the amino acid sequence of the peptide was determined to be FDPVNVNTR based on the nearly complete y ion series (except for y_8) observed in the spectrum. The proposed amino acid sequence was confirmed by the observation of an MS/MS spectrum of an overlapping chymotryptic peptide (DPVNVNTRY) (data not shown). This sequence informa-

tion was inserted into the heavy chain sequence frame based on homology with another heavy chain from a mouse antibody in which the region V_H (51–60) has an amino acid sequence similar to that of the peptide FDPVNVNTRY (NCBI accession number AAA38193). In the Kabat definition, the V_H CDR2 begins 15 residues after the end of V_H CDR1, usually being preceded by the sequence motif LEWIG (Fig. 1). The typical length is 16–19 residues [21]. Despite the efforts to obtain complete sequence coverage of the variable region, it was not possible to elucidate the amino acid composition of the region $V_H(61-80)$ using LC-MS/MS, assuming a length of 19 residues.

The sequence information for $V_H(99-107)$ CDR3 was determined from the MS/MS spectra of the doubly charged precursor ions of m/z 565.811 (Fig. 3) observed in the chymotryptic digest and m/z 864.952 observed in the trypsin digest (n). The MS/MS fragmentation of the $(M+2H)^{2+}$ ion of 565.811 is consistent with the peptide sequence CANVPLPGRF containing an alkylated cysteine, as determined by the observation of a nearly complete series of y ions (except for y_1). This amino acid sequence was not found in the database, but is similar to other heavy chain CDR3 amino acid sequences (accession number AA G25671). This information suggests that this peptide can be assigned as the CDR3 peptide. Usually, the V_H CDR3 begins three residues after a cysteine (a typical sequence motif is CAR) and is followed by the motif WGXXG (X can be any amino acid), while it may contain from three up to 25 residues [21]. According to this Kabat definition, the residues before the V_H CDR3 of the 6E10 are CAN, those

Fig. 2 Fragment mass spectrum of the precursor ion of m/z 531.278 ($2+$) from the heavy chain tryptic digest assigned by de novo interpretation to the CDR2 peptide $V_H(51-59)$, indicated at the top right. Internal fragments are indicated between square brackets. The asterisk denotes loss of ammonia

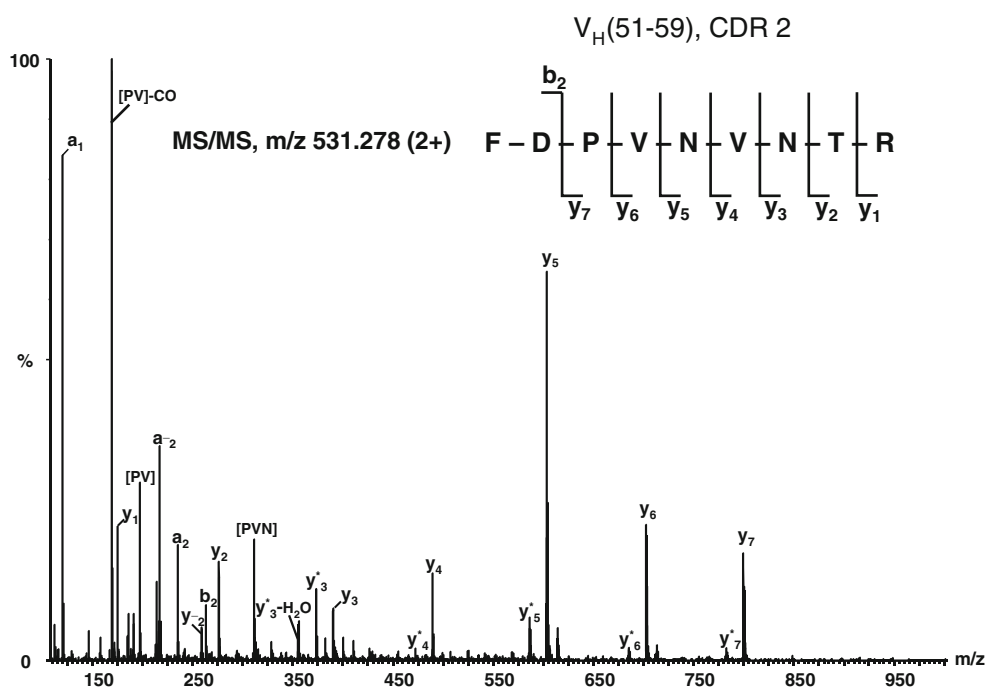
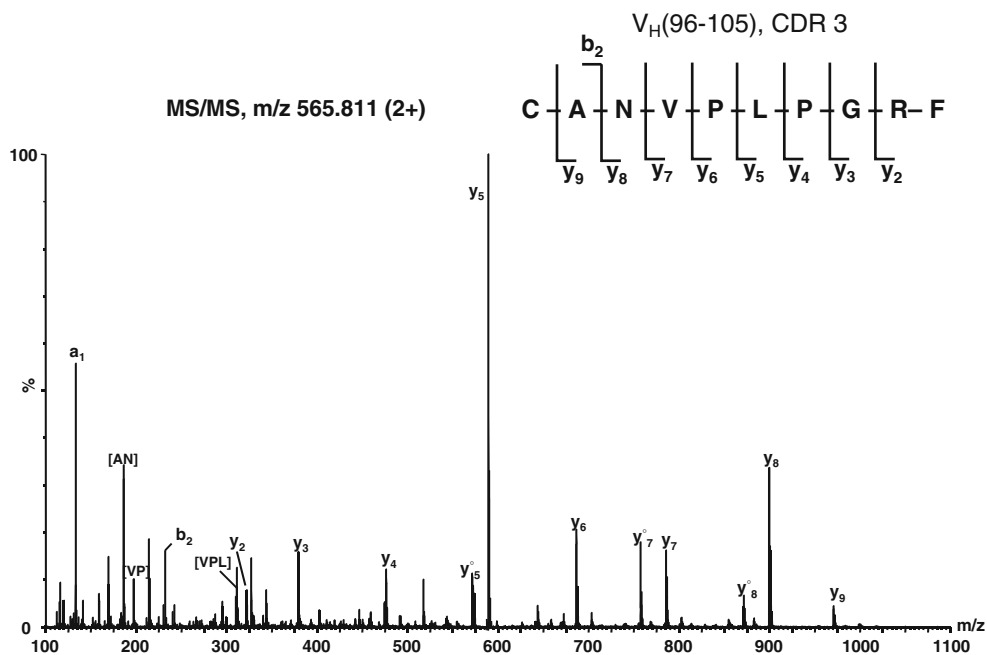


Fig. 3 Fragment mass spectrum of the precursor ion of m/z 565.811 (2+) from the heavy chain chymotryptic digest assigned by de novo interpretation to the CDR3 peptide $V_H(98-107)$, indicated at the *top right*. The cysteine was observed in alkylated form. Internal fragments are indicated between *square brackets*. \circ denotes loss of water

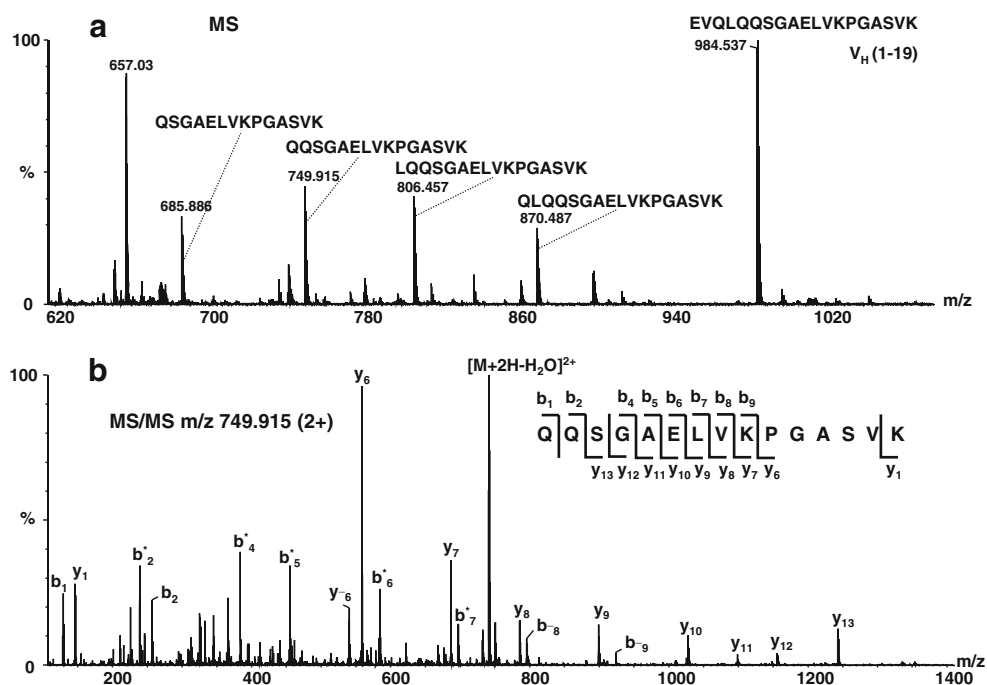


after are WGQG, and the determined length is of nine amino acids (Fig. 1).

In the tryptic digest of one lot of the 6E10 antibody, heterogeneity at the N-terminus of the heavy chain was observed. In addition to the full length N-terminal tryptic peptide (${}^1\text{EVQLQQSGAELVKPGASVK}^{19}$ m/z 984.537 (2+)), many other doubly charged ions were observed which correspond in mass to the successive loss of two (m/z 870.487), three (m/z 806.457), four (m/z 749.915), five (m/z 685.886), and six amino acids (m/z 621.857) from the heavy chain N-terminus. The averaged mass spectrum

comprising these ions and the peptides assigned to each ion are shown in Fig. 4a. The MS/MS spectrum for each of these ions is consistent with the assigned peptide structures. One example is illustrated in Fig. 4b and shows the fragment ion spectrum of the precursor ion of m/z 749.915 (2+). The specific product ions (b and y series) and the mass of the precursor ion enabled the unambiguous identification of the truncated peptide, ${}^5\text{QQSGAELVKPGASVK}^{19}$, which arises from the loss of the first four amino acids from the N-terminus of the heavy chain. These data suggest that, in addition to the well-documented C-terminal lysine clipping

Fig. 4 a Positive ion MS spectrum showing the doubly charged ions and the corresponding tryptic peptides assigned to each mass, which indicate N-terminal truncation of the heavy chain. **b** MS/MS of the precursor ion of m/z 749.915 (2+), confirming the indicated peptide sequence. The *asterisk* denotes loss of ammonia



[28], antibodies may possibly undergo N-terminal degradation as well. Since this phenomenon was observed only in one out of three antibody lots, which was not analyzed by Edman sequencing, the observed N-terminal truncations may arise from nonenzymatic degradation pathways which may be relevant if such a recombinant antibody were investigated as a potential therapeutic agent.

In addition to N-terminal truncation it was observed that 7% of the heavy chain had undergone terminal pyroglutamic acid (pyro-Glu) formation, based on the relative abundances of the ions observed for the native (*EVQLQQSGAELVKP GASVK*, m/z 984.537, 2+) and modified peptide (*pEVQLQQSGAELVKPGASVK*, m/z 975.532, 2+) (data not shown). The observation that the tryptic peptide containing the pyro-Glu residue elutes slightly later (39.55 min) than the native peptide (35.21 min) is in agreement with data published by other groups [29–31]. Pyro-Glu formation in antibodies and its identification by MS have been described previously [29–31] and is believed to be an artifact that arises via nonenzymatic pathways after prolonged storage of the samples at high temperature (45 °C) and low pH conditions (pH 4). In fact, pyroglutamate can form in almost all proteins containing glutamic acid at the N-terminus.

Primary structure determination of the 6E10 light chain

The amino acid sequence for the light chain of 6E10 is presented in Fig. 5, for which the calculated sequence coverage was 95%. Analogous to our approach to sequence identification for the heavy chain, the peptides from the 6E10 light chain were inserted into a template light chain sequence from the database (accession number AA A39162). Light chains usually contain 220 residues, comprised of the N-terminal variable region (V_L , approximately 110 amino acids) and a C-terminal κ or λ constant region (C_L , 110 amino acids). The complementary deter-

mining regions (CDRs) were matched using the Kabat rules [20, 21] and their positions are marked with black boxes in Fig. 5. The constant region C_H (113–220) was completely determined from MS/MS data, and showed that the light chain of 6E10 is a κ chain. A peptide from the variable region, corresponding to the region V_L (103–112) remained unelucidated, despite the use of alternative enzymes. The abundance of the ion m/z 628.803 (2+), assigned to the peptide V_L (69–81), was too low and provided insufficient MS/MS fragmentation; however, the observed C-terminal (y_1 , y_2) and N-terminal ions (b_1 , b_2) as well as the mass accuracy of 3 ppm of the experimental mass are consistent with the identification. This peptide is a feasible candidate for the region V_L (69–81), as the four amino acids at the C-terminus, TLTI (Fig. 5, and Table S2 in Electronic supplementary material), are overlapping with those observed at the N-terminus in the peptide V_L (78–89). MS analyses of the trypsin digest provided 100% sequence coverage of the constant region. Both enzymes provided complementary information for the variable region, indicating that the region V_L (68–102), determined from the MS/MS data of the chymotryptic digest, lacks potential trypsin cleavage sites. Upon analyzing the sequence of a large number of antibody light chains in the database, it was observed that the V_L (103–110) region may contain a considerable number of potential chymotrypsin cleavage sites and this could explain why no peptides of comparable length were observed in our LC/MS experiments. In addition, a single potential trypsin cleavage site is located at the C-terminal of V_L (103–110) (residue 109, Fig. 5) and this could explain the lack of the short region (110–113) in the MS. The amino acid residues spanning the V_L (25–36) CDR1 were not identified from the database search of the MS/MS data, suggesting that the corresponding protein sequence may not be contained in the NCBI nr protein database and/or the ion corresponding to this region was not selected for MS/MS analyses. Manual interpretation of

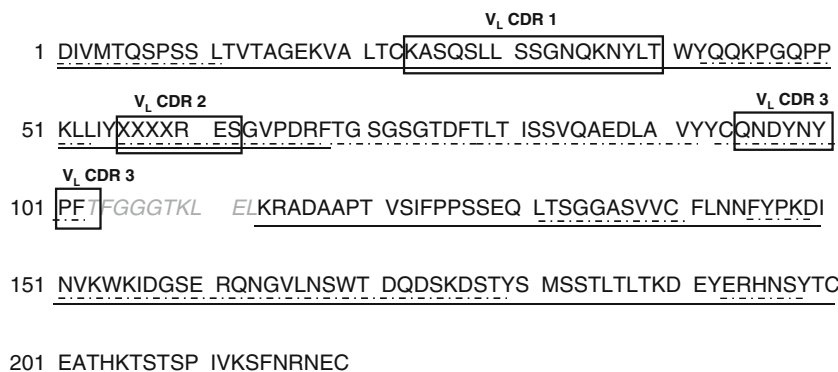
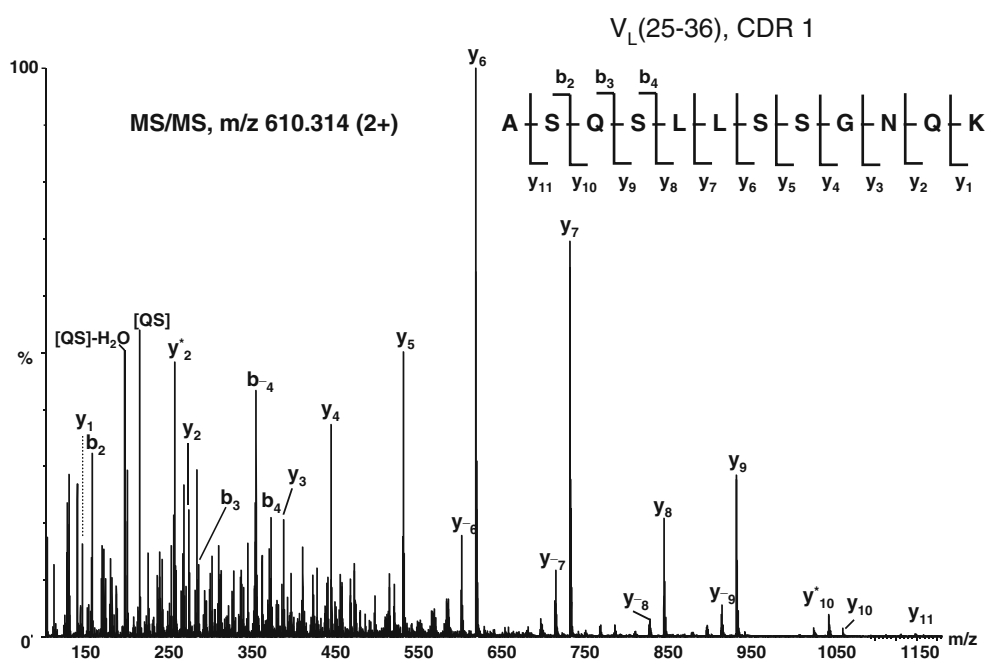


Fig. 5 Amino acid sequence determined for the 6E10 antibody light chain. The amino acids in *italics* were fitted in the light chain sequence based on homology with other antibody light chain sequences (database accession number AAA39162)

The amino acids in *italics* were fitted in the light chain sequence based on homology with other antibody light chain sequences (database accession number AAA39162)

Fig. 6 Fragment mass spectrum of the precursor ion of m/z 610.314 (2+) from the light chain tryptic digest assigned by de novo interpretation to the CDR1 peptide $V_L(25-36)$, indicated at the *top right*. Internal fragments are indicated between *square brackets*. The *asterisk* denotes loss of ammonia



the MS/MS data of the ion of m/z 610.314 (2+) (Fig. 6), however, revealed an amino acid sequence of a tryptic peptide resembling the CDR1 containing region $V_L(25-36)$ based on its homology to other mouse light chains (NCBI accession number ABK64007). These data enabled the determination of the peptide sequence ASQS(L/I)(L/I)SSGNQK based on the observation of a complete y ion series and a partial b ion series. Because Ile and Leu have the same mass, these amino acids could not be discriminated from the MS/MS spectrum. The amino acid sequence corresponding to the $V_L(97-102)$ CDR 3 was ascertained from the database search using the MS/MS spectrum of the $(M+2H)^{2+}$ ion of m/z 611.240, which is consistent with the peptide sequence CQNDYNYPF, where the Asn99 residue was assigned as deamidated and the cysteine at the N-terminus in alkylated form. According to the Kabat rules [21], the V_L CDR1 may have a length between 10 and 17 residues and is preceded by the first cysteine of the variable

region (Cys23) and followed by the sequence motif WYQ. The V_L CDR3 starts also after a cysteine and is followed by the residues FGXG, where X can be any amino acid. The V_L CDR1 of the 6E10 antibody contains the maximal number of amino acids according to the Kabat definition (see Fig. 5). The V_L CDR3 fits the definition as well; however, the identity of the amino acid at the C-terminus could not be confirmed from the MS/MS.

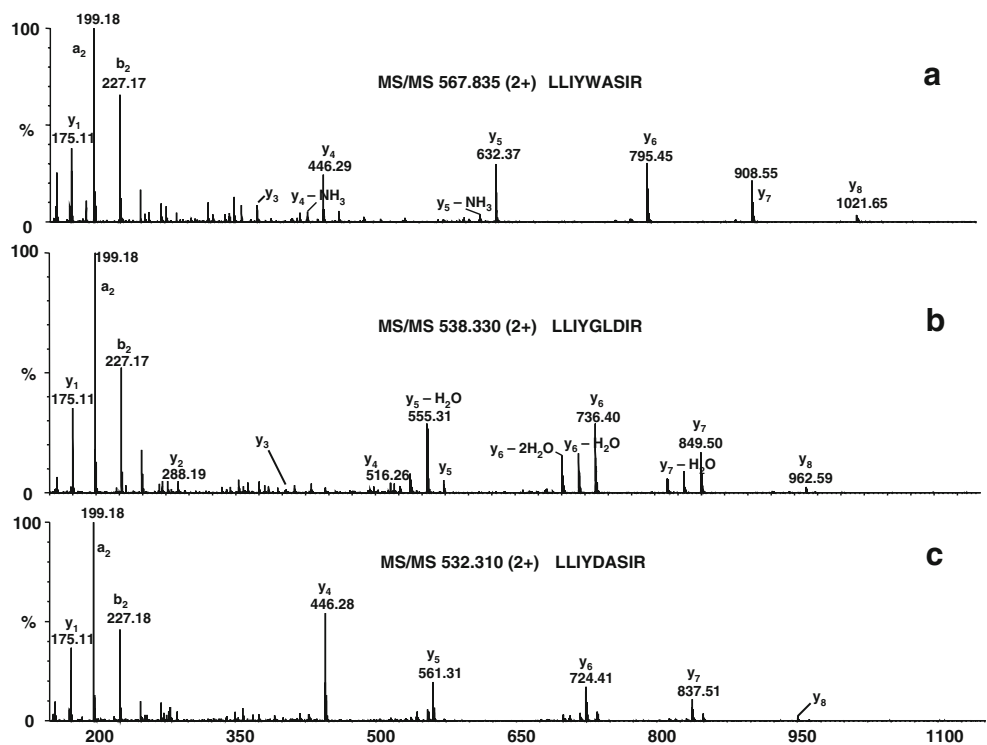
Interestingly, the MS analyses of the digests of the light chain of 6E10 revealed microheterogeneity in the second complementary determining region CDR2 corresponding to $V_L(52-60)$. Eight different tryptic peptides were observed by LC-MS/MS which contain the amino acid motif LLIYXXXXR, where X indicates the possible variations (Table 1). Representative MS/MS spectra of three of these ions are shown in Fig. 7. Because these peptides are identical at their N-termini (Leu-Leu-Ile-Tyr-), their fragmentation patterns are very similar with the most abundant

Table 1 Amino acid sequence of peptides from antibody light chain variable region V_L (52–60), identified in the LC-MS/MS analysis of the light chain tryptic mixture

Peptide no.	Peptide sequence ^a LLIYXXXXR	Observed mass (Da)	Calculated mass (Da)	Mass accuracy (ppm)
1	LLIYDASR	1,062.6046	1,062.6073	3
2	LLIYGLDR	1,074.6450	1,074.6437	1
3	LLIYWASTR	1,121.6174	1,121.6233	5
4	LLIYWASTR	1,133.6540	1,133.6597	5
5	LLIYWSSSTR	1,137.6058	1,137.6182	11
6	LLIYWVSTR	1,149.6492	1,149.6546	5
7	LLIYKVSYR	1,153.6800	1,153.6859	9
8	LLIYWDSR	1,165.6168	1,165.6132	3

^a Peptide sequences identified by database search performed with the MS/MS data; X represents the observed amino acid variations which are italicized within each peptide

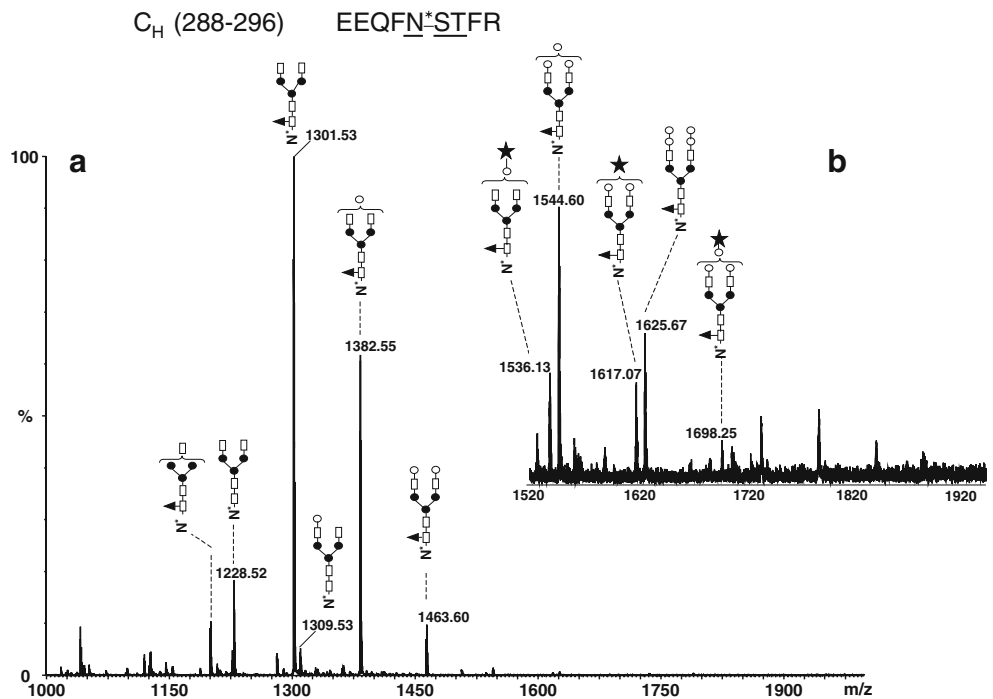
Fig. 7 Representative product ion spectra of the precursor ions: **a** m/z 567.835 (2+), **b** m/z 538.330 (2+), and **c** 532.310 (2+), confirming the amino acid sequence of three peptides spanning the light chain region $V_L(52-60)$



product ions belonging to the y series of ions. Two diagnostic product ions observed in the CID spectra of all these ions, m/z 199.175 (a_2) and m/z 227.176 (b_2), are specific for the Leu–Leu N-terminal sequence. The observation of these ions in the MS/MS data was used to identify as many variants as possible of different precursor ions. Ions which produced these fragment ions upon MS/MS analyses are shown in Table 1. Based on manual interro-

gation of the MS/MS data and/or a subsequent search of the NCBI nr database (taxonomy: all species), the peptide sequences were assigned. Of the determined amino acid sequence, the peptide sequence LLIYDASIR was found in human antibody light chains, but not in mouse antibodies. The same set of tryptic peptides was found in all three analyzed antibody batches, suggesting that this observed microheterogeneity was not an isolated random event. We

Fig. 8 Positive ion nHPLC/ESI/MS of the F_c glycopeptides EEQFN*STFR from the mouse monoclonal antibody 6E10. **a** MS spectrum averaged over the chromatographic window where glycopeptides eluted (30.6 min, average of 15 full MS scans). The glycan structures of the most abundant glycoforms are indicated above each ion. All ions are doubly charged. **b** The mass range m/z 1,520–1,920, showing low abundant di-, tri-, and tetragalactosylated glycoforms: empty square N -acetylglucosamine, black triangle fucose, black circle mannose, empty circle galactose, black star N -glycolylneuraminic acid



hypothesize that this structural diversity may be due to somatic hypermutations in Ig V_L genes observed at the protein level.

N-terminal Edman sequencing

The N-terminal sequences and homogeneity of the 6E10 antibody heavy and light chains were determined by Edman sequencing using a standard PVDF Edman sequencing protocol. Overall, the 22 N-terminal residues for the light chain and the 24 N-terminal residues for the heavy chain were unambiguously identified and these data were consistent with the sequence information derived from the MS/MS analysis. The N-terminal amino acids for both the heavy and light chain were determined to be unmodified, suggesting that no significant pyroglutamic acid formation had occurred. These data demonstrated the N-terminal homogeneity of the anti-A β (1–17) monoclonal antibody.

Characterization of the N-linked glycosylation of the heavy chain

The major glycoforms observed at the conserved N-glycosylation site of the heavy chain constant region Asn292, observed in the glycopeptide EEQFNSTFR, are shown in Fig. 8. The glycopeptides eluted earlier from the reversed-phase column than the unmodified peptides and were easily identified by LC-MS based on the mass interval of 81 amu between doubly charged glycopeptide ions, which is indicative of the presence of terminal hexose residues. A detailed analysis of the glycosylation pattern of the 6E10 antibody compared with other A β -antibodies is presented elsewhere (I. Perdivara et al., manuscript in preparation). The most abundant glycan populations decorating the heavy chain at Asn292 were found to be core-fucosylated, biantennary, complex-type structures containing zero to two galactose residues. Low amounts of core-fucosylated tri- and tetragalactosyl glycoforms, with the third and the fourth galactose residues probably α -linked to the GlcNAc- β (1–4)Gal moiety, as well as minimal amounts of N-glycolylneuraminic acid were observed. The glycosylation pattern observed for the 6E10 antibody is consistent with the structures previously described for various recombinant monoclonal antibodies [32].

Discussion

Characterization of the amino acid sequence and structural integrity of recombinant or hybridoma-derived monoclonal antibodies is becoming increasingly important, e.g., for the development of antibodies in the biotechnology and pharmaceutical industry. Modifications at the protein level

resulting from the use of a specific expression system, or after prolonged storage of an antibody, might alter its antigen specificity, effector functions, and/or in vivo half-life, factors which are of concern in developing a potential therapeutic agent. In addition, antigen–antibody crystal structures are only rarely available and thus the knowledge of the antigen-binding regions of an antibody essentially contribute to a better understanding of their interactions.

Primary structure determinations of recombinant immunoglobulins can be accomplished using mass spectrometric top-down or bottom-up approaches, because of the major advantage that the expected protein sequence can be derived from the cDNA information [29, 33]. However, the potential changes in protein sequences mentioned above can greatly complicate the data interpretation. For mass spectrometry, a particular challenge is the complete characterization of an antibody primary structure for which no cDNA information is available, i.e., when the antibody was produced in hybridoma. When genomic material can be obtained, the protein sequence can be derived from nucleotide sequencing [34]. However, in many cases this information is not available, e.g., if the antibody was obtained from mouse ascites.

In the present work, we have used liquid chromatography coupled to mass spectrometry in an effort to elucidate the amino acid sequence of an anti- β -amyloid (1–17) mouse monoclonal antibody. Of particular interest were the assemblies of amino acids in the antibody complementary determining regions, as they represent the key structures of the antigen–antibody recognition. The cDNA information for the commercially available 6E10 antibody was not provided. The success of such an experiment may depend on the MS methodology available, but also on the original amino acid sequence of the antibody.

The structural diversity of immunoglobulins derived from site-specific recombination and affinity maturation make the identification of a complete antibody sequence by peptide mass mapping and database searching extremely difficult. In addition, variable and constant regions of heavy and light chains sometimes have different database entry numbers. De novo sequencing based on MS/MS data is most often used to determine the amino acid sequences of peptides whose MS/MS data do not match known sequences. Therefore, we have undertaken characterization of the amino acid sequence of mAb 6E10 by MS/MS. High specificity of proteases, chromatographic resolution, and quality of MS/MS data were prerequisites for this approach.

Based on sequence homology, the peptides identified by database search of the MS/MS data were fit to known sequence frames for antibody heavy and light chains. For the light chain of 6E10, a coverage of 95% was achieved, while for the heavy chain the coverage was 82%. The incomplete assignment might be related to

several issues: (i) the protein sequence lacks potential cleavage sites for a specific enzyme, leading to formation of long peptides; or too many cleavage sites, leading to singly charged fragment ions which can be obscured by the chemical noise in the lower m/z range of the mass spectrometer; (ii) large peptides may be extracted from the gel after enzymatic digestion with poor efficiency; (iii) the division of ion abundance between multiple charge states reduces the abundance of any one charge state to the point that interpretable MS/MS data is not obtained; (iv) the peptide sequences generated by enzymatic digestion may not be contained in protein databases, and in this case sufficient ionization of the peptide, generation of specific backbone cleavages in CID, and high-quality MS/MS spectra are important prerequisites for a successful de novo interpretation.

Although it was not possible to obtain 100% sequence coverage for the 6E10 antibody, the determination of partial or complete sequence of the complementary determining regions was successful. Four of the six CDRs (V_H -CDR1, V_H -CDR3, V_L -CDR1, and V_L -CDR2) were completely elucidated, while only one amino acid remained unelucidated of the V_L -CDR3. Only partial sequence information could be disclosed for the CDR2 of the heavy chain. The peptides determined for heavy chain CDR1 and light chain CDR2 and CDR3 regions were identified directly by searching the MS/MS data against the NCBI protein database, while those spanning the heavy chain CDR2 and CDR3 and the light chain CDR1 were determined by means of manual de novo interpretation of the data and were not found in any other antibody sequence in the protein database.

Although the N-terminal Edman sequencing confirms the monoclonal nature of the 6E10 antibody for the first 25–30 amino acids and did not identify truncations, extended sequence microheterogeneities were determined for the V_L (52–60) region, comprising the partial CDR2 between the residues (56–60) (see Table 1). These were the only sequence microheterogeneities observed for both heavy and light chains, but it is possible that this phenomenon also occurs at other amino acid positions from the variable region. The microheterogeneity found in the monoclonal antibody light chain may result from point mutations which occur at the DNA level, termed somatic hypermutation (SHM) (reviewed in [23]). It is accepted that SHMs occur in mature B-cells during affinity maturation in secondary lymphoid organs (germinal centers) [35], and whether this mechanism might still occur in a later phase of B-cells, for example in hybridomas, is a point of controversy. Our data indicates that SHM might occasionally take place in hybridoma cells, but this hypothesis needs to be investigated in more detail. To our knowledge, this is the first mass spectrometric evidence of potential modifications at the DNA level directly from isolated protein data.

Conclusion

In conclusion, we demonstrate the utility of liquid chromatography–tandem mass spectrometry to successfully address the question of immunoglobulin sequence determination, and to reveal unexpected features of the primary structure of the 6E10 anti-A β -antibody.

Acknowledgements This research was supported in part by grants from the Deutsche Forschungsgemeinschaft, Bonn, Germany, and the University of Konstanz (M.P.), and the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (K.T.). The authors want to thank Dr. Marilyn Diaz for informative discussions throughout the project, as well as for the critical review of the manuscript.

References

- Schenk DB, Seubert P, Grundman M, Black R (2005) *Neurodegener Dis* 2(5):255–260
- Selkoe DJ (2001) *Physiol Rev* 81(2):741–766
- McLaurin J, Cecal R, Kierstead ME, Tian X, Phinney AL, Manea M, French JE, Lambermon MH, Darabie AA, Brown ME, Janus C, Chishti MA, Horne P, Westaway D, Fraser PE, Mount HT, Przybylski M, St George-Hyslop P (2002) *Nature Med* 8(11):1263–1269
- Check E (2002) *Nature* 415(6871):462
- Goldenberg MM (1999) *Clin Ther* 21(2):309–318
- Grillo-Lopez AJ, White CA, Varns C, Shen D, Wei A, McClure A, Dallaire BK (1999) *Semin Oncol* 26(5 Suppl 14):66–73
- Chung KF (2002) *Curr Opin Investig Drugs* 3(8):1157–1160
- Kohler G, Milstein C (1975) *Nature* 256(5517):495–497
- Terai K, Iwai A, Kawabata S, Tasaki Y, Watanabe T, Miyata K, Yamaguchi T (2001) *Neuroscience* 104(2):299–310
- Maddalena AS, Papassotiropoulos A, Gonzalez-Agosti C, Signorell A, Hegi T, Pasch T, Nitsch RM, Hock C (2004) *Neurodegener Dis* 1(4–5):231–235
- Tian X, Cecal R, McLaurin J, Manea M, Stefanescu R, Grau S, Harnasch M, Amir S, Ehrmann M, St George-Hyslop P, Kohlmann M, Przybylski M (2005) *Eur J Mass Spectrom* (Chichester, Eng) 11(5):547–556
- Przybylski M, McLaurin J, Cecal R-E, Kierstead ME, Tian X, Janus C, Horne P, Westaway D, Fraser PE, St George Hyslop P (2002) *J Peptide Sci* 8:73–75
- Van Regenmortel MH (1988) *Mol Immunol* 25(6):565–567
- Amit AG, Mariuzza RA, Phillips SE, Poljak RJ (1986) *Science* 233(4765):747–753
- Paterson Y, Englander SW, Roder H (1990) *Science* 249(4970):755–759
- Macht M, Fiedler W, Kurzinger K, Przybylski M (1996) *Biochemistry* 35(49):15633–15639
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) *Science* 246(4926):64–71
- Hillenkamp F, Karas M, Beavis RC, Chait BT (1991) *Anal Chem* 63(24):1193A–1203A
- Srebalus Barnes CA, Lim A (2007) *Mass Spectrom Rev* 26(3):370–388
- Kabat EA, Wu TT (1991) *J Immunol* 147(5):1709–1719
- Kabat EA, Wu TT, Reid-Miller M, Perry HM, Gottesman KS (1987) *Sequences of proteins of immunological interest*, 4th edn. Department of Health and Human Services, Bethesda

22. Tonegawa S (1983) *Nature* 302(5909):575–581
23. Di Noia JM, Neuberger MS (2007) *Annu Rev Biochem*
24. Higgins LM, McDonald SA, Whittle N, Crockett N, Shields JG, MacDonald TT (1999) *J Immunol* 162(1):486–493
25. Malmstrom V, Shipton D, Singh B, Al-Shamkhani A, Puklavec MJ, Barclay AN, Powrie F (2001) *J Immunol* 166(11):6972–6981
26. Pham V, Henzel WJ, Arnott D, Hymowitz S, Sandoval WN, Truong BT, Lowman H, Lill JR (2006) *Anal Biochem* 352(1):77–86
27. Johnson A (2002) *Molecular biology of the cell*, 4th edn. Garland Science, New York
28. Harris RJ (1995) *J Chromatogr A* 705(1):129–134
29. Beck A, Bussat MC, Zorn N, Robillard V, Klinguer-Hamour C, Chenu S, Goetsch L, Corvaia N, Van Dorsselaer A, Haeuw JF (2005) *J Chromatogr B Anal Technol Biomed Life Sci* 819(2):203–218
30. Chelius D, Jing K, Lueras A, Rehder DS, Dillon TM, Vizek A, Rajan RS, Li T, Treuheit MJ, Bondarenko PV (2006) *Anal Chem* 78(7):2370–2376
31. Yu L, Remmele RL Jr, He B (2006) *Rapid Commun Mass Spectrom* 20(24):3674–3680
32. Sheeley DM, Merrill BM, Taylor LC (1997) *Anal Biochem* 247(1):102–110
33. Hirayama K, Yuji R, Yamada N, Kato K, Arata Y, Shimada I (1998) *Anal Chem* 70(13):2718–2725
34. Shendure J, Mitra RD, Varma C, Church GM (2004) *Nat Rev Genet* 5(5):335–344
35. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T (1999) *J Biol Chem* 274(26):18470–18476