ELSEVIER

Journal of Chromatography A, 763 (1997) 91–103

JOURNAL OF CHROMATOGRAPHY A

Complete structure elucidation of a globular protein by particle beam liquid chromatography–Fourier transform infrared spectrometry and electrospray liquid chromatography–mass spectrometry
Sequence and conformation of β-lactoglobulin

Vincent E. Turula[a,1], Randall T. Bishop[a], Robert D. Ricker[b], James A. de Haseth[a,*]

[a]*Department of Chemistry, University of Georgia, Athens, GA 30602-2556, USA*
[b]*Rockland Technologies, Inc., 538 First State Boulevard, Newport, DE 19804, USA*

Abstract

The advantages to the use of both mass spectrometry (MS) and Fourier transform infrared spectrometry (FT-IR) in combination for the structural characterization of the tryptic digest of a model globular protein is demonstrated. HPLC has been interfaced to both spectroscopic techniques and has provided a high degree of structural detail for the target protein. β-Lactoglobulins A and B were digested with trypsin and chromatographed with narrow-bore, reversed-phase HPLC. As determined by LC–FT-IR spectrometry, the conformation of each form of intact β-lactoglobulin was randomized upon elution. The particle beam and the electrospray LC–MS interfaces enabled the acquisition of spectra for nanogram injection quantities. The primary structures were determined from the accurate molecular mass determinations of the digest fragments. Infrared spectra confirmed the presence of some amino acid functionalities.

Keywords: Detection, LC; Structure analysis; Proteins; Lactoglobulin

1. Introduction

Procedures for the determination of protein primary covalent structures involve the generation of a peptide map. This entails chemical or enzymatic cleavage of a target protein, separation of the peptide fragments produced by these treatments, and the sequential identification of the composite amino acids [1]. Identification of constituent amino acids can be made from several chemical, radiochemical or immunological techniques. Protocols that make use of gel filtration and sodium dodecyl sulfate–electrophoresis methods for the separation often involve a secondary detection scheme to obtain more exacting structural detail of the fragments (e.g. antibody immuno-staining as a secondary detection for polyacrylamide gel electrophoretic separations). The peptide map can also be generated from a reversed-phase HPLC separation [2]. The presence of certain amino acids can be confirmed with standard UV and fluorescence HPLC flow-cell detection (e.g., tyrosine at 280 nm or fluorescence at 300 nm) [3]; however, apart from the few residues with unique UV absorption behaviour, little structural information is con-

*Corresponding author.
[1]Present Address: Amvax Inc., 12103 Indian Creek Court, Beltville, MD 20705, USA.

PII S0021-9673(96)00960-0

veyed by UV alone. When HPLC is interfaced to spectroscopic detectors, such as infrared (IR) spectrometry, and mass spectrometry (MS), additional structural detail for fragment composition can be determined. By the combination of HPLC to both FT-IR spectrometry and MS, the digest peptide map can be unequivocally characterized.

The particle beam LC–FT-IR interface can be coupled to protein and peptide HPLC experiments for further elucidation of structure [4]. IR absorption spectra are not only sensitive to specific amino acid functional groups [5] but also to secondary structure [6]. Secondary structure (e.g. α-helix, β-sheet, random coil, turn) information is contained in the IR amide I, II and III absorption bands that arise from delocalized backbone peptide vibrations [7]. The majority of the energy distribution of the amide I band is associated with the C=O stretch and the bond character of the amide C=O is perturbed by the hydrogen-bonding of secondary structures. Therefore, the amide I band is sensitive to secondary structure and is also recognized as the most informative band for structure correlation. The protein amide I region is intrinsically broad because it consists of underlying component amide carbonyl bands that are characteristic for specific conformations. Particle beam LC–FT-IR spectrometry is an 'off-line' hyphenated technique, that is, after separation eluates are isolated from the solvent and then examined for structural information by spectroscopic interrogation. The particle beam functions by the conversion of an HPLC effluent stream into an aerosol. The more volatile mobile phase is rapidly evaporated from the eluate. Low temperature and pressure within the interface restrict conformational motion while the eluate is desolvated. Thus, when the eluates are proteins or other macromolecules, they never attain the activation energy necessary for either local or global unfolding during collection. Stripped of surrounding solvent, the protein or fragment is deposited onto an IR-transparent substrate. Upon completion of the collection, the substrate is removed from the interface and analyzed at a remote spectrometer. Dried protein deposits are examined in detail for secondary-structure content, and protein digestion fragments for specific functionalities with the use of FT-IR microscopy. In preliminary experiments it has been shown that no alteration to the structure or conformation of the protein occurred during particle beam operation [8]. Additionally, it was demonstrated that dried particle beam protein deposits retained the solution conformation that the protein possessed prior to nebulization. The ability to use the particle beam LC–FT-IR system to determine the conformation of proteins separated by HPLC has been recently demonstrated [9,10].

Electrospray (ES) LC–MS is used extensively to map and sequence peptides. The recent development of effective LC–ES-MS instruments [11] permits accurate measurements of fragment masses separated by HPLC. Because of the specificity and sensitivity that LC–ES-MS affords, it has emerged as the premier method for the determination of peptide composition [12]. Mass spectral analysis of HPLC peptide maps of the tryptic digests of βLG A and B from microbore liquid chromatography has been demonstrated [13]. In this study several ions were assigned to digestion fragments associated with the established primary structure. Sequence information was obtained with the use of LC–MS–MS with collision-activated dissociation. Optimization of the ES interface with mobile phases that contained trifluoroacetic acid (TFA) as an ion-pairing reagent has recently been achieved [14]. Nonetheless, LC–ES-MS has some shortcomings: not all of the theoretical fragments expected from digestion are always observed, and the solvent forms ion clusters that produce a considerable background or 'chemical noise' [13]. Losses on the column can be ascribed to either high solubility of hydrophillic components or insolubility of hydrophobic components [15]. The entire digest mixture can be statically analyzed without HPLC separation with the use of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [16]. In this technique, ions that pertain to each peptide within the mixture can be discriminated by the mass detector [17]. Thus, the MALDI-TOF-MS experiment provides an ancillary mass measurement of peptides complementary to those obtained by electrospray LC–ES-MS.

β-Lactoglobulin (βLG), the major protein found in the whey portion of bovine milk, was used in this study. At room temperature and approximately neutral pH, βLG is a dimer, but it undergoes dissociation at low pH (pH<3) [18]. βLG exists naturally in

two predominant variants, A and B [19]. Variant B has two substitutions of the A form; glycine at residue 64 for aspartic acid, and alanine at residue 118 for valine. The sequence is illustrated in Fig. 1 [20]. Dimers of mixed variants have been observed (e.g βLG AB) [19]. The approximate molecular mass of the monomers is 18 300. The βLG monomer is mostly antiparallel β-sheet, formed by 9 strands wrapped round to form a flattened cone. Residues that correspond to β-strands are illustrated in Fig. 2. The core of this motif is an eight-stranded β-barrel [21]. One three-turn α-helix occurs between residues 130 and 140. Two disulfide bridges exist between cysteine residues 66 and 160, and residues 106 and 119, respectively. One free cysteine thiol group is found at residue 121 [21].

In this study, βLG form A was digested with the proteolytic enzyme trypsin for subsequent separation by HPLC and spectroscopic detection. Specifically, the target protein structure was deduced from UV and IR absorptions and mass measurements of the peptide fragments. Trypsin cleaves the C-terminal lysine–X and arginine–X bonds [22]. When the adjacent residue, X, is an acidic residue, the rate of proteolysis is reduced, and if X is another lysine or arginine the rate is drastically reduced [3]. After an extensive digestion period, the mixture was separated by reversed-phase narrow-bore HPLC interfaced with infrared and mass spectrometers separately.

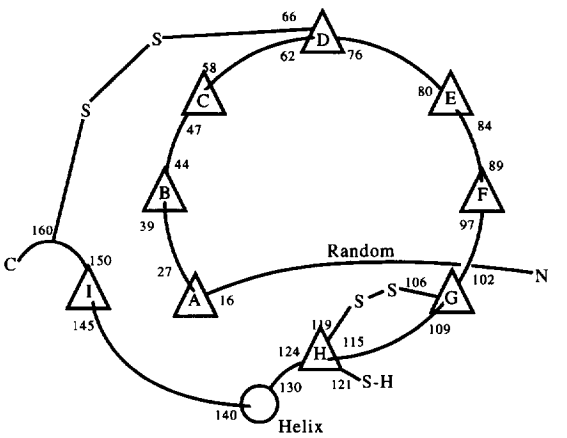

Fig. 2. Molecular topology for β-lactoglobulin. Triangles indicate β-sheet segments, and the circle the α-helix segment. The nine β-sheets are skewed inward and form a cone or β-barrel. The two disulfide bridges are placed in their approximate positions; after Papiz et al. [21].

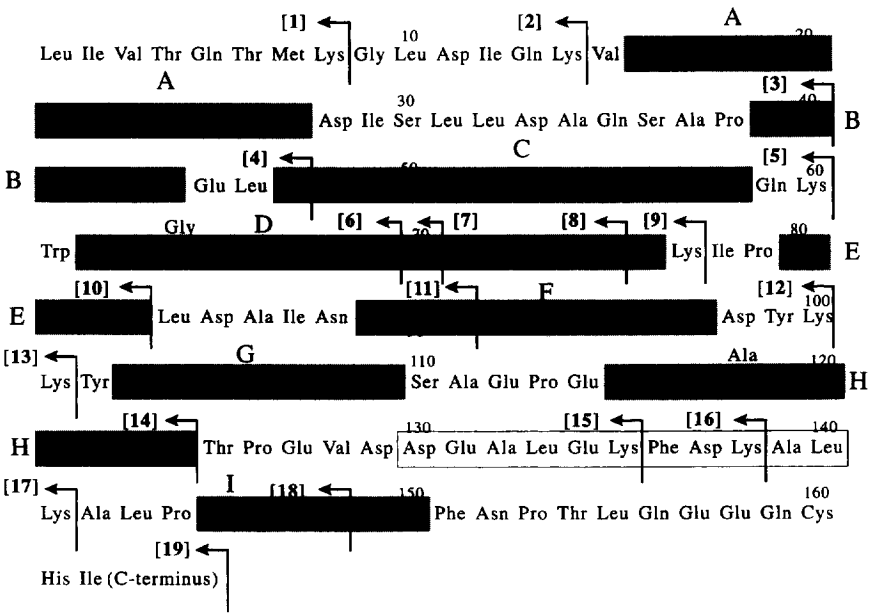

Fig. 1. Amino acid sequence of β-lactoglobulin A and B. βLG B has 2 substitutions: a glycine for aspartic acid at position 64 and valine for alanine at position 118. Arrows indicate theoretical cleavage points for tryptic digestion. Shaded portions indicate residues involved in β-sheet secondary structure, and the single open portion, residues 130–140, indicates the lone α-helix

Complementary mass measurements by MALDI-TOF-MS were made for verification of fragments assigned by the LC–ES-MS experiments. To the knowledge of these authors, this is the first report in which infrared spectrometry was used in the identification of fragments from tryptic digests.

2. Experimental

2.1. Materials

β-Lactoglobulins A, B and AB of the highest purity were purchased from Sigma (St. Louis, MO, USA) and used as received. Tosylamido-2-phenylethylchloromethyl ketone (TPCK) inhibited trypsin was purchased from Pierce (Rockland, IL, USA). A standard βLG A tryptic digest sample, used for the chromatographic work-up and for comparisons to actual digests, was purchased from PE Applied Biosystems (Foster City, CA, USA). TFA, ammonium bicarbonate (NH_4HCO_3), α-cyano-4-hydrocinnamic acid and 2-mercaptoethanol were purchased from Aldrich (Milwaukee, WI, USA). Urea, calcium chloride ($CaCl_2$) and acetonitrile were obtained from J.T. Baker (Phillipsburg, NJ, USA). Water was deionized to 18 MΩ with a Barnstead NANO ultrapure water system (Dubuque, IA, USA). Dithiotheritol, uracil and iodoacetamide were also purchased from Sigma.

2.2. Reduction and alkylation of β-lactoglobulin

10–15 mg of the target protein was incubated in 8 *M* urea and 0.5 *M* NH_4HCO_3 in polypropylene tubes for 30 min at 37°C. To reduce the disulfide bonds the sample was incubated in 5 m*M* dithiotheritol for 4 h at 37°C. To methylate the free sulfryl groups, 137 μl of 100 m*M* iodoacetamide was added and the solution was incubated for 15 min at room temperature. An excess 2-mercaptoethanol was added to consume any residual alkylating agent. To remove the protein from the reaction mixture, gel filtration chromatography was employed: Sephacryl S-100 HR (Pharmacia Biotech, Uppsala, Sweden), 20×1 cm column with 0.1 *M* NH_4HCO_3 mobile phase. The fractions (1–2 ml volume) that contained the protein were detected at 280 nm with a Beckman UV–visible DU-7 spectrophotometer (Irvine, CA, USA).

2.3. Tryptic digest

100 μl of the S-carboxymethylated βLG was diluted with 900 μl of the digestion buffer (0.1 *M* NH_4HCO_3/1 m*M* $CaCl_2$) in a polypropylene tube which was placed into a boiling water bath for 5–10 min. 10 to 20 μl of the 1 mg/ml trypsin solution was added to the βLG samples. The digest was thermostated in a heating block set at 37°C. The digestion required no longer than 8–10 h which depended upon the protein:enzyme ratio. The digestion was halted by the addition of dilute TFA (approximately 0.05%, v/v) followed by immediate submersion of the reaction tube into a dry ice–acetone bath. The digest was stored in a freezer at −77°C for 1 week.

2.4. Chromatography

The HPLC system consisted of a Perkin-Elmer Series 200 quaternary solvent pumping system (Norwalk, CT, USA), a Zorbax (MAC-MOD Analytical, Chadds Ford, PA, USA) 15×2.1 cm sterically bonded *n*-octadecyl column (C_{18}), a PE 235C diode array detector, and analytical interfaces. Column load varied between 10 and 30 μg for proteins and hundreds of ng for peptides. The column was thermostated at 27°C. A flow-rate of 0.25 ml/min was used throughout. Peptide bonds were monitored at 215 nm, and aromatics at 250 and 280 nm. Gradient elution was used in the following sequence: 95% A [0.1% (v/v) TFA], 5% B [0.07% (v/v) TFA in acetonitrile] to 61% A, 39% B in 30 min, isocratic hold for 10 min, then to 5% A and 95% B. The void volume was determined with uracil.

2.5. Particle beam FT-IR spectrometry

A complete description of the particle beam has been given previously [4,8]. The particle beam LC experimental setup is illustrated at the top of Fig. 3. The aerosol was produced with 25 μm I.D. fused-silica (Supelco, Supelco Park, PA, USA). For evaporation and sample deposition, a desolvation chamber of 22 cm in length was employed, mobile phase vapour was removed with a single momentum

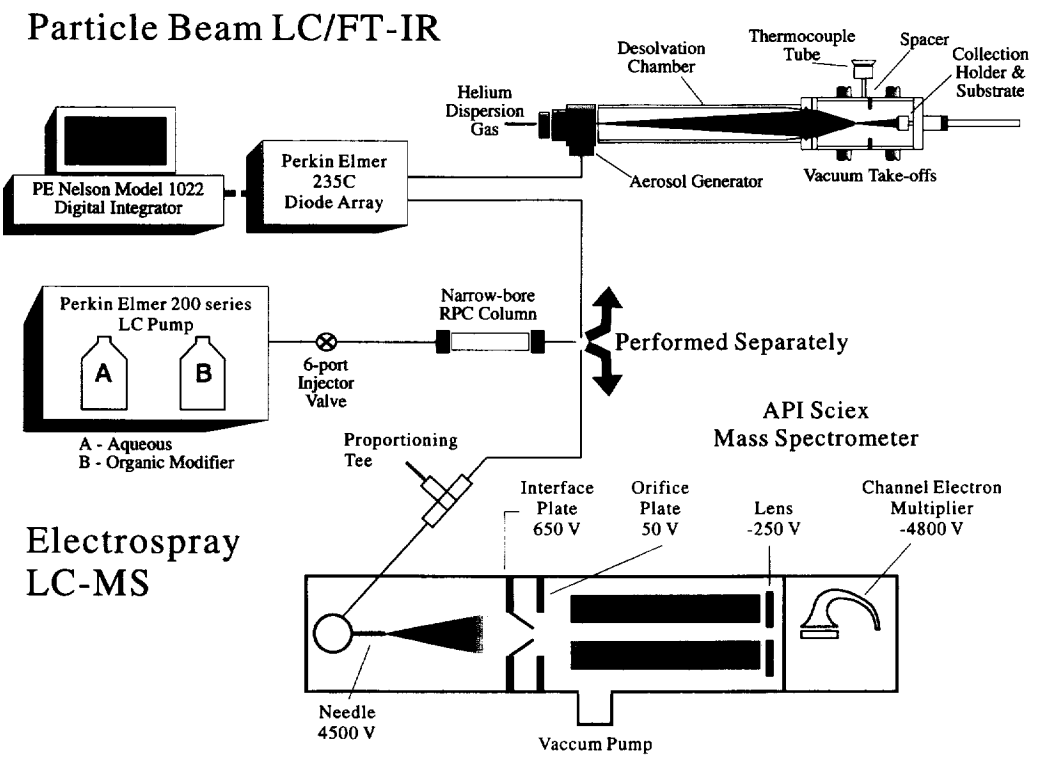

Fig. 3. Experimental diagram of the HPLC interface to FT-IR and MS. Connection to each spectrometric detector was performed separately.

separator, and the nozzle to substrate distance was maintained at 5 mm. Deposits with diameters of approximately 50 to 100 μm were produced under these conditions. Eluted digestion fragments were examined with a PE Spectrum 2000 FT-IR spectrometer equipped with a Series-I Microscope (Perkin-Elmer). A water insoluble calcium fluoride (CaF_2) window, 25 mm diameter×2 mm, was used as the collection substrate (International Crystal Labs., Garfield, NJ, USA). Spectra were produced from 500 to 1000 scans co-added at 8 cm^{-1} resolution. Strong Norton-Beer apodization was used. Reference IR spectra of both βLG and individual amino acids were acquired by the evaporation of an aqueous solution, slightly acidic or alkaline for certain amino acids, onto a CaF_2 window in a heatless-vacuum desiccator over Drierite desiccant (W.A. Hammond, Xenia, OH, USA) and scanned with the microscope in a fashion identical to that described for particle beam deposit spectra.

2.6. Electrospray LC–mass spectrometry

A PE Sciex API I single quadrupole mass spectrometer with ES introduction capabilities (Perkin-Elmer) was used for the LC–MS experiment, as shown at the bottom of Fig. 3. A split ratio of 10:1 allowed an approximate flow of 25 μl/min into the ion source. A mass range of 600 to 1200 u was monitored for all experiments. A +4500 V bias (20 μA) was placed on the discharge needle, and the electron multiplier was held at −4800 V. The following voltages were maintained on the ion optics: interface plate, +650 V; orifice plate, +50 V; electrostatic lens, −250 V. The chromatographic conditions were identical to those described above.

2.7. MALDI-TOF-MS

Static mass measurements on the βLG digest fragments were made with a Brüker MALDI-TOF-

MS instrument (Billerica, MA, USA) equipped with a nitrogen laser (337 nm). Spectra were prepared from 100 to 150 summed pulses. α-Cyano-4-hydrocinnamic acid was used as the MALDI matrix and samples were prepared from a concentrated solution of the digest in acetonitrile and 0.1% aqueous TFA (50:50, v/v).

3. Results and discussion

3.1. Chromatography of βLG

The chromatographic behaviour of βLG was examined to determine the affect of the chromatographic conditions on the conformation of the protein. The secondary structure upon elution was assessed by particle beam FT-IR spectrometry. The retention time of the whole intact protein was measured to identify any undigested βLG, should any have remained after digestion. For the forms βLG AB, A and B little difference in retention time was observed, as illustrated in the chromatograms in Fig. 4a–c. The βLG monomers, A and B, from the dissociation of the mixed AB dimer, were only partially resolved. The equal detector response indicated approximately equivalent quantities of both

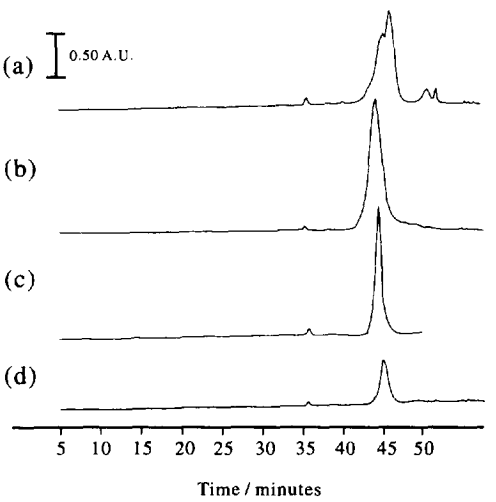

Fig. 4. Reversed-phase chromatography of β-lactoglobulin on an *n*-octadecyl narrow-bore sterically protected column at 215 nm. (a) Dimer of βLG AB, (b) Variant Monomer βLG A, (c) Variant Monomer βLG B, (d) S-carboxymethylated βLG A.

monomers. When the monomers were chromatographed separately, they had retention times of 44 (A) and 45 (B) min, respectively. In all cases, a small peak with poor recovery eluted consistently between 35 and 36 min. We suspect that this material is a mixture of folded and unfolded forms of the protein with less stationary phase interaction than the main peak. The sample of βLG A that underwent S-carboxymethylation had a retention time of 42.5 min, as illustrated in Fig. 4d, which was similar to those of native βLG A and B. This peak had a slightly higher degree of symmetry than those of the native, untreated forms, but was of lower intensity due to dilution during derivatization and filtration. A second sample of S-carboxymethylated βLG A was incubated for 8 h, the length of the digestion period, with conditions identical to those of the digest (37°C, pH~8) to determine if any thermal decomposition to the protein occurred during digestion. A single chromatographic peak, not illustrated, indicated that no significant thermally induced cleavage occurred, and was therefore not expected during actual digestion. It was subsequently learned that the intact protein eluted nearly 10 min after the last digest fragment. Thus, the parent protein did not interfere with the chromatography of the digest.

Particle beam collections were made on the separate chromatographic runs of βLG A and B. Representative IR reference and particle beam deposit spectra of βLG are illustrated in Fig. 5, in which the amide I and II regions have been enlarged. The particle beam reference spectra of βLG from water, Fig. 5a, clearly shows the asymmetry of the amide I band for both βLG A and B variants. The amide I contour of these spectra is indicative of a protein with a substantial amount of β-sheet segments [23]. Within the amide I region there is a band maximum at 1632 cm^{-1}, which is the perpendicular (⊥) component of the anti-parallel β-sheet mode, and the associated parallel (‖) component is found at approximately 1680 cm^{-1} as a weak shoulder. This IR amide contour has been observed with other high β-sheet proteins in previous studies [6,8]. In addition, the approximate position of the amide II band at 1540 cm^{-1} (NH peptide bending vibration) is consistent with a spectrum of a high β-sheet protein. The high degree of similarity between the spectra of forms A and B, which was also evident in second

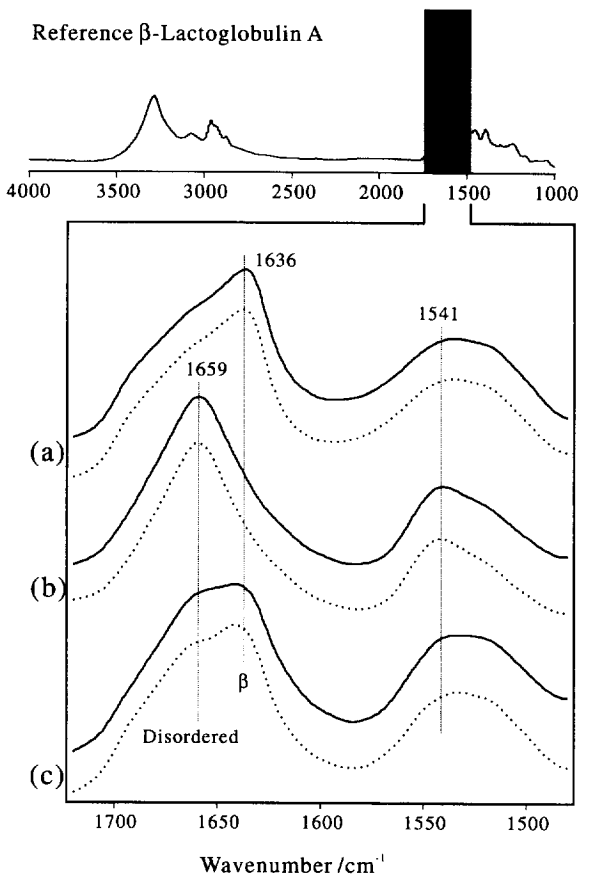

Fig. 5. Particle Beam LC–FT-IR spectra of whole β-lactoglobulin. Top reference spectrum is a full mid-IR scan of βLG AB dimer deposited from water. The expansion is of the amide I and II regions, 1750–1450 cm^{-1}. Solid lines ——— designate spectra of βLG A, and dotted lines · · · designate spectra of βLG B. (a) Reference evaporated film spectra of native forms deposited from water, (b) particle beam deposits after elution from the *n*-octadecyl column, and (c) particle beam deposits from, (b) after dissolution and reevaporation. Wavenumber lines show position of β-sheet maximum, 1636 cm^{-1} and disordered 1659 cm^{-1}. Amide II disordered band is at 1541 cm^{-1}.

derivative spectra (not shown), illustrated that the two forms assumed very similar conformational structures. The particle beam IR spectra of whole and untreated βLG A and βLG B collected upon elution from an *n*-octadecyl column with the use of gradient elution chromatography are shown in Fig. 5b. The substantial shift in the amide I maximum to 1659 cm^{-1}, as well as the high degree of symmetry observed, demonstrated that upon elution the secondary structure of βLG was randomized completely. A concomitant shift in the amide II band verifies this as well. When this particle beam deposit was redissolved in water, evaporated and re-scanned, a reversion to a spectrum with similar spectral features to that of the nature was observed. Presumably, upon column propagation and elution the covalent structure and the two disulfide bridges of the protein molecule remained intact. Therefore, upon post-column solvation the protein was able to reorganize back to a structure approaching that of the native protein. Second derivative spectra of Fig. 5c, not illustrated, show a significant recovery of β-sheet in comparison with Fig. 5b. Refolding was not complete, however, as a comparison between Fig. 5a and Fig. 5c illustrates. Significant amounts of disordered structure remain as evidenced by the band at 1659 cm^{-1}. Slight differences in the degree of refolding between βLG A and B are attributed to differences in refolding times, which were not strictly controlled, and not to differences in the intrinsic rates of refolding. Refolding behaviour after chromatographic analysis was observed with serine protease α-chymotrypsin [9] and bovine ribonuclease A [10] in previous studies.

3.2. Tryptic digestion of β-lactoglobulin

The proteolytic digestion of βLG with trypsin was carried out following general protocols [1,3,22]. Theoretically, the action of trypsin on βLG produces 19 peptide fragments, the composition of which varies between 1 and 26 amino acid residues. The analysis of the S-carboxymethylated form of βLG A is the focus of the remainder of this study.

The gradient procedure for the resolution of the tryptic digest of βLG was developed previously with a standard βLG A digest sample (chromatogram not illustrated). The chromatographic separation of the βLG A digest mixture that was used for both IR and MS collections is illustrated in Fig. 6a. Only 10 of the 13 peaks separated were examined, as designated in Fig. 6a, due in part to spatial limitations of the particle beam deposition substrate. Also, the small leading peak was too dilute for acquisition of a good spectrum, while the final two peaks nearly coelute with trypsin, the proteolytic agent, as shown in Fig. 6d. The aromatic residues tyrosine, tryptophan and

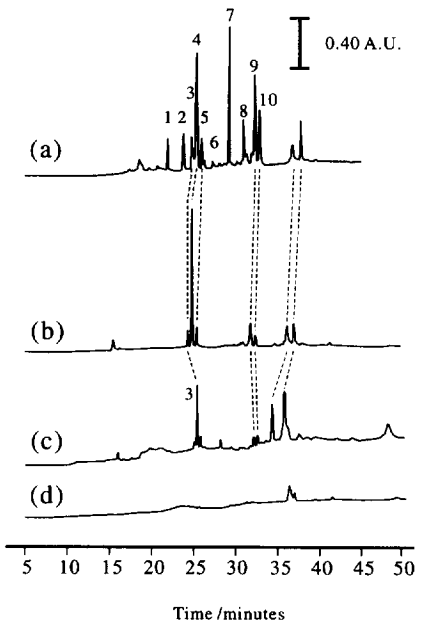

Fig. 6. Reversed-phase chromatography of tryptic digest of β-lactoglobulin at various wavelengths. 5 μg total was injected onto the column. Peak numbers are placed over the chromatogram and indicate the positions of both IR and MS collections. (a) tryptic digest of βLG A detected at 215 nm, (b) tryptic digest of βLG A detected at 280 nm, (c) tryptic digest of βLG A detected at 250 nm, and (d) trypsin/TPCK control chromatogram.

phenylalanine contain aromatic UV chromophores with an absorption envelope between 250 and 300 nm. These are $\pi-\pi^*$ transitions which have maxima of 275 nm for tyrosine and 280 nm for tryptophan [24]. The maximum for phenylalanine is around 257 nm [3]. Usually, peptides are detected at 215 nm at which the amide backbone exhibits absorption. The absorptions generated at 280 nm are selective to the presence of both tyrosine and tryptophan residues within the peptide fragments, as shown in Fig. 6b. The βLG A chromatogram at 280 nm suggested that these residues were present in chromatographic peaks 4, 5 and 9. The trace at 250 nm is not entirely selective to phenylalanine because the absorption envelope of the other aromatic residues extend to this wavelength. The drastic increase in absorption at 250 nm, as shown in Fig. 6c, however, strongly suggested the presence of phenylalanine in chromatographic peak 3. Slight shifts in peak retention times among the chromatograms are due to inadequate column temperature control. Lastly, a trypsin/TPCK control sample was incubated for 8 h, at concentrations (~100 ng) and conditions identical to those of the digest, to determine if auto-digestion of trypsin occurred and interfered with the chromatography (Fig. 6d). The low response of the UV trace at 215 nm is due to the low concentration and it is obvious that no auto-digestion occurred. It is believed the inclusion of the 1 m*M* $CaCl_2$ limited auto-digestion [22].

3.3. Mass spectrometry of the β-lactoglobulin digest

Accurate mass measurements of the peptides contained within the 10 major chromatographic peaks facilitated their assignment to segments of βLG A. Table 1 lists the expected, or theoretical, fragment masses for the tryptic digest of βLG A. The assignments of ions to specific chromatographic peaks from the tryptic digest of βLG A are presented in Table 2. Because of limitations in the single-quadrupole mass spectrometer data acquisition rate, a limited mass range of 600–1200 u was monitored. Therefore, masses above this range were detectable only as multiply charged ions, and fragments with masses below the range were undetectable. This selection of mass range allowed measurement of all but four expected digest products. The signal-to-noise ratio was very low for fragment 6 which made a reliable assignment on it impossible. A total of three peptide fragments, from chromatographic peaks 5, 8 and 10, was assigned to unexpected tryptic cleavage products. As previously stated, observation of a wide mass range was foregone to achieve increased time resolution with a narrowed mass range. Fortunately, all major chromatographic peak fragments from these digests were observable.

A static mass spectral analysis of the entire digest of βLG A by MALDI-TOF-MS verified each ion measured in the LC–ES-MS experiment. The MALDI mass spectrum of the digest, with assignments based on Table 1, is displayed in Fig. 7. The mass cluster patterns observed with each of the major ions were caused by reagent gas, water or sodium addition to the peptide. An abundant ion at 2848 u, that corresponds to residues 101–124, revealed an incomplete cleavage between lysine 101 and tyrosine 102. This fragment had an increase of

Table 1
Theoretical fragmentation of β-lactoglobulin from trypsin digestion

Theoretical fragment number	Amino acid residues	Sequence	Calculated molar mass/u[a]
T1	1– 8	LIVTQTMK	993.3
T2	9– 14	GLDIQK	673.4
T3	15– 40	VAGTWYSLAMAASDISLLDAQSAPLR	2707.4
T4	41– 47	VYVEELK	986.1
T5	48– 60	PTPEGDLEILLQK	1452.7
T6[b]	61– 69	WENDECAQK	1122.5
	61– 70	WENDECAQKK	1250.6
T7	70	K	147.1
T8	71– 75	IIAEK	573.4
T9	76– 77	TK	248.2
T10	78– 83	IPAVFK	674.4
T11	84– 91	LDAINENK	916.5
T12	92–100	VLVLDTDYK	1065.6
	92–101	VLVLDTDYKK	1192.7
T13	101	K	147.1
T14[b]	102–124	YLLFCMENSAEPEQSLVCQCLVR	2675.2
T15	125–135	TPEVDDEALEK	1245.6
T16	136–138	FDK	409.2
T17	139–141	ALK	331.2
T18	142–148	ALPMHIR	837.5
T19[b]	149–162	LSFNPTLQEEQCHI	1658.8

Monoisotopic masses of fragments MH^+ of β-lactoglobulin A. Cysteine residues are bold: Cysteine-C. Aromatic residues are underlined: Phenylalanine-F, Tyrosine-Y, Tryptophan-W.

[a] m/z expected for singly charged species.

[b] S-Carboxymethyated cysteine residue corresponds to mass increase of 43 u. Fragment 6 one S-carboxymethyated cysteine residue βLG A 61–69 1165.6 u, 61–70 1293.6 u. Fragment 14 three *S*-carboxymethyated cysteines βLG A 102–124 2804.2 u. Fragment 19 one S-carboxymethyated cysteine residue 149–162 1701.8 u.

Table 2
Electrospray LC–MS of tryptic digest of β-lactoglobulin A

Chromatographic peak	Retention time (min)	Mass to charge $(m/z)^{-1}$	Ion assigned	Fragment & residue number	Sequence
1	22.5	933.3	MH^+	[T1] 1–8	LIVTQTMK
2	23.2	818.4	Measured MH_2+2	[T15,T16] 125–138	TPEVDDEALEKFDK
		1635.8	Actual		
3	24.0	696.0	Measured Na^+	[T10] 78–83	IPAVFK
		674.4	Actual		
4	25.1	1065.3	MH^+	[T12] 92–100	VLVLDTDYK
5	25.8	690.0	MH^-	15–20	VAGTWY
7	27.8	836.5	Measured MH^-	[T18] 142–148	ALPMHIR
		837.0	Actual		
8	30.1	1015.5	Measured MH_2^{+2}	21–40	SLAMAASDISLLDAQSAPLR
		2030.6	Actual		
9	30.9	771.9	Measured MH_3^{+3}	[T4,T5] 41–60	VYVEELKPTPEGDLEILLQK
		2313.3	Actual		
10	31.5	762.0	Measured MH^+	27–33	SDISLLD
		762.4	Actual		

Actual: Calculated m/z of the ion of the singly charged species (MH^+).
Fragments with no number are unexpected from the digestion with trypsin.

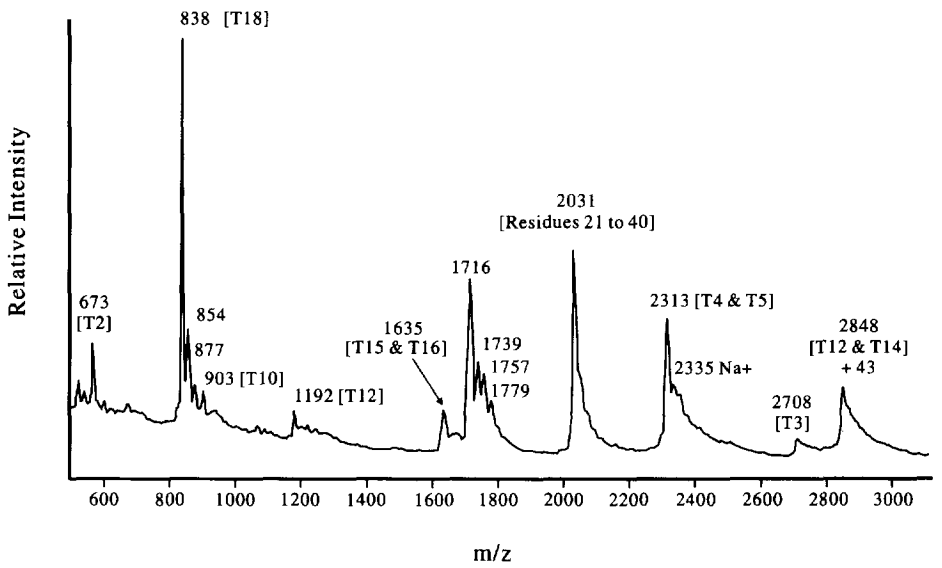

Fig. 7. Static MALDI-TOF-MS spectrum of the tryptic digest of β-lactoglobulin A. The ion locations are marked accordingly. The expected fragments or combinations of fragments associated with the ions are indicated in brackets.

43 u which is a result of the S-carboxymethylation of one of the three cysteine residues present in this fragment at positions 106, 119 and 121. It is plausible that only the free thiol, cysteine 121, was successfully alkylated and the disulfide bridges between residues 106 and 119 remained intact throughout digestion and separation. This ion was not observed by LC–ES-MS. The combined theoretical fragments T4 and T5, which were apparently uncleaved during digestion, give rise to the ion at 2313 u, and the Na^+ adduct at 2335 u. The ion at 2030 u was observed in chromatographic peak 9 as a triply charged ion (772 u) in the LC–MS experiment. A major group of peaks at 1716–1779 u could not be accurately identified. It was speculated that this grouping contained an unexpected proteolytic cleavage with perhaps incomplete cysteine alkylation. The most abundant ion, at 838 u, corresponds to theoretical fragment T18 which contains residues 142–148. This peptide fragment produced the most intense chromatographic peak in the HPLC run, chromatographic peak 7 at approximately 29 min. Smaller ions that were identified have been labelled on Fig. 7.

3.4. FT-IR spectrometry of the β-lactoglobulin fragments

The purpose of IR analysis in proteolytic digestion studies is to verify the presence of specific functional groups within fragments. In situations where the peptide fragment was large enough to contain the presence of secondary structure segments, the possibility of residual conformation was postulated. Fig. 8 presents the particle beam IR spectra of the peptide fragments that corresponded to the numbered chromatographic peaks in Fig. 6a. Although these spectra are somewhat similar in overall contour, and, thus, also in molecular framework, they do exhibit significant differences, particularly within the amide I region. The amide spectral region 5 of a peptide which is less than 5–6 residues in length is a composite that contains the amide carbonyl vibrations and significant absorptions from other functional groups (e.g., aromatic ring modes and −OH bending vibrations). Therefore, exact assignment of the amide I region was difficult for small peptides. Theoretical fragment T3 contained β-sheet segment A, and fragment T14 contained the β-sheet segment

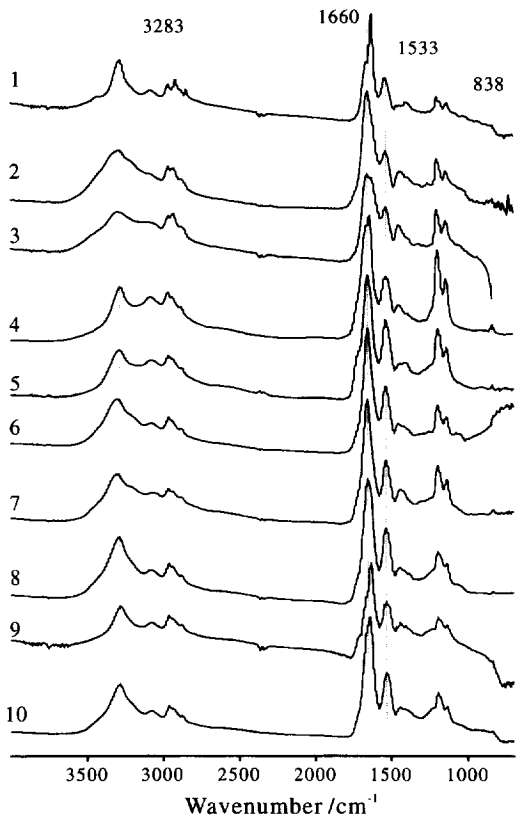

Fig. 8. Particle beam LC–FT-IR spectra from tryptic digest of β-lactoglobulin A. The ten peptide fragment spectra correspond to the peaks indicated in Fig. 6a. Lines indicate the general locations of the amide backbone NH stretching mode (3283 cm^{-1}), amide I carbonyl stretch (1660 cm^{-1}), the amide II NH bending mode (1533 cm^{-1}) and the *para*-disubstituted benzene ring C–H deformation (838 cm^{-1}).

G, as indicated in Figs. 1 and 2. Therefore, there is a possibility that some peptide fragments may retain residual secondary structure. From the initial work in this study with whole βLG, however, it was determined that ordered secondary structure was randomized during reversed-phase HPLC. Nonetheless, spectra 1, 3, 4, 5, 6, 8, 9 and 10 exhibit amide I absorption at 1630–1640 cm^{-1}, characteristic of β-sheets, to greater or lesser degrees and significant absorptions at ~1660 cm^{-1} due to disordered structure. Several factors, such as residual secondary structure, independent innate solution structure, mobile phase composition and aggregative behaviour, could be involved in the manifestation of peptide solution structure. For example, the β-sheet band in spectrum 3 suggests residual structure in the fragment that consists of residues 78–83. An entire β-sheet segment (E) of the native protein is contained within those residues. Spectra of peptide fragments not involved in native secondary structures, spectra 1, 9 and 10, however, also show β-sheet structure. Therefore, a more detailed structural characterization of peptides under reversed-phase conditions awaits further investigation. Because particle beam deposits are thoroughly dried, no residual water bands show in any of the ten spectra. Weak absorbance shoulders around 3400 cm^{-1} are most likely due to alcohol O–H stretching. The threonine and serine residues, which contain free –OH, can be identified. Spectra 1, 2, 4, 5, 8 and 10 contain one or both of these uncharged amino acids evident in the moderate absorptions at 3400–3500 cm^{-1}. Gem-dimethyl [$(CH)_2CH-$] C–H deformation modes are believed to be present in spectra 1, and 4–10. A weak bending band at 1380–1390 cm^{-1} and a corresponding C–C stretch evident as a moderate shoulder at 1182 cm^{-1} are ascribed to either valine and/or leucine. Only the aromatic amino acids and olefinic group of histidine exhibit C–H stretching at >3000 cm^{-1}. *para*-Disubstituted benzene in tyrosine exhibits a strong C–H out of plane deformation band at 840 cm^{-1}. From the chromatographic analysis of the digest detected at 280 nm it appeared that peaks 4, 5 and 9 contained tryptophan and/or tyrosine. Both a moderate shoulder at 3020 cm^{-1} and absorption at 840 cm^{-1} confirm the presence of tyrosine in the peptide fragment of all of the three aforementioned chromatographic peaks. The reversed-phase HPLC experiments were carried out at a pH well below that of the pK_a of the amino acids that contain ionizable acid side-chain groups. Therefore, these acidic residues were protonated during separation. An acid carbonyl C=O stretch above 1700 cm^{-1} would confirm the protonated glutamic and aspartic acid residues. In the spectra of chromatographic peaks 2, 4, 5, 8 and 9, shoulders well above 1700 cm^{-1} were evident in both mildly deconvolved and second derivative spectra, and thus revealed acid residues.

The acid COOH shoulder is less pronounced in fragments spectra 2, 8 and 9 because of the relatively larger number of non-acid amino acids present in these peptides.

To summarize the MS data, theoretical fragments T1–T5, and T10–T18 were observed either as properly cleaved fragments, combinations or products of further dissociation. Expected ions that relate to the interior segments of βLG were not observed. These unobserved segments correspond to theoretical fragments T6–T9, which is a region with the highest concentration of lysine residues. It was unlikely that the conical topology of the βLG molecule hindered the action of trypsin on these residues because the other β-segments were equally shielded by the conical geometry. Also, the action of the 8 *M* urea should have unfolded the protein molecule and exposed all potential cleavage points. Based on the MS patterns, it was concluded that reduction of the disulfide bridges was not complete. Consequences of incomplete denaturation and alkylation can be unexpected fragmentation and/or incomplete cleavage in regions where cleavage should be facile.

4. Conclusions

The sum of the analytical results are tabulated in Table 3 and it includes all relevant data for each chromatographic peak, with UV, IR and mass spectrometric data. This concisely illustrates how information obtained from the various techniques was correlated to determine structure segments of the digested βLG molecule. The use of FT-IR spectrometry in the identification of functional groups specific to certain amino acids was demonstrated. The particle beam LC–FT-IR interface can also be used to probe elution conformation of proteins. Detection limits favour MS when LC–FT-IR spectrometry and LC–MS are compared. Nonetheless, at given injection volumes and sample concentrations, quality IR spectra were obtained on nanogram levels. Future work involves improvement of this methodology with the use of similar digests of proteins with known structures and with the use of chemometric methods.

Acknowledgments

The authors gratefully acknowledge the help of Dr. Dennis S. Phillips of the Mass Spectrometry Facility at the Department of Chemistry, University of Georgia, Athens, GA, USA. Appreciation for assistance with gel filtration chromatography is extended to Dr. S. Bakthavatsalam of the Department of Biochemistry at the University of Georgia, Athens, GA, USA. Financial support was provided by the University of Georgia Research Foundation 1996.

Table 3
Summary of analytical results

Chromatographic peak	Retention time (min)	IR Bands (cm^{-1})	Measured ion $(m/z)^{-1}$	Aromatic UV (absorption/nm)	Sequence assignment
1	22.1–22.9	3284, 1662, 1630	933	–	(1–8) LIVTQTMK
2	22.9–23.9	3285, 1658	818	–	(125–138) TPEVDDEALEKFDK
3	23.3–24.6	3294, 1659, 1628	696	250	(78–83) IPAVFK
4	23.5–25.3	3278, 1669, 1640	1065	280, 250	(92–100) VLVLDTDYK
5	24.0–25.8	3283, 1720, 1657, 1635	690	280, 250	(15–20) VAGTWY
6	26.0–27.5	3296, 1657	–	–	–
7	27.8–29.4	3300, 1659	837	–	(142–148) ALPMHIR
8	29.9–31.2	3285, 1658, 1635	1016	–	(21–40) SLAMAASDISLLDAQSAPLR
9	30.9–32.0	3227, 1636	772	–	(9–15) GLDIQK
10	31.5–32.6	3285, 1662, 1641	762	–	(27–33) SLISLLD

Retention time range is the span of the variation of retention time observed during the course of all experiments.
Measured ion: ion observed in Electrospray LC–MS experiments.

References

[1] E.A. Carey, in T.E. Creighton (Editor), Protein Structure, IRL Press at Oxford University Press, Oxford, 1989, p. 117.
[2] W.S. Hancock and J.T. Sparrow, HPLC Analysis of Biological Compounds, Marcel Dekker, New York, 1984.
[3] R.A. Copeland, Methods for Protein Analysis, Chapman Hall, New York, 1994.
[4] V.E. Turula, Ph.D. Thesis, University of Georgia, Athens, GA, 1995.
[5] H.H. Mantsch and D. Chapman (Editors), IR Spectroscopy of Biomolecules, Wiley, New York, 1996.
[6] D.M. Byler and H. Susi, Biopolymers, 25 (1986) 469.
[7] H. Susi, in S.N. Timasheff and G.D. Fasman (Editors), Structure and Stability of Biological Molecules, Marcel Dekker, New York, 1969, p. 575.
[8] V.E. Turula and J.A. de Haseth, Appl. Spectrosc., 48 (1994) 1255.
[9] V.E. Turula and J.A. de Haseth, Anal. Chem., 68 (1996) 629.
[10] R.T. Bishop, V.E. Turula and J.A. de Haseth, Anal. Chem., 68 (1996) 4006.
[11] M.R. Emmett and R.M. Caprioli, J. Am. Soc. Mass. Spectrom., 5 (1994) 605.
[12] A.L. Burlingame and S.A. Carr (Editors), Mass Spectrometry in the Biological Sciences, Humana Press, Totowa, NJ, 1996.
[13] E.C. Huang and J.D. Henion, J. Am. Soc. Mass Spectrom., 1 (1990) 158.
[14] J. Eshraghi and S.K. Chowdhury, Anal. Chem., 65 (1993) 3528.
[15] A.L. Burlingame, R.K. Boyd and S.J. Gaskell, Anal. Chem., 68 (1996) 599.
[16] M. Mann and G. Talbo, Curr. Opin. Biotechnol., 7 (1996) 11.
[17] M.S. Bolgar and S.J. Gaskell, Biochem. Soc. Trans., 23 (1995) 907.
[18] M.J. Kelly and F.J. Reithel, Biochemistry, 10 (1971) 2639.
[19] G.W. Green, R. Aschaffenburg, A. Camerman, J.C. Coppola, P. Dunnill, R.M. Simmons, E.S. Komorowski, L. Sawyer, E.M.C. Turner and K.F. Woods, J. Mol. Biol., 131 (1979) 375.
[20] G. Braunitzer, R. Chen, B. Schrank and A. Stangl, Z. Physiol. Chem., 354 (1973) 867.
[21] M.Z. Papiz, L. Sawyer, E.E. Eliopoulos, A.C.T. North, J.B.C. Findlay, R. Sivaprasadarao, T.A. Jones, M.E. Newcomer and P.J. Kraulis, Nature, 324 (1986) 383.
[22] A.V. Flannery, R.J. Beynon and J.S. Bond, in R.J. Beynon and J.S. Bond (Editors), Proteolytic Enzymes, IRL, Oxford University Press, Oxford, 1990, p. 145.
[23] T. Miyazawa, in G. Fasman (Editor), Poly-α-Amino Acids, Marcel Dekker, New York, 1967, p. 69.
[24] D.J. Holme and H. Peck, Analytical Biochemistry, Longman, New York, 2nd ed., 1993.