ELSEVIER

Journal of Chromatography B, 662 (1994) 245–259

JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL APPLICATIONS

Primary structure control of recombinant proteins using high-performance liquid chromatography, mass spectrometry and microsequencing

François Frédéric Clerc[a], Bertrand Monégier[a], Didier Faucher[a], Françoise Cuiné[a], Christine Pourcet[a], John C. Holt[b], Sheng-Yuh Tang[c], Alain Van Dorsselaer[d], Jérôme Becquart[a], Marc Vuilhorgne[a,*]

[a]*Rhône-Poulenc Rorer S.A., Centre de Recherches de Vitry-Alfortville, 13 Quai Jules Guesde, B.P. 14, 94403 Vitry sur Seine Cedex, France*

[b]*Rhône-Poulenc Rorer Central Research, 680 Allendale Road, King of Prussia, PA, USA*

[c]*Rhône-Poulenc Rorer Central Research, 500 Arcola Road, Collegeville, PA, USA*

[d]*Laboratoire de Spectrométrie de Masse Bio-Organique, Faculté de Chimie, 1 rue Blaise Pascal, 67008 Strasbourg, France*

Abstract

The conformity of two recombinant proteins (a von Willbrand factor fragment and human serum albumin, consisting of respectively 289 and 585 amino acids) has been examined by HPLC combined with mass spectrometry and microsequencing, on both intact material and fragment peptides obtained by proteolytic cleavage. These studies confirmed that the primary structure of the recombinant proteins corresponds to that predicted from their gene, particularly the integrity of their N and C termini, and, in the case of albumin, the agreement between the observed disulfide bond pattern and the published model. Furthermore, the structure of an albumin-related compound could be elucidated. Application of LC–MS for batch-to-batch quality control is also under discussion.

1. Introduction

Recombinant DNA technology has enabled the production of proteins at the multigram level. Since these compounds are intended for therapeutic use, it is of utmost importance that the accuracy of their structure and their chemical integrity is ensured. Moreover, the expression host and/or the purification process may induce relatively minor modifications which are very difficult to detect analytically. Due to the complexity of such material, chemical integrity assessment can only be achieved by a combination of different methods including high-performance liquid chromatography (HPLC), electrophoresis, isoelectric focussing, microsequencing, amino acid analysis and mass spectrometry of the intact molecule or its fragments produced by protease digestion. Application of mass spectrometry in biotechnology has been reviewed in refs. 1 and 2. Besides, characterization methods have to be practical and quick in order to be applicable to batch-to-batch consistency assessment.

Mass spectrometry plays a crucial role in recombinant protein quality control. Its accuracy makes it possible to detect modifications as small

* Corresponding author.

0378-4347/94/$07.00
SSDI 0378-4347(94)00184-7

as one residue deletion in the case of proteins and disulfide formation or reduction in the case of peptides (2 amu). No other analytical method can detect and identify such modifications at this level of sensitivity (hundred picomoles of sample) within a few minutes. Three mass spectrometric methods have been widely applied in protein chemistry. Liquid secondary-ion mass spectrometry (LSI-MS) has permitted the molecular mass determination of compounds up to 10 000 Da with good accuracy (1 amu [3,4]). Unfortunately, most of the recombinant proteins produced in the pharmaceutical industry have molecular masses between 10 000 and 100 000 Da ; consequently, their characterization by LSI-MS has to go through proteolytic cleavage [5]. In the last 5 years, electrospray or ion-spray mass spectrometry (ES-MS or IS-MS) have enabled the determination of the molecular mass of proteins up to 100 000 Da [6]. This technique is both sensitive (down to the picomole level) and precise (accuracy of 0.01 to 0.001% with respect to the molecular mass measurement [7]). It improves both speed (only a few minutes are necessary to run a sample) and reliability. Modifications of large molecules such as C-terminal heterogeneity, N-terminal acetylation, few percent formation of covalent dimers, are easily detected, even within mixtures [8, 9]. Characterization of such post-translational modifications are much more difficult to achieve using classical biochemical methods. In the last two years, matrix-assisted laser-desorption-ionisation mass spectrometry (MALDI-MS) has proved to be very useful, especially for the analysis of higher molecular mass proteins (> 100 000 [10]) and for the analysis of oligosaccharide moieties associated with glycoproteins [11].

In the present study mass spectrometry, combined with peptide mapping and microsequencing, was applied to fully characterize two recombinant proteins: RG12986, a fragment of the von Willbrand factor (vWF) and recombinant human serum albumin (HSA). RG12986, fragment 445–733 of human vWF, a 2250 residue-blood protein, has been expressed in *Escherichia coli* [12]. This fragment encompasses the binding domain for the platelet receptor GP1b [13,14] and acts as an antagonist of vWF by the formation of an abortive complex with GP1b receptor. This polypeptide offers potential as an antithrombotic agent. RG12986 contains 7 cysteine residues which are alkylated by iodoacetamide to prevent oligomerization by disulfide bridging. The modified molecule retains full biological activity. For such a chemically modified protein it is of first importance to assess the completeness and the specificity of these modifications. Mass spectrometry was used to characterize the intact polypeptide and the proteolytic fragments produced by tryptic digestion.

HSA is the major protein component of human plasma. Its functions are the maintenance of the oncotic pressure and the transport of compounds such as fatty acids [15]. HSA is a protein of considerable clinical importance and is usually given in multigram quantities to replace blood after trauma or during surgery. Albumin is currently produced by fractionation of human blood and represents a major part of the blood product market. Although HSA can be effectively sterilized by heat treatment, the possible persistence of viral contaminants is the main reason that recombinant HSA can be considered a product with a considerable commercial future. HSA occurs as a single-chain protein of 585 amino acids which contains 17 disulfide bonds and a free cysteine as shown in Fig. 6 [16–18]. Recombinant HSA is expressed in yeast [19]. Assessment of the quality of the recombinant molecule is made more difficult by the lack of any *in vitro* activity. It relies therefore on high-resolution physico–chemical characterization. Mass spectrometric studies of the intact protein as well as of proteolytic fragments have been used to assess the chemical integrity of the molecule, the maturation of disulfide bonds, and to characterize process-induced HSA fragments.

2. Experimental

2.1. Mass spectrometry

LSI-MS spectra were obtained on a ZAB SEQ or a VG AutoSpec instruments. Thioglycerol or

metanitrobenzylalcohol were indifferentially used as the matrix. Instruments were tuned in the wide-scan mode and calibrated with CsI clusters. Data were acquired in the multi channel aquisition mode (MCA).

ES-MS and IS-MS spectra were recorded respectively on a VG BioQ (4000 amu mass range) and a Sciex API-III (2200 amu mass range). Samples were dissolved in water–methanol–acetic acid (50:50:1, v/v) or water–acetonitrile–acetic acid (50:50:1, v/v) and introduced into the ion source via a 100-μm I.D. silica capillary. The flow-rate for direct introduction was typically 5 μl/min and the sample concentration between 1 and 100 pmol/μl. The voltages at the tip of the capillary, used in the ionisation process at atmospheric pressure, were 4000 and 5000 V for ES-MS and IS-MS respectively. In IS-MS experiments, a nebulizer gas was added along the tip of the capillary (nitrogen, flow-rate 0.6 to 1.2 l/min). Data were acquired in the MCA mode. The VG BioQ was calibrated using multiply charged ions of horse myoglobin and the Sciex API-III with polypropyleneglycol, in order to get a one mass unit resolution in the mass range scanned.

LC–MS experiments were carried out using the Sciex API-III (since this instrument allows routinely flow-rates up to 200 μl/min and therefore permits to use a lower split ratio at the outlet of the column); the flow-rate injected into the ion source was 40 μl/min. Acquisition was done in the profile mode (mass range : 400 to 1800 amu in 3 s/scan with a step size of 0.1 amu). LC–MS data could be presented as plots of m/z against retention time or total ionic current (TIC) chromatograms. Mass spectra of peptides were reconstituted by averaging the scans (*ca.* 10, 30 s), which are characteristic of peptide retention time while the background from selected regions (*ca.* 10 scans) of both sides of the peaks was subtracted.

2.2. Microsequencing

Amino terminal sequences of peptides and proteins were determined using an Applied Biosystems 477A instrument. Ten to 300 pmol samples were used for each analysis.

2.3. Trypsin digestion of RG 12986

RG12986 was solubilized at a 1 mg/ml concentration in 67 m*M* ammonium bicarbonate pH 7.8 and trypsin (sequencing grade, Boehringer Mannheim) was added to 1.5 wt.% of the protein. The mixture was incubated overnight at 37°C and then evacuated *in vacuo*. The sample was solubilized in 6 *M* aqueous guanidinium chloride at a protein concentration of 2 mg/ml to avoid the presence of insoluble material. The sample was analyzed by reversed-phase HPLC on a Vydac C_{18} 300 Å (250 × 2.1 mm I.D.) column eluted with a linear gradient from 0 to 35% acetonitrile in water containing 0.07% trifluoracetic acid (TFA) in 140 min at a flow-rate of 0.2 ml/min. Peptides were detected by their absorbance at 215 nm. Peptide containing fractions were evacuated *in vacuo* and solubilized in 40 to 200 μl of water–methanol–acetic acid (50:50:1, v/v) on the basis of peak heights, 20 μl were analyzed by ES-MS and 20 μl by microsequencing.

2.4. V8 protease digestion of HSA

Recombinant and natural (Institut Mérieux) HSA were solubilized at a concentration of 1 mg/ml in 50 m*M* ammonium acetate pH 4.0 buffer; under such conditions, V8 protease is known to cleave only Glu–Xaa bonds [20]. The protein solution was incubated at 37°C and equal amounts of V8 protease (Boehringer Mannheim) were added at $t = 0$, 3, 6, 9, 24, 27, 30 and 48 h to yield a final protease/HSA ratio of 1/5 by weight. The digestion was stopped at $t = 52$ h by addition of a final concentration of 5% TFA. The samples were evacuated *in vacuo* and solubilized in 6 *M* aqueous guanidinium chloride at a protein concentration of 2 mg/ml to avoid the presence of insoluble material. The samples were analyzed by reversed-phase HPLC on a Waters Delta-Pak C_{18} 300 Å (150 × 4 mm I.D.) column eluted with a linear gradient from 0 to

40% of acetonitrile in water in the presence of 0.07% TFA in 120 min at a flow-rate of 1.0 ml/min. Peptides were detected by their absorbance at 220 nm. For the combined mass spectrometry and microsequencing experiments, 400-μg aliquots of digested material were injected and 1-min fractions were collected. Peptide containing fractions were analyzed by microsequencing and ES-MS as described above. When LSI-MS experiments were conducted, fractions were reconstituted in a minimal volume of 10% aqueous TFA (usually 5 μl) and deposited on the instrument probe. For LC-MS experiments, 600 μg of digested material were injected and the column was washed for 20 min with water containing 0.07% TFA at a flow-rate of 1.0 ml/min to elute guanidinium chloride. The outlet of the column was then connected to a 1/25 splitting device and the splitted flow was injected into the ion source of the mass spectrometer. The separation was then conducted using the gradient described above. In these conditions, 10 to 50 pmol of each individual peptide reached the spectrometer.

2.5. Cyanogen bromide cleavage of HSA

Recombinant HSA or the 45-kDa HSA fragment were solubilized at a 2 mg/ml concentration in 70% aqueous formic acid containing 20 mg/ml of CNBr (Eastman Kodak). Under these conditions, CNBr cleaves proteins at the C-terminus of methionine residues, transforming methionine into homoserine lactone [21]. The CNBr solution has to be freshly prepared in nitrogen bubbled aqueous formic acid. The reaction was left to proceed overnight at room temperature. The reaction was then quenched by a 5-fold dilution in water and the samples were evacuated *in vacuo*. Samples were then solubilized in 6 *M* guanidinium chloride, 1 *M* dithiothreitol, 0.25 *M* ammonium bicarbonate pH 7.8 and heated at 95°C for 15 min in order to reduce disulfide bonds and cleave the formiate adducts formed during formic acid incubation. The samples were then analyzed by reversed-phase HPLC on an Applied Biosystems C_4 300 Å column (30 × 4.6 mm I.D.) eluted by a gradient of acetonitrile in water in the presence of 0.07% TFA at a flow-rate of 1.0 ml/min. This gradient was constituted of linear stretches from 0 to 15% acetonitrile in 10 min, 15 to 33% in 30 min, and 33 to 38% in 30 min. One-minute fractions were collected and analyzed by ES-MS and microsequencing as described above.

3. Results and discussion

3.1. Analysis of RG12986

Molecular mass measurement of the intact molecule

A typical ES-MS spectrum of the cysteine alkylated vWF fragment RG12986 is shown in Fig. 1. The measured molecular mass for this compound is 32 522 for a calculated value of 32 518. Nevertheless, the spectrum is not totally resolved and the observed peaks are too broad to exclude microheterogeneities. The narrow scan analysis of the molecule (data no shown) does not enable to assign the observed peak broadening to proteinaceous heterogeneities or ionic adducts. This sole determination is not sufficient to fully characterize the molecule.

Tryptic map of RG12986

The chromatogram obtained after digestion of RG12986 with trypsin is shown in Fig. 2 and the position of the fragments in the RG12986 sequence is shown in Fig. 3. All the peaks were analyzed by ES-MS and microsequencing. From the data 99% of the primary structure of the molecule could be checked. The tripeptide corresponding to the amino terminus of RG12986 has not been detected in this experiment; this fragment is very hydrophillic and should elute in the injection peak. Nevertheless the presence of the 3 amino terminal residues has been confirmed by the sequencing of the intact molecule. The pentapeptide T34–T35 (RNSMV) which corresponds to the carboxy terminus of RG12986 has been detected and correctly characterized by microsequencing in three different fractions

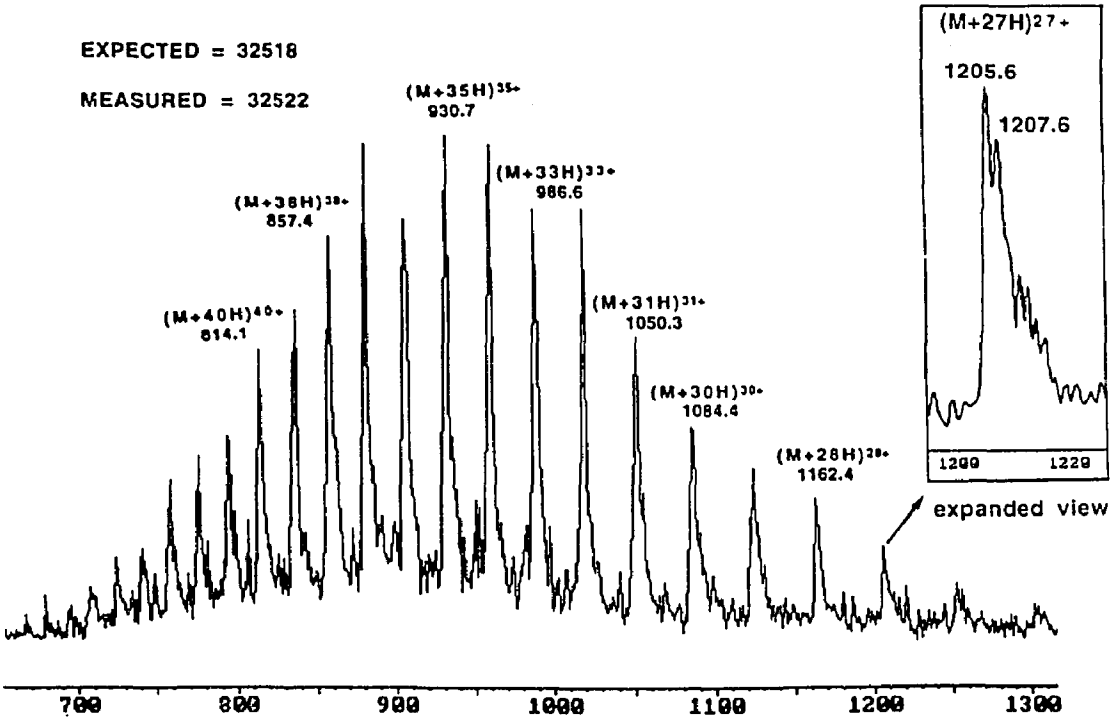

Fig. 1. ES-MS spectrum of a HPLC purified batch of Rg12986.

(peaks 7, 8 and 12). By ES-MS, the expected molecular mass (605.7) has been detected in peak 12 but the masses measured for peaks 7 and 8 (668 and 399 respectively) could not be interpreted. Some of the molecular mass measurements showed adducts of 16 and 32 mass units (T6: peaks 17 and 22, T33: peaks 36 and 37), suggesting oxydation of methionine residues into sulfoxydes or sulfone. Since RG12986 has no natural counterpart that can be used as a reference, the origin of these putative oxydations cannot be identified. They could occur during product processing as well as during trypsin fragmentation. Since the analysis of the intact molecule by ES-MS yielded the expected molecular mass, it can be concluded that RG12986 is processed mainly in the non-oxydized form, and that the oxydation of methionine residues occurs during trypsin fragmentation and work up of the fragments. This analysis made it possible to verify that all the cysteine residues were alkylated.

3.2. Analysis of recombinant HSA

Molecular mass measurement of the intact molecule

Spectra of several batches of natural and recombinant HSA of different origins have been obtained. A typical example is shown in Fig. 4 for recombinant HSA. All spectra exhibit a similar pattern with broad peaks even for highly purified material. It should be noted that the quality of spectra was correlated with the purity of the analyzed material, and that crude recombinant HSA yielded unworkable spectra. Most of the time, spectra exhibit strong heterogeneity showing adducts with masses up to 200–300 amu higher than the expected value. This behavior can be at least partially attributed to cation adducts since albumin is known to have a high affinity for these ions [15] and the presence of adducts on the free cysteine of HSA. Another source of peak broadening could be the presence of phosphate and sulfate adducts as noted by

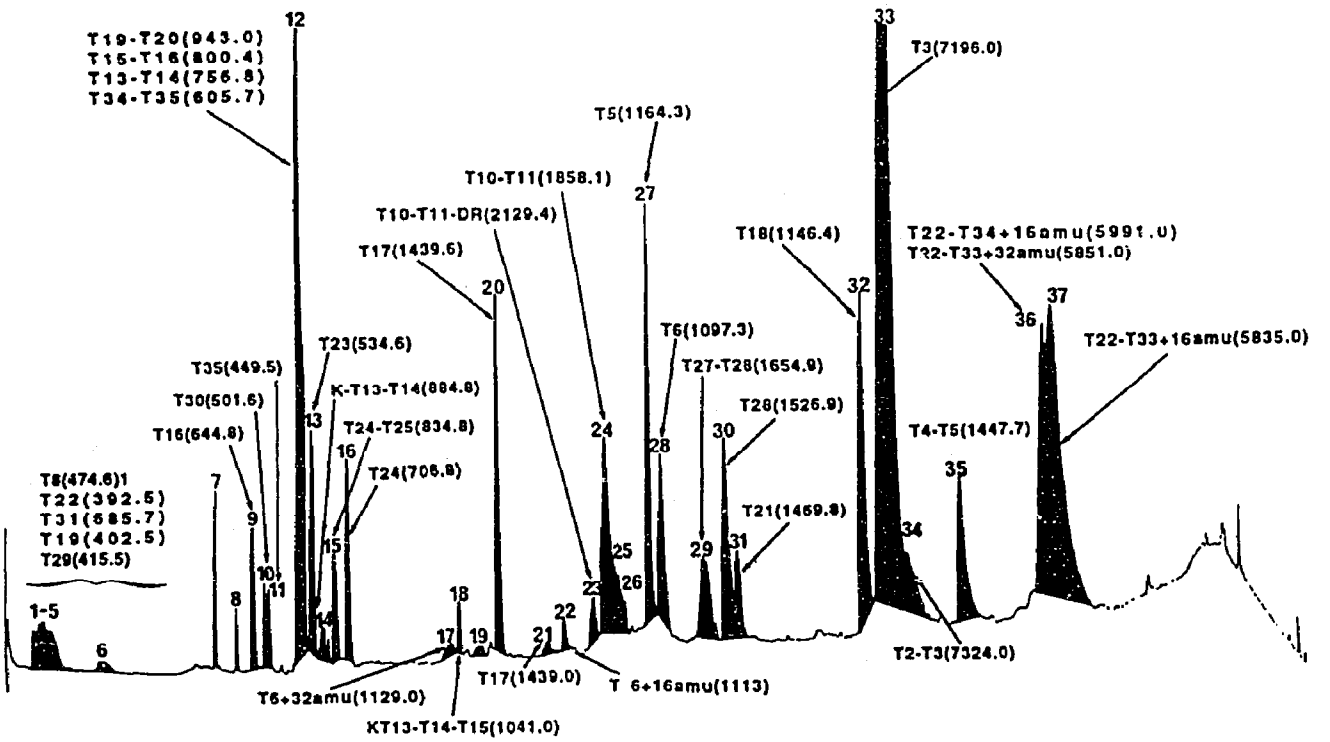

Fig. 2. Reversed-phase HPLC separation of tryptic peptides of Rg12986. Measured mass and sequence position (with respect to Fig. 3 nomenclature) of the fragments are indicated on each peak.

others for numerous proteins [22]. The higher resolution that can be obtained with a quadrupole analyser is far too low to fully interpret the spectra obtained. Therefore a more detailed characterization of HSA has to go through proteolytic cleavage.

V8 protease map of natural and recombinant HSA

Recombinant and natural HSA have been cleaved with V8 protease as described in the Experimental section. Chromatograms of both are compared in Fig. 5. The combined analysis of peptidic fractions by microsequencing and mass spectrometry enabled us to check 96% of the primary structure of the molecule with an accuracy of 1 amu (see Fig. 6 for the structure of the identified fragments). Only one tripeptide (Val_{493}–Glu_{495}) was not identified during this study. As for RG12986, this peptide should be too hydrophillic to be retained on the reversed-phase column. The peptide Ile_{25}–Glu_{37}, bearing

SGK K VTLNPSDPEHCQICHCDVVNLTCEACQEPGGLVVPPTDAPVSPTTLYVEDISEPPL
T1 T2 T3

HDFYCSR LLDLVFLLDGSSR LSEAEFEVLK AFVVDMMER LR ISQK WVR
T4 T5 T6 T7 T8 T9

VAVVEYHDGSHA YIGLK DRK R PSELR R IASQVK YAGSQVASTSEVLK
T10 T11 T12 T13 T14 T15 T16 T17

YTLFQIFSK IDR PEASR IALLLMASQEPQR MSR NFVR YVQGLK K K
T18 T19 T20 T21 T22 T23 T24 T25 T26

K VIVIPVGIGPHANLK QIR LIEK QAPENK AFVLSSVDELEQQR
T27 T28 T29 T30 T31 T32

DEIVSYLCDLAPEAPPPTLPPDMAQVTVGPGLLGVSTLGPK R NSNV
T33 T34 T35

Fig. 3. Amino acid sequence and tryptic cleavage fragments of Rg12986.

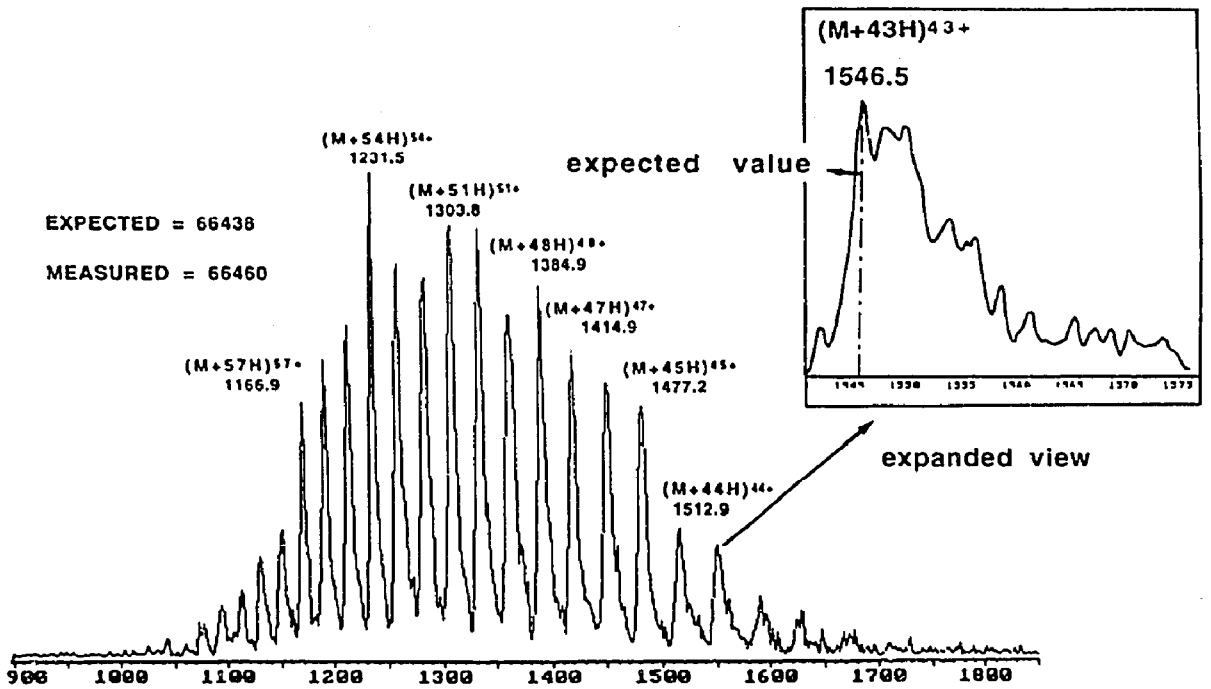

Fig. 4. ES-MS spectrum of a HPLC purified batch of recombinant HSA.

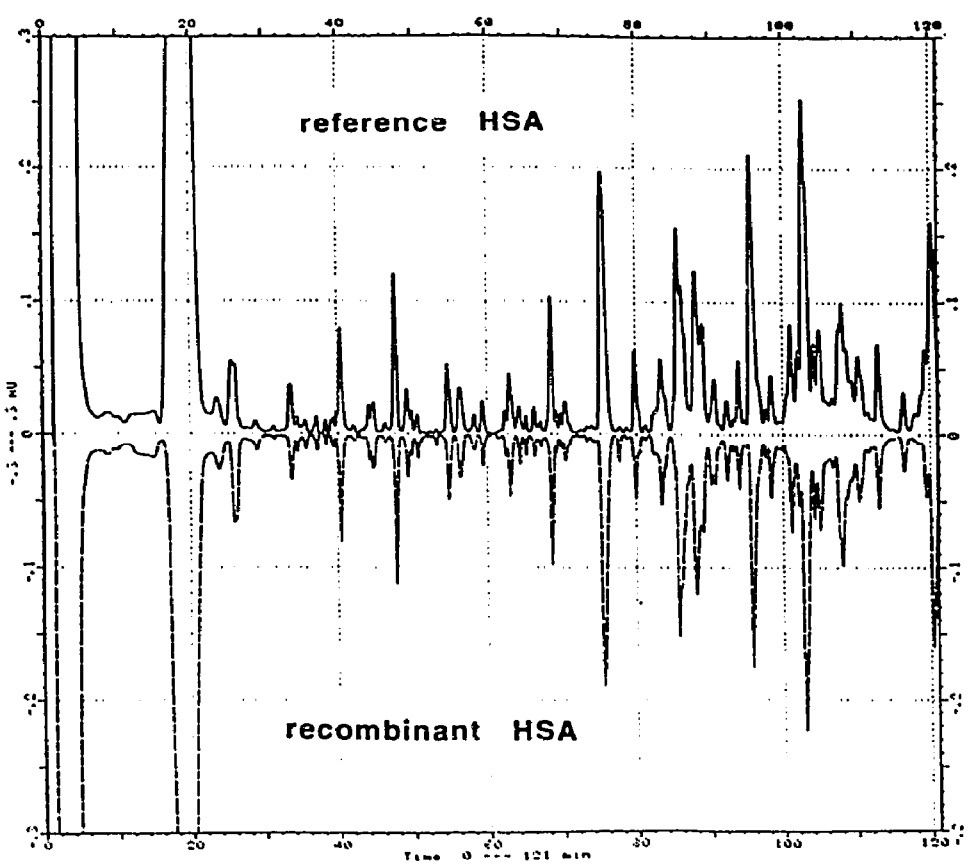

Fig. 5. Comparison of the V8 protease peptide map (reversed-phase C_{18} column) of reference and recombinant HSA.

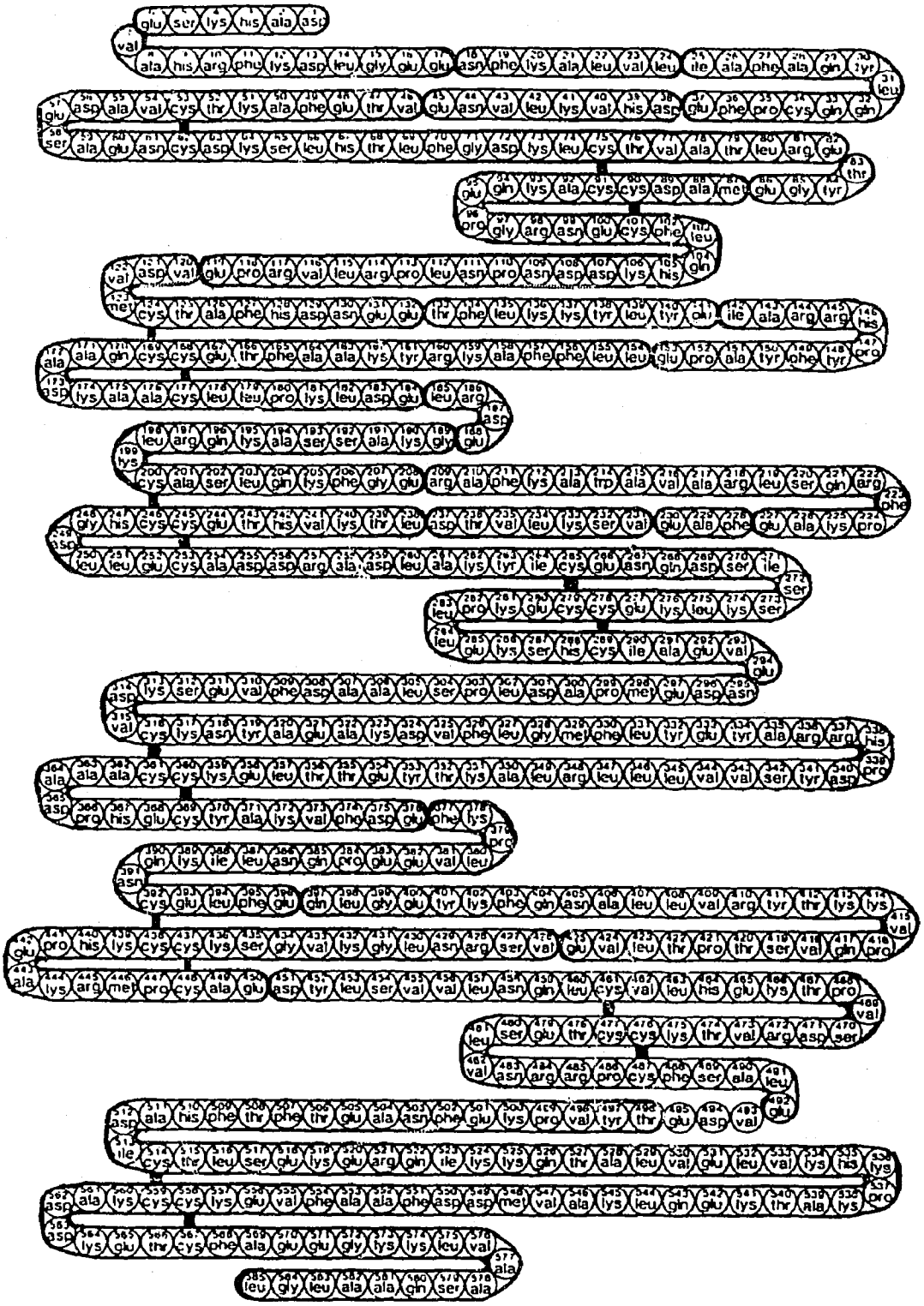

Fig. 6. Primary sequence covered for reference and recombinant HSA, combining Edman degradation and mass spectrometry. It represents 99% of the protein.

the free cysteine, has been fully characterized by microsequencing (retention time = 90 min in Fig. 5) but could neither be detected by ES-MS nor LSI-MS of the corresponding fraction. Taking into account microsequencing data, more than 99% of the primary structure of recombinant HSA has been verified. All the disufide-containing peptides have been detected with the molecular mass of their oxidized form. Unexpected proteolytic cleavages of the molecule have been observed. For example the cleavage of Asp_{237}–Leu_{238} bond occurs even though acidic conditions have been used for proteolysis; under these conditions, V8 protease is known to cleave only Glu–Xaa bonds and not Asp–Xaa bonds [20]. The cleavage of Leu_{24}–Ile_{25} bonds can be attributed to the activity of a contaminating protease in the V8 enzyme preparation since it corresponds to a different enzymatic specificity. Since the same unexpected fragments have been observed for the reference natural HSA, these observations do not question the integrity of the recombinant product.

The peptide map of recombinant HSA was also analyzed by LC–MS (Fig. 7). In this experiment, all the fragments detected by ES-MS analysis of chromatographic fractions have been characterized with a 1-amu accuracy. Knowing the retention time and structure of the HSA fragments from the previous experiment, 96% of the primary structure of the recombinant HSA could be verified in a single 2-h experiment. Moreover, a plot of m/z against retention time gives a characteristic finger print of the protein.

Assignment of disulfide bonds

A feature specific to albumin is its complex disulfide bond pattern. In the course of the characterization of recombinant HSA, the correct maturation of these disulfide bonds with respect to the published model [17] has to be assessed. This verification can be made on the V8 protease map of recombinant HSA. Comparison of the retention time of each individual fragment with those of the reference HSA is a first criterion. By sequence analysis disulfide bonded fragments are identified by the presence of two or three individual sequences of similar intensity. The mass spectrometric analysis of these fragments has been conducted using ES-MS and LSI-MS techniques, as shown in Fig. 8 for the fragment $[Asp_{451}–Glu_{465}]–(Cys_{461}–SS–Cys_{477})–[Lys_{466}–Glu_{492}]$. In the LSI-MS experiment, the oxydized peptide is detected together with characteristic fragmentation corresponding to the cleavage of the disulfide bond (Fig. 8A) while the ES-MS experiment detects only the oxydized form of the peptide (Fig. 8B). On the other hand, the reduction of the peptide can be observed by ES-MS after a 1-h treatment of the oxydized peptide with 50 to 100 m*M* dithiothreitol in ES-MS buffer without acetic acid prior to MS analysis (data not shown). However, characteristic for the albumin disulfide bond pattern (Fig. 6) is the presence, at several points in the sequence, of two adjacent cysteine residues, both of which participate in the formation of separate disulfide bonds (the so-called double Cys bridge, [16]). Using the approach described here, it is not possible to assign the position of the disulfide bonds to one or other adjacent cysteines. The exact assignment of these bonds could be done only after a new digestion of the peptides at the level of the Cys–Cys peptide bonds using for example thermolysin protease [23].

Characterization of a fragment of recombinant HSA

Some batches of recombinant HSA contained, as an impurity, an additional protein with an apparent molecular mass of 45 kDa. By microsequencing, this impurity showed the same N-terminal sequence as HSA. The purified contaminant was studied directly by ES-MS; however, this gave an unworkable spectrum suggesting an heterogeneous product. For further characterization, the purified contaminant was cleaved with cyanogen bromide as described in the Experimental section in parallel with recombinant HSA as a control. The chromatograms obtained are shown in Fig. 9. The peptide-containing fractions were analyzed by microsequencing and ES-MS. For the 45 kDa impurity all the fractions contained CNBr fragments of HSA except fraction 9 (see Fig. 10). These CNBr

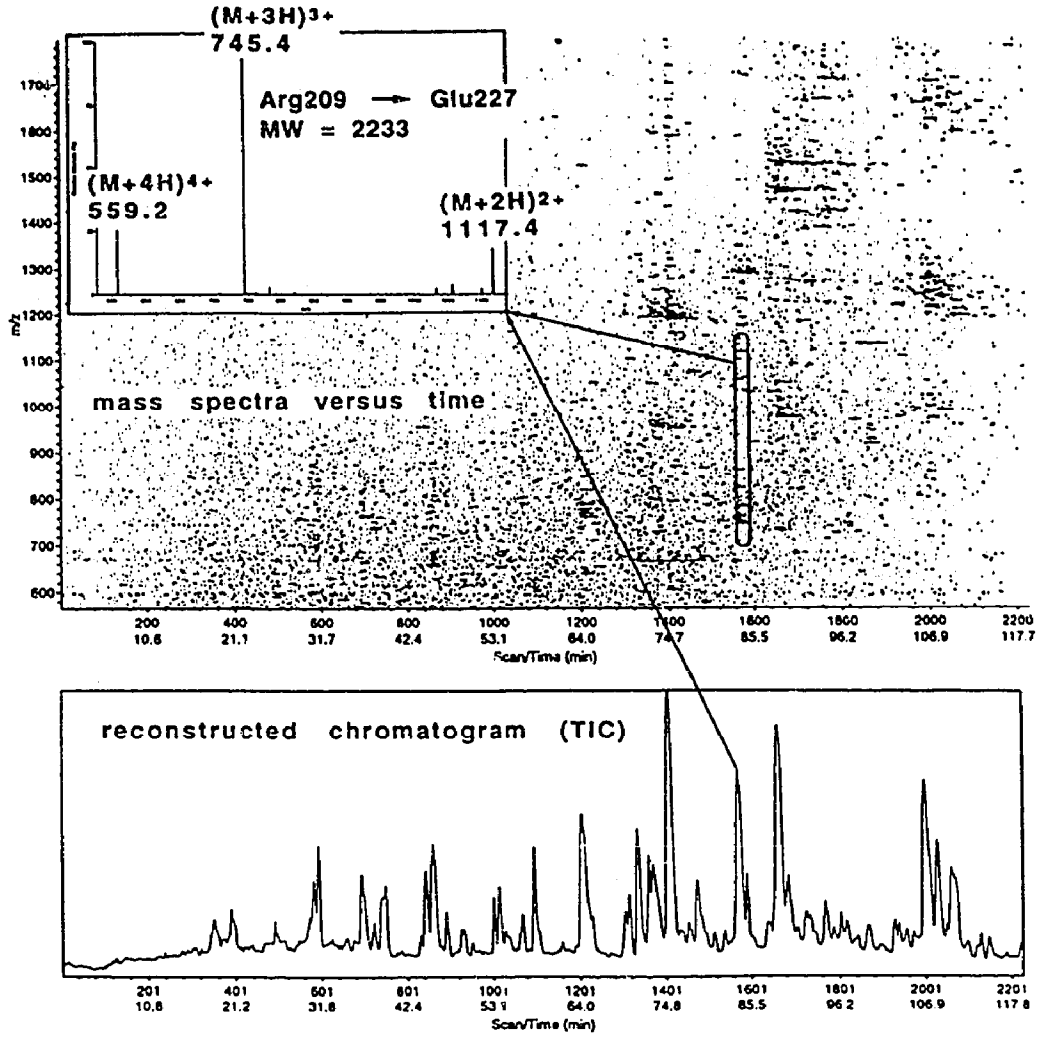

Fig. 7. Typical V8 protease peptide map obtained by LC–MS on HSA (recombinant HSA in this particular case). The upper representation shows the raw data obtained by plotting the *m/z* values of each scan *versus* the time (the contrast is correlated to the intensity of the signal detected for each *m/z* value). The corresponding total ionic current (TIC) is shown below. As an example, the spectrum obtained at a retention time of 83.5 min, which has permitted the determination of the sequence Arg_{209}–Glu_{227} is presented in the upper left frame.

fragments were detected as the homoserine form of their C-terminal methionine, the lactone being hydrolyzed during alkaline treatment of the hydrolysate. For fraction 9, a homogeneous N-terminal sequence has been determined corresponding to Phe_{330} of HSA as a result of the CNBr cleavage of the Met_{329}–Phe_{330} bond. When analyzed by ES-MS, fraction 9 showed the presence of three main products Phe_{330}–Tyr_{401}, Phe_{330}–Lys_{402}, Phe_{330}–Ala_{406} (Fig. 10). Since the cleavage at Tyr_{401}, Lys_{402} and Ala_{406} could not have been caused by CNBr action and have not been found in the CNBr hydrolysate of recombinant HSA, it can be concluded that the 45 kDa impurity is a mixture of at least 3 fragments of HSA (1-401, 1-402 and 1-406). Others [24] have also detected 45 kDa fragments in recombinant HSA expressed in yeast and have located their C-termini in the same region of the HSA sequence. Further evidence is afforded by the identification of the fragments 549–585 and 447–548 located downstream of the C-terminus

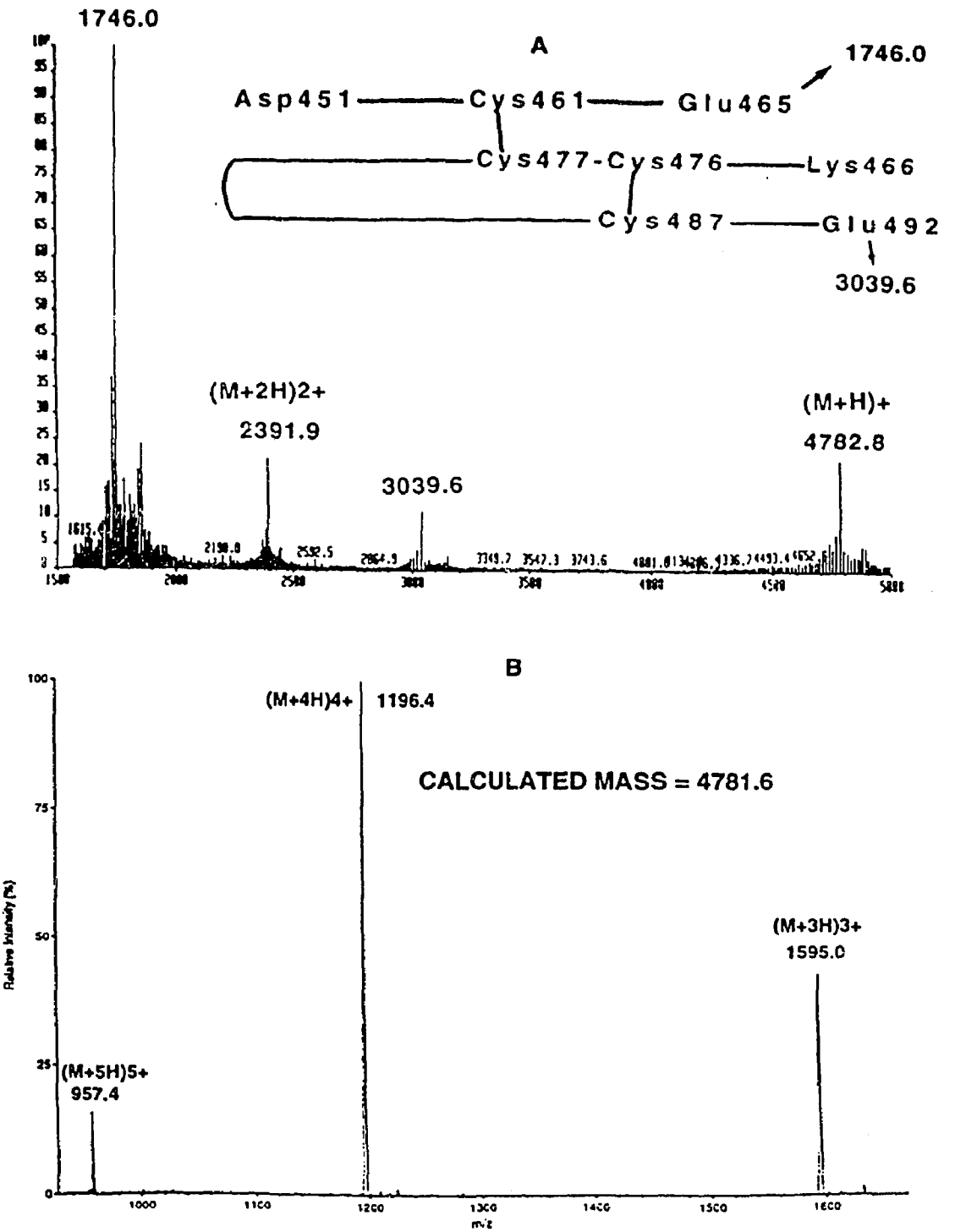

Fig. 8. Comparison of LSI-MS (A) and ES-MS (B) spectra of a peptide containing two disulfide bonds.

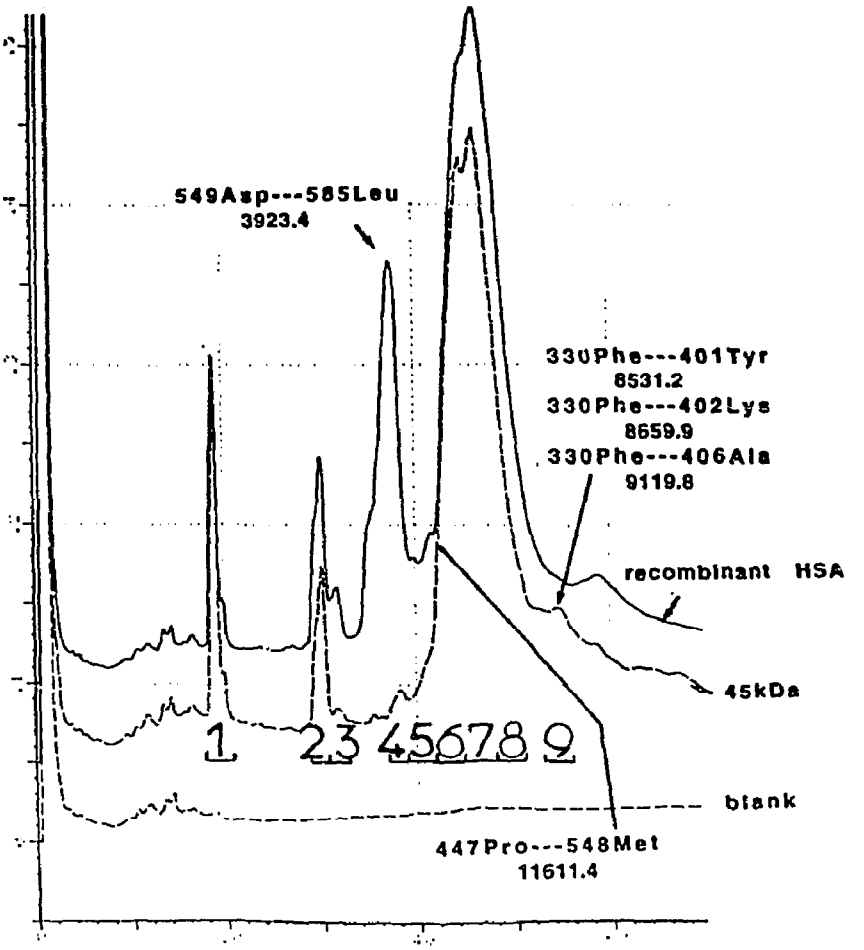

Fig. 9. CNBr peptide map of recombinant HSA and the 45 kDa impurity. The molecular masses of the differences observed, as well as the amino acid sequences, are indicated.

of the 45 kDa impurity. These fragments are present in the CNBr hydrolysate of recombinant HSA (fraction 4 and 6 respectively in Fig. 9) and have not been detected in the hydrolysate of the 45 kDa impurity. This characterization suggests that the recombinant HSA is cleaved during the course of its secretion by yeast endoproteases. The heterogeneity of the C-terminus could be the result of the action of the carboxypeptidase Y. In fact this exopeptidase follows the same secretion pathway as recombinant HSA. Especially, they are both present at high local concentration in the vacuoles [19].

In addition the fragment 447–548 contains the tripeptide 493–495 and the fragment 1–87 (molecular mass 9809, fraction 7 in Fig. 9) encompasses the peptide 25–37. These 2 peptides have not been detected in the V8 protease map of recombinant HSA. By combination of the two peptide maps (V8 protease and CNBr), 100% of the primary structure of recombinant HSA have been verified.

4. Conclusions

The detailed study of the primary structure of RG12986 and recombinant HSA confirmed their integrity. The completeness of the alkylation of cysteine residues of RG12986 was verified by mass spectrometry. In the case of recombinant HSA, our experiments confirmed the correct

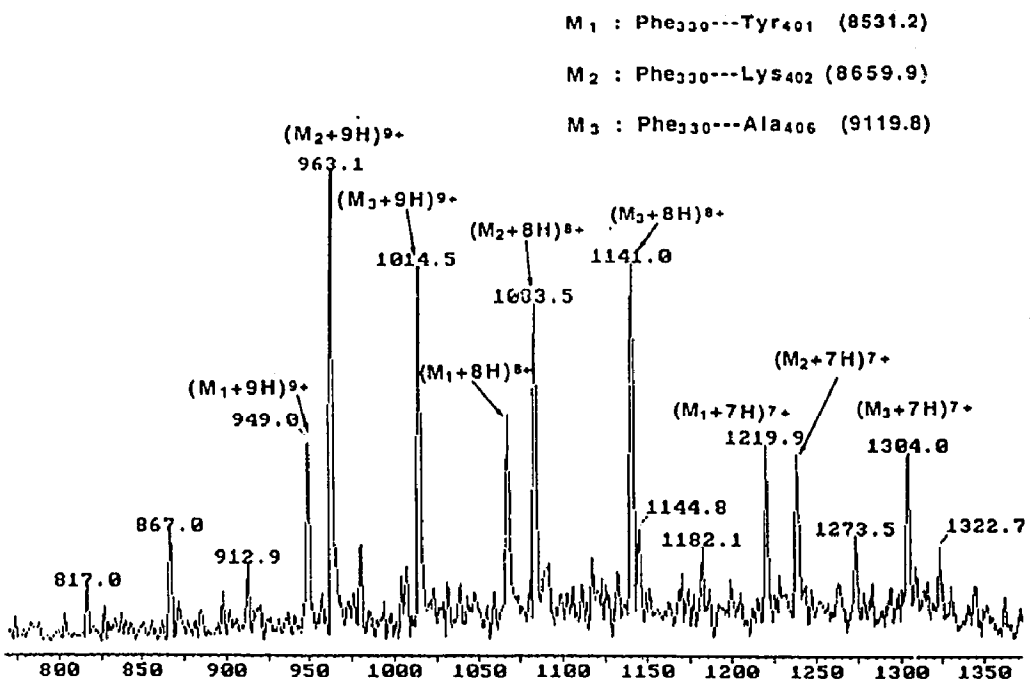

Fig. 10. ES-MS spectrum of fraction 9 of the 45 kDa impurity CNBr cleavage (see Fig. 10). It clearly shows the presence of three main compounds which have been identified as Phe_{330}–Tyr_{401}, Phe_{330}–Lys_{402} and Phe_{330}–Ala_{406}.

maturation of the 17 disulfide bonds of the molecule and led to the characterization of the structure of an albumin 45 kDa fragment produced as a contaminant. The characterization of this impurity will guide genetic work to reduce its occurrence. For each of the two proteins 100% of the primary structure has been verified by combination of all experiments and techniques used. In the case of RG12986, a single tripeptide was not detected in the tryptic map, but was identified by sequence analysis of the intact protein. In the case of recombinant HSA, two peptides could not be detected in the V8 protease map, but these peptides were included in CNBr fragments that have been identified by ES-MS.

Mass spectrometric analysis of RG12986 and HSA was carried out on intact molecules as well as on proteolytic fragments produced by enzymatic or chemical treatments. Although analysis of intact proteins can easily be done by ES-MS and is useful for batch-to-batch control, this global measurement has too low a resolution to assess the chemical integrity of a compound with a molecular mass exceeding 40 000 Da. Mass spectrometric analyses of proteolytic fragments separated by HPLC can fulfil this purpose. In our studies all the proteolytic fragments analyzed were characterized with an accuracy of 1 amu. During the course of this work, microsequencing appeared to be indispensable for the interpretation of mass spectra. On one hand, it helped identification of unexpected fragmentations resulting from deviations of the protease specificity or from contaminating proteases. On the other hand, for large proteins like HSA, there could be more than one possible HSA fragment structure for a given mass ± 1 amu. This fact is related to the complexity induced by the disulfide bond pattern. Sequence information can also be obtained by tandem mass spectrometry [25–27], but this technique does not always yield data easy to interpret (for instance in the case of large disulfide-linked peptides). Considering the interpretation of proteolytic fragments, the availability of a natural reference product (HSA purified from human blood in our case) is of utmost importance. In the case of RG12986 which has

no natural counterpart, the origin of the observed oxydation of methionine residues cannot be investigated with precision. A crucial point in this study was to set up proteolytic cleavage protocols that were consistent over time and samples. For this reason, kinetic studies have to be performed in order to optimize the method. Even with an optimized protocol, the proteolysis is not always complete. In the V8 protease map of HSA, some Glu residues were not cleaved (see Fig. 6).

However, the procedure consisting of production of proteolytic fragments, HPLC separation and analysis by microsequencing and mass spectrometry can not be used for batch-to-batch control of recombinant proteins. In the case of HSA, more than 80 individual chromatographic fractions were analyzed. This took hours of mass spectrometry and weeks of sequencing work. But once the structure and retention time of each individual fragment are characterized, a single LC–MS experiment can produce the same information within two hours and can then be used for batch-to-batch control. Owing to its versatility, LC–MS could also be used for the study of the polymorphism of proteins. When compared to a reference wild type, mutants could be easily detected on plots of m/z against retention time and the mutation could be located on a particular proteolytic fragment. Indeed, software has to be developed for the automation of LC–MS data interpretation.

The complete characterization of the primary structure of a recombinant protein can not always be achieved with a single proteolytic fragmentation. These reactions can produce peptides that are too hydrophillic for the HPLC separation (*e.g.* fragment 1–3 of RG12986) or that can not be correctly ionized (*e.g.* fragment 25–37 of HSA). The use of two complementary cleavage methods (V8 protease and CNBr in this study) considerably lowered the probability that a given fragment could not be detected in one of the two different peptide maps.

Considering the mass spectrometric techniques used, ES-MS appears to be most suitable for protein and peptide analysis, since it is *ca.* a 100-fold more sensitive than LSI-MS. LSI-MS affords more structural information especially in the case of disulfide bonded peptides, but similar information can be obtained by reduction of disulfide bridges prior to ES-MS analysis. LSI-MS appears to be appropriate for high resolution analysis [4] when sensitivity is not the main concern.

In conclusion, the combination of mass spectrometry with "classical" primary structure determination techniques (microsequencing and peptide mapping) is a key methodology in the characterization of recombinant proteins and has become a standard in this field. The accuracy of mass spectrometry is unequalled by other techniques. Nevertheless, primary structure determinations have to be completed by other biochemical and spectroscopic methods to investigate protein tridimensional structure and detect trace amount of contaminants.

Acknowledgements

This work was supported by the BioAvenir program (Ministère de la Recherche et de l'Espace, Ministère de l'Industrie et du Commerce Extérieur) and by the Institut Mérieux for the work on HSA. B. Green (VG Biotech., Altrincham, UK) is warmly acknowledged for fruitful discussions in designing some MS experiments.

References

[1] M.J. Geisow, *TIBTECH*, 10 (1992) 432–441.
[2] C. Fenselau, M.M. Vestling and R.J. Cotter, *Curr. Opin. Biotechnol.*, 4 (1993) 14–19.
[3] N. Barber and B.N. Green, *Rapid Commun. Mass Spectrom.*, 1 (1987) 80–83.
[4] A. Van Dorsselaer, P. Lepage, F. Bitsch, O. Whitechurch, O. Riehl-Bellon, D. Fraisse, B. Green and C. Roitch, *Biochemistry*, 28 (1989) 2949–2956.
[5] S.A. Carr, M.E. Hemling, G. Folena-Wasserman, R.W. Sweet, K. Anumula, J.R. Barr, M.J. Huddleson and P. Taylor, *J. Biol. Chem.*, 264 (1989) 21286–21295.
[6] B. Monégier, F.F. Clerc, A. Van Dorssealer, M. Vuilhorgne, B. Green and T. Cartwright, *BioPharm*, 2 (1990) 26–35.
[7] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802–2824.

[8] A. Van Dorsselaer, F. Bitch, B. Green, S. Jarvis, P. Lepage, R. Bischoff, H.V.J. Kolbe and C. Roitch, *Biomed. Environ. Mass Spectrom.*, 19 (1990) 692–704.
[9] P.A. Schindler, A. Van Dorsselaer and A.M. Falick, *Anal. Biochem.*, 213 (1993) 256–263.
[10] M.M. Siegle, I.J. Hollander, P.R. Hamann, J.P. James, L. Hinman, B.J. Smith, A.P.H. Farnsworth, A. Phipps, D.J. King, M. Karas and F. Hillenamp, *Anal. Chem.*, 63 (1991) 2470–2481.
[11] K.K. Mock, M. Dave and J.S. Cottrel, *Biochem. Biophys. Res. Commun.*, 177 (1991) 644–651.
[12] C. Prior, V. Chu, J. Holt, V. Windisch, T. Lee, J. Mitschelen, J. Newman, G. Ricca, C. Tarr and M. Hrinda, *Bio/Technology*, 10 (1992) 66–72.
[13] H. Mohri, Y. Fujimura, M. Shima, A. Yashioka, R. Houghten, Z. Ruggeri and T. Zimmerman, *J. Biol. Chem.*, 263 (1988) 17901–17904.
[14] H. Mohri, A. Yashioka, T. Zimmerman and Z. Ruggeri, *J. Biol. Chem.*, 264 (1989) 17361–17367.
[15] T. Peters, *Adv. Protein Chem.*, 37 (1985) 161–245.
[16] J.R. Brown, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 35 (1976) 2141–2144.
[17] A. Dugaiczyk, S.W. Law and O.E. Dennison, *Proc. Natl. Acad. Sci. USA*, 79 (1982) 71–75.
[18] B. Meloun, L. Mararek and V. Kostka, *FEBS Lett.*, 58 (1977) 134–137.
[19] R. Fleer, P. Yeh, A. Amellal, C. Maury, A. Fournier, F. Bachetta, P. Borduel, G. Jung, J. Becquart, H. Fukuhara and J.F. Mayaux, *Bio/Technology*, 9 (1991) 968–975.
[20] J. Houmard and G.R. Drapeau, *Proc. Natl. Acad. Sci. USA*, 69 (1972) 3506–3509.
[21] E. Gross, *Methods Enzymol.*, 11 (1967) 238–255.
[22] S.K. Chowdhury, V. Katta, R.C. Beavis and B. Chait, *J. Am. Soc. Mass Spectrom.*, 1 (1990) 382–388.
[23] P. Lepage, F. Bitsch, D. Roecklin, E. Keppi, J.L. Dimarcq, J.M. Reichhart, J.A. Hoffmann, C. Roitsch and A. Van Dorsselaer, *Eur. J. Biochem.*, 196 (1991) 735–742.
[24] D. Sleep, G.P. Belfield and A.R. Goodey, *Bio/Technology*, 8 (1990) 42–46.
[25] D.F. Hunt, H. Michel, T.A. Dickinson, J. Shabanowitz, A.L. Cox, K. Sakaguchi, E. Appella, H.M. Grey and A. Sette, *Science*, 256 (1992) 1817–1820.
[26] C.H.L. Shackelton, A.M. Falick, B. Green and Witkowska, H.E., *J. Chromatogr.*, 562 (1991) 175–190.
[27] K. Biemann, *Ann. Rev. Biochem.*, 61 (1992) 977–1010.