Markwardt, F. (1957) *Hoppe-Seyler's Z. Physiol. Chem. 308*, 147–156.
Markwardt, F. (1970) *Methods Enzymol. 19*, 924–932.
Markwardt, F., & Walsmann, P. (1958) *Hoppe-Seyler's Z. Physiol. Chem. 312*, 85–98.
Markwardt, F., Nowak, G., & Hoffmann, J. (1977) *Thromb. Res. 11*, 275–283.
Markwardt, F., Hauptmann, J., Nowak, G., Klessen, Ch., & Walsmann, P. (1982) *Thromb. Haemostasis 47*, 226–229.
Markwardt, F., Nowak, G. Stürzebecher, J., Griessbach, U., Walsmann, P., & Vogel, G. (1984) *Thromb. Haemostasis 52*, 160–163.
Mürer, E. H. (1972) *Biochim. Biophys. Acta 261*, 435–443.
Naylor, S., Findeis, F., Gibson, B. W., & Williams, D. H. (1986) *J. Am. Chem. Soc. 108*, 6359–6363.
Nowak, G., & Markwardt, F. (1980) *Exp. Pathol. 18*, 438–443.
Petersen, T. E., Roberts, H. R., Sottrup-Jensen, L., Magnusson, S., & Bagdy, D. (1976) *Protides Biol. Fluids 23*, 145–148.
Richter, W. J., Raschdorf, F., & Maerki, W. (1985) *Mass Spectrom. Health Life Sci. 24*, 193–206.
Roitsch, C., & Barnes, M. (1985) *Immunol. Methods 3*, 85–110.
Singh, A., Lugovoy, J. M., Kohr, W. J., & Perry, L. J. (1984) *Nucleic Acids Res. 12*, 8927–8938.
Stone, S. R., & Hofsteenge, J. (1986) *Biochemistry 25*, 4622–4628.
Sukumaran, D. K., Clore, G. M., Preuss, A., Zarbock, J., & Gronenborn, A. M. (1987) *Biochemistry 26*, 333–338.
Tarr, G. E. (1986) in *Methods of Protein Microcharacterization: A Practical Handbook* (Shively, J. E., Ed.) pp 155–194, Human Press, Clifton, NJ.
Tripier, D. (1988) *Folia Haemetol. 115*, 530–35.
Van Dorsselear, A., Lepage, P., Bitsch, F., Whitechurch, O., Riehl-Bellon, N., Fraisse, D., Green, B., & Roitsch, C. (1989) *Biochemistry* (following paper in this issue).
Vlasuk, G. P., Bencen, G. H., Scarborough, R. M., Tsai, P. K., Whang, J. L., Maack, T., Camargo, M. J. F., Kirsher, S. W., & Abraham, J. A. (1986) *J. Biol. Chem. 261*, 4789–4796.
Vogel, G., & Markwardt, F. (1987) *Thromb. Res. Suppl. 7*, 42.
Walsmann, P., & Markwardt, F. (1981) *Pharmazie 10*, 653–659.
Zsebo, K. M., Lu, H. S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K., Suggs, S. V., Lai, P. H., & Bitter, G. A. (1986) *J. Biol. Chem. 261*, 5858–5865.

Mass Spectrometry Analyses of Recombinant Hirudins (7 kDa)

Alain Van Dorsselaer,*,‡ Pierre Lepage,§ Francis Bitsch,‡ Odile Whitechurch,§ Nadine Riehl-Bellon,§ Daniel Fraisse,∥ Brian Green,⊥ and Carolyn Roitsch*,§

Laboratoire de Chimie Organique des Substances Naturelles, Centre de Neurochimie, UA31 CNRS, 5 Rue B. Pascal, 67084 Strasbourg Cedex, France, Transgène S.A., 11 Rue de Molsheim, 67000 Strasbourg, France, SCA-CNRS, 69390 Vernaison, France, and VG Analytical Ltd., Floats Road, Wythenshawe, Manchester M23 9LE, England

Received June 1, 1988; Revised Manuscript Received September 29, 1988

ABSTRACT: The use of liquid secondary ion mass spectrometry (LSIMS) in the characterization of related recombinant 7-kDa peptides illustrates the adequacy of average mass measurement by scanning at low resolution. The difficulty in using the high-resolution technique in the case of poor LSIMS sensitive peptides is discussed, as well as the fact that it does not give, for these molecular weights, any real advantage. The average (or chemical) molecular weights of three recombinant hirudin molecules, hirudin variant 2 (rHV2, 6892.4 Da), hirudin variant 2-Lys47 (rHV2-Lys47, 6906.5 Da), and hirudin variant 2-Arg47 (rHV2-Arg47, 6934.5 Da), ≤10 μg each, have been measured with an accuracy ≤0.3 Da in the narrow-scan mode and ≤0.5 Da (from the protonated molecular ion) in the wide-scan mode within 10–15 min; this allows easy distinction of the three 65 amino acid proteins, which differ by a single amino acid. These three molecules could also be distinguished from one another in a mixture. Mass spectrometry and limited sequence characterization of several minor, similarly isolated peptides identified them to be N-terminal additions and/or C-terminal deletions of rHV2-Lys47. LSIMS analysis is consistent with there being no covalent dimer of rHV2-Lys47 as a narrow scan of the 7-kDa molecular ion cluster at high resolution shows it not to be a doubly charged ion.

Natural hirudin is a potent polypeptide (65–66 amino acids) thrombin inhibitor isolated from the salivary glands of the medicinal leech, *Hirudo medicinalis* [for a review, see Markwardt (1985)]. Pharmacological studies in animal models have suggested the potential of natural hirudin as a pharmaceutical agent in both venous and arterial thrombosis (Markwardt et al., 1982).

As hirudin is found only in low quantities in leeches, production through recombinant DNA technology has been undertaken (Bergmann et al., 1986; Dodt et al., 1986; Fortkamp et al., 1986; Harvey et al., 1986). Three recombinant hirudin variants, with either asparagine (rHV2), lysine (rHV2-Lys47), or arginine (rHV2-Arg47) at position 47, have been produced by using a *Saccharomyces cerevisiae* host and a secretion vector based on the α-pheromone export signals (Loison et al., 1988; Riehl-Bellon et al., 1989). All three of these recom-

* Address correspondence concerning mass spectrometry to A.V.D. and correspondence concerning hirudin to C.R.
‡ Centre de Neurochimie, CNRS.
§ Transgène S.A.
∥ SCA-CNRS.
⊥ VG Analytical Ltd.

0006-2960/89/0428-2949$01.50/0

binant hirudins are effective in blocking thrombosis formation in rat and rabbit model systems (Degryse et al., 1989).

The potential of liquid secondary ion mass spectrometry (LSIMS)[1] to give an accurate value of the molecular weight of large peptides (≥4 kDa) has been indicated by several workers since 1984. Use of the term LSIMS instead of FAB (fast atom bombardment) has been proposed by Burlingame et al. (1986). The introduction of a cesium ion gun (Aberth et al., 1982; McEwen, 1983) followed by higher energy (30 keV) versions (Elliott et al., 1986; Barber & Green, 1987) in place of an atom gun, to provide the bombarding particles, has sufficiently improved the sensitivity for large peptides that routine molecular weight determinations up to 15 kDa can now be accomplished within 10–15 min by use of the low-resolution method. Nevertheless, very few examples of mass measurements for 5–10-kDa peptides have been described, and the accuracy that one can expect from this method is still undetermined.

Two different methods have been employed. In one method, the mass spectrometer was operated at a resolution sufficiently high to separate the isotope peaks in the molecular ion cluster, and the molecular weight was measured as the monoisotopic value. This was illustrated on growth hormone releasing hormone (m/z 5077) and human insulin (m/z 5804) (Grotjahn & Steinert, 1985), several animal insulins (Barber et al., 1984), and bovine proinsulin (m/z 8676) (Cottrell & Franck, 1985). In the second method, the molecular cluster was reduced to a single peak by operating the mass spectrometer at low resolution (about 1000). In this case, the average or chemical molecular weight was measured, though few examples have been described for peptides ≥5 kDa (Hyver et al., 1985; Richter et al., 1985; Green & Bordoli, 1986; Hirayama et al., 1987; Johnson & Biemann, 1987; Boyot et al., 1988).

On peptides of ≤5 kDa, provided adequate sample is available, both methods (high and low resolution) are usually considered sufficiently accurate in the mass measurement to differentiate, for example, a carboxyl function from an amide function or to detect the formation of disulfide bonds. However, there are so few ≥5-kDa peptides studied to date that few generalities can be made concerning the expected accuracy. Advantages and disadvantages of the two methods are discussed in some detail by Green and Bordoli (1986). In essence, not only does the low-resolution method offer higher sensitivity, and hence a more rapid analysis over a wider mass range, but it also tends to be more reliable for peptides of more than 5 kDa for the reasons we develop below.

Indeed, the high-resolution method becomes more difficult to employ for molecules of increasing mass. One reason is that the observed and expected isotope patterns become harder to match because the number of ions collected in each peak falls and the data become statistically less accurate. For a peptide of 7 kDa, such as hirudin, the monoisotopic molecules represent less than 1%, and the corresponding peak is difficult to identify. The mass measurement of monoisotopic molecular species then reaches a physical limit (Yergey et al., 1983; Green & Bordoli, 1986). In this case, the low-resolution method has to be used since it is not possible to compensate for this natural low abundance of monoisotopic ions by an accumulation of many scans. Monoisotopic molecular masses have been previously measured on compounds available at high purity and in relatively large amounts (100–200 μg). These measurements were obtained after optimization with unit mass resolution and by accumulation of up to 80 scans, generated from several consecutive introductions of sample, to improve the signal/noise ratio (Grotjahn & Steinert, 1985). On small amounts of substance, or with larger molecules, the sensitivity has to be increased at the expense of resolution, and only an average mass can be determined.

In this work, we evaluate the use of LSIMS in characterizing several related recombinant hirudin molecules (6.9 kDa) individually or in a mixture. A wide scan, up to 15 kDa, is used to search for the possible appearance of a hirudin dimer. An assessment of the speed, reliability, and accuracy of average mass measurement on hydrophilic peptides (≤10 μg) with masses of approximately 7 kDa is given.

Materials and Methods

Sample Origins. The three full-length hirudin recombinants (rHV2, rHV2-Lys47, and rHV2-Arg47) were isolated as the major active forms in their respective culture supernatants from the yeast expression system previously described (Loison et al., 1988). They were purified by ion-exchange and reverse-phase HPLC (Riehl-Bellon et al., 1989). Final pools displayed single symmetrical peaks (≥95% pure at 205 nm) from 5 μm, C-8 columns (see HPLC). These protein peaks, which coincided with antithrombin activity, have subsequently been sequenced by Edman degradation of their tryptic peptides, and the observed primary structures have been found to agree precisely with those predicted from the cDNA (Riehl-Bellon et al., 1989; our own unpublished observations). Comparable purity rHV2-Lys47 is available from Sigma (St. Louis, MO).

Six minor fractions containing antithrombin activity were isolated by ion-exchange and reverse-phase HPLC, but displayed distinctly different retention properties on analytical reverse-phase HPLC (see below). The purity of these samples as measured at 205 nm was approximately 80–90% except for one sample containing an approximately equimolar mixture of peptides with retention times 10.07 and 10.36 min. Protein quantity was estimated by integration of the 205-nm absorbance peak using an external hirudin standard quantitated by amino acid analysis.

High-Performance Liquid Chromatography (HPLC). The retention time and purity estimation of highly purified recombinant hirudins (rHV2, rHV2-Lys47, and rHV2-Arg47) were determined by reverse-phase HPLC on an analytical HP 1090 liquid chromatograph (Hewlett-Packard, Waldbronn, FRG) equipped with a 5-μm Nucleosil C-8 column (0.46 × 10 cm). Elution was performed by a linear gradient from 15% solvent B to 30% solvent B (solvent A = 0.1% TFA/water and solvent B = 0.1% TFA/acetonitrile) in 15 min with a flow rate of 1 mL/min. Six active hirudin-related samples were similarly characterized by using a 3-μm Nucleosil C-8 reverse-phase column (0.46 × 10 cm) with identical elution conditions.

Analytical runs designed to resolve the mixture of the three recombinant hirudins were performed on a 7-μm C-8 Brownlee Aquapore RP300 microbore column (2.1 × 30 mm) with a flow rate of 0.1 mL/min. Solvent A was 0.1% TFA/water. Solvent B was 0.1% TFA/acetonitrile. One linear gradient was run from 5% solvent B to 15% solvent B in 10 min and from 15% solvent B to 35% solvent B in 35 min. A second linear gradient was run from 5% solvent B to 15% solvent B in 10 min and from 15% solvent B to 35% solvent B in 50 min.

Amino Acid Analysis. Quantitation of reverse-phase-purified proteins was accomplished following total acid hydrolysis

[1] Abbreviations: Da, Dalton; m/z, mass to charge ratio; $[MH]^+$, protonated molecular ion; $[MH_2]^{2+}$, doubly protonated molecular ion; HPLC, high-performance liquid chromatography; LSIMS, liquid secondary ion mass spectrometry; TFA, trifluoroacetic acid; P, monoisotopic peak; P + n, monoisotopic peak plus unit mass values, n.

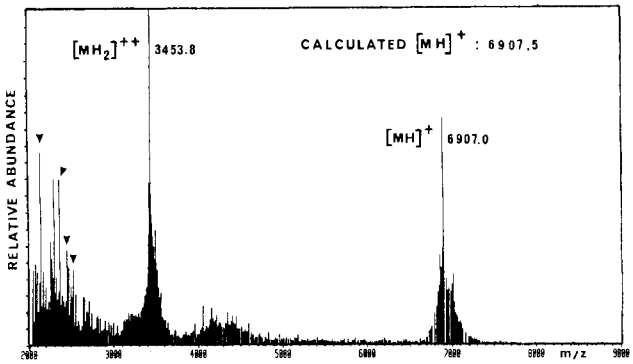

FIGURE 1: Chemical mass of hirudin rHV2-Lys47. Survey (wide scan) mass spectrum of rHV2-Lys47 (10 μg) recorded at 1000 resolution by scanning the magnetic field from 11 to 2 kDa (the zone from 9 to 11 kDa is not shown since no peaks were detected above 7 kDa). The protonated molecular ion appears at m/z 6907.8 (expected $[MH]^+$ = 6907.5). Data acquisition time was 2 min. Four of the largest peaks (marked by an arrow) in the mass range 2–3 kDa are at m/z 2147, 2372, 2474, and 2531 and correspond to C-terminal fragments of rHV2-Lys47. Analogous fragments with similar intensities are observed for rHV2 and rHV2-Arg47 (see Results, Table IV, and Figure 5).

(6 N HCl, 106 °C, 24 h) of duplicate samples using the PICO-TAG system of Waters Associates (Waters Millipore, Milford, MA).

N-Terminal Sequence Analysis. Automated Edman degradation was performed by using an Applied Biosystems (Foster City, CA) gas-phase (Model 470A) or liquid-phase (Model 477A) sequencer with an on-line PTH amino acid analyzer (Model 120A).

Mass Spectrometry. Positive ion mass spectra were obtained on a VG ZAB-2SE double-focusing instrument (mass range 15 kDa at 8-keV ion energy) and recorded on a VG 11-250 data system (VG Analytical Ltd., Manchester, England). Ionization of the sample was performed with about 1 μA of 30-keV energy cesium ions by using the cesium ion gun normally provided by the supplier.

The underivatized peptides were dissolved in 5% acetic acid/water at a concentration of about 10 μg/μL. 1-Thioglycerol containing 1% trifluoroacetic acid was used as the matrix. Typically 2 μL of matrix was deposited on the target, and then 1 μL of the peptide solution was added and mixed with the matrix by using the tip of the syringe needle. Hirudin is not a particularly LSIMS-sensitive compound compared to other ≥5-kDa molecules (our unpublished observations).

The survey, or wide-scan, spectra were generated by classical exponential magnetic scanning over the mass range 11–2 kDa at an instrumental resolution of 1000. Narrower scans were made over the molecular ion region at 1000 resolution to obtain a more accurate molecular weight assessment, either by magnetic scanning over a scan width of about 1 kDa (to include four or five consecutive cesium iodide clusters) or by voltage scanning over 300 Da (to include two consecutive cesium iodide clusters). The high-resolution spectra were also generated by voltage scanning over 300 Da but at 6000 resolution. In all cases, the scan rate was chosen so that each scan took 15–20 s, and several scans were summed by operating the data system as a multichannel analyzer (MCA), generally over a total acquisition time of 1–2 min. The exponentially scanned data were processed by standard routines supplied with the data system. Semiautomatic procedures smoothed and converted the MCA profile data into line spectra after peak detection. The background was reduced by tracking base line during the peak detection procedure. Calibration and mass measurement of the processed data were then carried out by using routines similar to those used at low mass, employing spectra obtained from a separate introduction of cesium iodide. The total time to acquire and process the data was 10–15 min/sample.

Table I: Molecular Weight Determinations of Three Hirudin Variants from Wide Scans (2–11 kDa)

peptide	expected mol wt	measured mol wt from $[MH]^+$ (error)	measured mol wt from $[MH_2]^{2+}$ (error)
rHV2	6892.4	6891.9 (−0.5)	6891.8 (−0.6)
rHV2-Arg47	6934.5	6934.2 (−0.3)	6934.0 (−0.5)
rHV2-Lys47	6906.5	6906.0 (−0.5)	6905.6 (−0.9)
rHV2-Lys47	6906.5	6906.8 (+0.3)	6907.4 (+0.9)

Table II: Molecular Weight Determinations of Three Hirudin Variants from Narrow Scans (Da)

peptide	expected mol wt	measured mol wt from $[MH]^+$ (error)
rHV2	6892.4	6892.5 (+0.1)
rHV2-Arg47	6934.5	6934.7 (+0.2)
rHV2-Arg47	6934.5	6934.8 (+0.3)
rHV2-Arg47	6934.5	6934.7 (+0.2)
rHV2-Arg47	6934.5	6934.7 (+0.2)
rHV2-Lys47	6906.5	6906.7 (+0.2)
rHV2-Lys47	6906.5	6906.6 (+0.1)
rHV2-Lys47	6906.5	6906.3 (−0.2)
rHV2-Lys47	6906.5	6906.3 (−0.2)

RESULTS

Average Molecular Weight Measurement by Wide Scans on rHV2, rHV2-Lys47, and rHV2-Arg47. A typical survey spectrum from 10 μg of rHV2-Lys47 after centroiding and mass measurement is shown in Figure 1. Characteristic features in these spectra are prominent peaks that can be assigned to the singly and doubly charged protonated molecular ions. In addition, intense fragment peaks, particularly in the mass range 2–3 kDa, were observed (see below). No other significant peak that could be interpreted as a molecular ion was apparent in any of the spectra, indicating the absence of major impurities.

Table I summarizes the average molecular weights determined from both the singly and doubly charged ions. The majority of the measurements are within 0.6 Da and all are within 1 Da of the expected value for the completely oxidized form. Thus, survey spectra provide an initial confirmation of the molecular weight together with a limited measure of the purity.

A survey scan on rHV2-Lys47 covering the mass range 2–15 kDa produced peaks predominantly associated with the mass of the monomer. The lower part of the spectrum was, as expected, identical with Figure 1. The upper part presented only a very small peak (≤0.1% of $[MH]^+$ at m/z 6906.5) in the region of 13.8 kDa close to the expected value of a possible dimer. This signal was at least a factor of 1000 smaller than would be expected from a peptide with a dimer molecular weight and is considered to be the result of a molecular–ion reaction (Lepage et al., 1989).

Average Molecular Weight Determination of rHV2, rHV2-Lys47, and rHV2-Arg47 Using Narrow Scans. Measurements of the molecular weights obtained by using the narrow-scan mode were more accurate than those obtained by using a wide scan. Figure 2 shows a typical narrow-scan spectrum as raw data from 10 μg of rHV2-Arg47. To evaluate the reproducibility of the method, four separate measurements were made for two of the three variants. The results are summarized in Table II and show that the average molecular

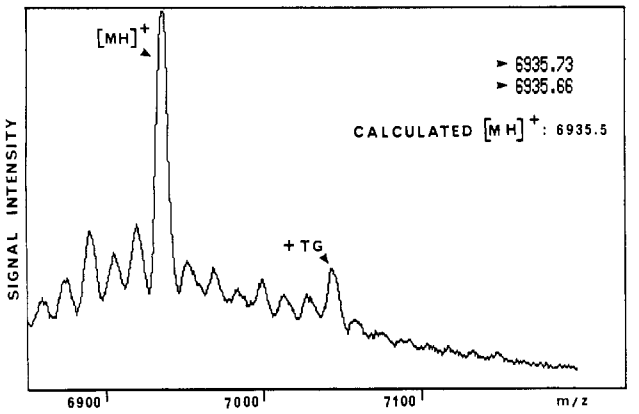

FIGURE 2: Chemical mass of hirudin rHV2-Arg47. Narrow scan over the molecular ion region of rHV2-Arg47 (10 μg). The magnetic field was maintained constant and the accelerating voltage scanned between the two closest cesium iodide clusters (m/z 6887.96 and 7147.77). Two centroid measurements (10 and 20% from the base line) were made on the largest protonated molecular ion peak giving m/z = 6935.73 and 6935.66 (expected $[MH]^+$ = 6935.5). Peak TG corresponds to a thioglycerol adduct $[MH + TG]^+$ centered at m/z 7043.8. The six peaks between $[MH]^+$ and the thioglycerol adduct are interpreted to be fragments from $[MH + TG]^+$. Similarly, peaks below 6935 are interpreted as being fragments of $[MH]^+$.

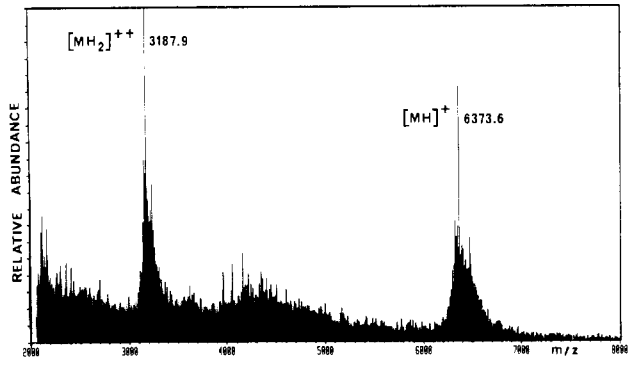

FIGURE 3: Incorrectly processed form of hirudin rHV2-Lys47. Survey mass spectrum of a truncated form (5 μg) recorded in 3 min showing singly and doubly charged ions at m/z 6373.6 and 3187.9. The measured mass of the singly charged protonated ion can be attributed to an rHV2-Lys47 lacking the last four amino acids (expected $[MH]^+$ = 6373.9).

weight was always measured to within 0.3 Da of the expected value. Note that the base width of the molecular peak covers about 10 Da, which corresponds roughly to the number of isotope peaks.

Identification of Minor Incorrectly Processed Forms of rHV2-Lys47. During the purification of rHV2-Lys47, minor antithrombin activity peaks were detected and isolated for primary structure characterization. Six such active protein fractions, characterized by automated N-terminal sequencing, were distinguished in all cases by reverse-phase HPLC analysis and LSIMS measurements scanning over the range 2–11 kDa. Figure 3 presents the mass spectrum from 5 μg of the peptide eluted in 10.11 min. Together with the N-terminal sequence information and the exact measured molecular weights, the seven component peptides were readily characterized as incorrectly processed forms of rHV2-Lys47 lacking 1–4 C-terminal amino acids and in some cases also containing 1, 2, or 11 extra N-terminal amino acids. Table III summarizes the results. As with the wide-scan measurements on the correctly matured hirudins (Table I), it can be seen that the majority of the measured molecular weights are within 0.6 Da of the calculated value, and all are within 1 Da.

The elongated N-terminal sequences correspond to a segment of the yeast α-factor precursor. To produce and secrete

Table III: Molecular Weight Determinations of rHV2-Lys47-Related Forms from Wide Scans (2–11 kDa)

HPLC[a] T (min)	rHV2-Lys47 forms	expected mol wt	measured mol wt from $[MH]^+$ (error)	measured mol wt from $[MH_2]^{2+}$ (error)
16.02	1–64	6778.4	6778.4 (0.0)	6778.6 (+0.2)
12.35	1–63	6665.2	6665.9 (+0.7)	6665.6 (+0.4)
10.07	1–62 +	6502.0	6501.3 (–0.7)	ND[b]
10.36	Arg-(1–62)[c]	6658.2	6658.2 (0.0)	6658.8 (+0.6)
10.11	1–61	6372.9	6372.6 (–0.3)	6373.8 (+0.9)
12.13	Lys-Arg-(1–63)	6949.6	6949.5 (–0.1)	ND[b]
12.88	11AA-(1–63)[d]	7878.6	7878.0 (–0.6)	7878.2 (–0.4)

[a] rHV2-Lys47 (1–65) is eluted at 14.80 min in this system (corresponding to 23% CH_3CN). [b] ND = not determined. [c] 50/50 mixture; see Materials and Methods. [d] Ala-Lys-Glu-Glu-Gly-Val-Ser-Leu-Asp-Lys-Arg-(1–63).

hirudin from yeast, a chimeric gene has been constructed where the prepro part of the α-factor precursor has been fused in frame with the mature form of hirudin (Loison et al., 1988). In this construct processing signals for prepro α-factor are retained. These involve endoproteolytic cleavage of the precursor by protease yscF and N-terminal maturation of intermediates by dipeptidyl amino peptidase yscIV (Bussey, 1988). The peptides isolated are indicative of incomplete processing of the prepro sequence by the yeast. Loss of C-terminal amino acids is probably due to the action of carboxypeptidases.

This series of analyses has allowed the rapid identification of improperly processed hirudin forms using quantities as low as 5 μg. Characterization of these minor byproducts is important in monitoring their removal in preparations destined for pharmaceutical use.

Monoisotopic Mass Measurement on rHV2-Lys47. For the high-resolution molecular weight determination of rHV2-Lys47, at a resolution of 6000, in the narrow-scan mode, three scans were accumulated; the unsmoothed data are presented in Figure 4. By comparison with the theoretical isotope pattern the highest peak of the molecular cluster is identified as a [P + 4] molecular isotope peak; the protonated monoisotopic peak is therefore also identified and is located at m/z 6903.04, in good agreement (error = 0.06 Da) with the calculated value of 6902.98. The good correlation between the theoretical and the experimental isotope pattern (Larsen et al., 1985) confirms the complete oxidation of the disulfide bonds.

Fragments Observed in the Spectra of rHV2, rHV2-Arg47, and rHV2-Lys47. Although peptides with molecular weights as high as 7 kDa have not previously been reported to produce fragments by LSIMS, we have observed intense signals in the spectra of recombinant hirudins in the mass regions 2–3 and 4–5 kDa (see Figure 1). Furthermore, the fragments in the mass range 2–3 kDa have provided limited structural information for the three recombinant proteins differing only by the amino acid in position 47. The observed masses of the fragments of all three recombinant variants are summarized in Table IV, which uses the nomenclature proposed by Roepstorff and Fohlman (1984). A schematic diagram of the fragmentation pattern demonstrating the correlation of the observed mass fragments with the positions of cleavage along four positions of the primary rHV2 sequence is shown in Figure 5.

All the hirudin spectra show the same prominent peak at m/z 2147 (Y18″), which corresponds to the C-terminal fragment of the molecules containing identical residues 48–65.

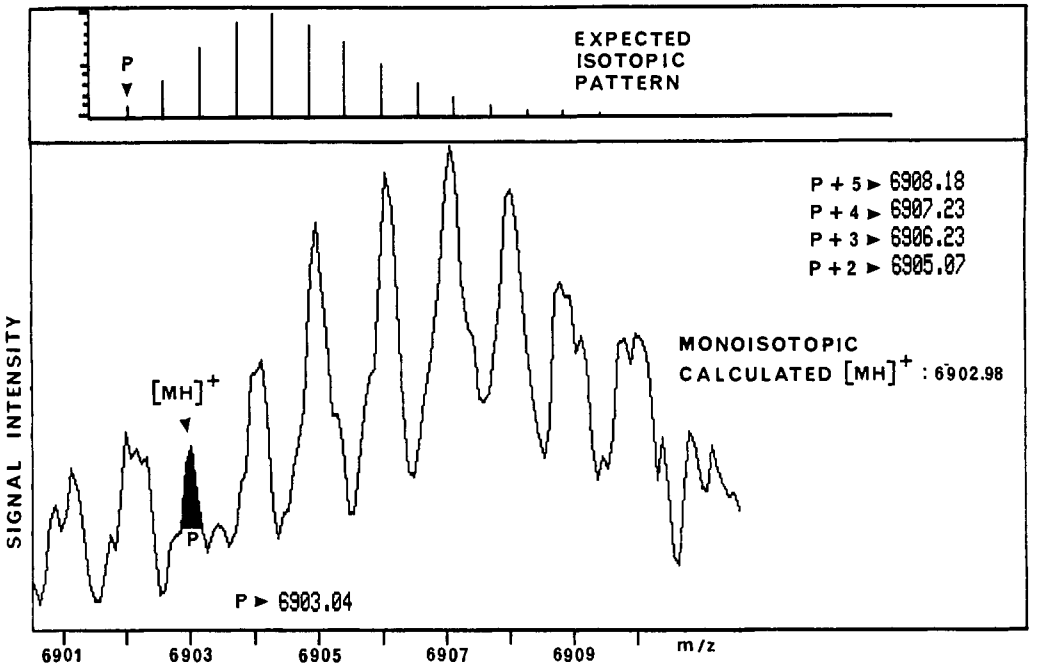

FIGURE 4: Monoisotopic mass measurement of hirudin rHV2-Lys47. Accurate mass measurement of the molecular zone at resolution 6000 on 10 μg. Other conditions are identical with those given in Figure 2. Three scans were accumulated. By use of the calculated theoretical isotopic pattern the highest peak is identified as a P + 4 molecular species, and therefore the position of the monoisotopic protonated peak is known. Its mass is measured at m/z 6903.04, which is in good agreement with the expected value (m/z 6902.98). Identification of the monoisotopic peak in the molecular cluster is difficult in this case and will be even more difficult for larger molecules.

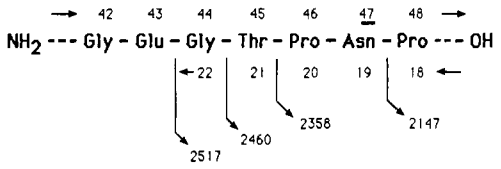

FIGURE 5: Fragments of hirudin rHV2. Interpretation of the C-terminal fragments observed for rHV2 in the mass range 2–3 kDa. Analogous fragmentation is observed for rHV2-Lys47 and rHV2-Arg47, as given in Table IV.

Table IV: Fragments Observed in the Spectra of the Three Hirudin Recombinants[a]

fragment type	m/z for rHV2	m/z for rHV2-Arg47	m/z for rHV2-Lys47
Y18″	2147	2147	2147
Y20″ = Y18″ + X[b] + Pro	2358	2400	2372
Y21″ = Y20″ + Thr	2460	2502	2474
Y22″ = Y21″ + Gly	2517	2559	2531

[a] Masses are measured at 1000 resolution; therefore, the values are average masses of fragments. These values correspond to the intense peaks observed for the three recombinants in the mass range 2–3 kDa; see Figure 1 for rHV2-Lys47. [b] X = Asn, Arg, or Lys.

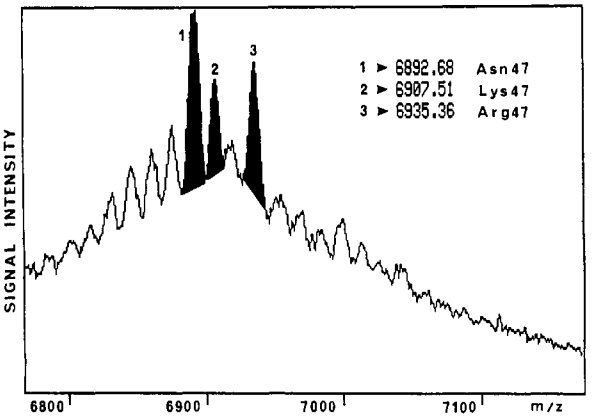

FIGURE 6: Mixture of hirudins. Mass spectrum of a 1:1:1 mixture of recombinant hirudins. The narrow-scan mode was used between the two closest CsI clusters (m/z 6887.96 and 7147.77) at 1000 resolution. Peaks are identified as rHV2 (peak 1, expected $[MH]^+$ = 6893.4), rHV2-Lys47 (peak 2, expected $[MH]^+$ = 6907.5), and rHV2-Arg47 (peak 3, expected $[MH]^+$ = 6935.5). Minor peaks can be attributed to fragments of the three hirudins and to thioglycerol adducts as described in Figure 2.

The next distinct higher mass peaks, which differ for the three recombinants, can be assigned by the addition of amino acids 46 plus 47 (Y20″). These mass values agree well with the predicted value for the C-terminal fragments containing amino acids 46–65, thus demonstrating the presence of three different amino acids in either position 47 or 46 for the three variants. As no fragmentation corresponding to cleavage between amino acids 46 and 47 was observed, the exact position of the modification cannot be determined more accurately.

Additional C-terminal fragments, ions containing residues 45–65 (Y21″) and 44–65 (Y22″), were detected in the spectra of the three recombinants at expectedly different values since these three proteins have amino acid changes at position 47. The intense nature of these fragments could be due to the presence of two prolines in positions 46 and 48, as prolines have been observed to initiate intense fragmentation in at least three other proteins (personal observations; Boyot et al., 1988).

Chemical Mass Measurement on a Mixture of rHV2, rHV2-Lys47, and rHV2-Arg47. A mixture of the three recombinant hirudins was analyzed by HPLC using an Aquapore RP 300 (C-8) column eluted with a TFA/acetonitrile/water system as described under Material and Methods. As no separation of the three recombinants was obtained by using a gradient of 15–35% acetonitrile performed in 35 min, a second more shallow gradient was tried with the same gradient performed in 50 min. With both of these gradient conditions, a single protein peak was obtained, demonstrating the difficulty in separating these three active hirudin forms differing by only one amino acid on reverse-phase HPLC (differences of 14.1, 28.0, or 42.1 Da). In contrast, the mass spectrum of the mixture (Figure 6) clearly shows the presence of peaks at m/z

6892.7, 6907.5, and 6935.4, which respectively correspond to rHV2 (calculated $[MH]^+$ 6893.4), rHV2-Lys47 (calculated $[MH]^+$ 6907.5), and rHV2-Arg47 (calculated $[MH]^+$ 6935.5). The heterogeneity of the sample is obvious when Figure 6 is compared with the spectrum of a pure recombinant (Figure 2). These results show the capability of mass spectrometry to distinguish easily three proteins whose molecular weights differ by as little as 14 Da.

The correlation between the calculated (Table I) and the experimental values of Figure 6 is good for rHV2-Lys47 (error 0.0 Da) and rHV2-Arg47 (error 0.1 Da). The slightly larger error of 0.7 Da for rHV2 could be attributed to the fact that fragments from the two heavier molecules affect the accuracy of the measurement of the protonated rHV2 molecular ion.

Discussion and Conclusion

Accuracy of Molecular Weight Measurements Obtained with Different Methods. For each of the three highly purified variants of recombinant hirudin, an experimental average molecular weight was determined by mass spectrometry, which was within ±0.3 Da of the values calculated from the sequences predicted by the cDNA (three cystines) when the measurements were performed in the narrow-scan mode at low resolution. This accuracy, identical with that reported for eglin C (Richter et al., 1985), remains one of the lowest published values for this type of determination and is the limit of what one can expect from such measurements due to natural variation in the isotopic abundances of ^{12}C and ^{13}C (Cottrell & Franck, 1985). Analogous experiments performed in the wide-scan mode gave an accuracy of ±0.5 Da (from m/z $[MH]^+$). In each case, the measured average molecular weight agrees with the values calculated for the oxidized form, where six cysteines form three disulfide bonds. Since all three recombinant hirudin molecules exhibited high specific antithrombin activities, it is likely that the disulfide bridges have been formed as shown with previous natural hirudin studies (Dodt et al., 1985). These results not only reinforce the primary structural determination of these peptides but also confirm the reliability of chemical mass measurement for this type of analysis.

The mass of the monoisotopic protonated molecular ion of rHV2-Lys47 was measured with good accuracy (error ±0.06 Da) by using a resolution of 6000. The monoisotopic peak was identified by comparing the observed and the theoretical isotopic patterns (Green & Bordoli, 1986). However, the difficulty of identifying the monoisotopic peak, P, within the molecular ion cluster is obvious. If the wrong peak is chosen as being P because of poor signal to background ratio, the error made is in integers of 1 mass unit.

This type of misinterpretation is an unacceptable limitation of the monoisotopic species mass measurement. Other limitations of the high-resolution method have been described elsewhere (Yergey et al., 1984). Though both the high- and low-resolution methods have given accurate values for recombinant hirudins, it is clear from this study that the low-resolution method with ≤0.3-Da accuracy is the more reliable, particularly because of the poor sensitivity of hirudin. The primary advantage of the nominal resolution is not necessarily the accuracy of the mass measurement but its ability to distinguish the monomer from the dimer possibility as discussed below.

Mass measurement accuracy of ≤0.3 Da allows the distinction of cysteines from cystine where one expects a difference of 2 Da. In a cysteine-containing peptide, if the observed average mass is 1 or 2 Da higher than expected from the sequence, it is of great importance to be able to attribute this difference either to partially reduced disulfide bridges as described for Paim I (Hirayama et al., 1987) or to an error in the proposed primary structure. Any mixture of peptides containing modifications that could lead to mass differences smaller than ca. 10 Da (i.e., a small fraction of deamidated or reduced molecules) will produce a shift in the mass of the unresolved molecular ion cluster and give a false average mass measurement. To avoid any inaccuracies due to possible mixtures of peptides, it is therefore of utmost importance that the average mass measurements be made only on highly purified peptides. With the ≤0.3-Da accuracy in the average mass measurements for the three highly purified recombinants (≥95% pure) compared to those predicted from the nucleic acid sequence, there is little chance of alterations existing in the protein sequences of the molecules analyzed. In addition, as it is, in this case, possible to obtain a high-resolution spectrum in agreement with the low-resolution analysis, the opportunity was taken to have increased accuracy.

Monomer/Dimer. For many years various molecular weight values have been reported for natural hirudin, varying from 9.1 to 16.0 kDa [for summary table, see Bagdy et al. (1976)]. While these values suggest that under certain conditions natural hirudin may exist as a dimer, the existence of a covalent dimer has, however, been generally excluded from consideration (Bagdy et al., 1976). Nonetheless, the physical demonstration of the monomeric nature of hirudin has proven somewhat more difficult for either the natural or for the recombinant molecules (Loison et al., 1988), making mass spectral analysis a critical molecular weight evaluation.

LSIMS has proven its ability to generate singly charged molecular mass ions from small proteins from 11 to 24 kDa (Johnson & Biemann, 1987; Barber & Green, 1987). Moreover, peptides of ~14-kDa molecular weight have been shown to give intense multiply charged ions as, for example, lactalbumin, 14 175 Da (Barber & Green, 1987), that gives both doubly and triply charged species. Thus, the presence of such multiply charged ions would be additional evidence for the presence of a covalent hirudin dimer. While the corresponding doubly charged ions would be masked by the molecular ion of the monomer, the triply charged ions would, however, be easily detected at m/z 4605. In scanning up to 15 kDa for rHV2-Lys47, no signal corresponding either to a covalent dimer or to triply charged ions could be recognized. Finally, the molecular ion pattern of rHV2-Lys47 at unit resolution (Figure 6) shows peaks separated by 1 m/z unit. Should the monoisotopic cluster at 6.9 kDa be the result of ions produced from a doubly charged dimer, it would be expected to exhibit doubly charged ion peaks separated by 0.5 m/z units. From these observations, we conclude that the peak at 6.9 kDa must correspond to a monomer and cannot correspond to a dimer.

Clearly, the distinction between singly and doubly charged ions could have also been made at somewhat lower resolution since even with poorly resolved peaks it is possible to distinguish a figure where peaks are separated by 1 Da (singly charged) from a figure where peaks are separated by 0.5 Da (doubly charged).

Mass Spectrometry as a Complementary Criterion of Purity. The general use of LSIMS as a criterion of purity for peptides is hampered by the fact that suppression effects, where one component of a mixture can be prevented from being ionized by the presence of other components, are well-known with this technique (Naylor et al., 1986). In addition, the range of impurities that can be detected is limited to the molecular weight region scanned.

However, we believe that the data presented here give good complementary information concerning the homogeneity of the different hirudin recombinant variants. The three pure recombinants (rHV2, rHV2-Arg47, and rHV2-Lys47) not only display the same biological activity but are coeluted on reverse-phase HPLC. Nonetheless, mass spectrometry can rapidly distinguish these three molecules individually or in a mixture. In a similar way a 50/50 mixture of rHV2-Lys47-(1–62) and Arg-rHV2-Lys47-(1–62) exhibits in its mass spectrum the molecular peaks of both peptides (Table III) with similar intensity. These hirudin-related peptides of similar lipophilicity are not discriminated by suppression effects. Thus, multiple components having similar lipophilicity as isolated from a symmetrical reverse-phase peak can be distinguished by mass spectroscopy, indicating that this technique can be considered to be a powerful method to complement reverse-phase HPLC for the analysis of purity in a limited mass range.

Fragment Ions Obtained in the Mass Spectra. The fragments observed in the spectra of the three 7-kDa hirudin recombinants are worthy of note as they provide complementary structural information due to the fortuitous locality of the single amino acid difference in these peptides at position 47. These results are also encouraging in the development of tandem mass spectrometry (MS-MS) (Biemann & Scoble, 1987) in the mass range 3–10 kDa, which is a current challenge for use in the partial or total sequence determination of large peptides.

Rapid Identification of Hirudin-Related Forms. LSIMS has been used to identify seven hirudin-related peptides purified as minor contaminants from the yeast culture medium containing rHV2-Lys47. In four cases the average molecular weight, rapidly obtained from ≤ 10 μg of the peptides, agreed with the molecular weights of truncated forms lacking one, two, three, or four of the C-terminal amino acids and in the remaining three cases, with truncated forms bearing one or more additional N-terminal amino acids. With the N-terminal amino acids of the reverse-phase-purified molecules clearly identified by Edman degradation, the molecular weight measurements (always ± 0.9 Da when the wide-scan mode was used), can be considered as sufficient for identification. Mass spectrometry is particularly convenient and fast for the determination of modifications at the C-terminal end of molecules, whereas classical enzymatic methods often require much larger quantities of protein and are more tedious. Edman degradation of a chemically or enzymatically derived C-terminal peptide is frequently used but necessarily requires the time and sufficient protein to perform a peptide map, making it much more time- and substance-consuming than mass spectrometry.

In conclusion, we have shown that the determination of the average or chemical mass of large peptides ≥ 7 kDa by mass spectrometry can provide precise structural information. In addition to the exact mass value, information concerning the presence or absence of certain amino acids, the oxidation state of cysteines, and molecular homogeneity within a limited hydrophobicity and molecular weight range can all be obtained from 10-μg hirudin samples in a few minutes. Results from three mass spectrometry methods, wide scan (low resolution) and narrow scans (low and high resolution), demonstrate that mass spectrometry is now a rapid, sensitive, reliable, and necessarily complementary identification technique in the field of large peptide analysis.

Acknowledgments

The pharmaceutical development of recombinant hirudin is being pursued within a partnership agreement between Transgène, 16 Rue Henri Regnault, 92411 Courbevoie Cedex, France, and Sanofi, 40 Avenue Georges V, 75008 Paris, France. We thank J.-P. Lecocq and E. Eisenmann for their continued encouragement in the development of this project and Profs. P. Chambon and P. Kourilsky for helpful discussions. The contributions of C. Teutsch and D. Clesse for the purification of the various hirudin preparations and D. Roecklin and M. Acker for the protein sequence characterization have been invaluable. Our gratitude goes to H. Major for her expertise in the VG Analytical data system handling. We are indebted to Prof. G Ourisson for his counsel and guidance in the preparation of the manuscript. Our appreciation also goes to Drs. T. Achstetter, R. Bischoff, M. Courtney, and E. Degryse for critical evaluation of the manuscript. We also thank N. Poujol for secretarial assistance.

Registry No. Hirudin, 8001-27-2.

References

Aberth, W., Straub, K. M., & Burlingame A. L. (1982) *Anal. Chem. 54*, 2029–2034.

Bagdy, D., Barabas, E., Graf, L., Petersen, J. E., & Magnusson, S. (1976) *Methods Enzymol. B 45*, 669–678.

Barber, M., & Green, B. N. (1987) *Rapid Commun. Mass. Spectrom. 1*, 80–83.

Barber, M., Bordoli, R. S., Elliott, G. J., Tyler, A. N., Bill, J. C., & Green, B. N. (1984) *Biomed. Mass Spectrom. 11*, 182–186.

Bergmann, C., Dodt, J., Köhler, S., Fink, E., & Gassen, G. (1986) *Biol. Chem. Hoppe-Seyler 367*, 731–740.

Biemann, K., & Scoble, H. A. (1987) *Science 237*, 992–998.

Boyot, P., Trifilieff, E., Van Dorsselaer, A., & Luu, B. (1988) *Anal. Biochem. 173*, 75–85.

Burlingame, A. L., Baillie, T. A., & Derrick, P. J. (1986) *Anal. Chem. 58*, 165R–211R.

Bussey, H. (1988) *Yeast 4*, 17–26.

Cottrell, J. S., & Franck, B. H. (1985) *Biochem. Biophys. Res. Commun. 127*, 1032–1038.

Degryse, E., Acker, M., Defreyn, G., Bernat, A., Maffrand, J. P., Roitsch, C., & Courtney, M. (1989) *Protein Eng.* (in press).

Dodt, J., Seemüller, U., Maschler, R., & Fritz, H. (1985) *Biol. Chem. Hoppe-Seyler 366*, 379–385.

Dodt, J., Schmitz, T., Schafer, T., & Bergmann, C. (1986) *FEBS Lett. 202*, 373–377.

Elliott, G., Cottrell, J. S., & Evans S. (1986) presented at the 34th ASMS Conference on Mass Spectrometry and Allied Topics, June 8–13, Cincinnati, OH.

Fortkamp, E., Rieger, M., Heisterberg-Moutses, G., Scheitzer, S., & Sommer, R. (1986) *DNA 5*, 511–517.

Green, B. N., & Bordoli, R. S. (1986) in *Mass Spectrometry in Biomedical Research* (Gaskell, S. J., Ed.) pp 235–250, Wiley, Chichester, U.K.

Grotjahn, L., & Steinert, H. (1985) presented at the 33rd ASMS Conference on Mass Spectrometry and Allied Topics, San Diego, CA.

Harvey, R. P., Degryse, E., Stefani, L., Cazenave, J.-P., Courtney, M., Tolstoshev, P., & Lecocq, J.-P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1084–1088.

Hirayama, K., Takahashi, R., Akashi, S., Fukuhara, K., Oouchi, N., Murai, A., Arai, M., Murao, S., Tanaka, K., & Nojima, I. (1987) *Biochemistry 26*, 6483–6488.

Hyver, K. J., Campana, J. E., Cotter, R. J., & Fenselau, C. (1985) *Biochem. Biophys. Res. Commun. 130*, 1287–1293.

Johnson, R. S., & Biemann, K. (1987) *Biochemistry 26*, 1209–1214.

Larsen, B. S., Yergey, J. A., & Cotter, R. J. (1985) *Biomed. Mass Spectrom. 12*, 586-587.
Lepage, P., Roecklin, D., Acker, M., Roitsch, C., Bitsch, F., Green, B., & Van Dorsselaer, A. (1989) in *Advances in Mass Spectrometry 1988, Bordeaux* (Longevialle, P., Ed.) Heyden and Son Ltd., London (in press).
Loison, G., Findeli, A., Bernard, S., Nguyen-Juilleret, M., Marquet, M., Riehl-Bellon, N., Carvallo, D., Guerra-Santos, L., Brown, S. W., Courtney, M., Roitsch, C., & Lemoine, Y. (1988) *Biotechnology 6*, 72-77.
Markwardt, F. (1985) *Biomed. Biochem. Acta 44*, 1007-1013.
Markwardt, F., Hauptmann, J., Nowak, G., Klessen, Ch., & Walsmann, P. (1982) *Thromb. Haemostasis 47*, 226-229.
McEwen, C. N. (1983) *Anal. Chem. 55*, 967-968.
Naylor, S., Findeis, F., Gibson, B. W., & Williams, D. H. (1986) *J. Am. Chem. Soc. 108*, 6359-6363.
Richter, W. J., Raschdorf, F., & Maerki, W. (1985) *Mass Spectrom. Health Life Sci. 24*, 193-206.
Riehl-Bellon, N., Carvallo, D., Acker, M., Van Dorsselear, A., Marquet, M., Loison, G., Lemoine, Y., Brown, S. W., Courtney, M., & Roitsch, C. (1989) *Biochemistry* (preceding paper in this issue).
Roepstorff, P., & Fohlman, J. (1984) *Biomed. Mass Spectrom. 11*, 601.
Yergey, J. A., Heller, D., Hansen, G., Cotter, R. J., & Fenselau, C. (1983) *Anal. Chem. 55*, 353-356.
Yergey, J. A., Cotter, R. J., Heller, D., & Fenselau, C. (1984) *Anal. Chem. 56*, 2236-2264.

Preparation of Biologically Active Platelet-Derived Growth Factor Type BB from a Fusion Protein Expressed in *Escherichia coli*

J. Hoppe,* H. A. Weich, and W. Eichner

Department of Cytogenetics, GBF—Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-3300 Braunschweig, FRG

Received April 14, 1988; Revised Manuscript Received November 15, 1988

ABSTRACT: Preparations of the mitogen platelet-derived growth factor (PDGF) from human platelets contain two related polypeptides termed A chain and B chain. PDGF-B is highly homologous to a portion of $p28^{v\text{-}sis}$, the transforming protein of *simian sarcoma virus*. We have studied the mitogenic potential of a PDGF-BB-like homodimer by expressing the sequence coding for the mature part of PDGF-B in *Escherichia coli*. Expression was achieved as cro-β-gal–PDGF-B fusion protein which was exclusively found in the "inclusion bodies". A monomeric PDGF-B fragment shortened by 12 amino acid residues from the NH_2 terminus was excised from the fusion protein by CNBr cleavage. After protection of thiols by S-sulfonation, this fragment was purified by gel permeation chromatography and reversed-phase high-performance liquid chromatography. This monomeric protein was dimerized in the presence of a mixture of reduced and oxidized glutathione to yield biologically active rPDGF-BB with an overall yield of ≈0.7 mg of rPDGF-BB/L of culture. *Escherichia coli* rPDGF-BB stimulated [^{3}H]thymidine incorporation into AKR2B fibroblast at concentrations of about 1 ng/mL.

Platelet-derived growth factor (PDGF)[1] is a major mitogen in serum which promotes the proliferation of fibroblasts and smooth muscle cells in vitro (Heldin et al., 1984; Deuel et al., 1985; Ross et al., 1986). PDGF elicits its mitogenic effect at low concentrations (1 ng/mL) by binding with high affinity to a specific cell surface receptor. This single polypeptide of M_r 180 000 is stimulated to autophosphorylate tyrosine residues. So far, in vivo substrates for this cytoplasmic tyrosine-specific protein kinase have not been identified unequivocally.

In vivo PDGF is stored in the α-granules in platelet from which it is released following platelet activation. Purified PDGF is a cationic glycoprotein of M_r 30 000. It exhibits considerable size heterogeneity with species between M_r 27 000 and 31 000 which are thought to result from different extents of glycosylation, processing, and the presence of two forms, A and B. All these forms exhibit identical biological activities which are destroyed by chemical cleavage of disulfide bridges. Since it was not possible to resolve the different species into homodimers of A chains or B chains, it is believed that biologically active PDGF is built up of two different polypeptide chains of $\approx M_r$ 14 000 linked together by disulfide bridges (Johnson et al., 1982, 1984).

Amino acid sequencing data in combination with cDNA sequencing data have revealed that the A and B chains are partially homologous. Interestingly enough, the B chain of PDGF was highly homologous to a portion of the predicted amino acid sequence of $p28^{v\text{-}sis}$, the transforming product of the *simian sarcoma virus* (SSV) (Doolittle et al., 1983; Waterfield et al., 1983). The relationship between PDGF-B and $p28^{v\text{-}sis}$ has raised the possibility that $p28^{v\text{-}sis}$ might act as

* Address correspondence to this author at the Institute for Physiological Chemistry, University of Würzburg, Koellikerstrasse 2, 8700 Würzburg, FRG.

[1] Abbreviations: HBS, Hepes-buffered saline; PBS, phosphate-buffered saline; PDGF, platelet-derived growth factor; rPDGF-BB, recombinant platelet-derived growth factor type BB; $p28^{v\text{-}sis}$, transforming protein from simian sarcoma virus; cro-β-gal, fusion protein from cro repressor and β-galactosidase; DTNB, dithionitrobenzoic acid (Ellman's reagent); FCS, fetal calf serum.

0006-2960/89/0428-2956$01.50/0