

---

---

# Four Decades of Structure Determination by Mass Spectrometry: From Alkaloids to Heparin

Klaus Biemann

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

---

The early (1950's and 1960's) use of mass spectrometry in natural products chemistry and its evolution to the present significance in biochemistry is recounted. This methodology allowed the facile and speedy determination of the structure of a number of indole alkaloids, such as sarpagine, quebrachamine, and two groups isolated from the roots of *Aspidosperma quebracho blanco*. At the same time, the first strategy for the sequencing of small peptides by mass spectrometry was demonstrated. It slowly advanced, over a period of two decades, to an important alternative of the ubiquitous automated Edman degradation. Further advances in methodology and instrumentation established mass spectrometry as today's indispensable tool for the characterization of proteins in biochemistry and biology. A new concept of the ionization of highly acidic compounds as the protonated complexes with basic peptides, which allows the accurate determination of the molecular weights of the former, a highly sensitive method for the sequencing of heparin fragments and related sulfated glycosaminoglycans was developed more recently. (J Am Soc Mass Spectrom 2002, 13, 1254-1272) © 2002 American Society for Mass Spectrometry

---

---

In the 1950's the most vigorously pursued research efforts in organic chemistry concerned the study of the structure and chemistry of natural products and the elucidation of the detailed mechanism of chemical reactions. Both were fostered by developments before, during, and after World War II. At that time rational drug design was still a relatively unknown concept. Instead, a search for pharmacologically and medically useful plant products was on all over the world, especially in the tropics. Many of these were alkaloids, because they could be most easily isolated by aqueous acid from an otherwise complex plant extract.

An early success was an alkaloid isolated in 1932 from the root of *Rauwolfia serpentina* and named reserpine [1]. Its structure was not determined until 12 years later [2]. Reserpine was one of the first antihypertensive drugs and became a huge financial success for CIBA (Basel, Switzerland) in the early 1950's. Research laboratories in the pharmaceutical industry and academia raced to duplicate this achievement and shelves began to fill with alkaloids, which had names and melting points, but unfortunately, mostly lacked useful pharmacological activity.

The determination of the structure of these complex molecules (aromatic-alicyclic, about 20 carbon atoms, two nitrogens and a few oxygens) was tedious, time and material-consuming. At that time (1950's) it generally required degradation or conversion to known com-

pounds and establishing their identity using elemental analysis (by combustion) and "mixed melting point" [the melting point of a mixture of two identical compounds is the same as that of each of the components, while it is depressed if two different compounds of same melting point are mixed], the gold standard of the time. The understanding of reaction mechanisms in vitro had led to the formulation of biosynthetic pathways in vivo, which aided in the postulation of structures that then could be proven by conventional means. Ultraviolet spectra were used routinely for the detection of chromophores, particularly aromatic substitution patterns and infrared spectra came slowly into use.

## Alkaloids

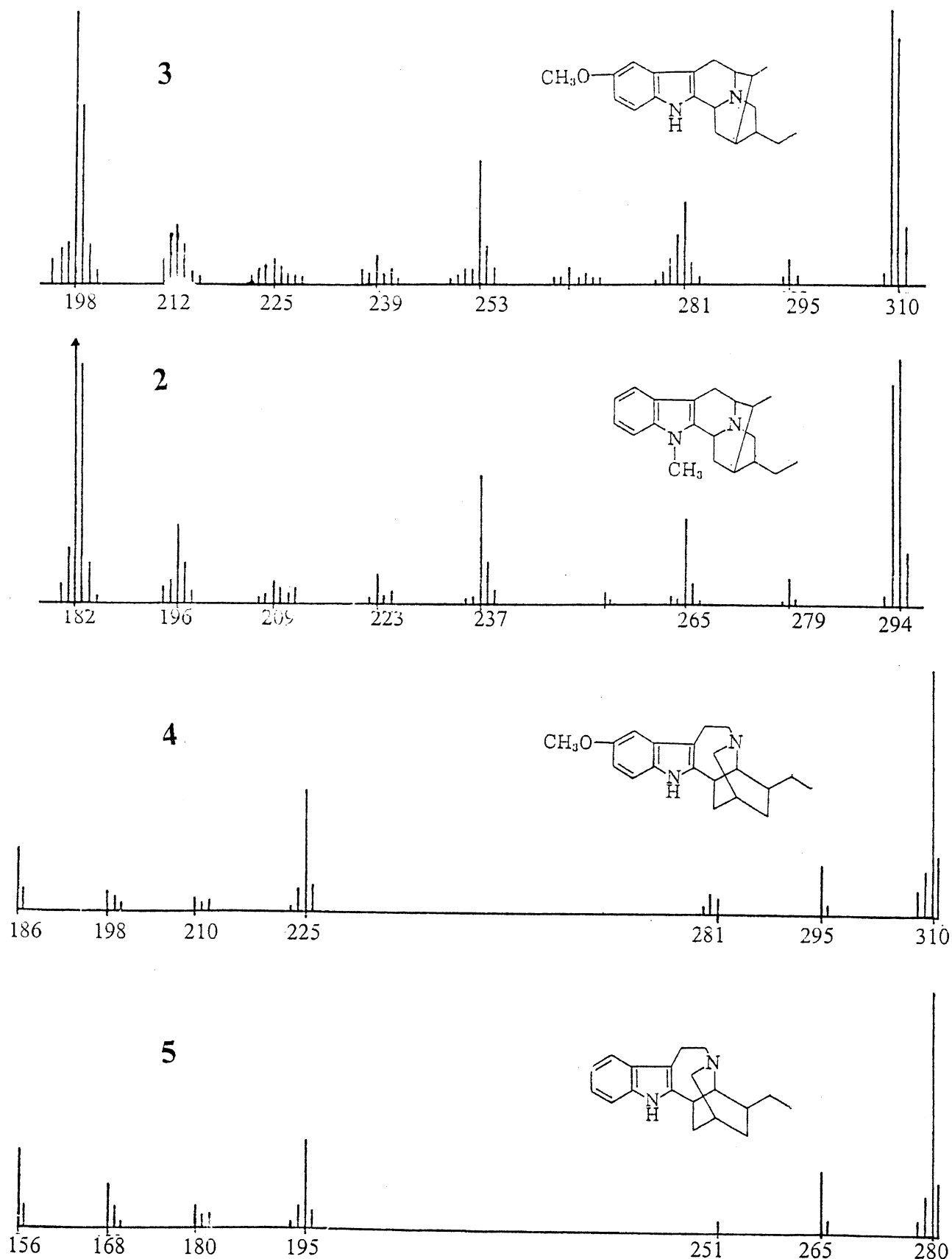
The circumstances that led me to choose mass spectrometry—a field at that time well established in physics, the petroleum industry, and then analytical chemistry, but virtually unknown among organic chemists—as my area of research have been described earlier [3]. In 1957, when I made the transition from a synthetic organic chemist to instructor (in those days the bottom rung of the tenure-track at MIT) in analytical chemistry, I had been considering the determination of the structure of sarpagine, an indole-alkaloid also isolated from *R. serpentina*.

However, before I could design a proper strategy, three research groups independently suggested [to limit the number of references listed in this historical paper, the interested reader is referred to further details quoted in those cited here] structure **1**. The aromatic

---

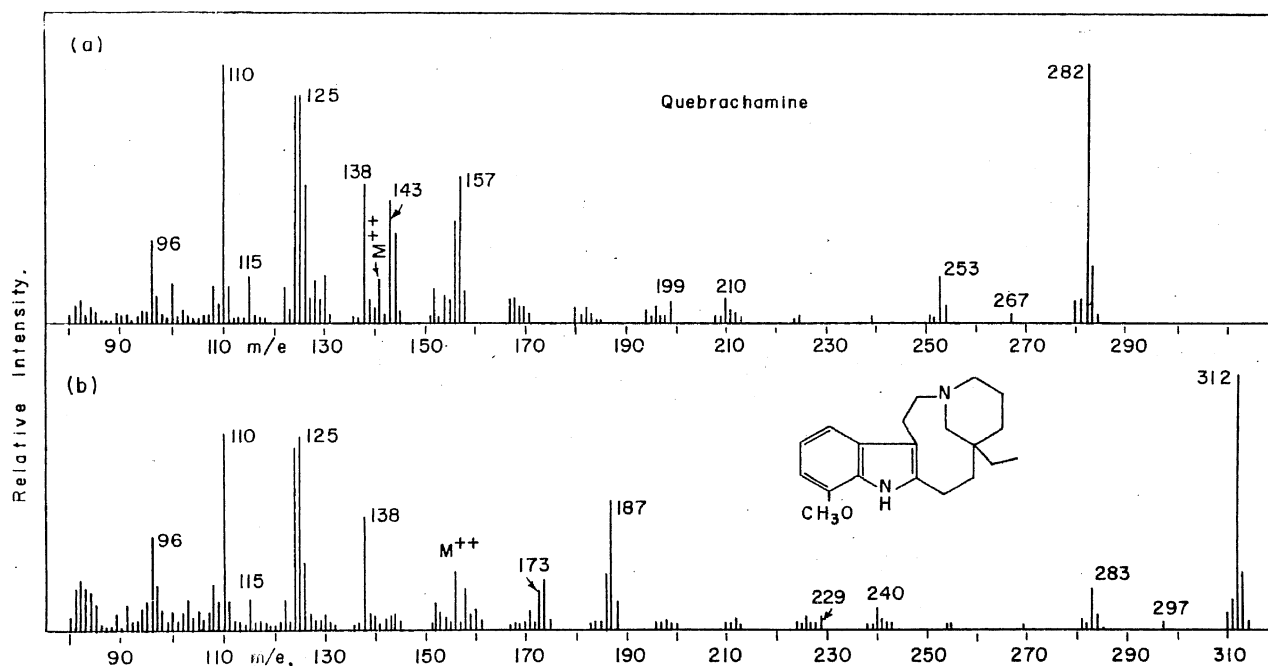
Published online September 17, 2002

Address reprint requests to Dr. K. Biemann, Department of Chemistry, MIT, Cambridge, MA 02139-4307, USA. E-mail: kbiemann@mit.edu



**Figure 1.** Mass spectra of 3 from sarpagine, 2 from ajmaline, and of ibogaine (4) and ibogamine (5) (reproduced from Reference [6], copyright 1960, with permission of Elsevier Science).





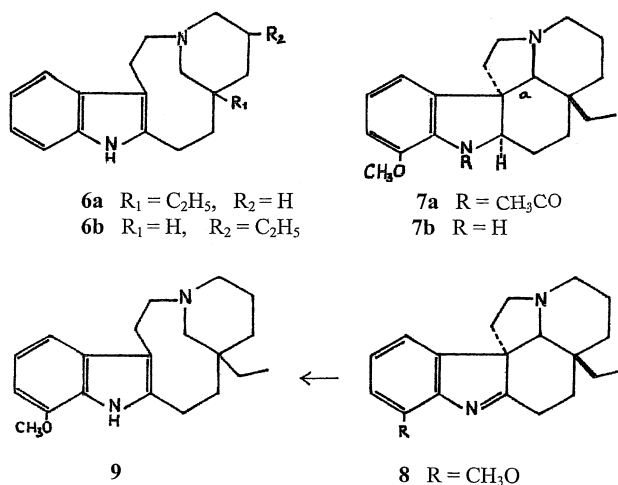
**Figure 2.** Mass spectra of (a) quebrachamine and (b) its methoxy analog derived from desacetylaspidospermine (reproduced with permission from Reference [13], copyright 1962, American Chemical Society).

thoxy analog of **6a** enabling correlation of the two structures by the mass spectrometric shift technique.

This conversion was easily accomplished by dehydrogenation of **6b** to the indolenine **8** which could be reduced with sodium borohydride to the ring-opened form **9** [13]. The mass spectra (Figure 2) of quebrachamine and **9** clearly exhibited the same pattern, considering the identity of the fragments representing the alicyclic system and the shift by 30 u for the aromatic ones. The relative abundance of these fragments would have been significantly altered for the less highly branched isomer **6b**.

The work described above produced not only analogs varying in aromatic substituents ("substituent labeling") but also isotope labeling by the use of catalytic deuteration,  $\text{LiAlD}_4$  and  $\text{NaBH}_4$ . Their mass spectra made it possible to delineate the fragmentation pathways of these rather complex radical molecular ions [13–15, 18].

When this work was presented at the IUPAC Conference on Natural Products held in Melbourne, Australia, in August 1960 it created much attention among the leaders of this field and put mass spectrometry on the map of organic chemistry. Indeed, right after the lecture, Professor Carl Djerassi came up to invite me to Stanford University to help with the installation of a mass spectrometer in his laboratory and to train a postdoc (who turned out to be Herbert Budzikiewicz) in its operation and the interpretation of the spectra. He recounted his impression 32 years later: "... It was the elegant rationalization by Biemann et al. of the mass spectral fragmentation behavior of alkaloids of the



aspidospermine class that stimulated a serious effort at Stanford on organic chemical applications of mass spectrometry." [16].

That elucidation of the fragmentation (Scheme 1) of desacetylaspidospermine (**7b**), which leads to the mass spectrum shown in Figure 3, prompted us to search for other alkaloids produced by *Aspidosperma quebracho blanco*. Upon injection of an extract of the bark of this plant [because of the demand by pharmaceutical research laboratories at that time, tropical plant materials could be obtained commercially (in this case from S. P. Penick and Co.)] into a packed GC column held at 265 °C, a chromatogram (Figure 4) was obtained, which

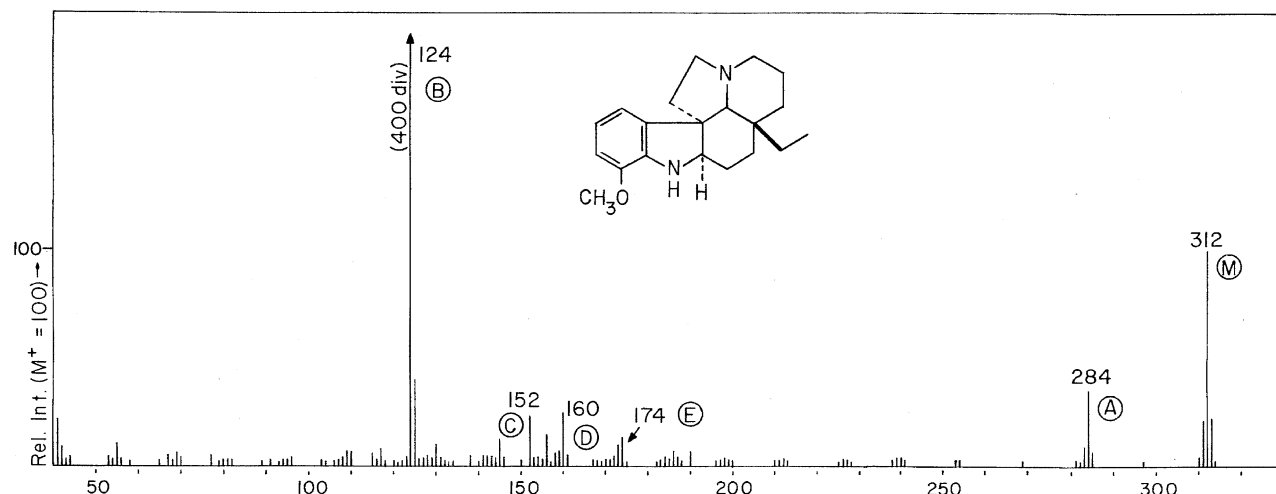


Figure 3. Mass spectrum of desacetylaspidospermine (7b). (Reproduced with permission from Reference [18], copyright 1963, American Chemical Society.)

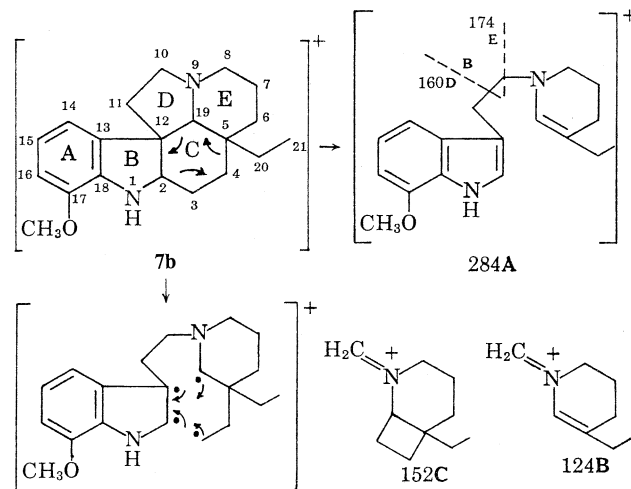
looks awful by today's standards, but at that time it was terrific. Fractions were collected individually by sticking a melting point capillary over the exit of the column and cooling it with a piece of dry ice. The capillary was then placed into the heated inlet system of the CEC 21-103C mass spectrometer. The spectra revealed the presence of at least 17 different alkaloids. Judging from the easily discerned molecular weights, most of them were new. It was clear from the spectra that there were two types of carbon skeletons present. One (Group A), recognized by the strong signal at  $m/z$  124, corresponded to the aspidospemine type (see Figure 3), but the other one (group B), dominated by a peak at  $m/z$  136 had to represent a different structural type.

When a larger portion of the extract was separated on an alumina column using gravity chromatography, sufficient pure material was obtained to record good, individual mass spectra. Based on the mass and UV spectra it was possible to assign structures to all the components belonging to the aspidospermine class (Group A, Figure 5). From what had been learned about the spectra of several indole alkaloids and using the mass shifts of the various components (Figure 6), it was possible to assign Structure 10 to alkaloid 266B (Scheme 2) and its five congeners (Figure 7). The positions of substituents on the benzene ring of the indole moiety were again deduced from UV data. Quite a few of the components were sufficiently abundant to crystallize so that their melting points could also be determined. The value for 338B (157–9 °C) compared well with 162 °C of a substance isolated 80 years earlier [8] from the same plant and named aspidospermatine. The identity of the two compounds was further corroborated by elemental analysis and optical rotation, the only other data reported by Hesse who had assigned it a molecular formula  $C_{22}H_{28}N_2O_2$ , i.e., a  $CH_2$  too much.

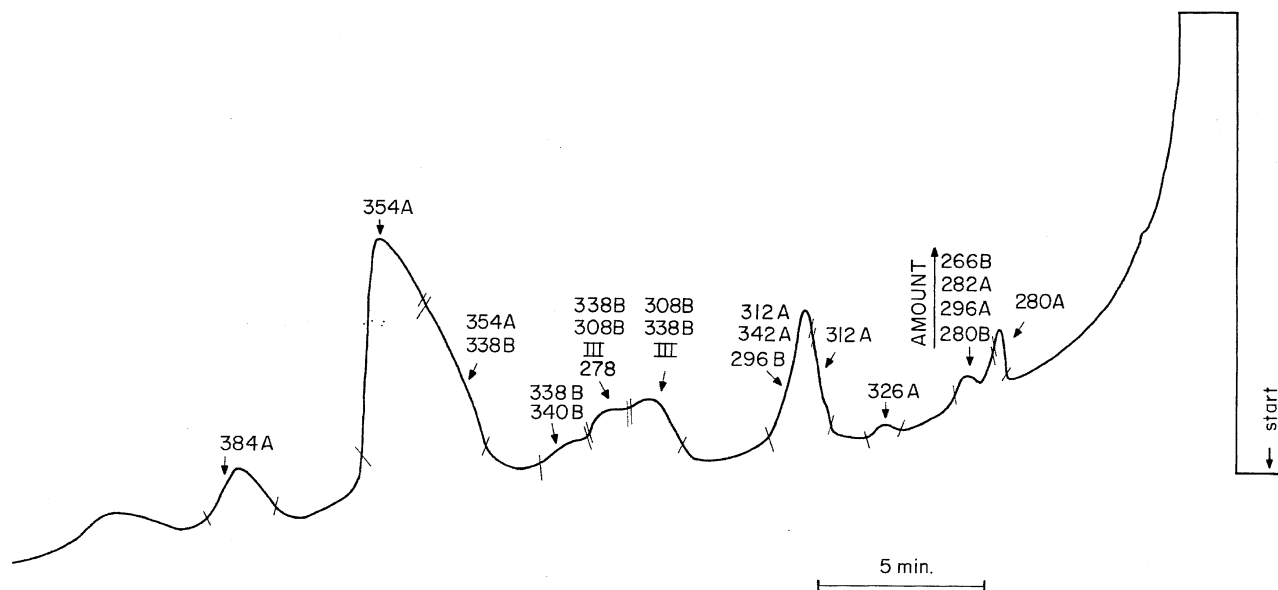
In the past it was common practice to report newly discovered natural products after characterizing them by their elemental composition (determined by com-

bustion) and melting point. A name was assigned, generally based on a part of the botanical name of the plant from which the product had been isolated. Since we determined the structure of the new compounds right after their isolation, we did not need to invent a trivial name for each one of them but simply characterized them first by their molecular weight and structural type. After the individual structures were determined, we could derive a chemical terminology based on the historical names aspidospermine and aspidospermatine and using the common insert -idi- for the unsubstituted analog.

When submitting the full paper for publication in JACS, we ran into a problem with the Assistant Editor, A. L. Autrey, who remarked in a letter dated October 9, 1962: "... More important is the need for pointing out the restricted usefulness of the work in its present form ... the samples you have applied it [mass spectrometry] to are open to question as to their homogeneity. You have detected 20 compounds, but we question that you



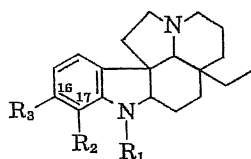
Scheme 1



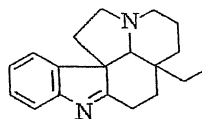
**Figure 4.** Gaschromatogram of the extract from the bark of *Aspidosperma quebracho blanco*. Slanted lines bracket fraction collection. Code numbers indicate molecular weight and pattern type (reproduced with permission from Reference [18], copyright 1963, American Chemical Society).

have "isolated" 16 of them. There is no demonstration, other than by gas chromatography, of their purity. None of the usual criteria of purity have been applied, none have been characterized by classical means: No melting points, no [combustion] analyses, few UV and no IR spectra . . . It is quite possible that in the future natural product chemists will take the turn your work has indicated, and that investigators will obtain materials, name them, and deduce their structures without characterizing them in presently accepted ways. However, until this is more common, a clear indication of the extent of your departure from past methods is necessary."

In his ardent defense of the status quo Dr. Autrey



- 282A ( $R_1 = R_2 = R_3 = H$ ): aspidospermidine  
 296A ( $R_1 = CH_3, R_2 = R_3 = H$ ): 1-methylaspidospermidine  
 312A ( $R_2 = OCH_3, R_1 = R_3 = H$ ): deacetylaspidospermine  
 326A ( $R_1 = CH_3, R_2 = OCH_3, R_3 = H$ ): 1-methyldeacetyl-aspidospermine  
 342A ( $R_1 = H, R_2 = R_3 = OCH_3$ ): deacetylpyrifolidine  
 354A (= II;  $R_1 = CH_3CO, R_2 = OCH_3, R_3 = H$ ): aspidospermine  
 384A ( $R_1 = CH_3CO, R_2 = R_3 = OCH_3$ ): (-)-pyrifolidine



280A: 1,2-Dehydroaspidospermidine

**Figure 5.** Structures of alkaloids belonging to group A (aspidospermine type).

had apparently also overlooked the fact that we had determined the melting points of the more abundant new compounds, which crystallized, and even burned a few milligrams of 338B to compare the elemental analysis with Hesse's data of 1882. It took a three-page rebuttal and a copy of my book [17], which had just appeared, to get the paper finally published [18]. Fortunately, natural products chemists quickly accepted and adopted our "departure from past methods".

The need to establish elemental compositions by combustion analysis, which required a few milligrams, much more than needed for a mass spectrum and other non-destructive spectral data (UV, IR, and later NMR) was soon eliminated by the advent of high resolution mass spectrometry. Inspired by J. Beynon's use of the "peak matching" method [19] for accurate mass measurements with a Nier-Johnson double-focusing mass spectrometer (MS-8, which became the commercial MS-9 of AEI), we recorded complete high resolution spectra on photographic plates exposed in a Mattauch-Herzog instrument [17]. The ability to deduce the elemental composition not only of the intact molecule, but also of each one of the fragments in a single experiment greatly facilitated the determination of structures from that time on.

An example where high resolution data became crucial was the determination of the structure of Vinblastine, a "dimeric" alkaloid of oncolytic activity. It had been established previously that it consists of two indole alkaloids, Velbanamine and Vindoline, but their connectivity was unknown. Because of the size of this molecule, elemental analyses were inconclusive, but the exact mass of 810.4219 (a world record at the time) established a composition of  $C_{46}H_{58}N_4O_9$  [20]. The composition of a number of key fragments allowed the

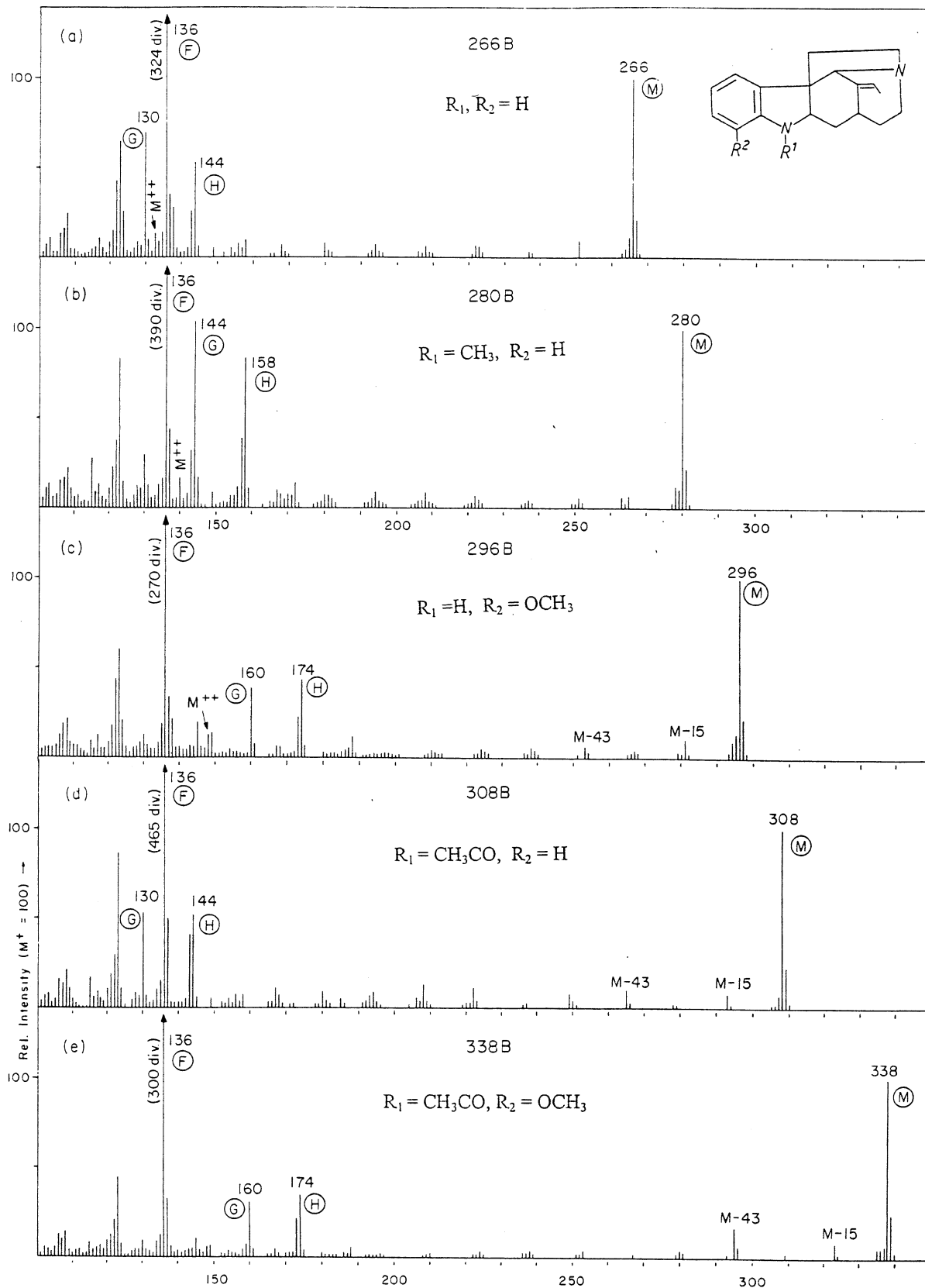
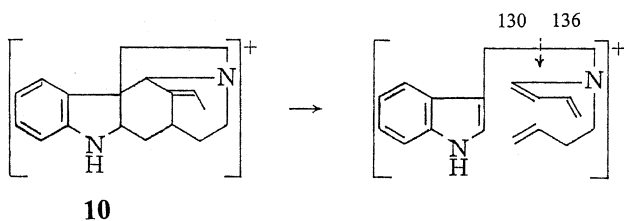
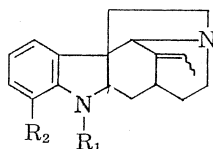


Figure 6. Mass spectra of alkaloid type B (reproduced with permission from Reference [18], copyright 1963, American Chemical Society).

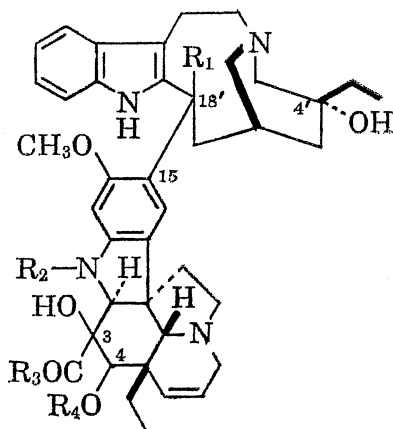


Scheme 2



- 266B ( $R_1 = R_2 = H$ ): aspidospermatidine  
 280B ( $R_1 = CH_3, R_2 = H$ ): 1-methylaspidospermatidine  
 296B ( $R_1 = H, R_2 = OCH_3$ ): deacetylaspidospermatine  
 308B ( $R_1 = CH_3CO, R_2 = H$ ): 1-acetylaspidospermatidine  
 338B ( $R_1 = CH_3CO, R_2 = OCH_3$ ): aspidospermatine  
 340B: 14,19-dihydroaspidospermatine

**Figure 7.** Structures of alkaloids belonging to group B (aspidospermatine type).



- 11**  $R_1 = COOCH_3; R_2 = CH_3;$   
 $R_3 = OCH_3; R_4 = COCH_3$   
**12**  $R_1 = COOCH_3; R_2 = CHO;$   
 $R_3 = OCH_3; R_4 = COCH_3$

determination of the complete structure of Vinblastine (**11**) as well as that of the related Vincristine (**12**) [21].

This work continued until the end of the 1960's when we at MIT and Djerassi's group at Stanford, as well as others, had established the structures of close to hundred alkaloids [22], in part or entirely by using mass spectrometry in a way first demonstrated only a decade earlier [6]. Now the shelves were bare and the quest for new natural products of that category had subsided, chiefly because none of them had become a block-buster drug. A minor exception was vinblastine, for quite some time the most useful medication for a relatively rare type of leukemia, Hodgkin's disease.

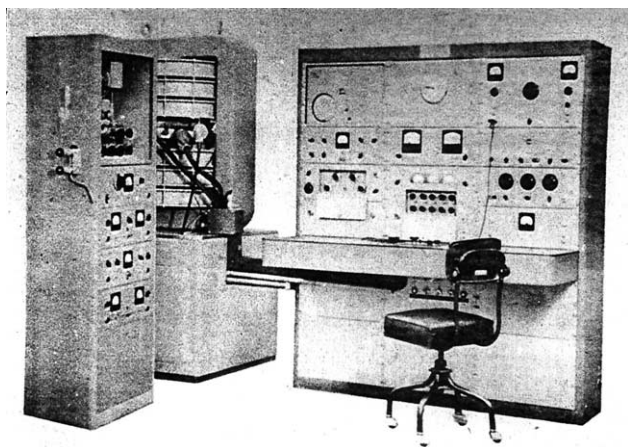
## Early Instrumentation

In the light of today's compact, microprocessor controlled mass spectrometers incorporating sophisticated data systems and sometimes even robotic sampling devices, it may be educational, at least entertaining, to look back almost half a century. Much of what is described briefly in the following is outlined in detail in Reference [17].

### The Mass Spectrometer

At the time I planned to use mass spectrometry for the determination of the structure of natural products, the choice of commercially available instruments was practically limited to a single one: The Consolidated Electro-dynamics Corporation (CEC) model 21-103C (Figure 8). This was because of its dominance in the petroleum industry and patent protection, which kept the fledgling British AEI (Associated Electrical Industries, now split into Kratos and Micromass, Manchester, UK) MS-2 and German ATLAS Werke (now Finnigan MAT, Bremen, Germany) CH-4 out of the United States. A further selling point was the American Petroleum Institute's (API) collection of mass spectra, mainly of hydrocarbons and simple monofunctional organic compounds, most of which had been recorded on instruments of the 21-103 series. The dominance of quantitative, analytical applications of MS made this database very valuable and provided good public relations for CEC.

Obtaining a mass spectrum was quite different from the highly automated process of today. The sample had to be vaporized into an all-glass inlet system held at up to 200 °C. Enough sample (0.1–1 mg) was needed to generate  $10^{-2}$  mm Hg pressure in the reservoir from which it leaked through a multi-hole molecular leak (CEC's dominating patent!) into the ion source. That



**Figure 8.** The Consolidated Electro-dynamics Corp. Model 21-103C mass spectrometer. Left: Inlet system cabinet (in the author's laboratory, the top portion was enclosed and heatable to 250 °C); center: 3-Coil electromagnet, the flight tube in the middle, main pumping system housed below; right: Operator console (note the round door in the upper left for retrieval of the photographic recording paper).



large amount was needed so that the pressure remained relatively constant during the ca. 20 min. required to record the spectrum. Scanning was by decreasing the accelerating voltage from 3 kV to 300 V at constant magnetic field, because that was easier and much more reproducible than scanning the latter. The penalty was a limited (1:10) mass range, and one thus had to record a spectrum in two pieces (e.g.,  $m/z$  10–100 and say, 50–500) at two different magnetic field settings. Each scan took about 5–10 min., limited by the recording device, a 5-mirror galvanometer that deflected a light beam onto a roll of photographic paper. At the end of the scan, the paper had to be retrieved using a black silk sleeve, taken to the darkroom, developed, fixed, and dried. A quick inspection of the 1–2 m long record showed whether the spectrum was any good and the sample could be pumped out from the inlet system. While the wet processing of the record was cumbersome, it had the advantage of producing a permanent record with a black-on-white trace. In that respect, the self-developing UV-sensitive paper that came into use somewhat later was a regression because the trace was of low contrast and faded quickly in daylight. Attempts to use a copy machine turned the record completely black.

While there was a "mass marker" (triggered by the electric field), which put a dot on the bottom of the record for every 10 mass units, it was not reliable enough to establish the mass scale. Therefore, one had to recognize the ever-present peaks at  $m/z$  18, 28, 32, and 44, due to unavoidable air leaks, and then manually count up from there. The high dynamic range of the instrument made this quite easy, because the five galvanometers had deflection ratios of 1:3:10:30:100, providing a range of better than 1:100,000 from full-scale on the least sensitive trace. Once the mass scale was established on the low mass scan, the bottom end of the high-mass scan had to be matched and the counting continued to the upper mass end (Figure 9). A good checkpoint was always the characteristic pattern of the mercury isotopes at  $m/z$  196–204 (a free benefit provided by residual back-streaming from the Hg-diffusion pumps used).

After establishing the  $m/z$  value of each peak, its height had to be measured with a ruler and converted to its absolute value by multiplication with the appropriate factor. Needless to say that it took quite some time and effort until one could draw the type of "bar-graph" nowadays instantly provided by the instrument's data system. The tediousness of this process had, however, two great advantages: First, it assured that only very few but significant mass spectra were recorded; and second, that the investigator had a lot of time to think about each spectrum, and what it may mean. I personally did almost all my thinking and preliminary interpretations during these otherwise boring tasks.

The requirement to achieve a sample pressure of about  $10^{-2}$  mm Hg in a relative large sample reservoir was dictated by the use of the mass spectrometer for the

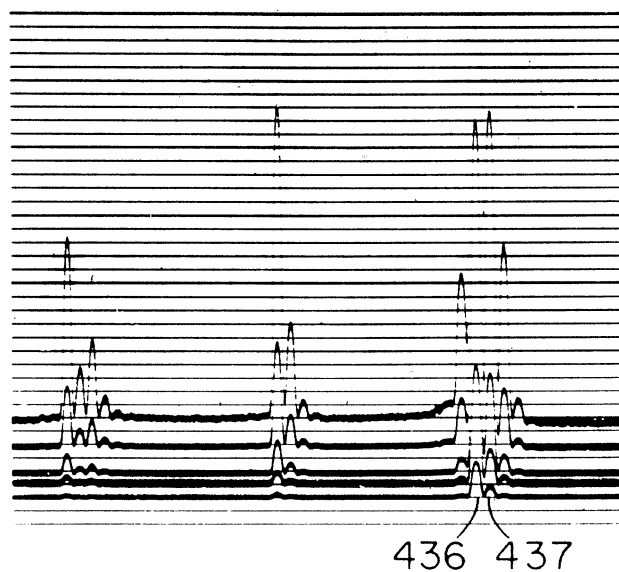


Figure 9. Part of the oscillograph record at high mass. Most sensitive trace on top, least sensitive on bottom.

quantitative analysis of hydrocarbon mixtures in the petroleum industry. The procedure was, however, wasteful as far as the required amount of sample was concerned, and limited the methodology to compounds that can be vaporized at  $<250$  °C to achieve such a pressure without decomposition. For our purposes the precise reproducibility of relative intensities was not important and we therefore tried to circumvent the entire inlet system by vaporizing the sample directly into the ion source, which required only a vapor pressure of  $10^{-6}$  mm Hg and to be maintained only in the small volume of the ion source housing.

Preliminary experiments were carried out with a Bendix time-of-flight (TOF) mass spectrometer. It was equipped with a pyrolysis probe consisting of a resistance-heated filament that could be placed below the relatively open ion source through a vacuum lock. By replacing that filament with a heating coil holding a short piece of melting point capillary containing the sample close to the ionizing electron beam, we were able to obtain the mass spectra of intact, underivatized nucleosides and amino acids [23]. Because of the inferior resolution of the TOF spectrometer, we then constructed similar direct introduction probes for the CEC 21-103C and finally for the high-resolution mass spectrometer briefly outlined in the following.

As mentioned above in the section on dimeric indole alkaloids, the determination of the elemental composition of the intact molecule and all the fragments thereof via the measurement of exact mass became an important tool in the beginning of the 1960's. We chose the CEC 21-110 model, because it was based on the Mat-tauch-Herzog geometry, which allowed for the simultaneous focusing of the entire spectrum. By placing a  $340 \times 50$  mm photographic plate into the focal plane, one could record a complete high-resolution spectrum during an exposure time of less than one minute. With



**Figure 10.** The author (standing left) discussing a photographic plate with Walter McMurray and Peter Bommer (seated) in front of the operator panel of the CEC 21-110 high-resolution mass spectrometer (photo ca. 1982/1983).

a 1 mm high slot in a mask ahead of the focal plane, one could accumulate up to 30 spectra onto one plate before it had to be replaced through a vacuum lock (Figure 10).

Processing of the large amount of data represented by each exposure required a number of steps: Measuring the exact position of each of the often more than hundred narrow lines along the plate by using a microdensitometer (also called "comparator"); conversion of position to mass, based on the positions of lines from perfluorokerosene always added as an internal mass standard; and calculation of all possible elemental compositions that would fit within a specified limit (e.g.,  $\pm 0.003$  u). These principles were well known from spark-source mass spectrometry, used for qualitative and quantitative analysis of metals, alloys etc. In those applications only a few lines were present and even fewer needed to be measured, which could easily be accomplished manually. In our case, hand measurements and calculations sufficed to provide a feel for the accuracy attainable, but not for general use.

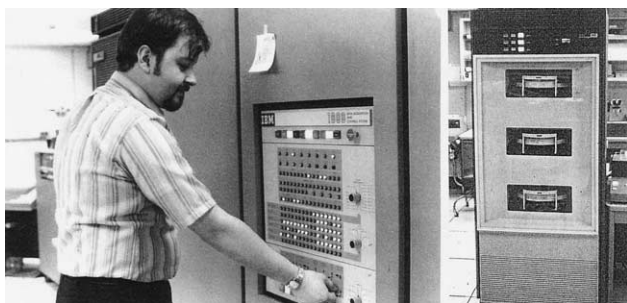
### Computers

Fortunately, MIT had begun to make its computer center's IBM 709 available to research in other departments, as long as one provided a FORTRAN program for the task and the data to be processed, everything punched into IBM cards. The results could be picked up the next day, printed on reams of paper as well as punched on cards for any further use. Punching the measured distances onto the cards without mistakes soon became too tedious and we began to automate the entire process. A variable speed motor replaced the

hand-crank of the comparator, a numerical position encoder mounted on its precision screw and another one on the photomultiplier output. These two signals were fed into a card punch operated in the continuous mode. To avoid recording the useless data (>95%) of the baseline between the lines from one nominal mass to the next, the operator slowed down the scanning speed when a line (or multiplet of lines) approached and pushed a button to activate the card punch. It was part of the data processing algorithm to find the center(s) of the line(s) from the profile recorded.

This system worked well for a while, but the number of boxes of IBM cards that had to be moved back and forth steadily increased and MIT's computer center started to charge ever higher user fees. When, in 1964, NIH asked me (yes, at that time they sometimes *asked*) to apply for one of the newly created National Research Facility grants (now NCRF) to make high-resolution mass spectrometry available to the biomedical community, and also for a training grant in the same area, the opportunity had come to set up my own computer system. Again, the choice was very limited: The only system then on the market and capable of recording continuous data "on the fly" was the IBM 1800. Its original purpose was to serve as a process monitor for industrial plants. To put it into perspective with today's MS data processors, which in addition also control much of the mass spectrometer's operation and are hidden within the instrument cabinet, the 1800 stood about 6 feet high. The central processing unit (CPU) operated with a core memory of 32,767 16-bit words, from which data could be dumped on one of three magnetic disks (12,000 words each) housed in a separate cabinet (Figure 11). Thus, the total space to work with was 135,534 bytes, less than one-tenth the capacity of one of today's 3 1/2 inch floppy disks! Final storage was on digital tape. The entire system was first leased for \$50k annually, purchased for \$250k two years later, when its usefulness had been established.

Interfacing a new, specially designed comparator (D. Mann, Burlington, MA) to the IBM 1800 system took care of all the data generated by the high-resolution mass spectrometers, of which we had two by 1965. It also provided us with the opportunity to develop methodologies to explore and utilize the vast amount of information generated by continuously scanning low-resolution mass spectrometers, especially when interfaced with a gas chromatograph (GCMS) (see next section). We had already begun this work by recording data on analog and digital tapes, but the installation of the IBM 1800 made it possible to accomplish this for the first time "on-line". The process began with digitizing the electron multiplier output of the computer-controlled, continuously scanning (3 s/scan,  $m/z$  40–600) Hitachi RMU-6D mass spectrometer, defining all peak centers, assigning its  $m/z$  value and storing all resulting mass spectra on disk. For further processing other concepts, most notably "mass chromatograms", automated identification by "library searching", etc. were



**Figure 11.** The IBM 1800 computer. Left: The CPU cabinet; the banks of lights tell the operator (here Ed Ruiz) the status of processing. Right: The cabinet holding three disks.

developed [24]. To facilitate inspection and interpretation of the entire data set “off-line” (“multi-tasking” of the CPU was not possible at that time) each GCMS data set (all spectra and all mass chromatograms) was automatically put on a roll of microfilm by using a 16 mm Bolex movie camera focused on a CRT screen, which automatically displayed one spectrum and one mass chromatogram after the other. The investigator then could look at the data leisurely on any one of the film readers we had around our laboratory.

While this system was never duplicated elsewhere, instrument manufacturers began to incorporate the concepts into their products as computer technology rapidly advanced. Size and cost of processors decreased almost exponentially. In retrospect, NIH’s money handsomely paid its dividends for the benefit of the biomedical (and chemical) community.

### *Gas Chromatography*

In the mid 1950’s Professor E. R. H. Jones, University of Manchester, UK, spent a summer at MIT and, almost accidentally, told us about the usefulness of gas chromatography, which had just been invented by James and Martin [25]. As an analytical method, it had not immediately attracted the attention of organic chemists, but Professor Jones’ examples of its use in the organic synthesis laboratories of Imperial Chemical Industries (ICI) in the UK made us aware of the unique significance of this methodology. As a consequence, our machine shop cranked out the necessary parts so that each laboratory could put together its own “preparative” GC suited not only for the quantitative analysis of reaction mixtures, but also for the isolation of their components.

Compared to today’s gas chromatographs, the contraption was very simple, one may say primitive: It employed a U-shaped glass tube packed with finely ground firebrick (capillary columns were invented much later) coated with Apiezon (a commonly available purified stopcock grease). Helium (at about 30 mL/min) passed first over a thermistor, then through the column and again over another thermistor to exit. Both

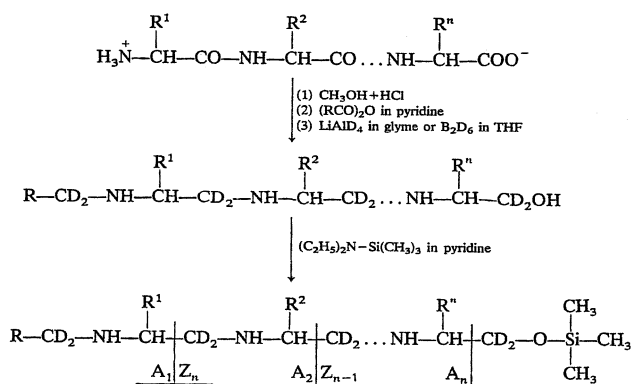
were part of a Wheatstone bridge circuit and the difference in current (representing differences in thermal conductivity of the gas mixture) was displayed on a strip-chart recorder, the only expensive part of the contraption. The column was kept at constant temperature (easy to implement, compared to the temperature programming used much later).

Packed columns required, but also could accommodate, large (milligram) samples, an advantage for the organic chemist. Constant temperature meant that volatile components eluted fast at the beginning and less volatile ones came off as broad peaks (see Figure 4). But, if necessary, the mixture could be injected at two different column temperatures to optimize resolution. Most importantly, the detector was non-destructive, i.e., the effluent could be collected and used for further experiments (record a mass spectrum or other data) or re-injected for better resolution.

Collecting fractions was tedious and we, as well as others, thought to overcome that step by interfacing the exit port of the gas chromatograph directly to the ion source of the mass spectrometer. This problem was finally solved in a practical way simultaneously by Ragnar Ryhage [26] and in our laboratory [27]. The former used the diffusion principle, a jet orifice in front of an aperture (modeled after processes previously used for the separation of uranium isotopes as their hexafluorides), while we used the faster effusion of small molecules (helium) over large molecules through the wall of a tube of fritted glass. Because the “Ryhage separator” was patented, it could only be purchased along with an entire LKB GCMS system, while the “Watson-Biemann” separator (not patented) was freely used by the other instrument manufacturers and without royalties. These “carrier gas separators” were needed because of the high flow rates required by packed columns and the limited pumping capacity of the mercury diffusion pumps of earlier mass spectrometers. They became obsolete with the advent of capillary columns, the commercial use of which was delayed in the United States (compared to Europe) by a patent held by the Perkin Elmer Corp. Efficient oil diffusion pumps also could easily handle the lower gas flow rate of these columns.

### **Peptides and Proteins**

It took a little longer to convince the biochemists and biologist of the usefulness of mass spectrometry although my very first paper published in this field in 1959 outlined a strategy for the determination of the amino acid sequence of small peptides [28]. Methodologies for peptide sequencing were still in their infancies. Fred Sanger’s N-terminal labeling of di- and tripeptides with a dinitrophenyl group, followed by total hydrolysis and identification of the labeled and unlabeled amino acid(s) by paper chromatography, enabled him to establish the first structure of a protein, insulin in 1953 [29]. At about the same time Pehr Edman began to



R = CH<sub>3</sub> or CF<sub>3</sub>

Scheme 3

develop the stepwise degradation of proteins from their N-terminus [30].

I had planned to develop a chemical approach for the labeling of the C-terminus of small peptides, to complement Sanger's method, but as outlined earlier [3] abandoned that idea in favor of mass spectrometry. I realized that peptides should be ideally suited for this methodology because they represent linear molecules of repeating backbone units substituted with a limited set of side chains, all of which differ in mass with the exception of the isomeric pair leucine and isoleucine, and the isobaric glutamine and lysine. The main obstacle was, of course, the utter involatility of these zwitterionic molecules, which prevented their vaporization so necessary for electron ionization. Here again, training in organic chemistry came in handy, as I knew that this could be overcome by acylation of the amino group(s) and esterification of the carboxyl group(s). Furthermore, the polar amido groups could be converted to the much less polar secondary amines (first step of Scheme

3). These reactions also caused the modified side chains of glutamine and lysine to differ in mass. The use of lithium aluminum deuteride instead of the hydride avoided the side chain of aspartic acid to become identical in mass to that of threonine. As I had predicted—or at least hoped—the mass spectra of the products, ethyl-oligoethylenediamino alcohols were extremely simple due to their specific cleavage at the ... NH-CHR ... bonds, resulting in sequence specific fragments (Figure 12).

It was clear from the outset that any viable sequencing method had to be applicable to the complex peptide mixtures resulting from the chemical or enzymatic degradation of the protein of which the structure is to be determined. Gas chromatography seemed to be the method of choice. Compared to paper chromatography, it had superior resolving power and loading capacity, but made it even more important to use derivatives that could be vaporized at atmospheric pressure, not just in the vacuum of the mass spectrometer. The earliest experiments already demonstrated [31] that the sequences of five peptides could be determined by mass spectrometry after conversion of the mixture to amino-alcohols, separation by GC, and collection of the components as they eluted (similar to our alkaloid studies described above) (Figure 13).

The development of mass spectrometric methods for peptide and protein chemistry has been chronicled in more detail elsewhere [32]. It took more than 15 years from the original experiments until the first sequencing of a small protein, subunit I (44 amino acids long) of monellin, was accomplished by mass spectrometry [33]. For that work, the protein had to be subjected to partial acid hydrolysis to generate the complex mixture of many overlapping di- to hexapeptides. By that time we had continuously improved the methodology through the replacement of acetylation with trifluoroacetylation

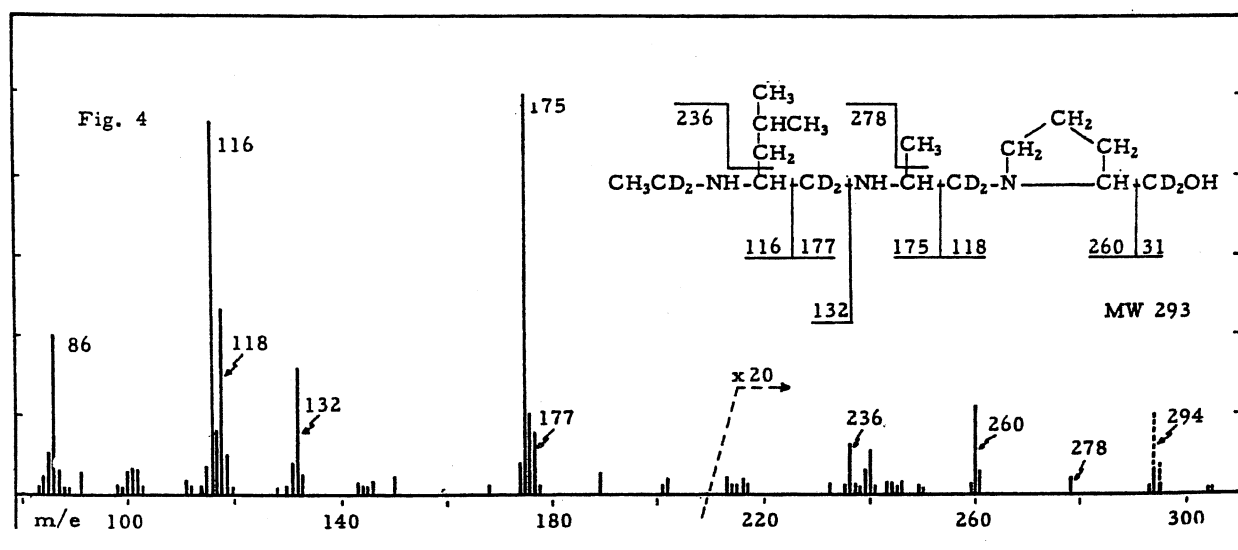
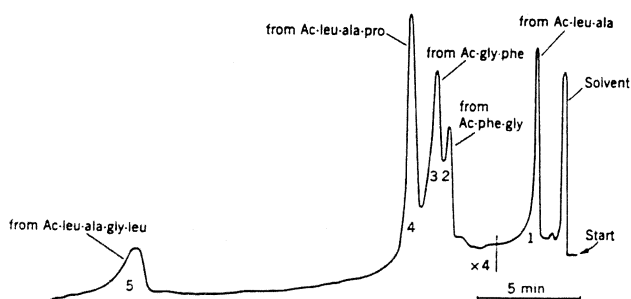


Figure 12. The mass spectrum of the reduction product of N-acetyl-isoleucyl-alanyl-proline methyl ester (from Reference [31], copyright 1960, Elsevier Science [U.S.A.], reproduced with permission).



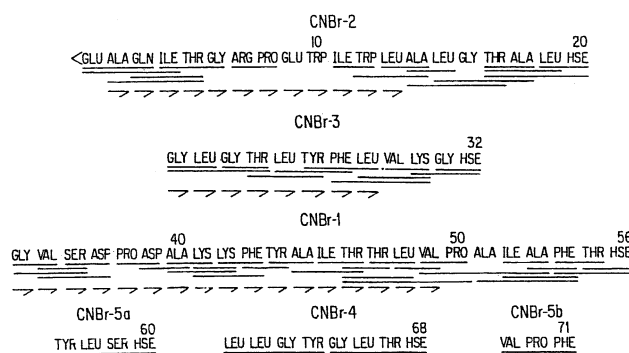
**Figure 13.** Gas chromatogram of the reduction product of a mixture of five peptides (from Reference [31], copyright 1960, Elsevier Science [U.S.A.], reproduced with permission).

[34] (second step of Scheme 3) and O-trimethylsilylation [35] to increase volatility, which extended applicability to these larger peptides. Direct coupling of the gas chromatograph with the mass spectrometer [27] made it possible to identify 61 peptides in the hydrolyzate of monellin-I in a single experiment.

During the 1960's the stepwise Edman degradation had been essentially automated and had become the mainstay of protein sequencing. Its major deficiency, the inability to work on blocked N-terminal amino acids, provided a niche for mass spectrometry which on individual small peptides could not only deduce their sequence but also identify the blocking group. Post-translationally modified amino acids which were not amenable to the Edman procedure, such as  $\gamma$ -carboxyglutamic acid, could be identified by mass spectrometry. Osteocalcin, a calcium-binding protein rich in this component, could thus be successfully sequenced by GCMS [36].

Furthermore, stretches of hydrophobic amino acids, particularly toward the C-terminus, led to high losses of material from the reaction vessel of the Edman apparatus because of "wash-out". This had been one of the difficulties in the Edman sequencing of monellin [37], which has a very hydrophobic C-terminal sequence -Gly-Pro-Val-Pro-Pro.

Extreme hydrophobicity is also a common characteristic of trans-membrane proteins such as bacteriorhodopsin, the amino acid sequence of which could only be determined correctly by a symbiosis of Edman degradation and GCMS [38a, 38b]. This protein, which turned out to be 248 amino acids long, was very resistant to enzymatic cleavage because of its insolubility in water. In aqueous suspension it was slowly attacked at a single bond by chymotrypsin, producing two large peptides, C-1 and C-2. These in turn could only be dissolved in 70% formic acid, the medium in which cyanogen bromide (CNBr) cleaves at methionine, which is converted to homoserine in the process. C-2, which contained five methionines resulted in six peptides, CNBr-1 through CNBr-5a and b. About two-thirds of the sequences of CNBr-1 to CNBr-4 could be determined by the automated Edman degradation, but excessive washout precluded reaching the C-termini. CNBr-5a and b were too short for this method.



**Figure 14.** The amino acid sequence of cyanogen bromide peptides of C-2 of bacteriorhodopsin (CNBr-1 through 5a, b correspond to elution sequence from liquid chromatograph). Arrows indicate Edman data, underlining denotes sequences determined by GCMS.

The GCMS experiments carried out in parallel not only confirmed the Edman data (Figure 14), but also provided the information necessary to assemble the complete sequence of C-2: (1) Establish the missing C-terminal sequences; (2) sequence the short peptides CNBr-5a and b; (3) identify the blocked N-terminus of CNBr-2; and (4), most importantly, permit the proper alignment of the six peptides to a single sequence. For this, the amino acids adjacent to each of the five methionines of C-2 had to be determined. This was accomplished by partial acid hydrolysis of C-2, converting the complex mixture of peptides to the O-TMS polyamino alcohol derivatives and injecting it into the GCMS. The entire set of 250 scans was then searched for ions specific for a sequence A-Met . . . (where A is any amino acid), the minimum sequence for a peptide defining an overlap. Using the principle of coinciding mass chromatograms [24], the mass spectra of all methionine-containing derivatives could be extracted from the mass of data and their sequence determined [38a]. The same strategy was then successfully applied to the larger C-1, thus completing the structure of bacteriorhodopsin [38b].

By 1980 it had become possible to sequence the gene coding for a protein. However, the "reading" of the gels was still beset with errors. Combination with an entirely different approach employed simultaneously seemed to be the most efficient strategy. Because of its high sensitivity, our GCMS methodology for peptide sequencing was used to aid in the determination of the amino acid sequence of very large proteins (some close to 1000 amino acids long), such as aminoacyl-tRNA synthetases [39]. By matching random, short amino acid sequences obtained by GCMS continuously to the DNA data as they were acquired, it was instantly possible to see whether the latter were correct. A missing or erroneously inserted nucleotide causes a "frame shift", which then translates into a fictitious amino acid sequence that would not appear in any of the GCMS data produced from the real gene product, the protein. That

information made it possible to detect the error and to correct it.

During these two decades since our first publication [28] there were very few laboratories, in addition to ours, which explored or used mass spectrometry for peptide sequencing. Foremost among them was the research group of Howard Morris (Imperial College, London), who used N, O-methylated N-acetyl methyl esters of small peptides, which could be crudely fractionated directly into the ion source of the mass spectrometer [40]. Somewhat later D. F. Hunt used similar derivatives for chemical ionization and a triple-quadrupole mass spectrometer in conjunction with a high-performance liquid chromatograph to sequence mixtures of peptides [41]. All these strategies required considerable expertise to be used at the sensitivity required and thus were practiced almost exclusively in the laboratories where they had been developed. But, as a consequence, biochemists and biologists had become aware of the potential of mass spectrometry in protein chemistry.

The situation started to change dramatically in the early 1980's. The late Michael Barber had discovered the ionization of comparatively large, polar molecules by "fast atom bombardment" (FAB) [42]. The spectrum of an undecapeptide ([Met-Lys]-bradykinin) of mol. wt. 1318 opened the way to new strategies in peptide and protein characterization. It now became possible to employ proteolytic enzymes of very high structural specificity, which by necessity produced relatively large peptides. While our GCMS methodology became almost instantly obsolete, we could quickly adapt the field desorption ion source of our MAT 731 Mattauch-Herzog type high resolution mass spectrometer to FAB by installing an argon gun [43].

As a "chemical" ionization method FAB generated abundant protonated molecule ions of low internal energy and thus low tendency to fragment. While not providing much sequence information, it allowed the reliable measurement of molecular weights of all components of a mixture of relatively large peptides without prior separation and at the nanomole level. Thus FAB-MS made it now possible to solve many problems quickly, such as the identification of mutant proteins, detection and corrections of errors in previously suggested sequences, or conversely, the identification of a protein the sequence of which is known (reviewed in [44]).

The lack of sequence information, caused by the low tendency of the protonated molecules generated by FAB to fragment, was overcome by their collisionally induced dissociation (CID) in a four-sector magnetic or triple-quadrupole tandem mass spectrometer. The former permitted high-energy collisions producing simple, sequence-specific fragmentation [45]. We used this approach for the first time in the sequencing of a number of glutaredoxins [46] and thioredoxins [47], proteins about 100–110 amino acids in length. In the course of this work, a computer algorithm for the

interpretation of high energy CID spectra of peptides (SEQPEP) was developed [48].

Novel ionization techniques continued to revolutionize mass spectrometry in general and protein chemistry in particular. John Fenn had demonstrated that exposure of fine droplets of a solution to a high electric field generates highly charged ions of even very large molecules. Not much attention was paid to his work until he reported the successful ionization of intact proteins at the 1988 ASMS conference in San Francisco [49]. Then people took notice and the measurement of molecular weights by mass spectrometry became commonplace. One of the major advantages of this method was its compatibility with the then ubiquitous quadrupole mass spectrometers, because the high charge state of the multiprotonated molecules made limited mass range not an issue. As the analyte had to be in aqueous solution, direct interfacing with a liquid chromatograph was ideal for the analysis of complex mixtures by this method, termed electro-spray ionization (ESI).

In the same year, Hillenkamp and Karas developed matrix assisted laser desorption ionization (MALDI), also capable of ionizing intact proteins [50]. This was even simpler and required less material, but at least at that time, could only be implemented on a time-of-flight (TOF) mass spectrometer. As mentioned in the Instrumentation section above, that instrument flourished briefly in the early 1960's, but had not been able to keep up with the ever improving performance of magnetic mass spectrometers and the quadrupoles that followed.

The simplicity of MALDI and its potential got us interested in its application to the determination of protein structure. Hillenkamp and Karas had carried out their experiments by modifying a TOF instrument designed for laser ionization of inorganic materials. Rather than duplicating that, we contracted with Vestec Corporation (Houston, TX) to build a MALDI-TOF mass spectrometer (following a design of Brian Chait at Rockefeller University) more suited for peptides and proteins. The instrument was delivered in October of 1990. This prototype (designated Model 2000) evolved, after the acquisition of Vestec by PerSeptive Biosystems (then by Perkin-Elmer which is now Apelera Corp.) into the Voyager Elite.

The commercial availability of such a powerful instrument and methodology at a reasonable price enabled many other laboratories to enter this field of research. The measurement of the molecular weight of a protein, such as myoglobin, to better than 1 Da [51] or the unraveling of the (inhomogeneous) pattern of glycosylation at 13 sites in invertase from yeast [52] are only a few examples from our own work besides that of many others. The development of powerful computer techniques for the identification of known or DNA derived protein sequences originally conceived by Henzel et al. at Genentech [53] and expanded by others a few years later is particularly noteworthy.

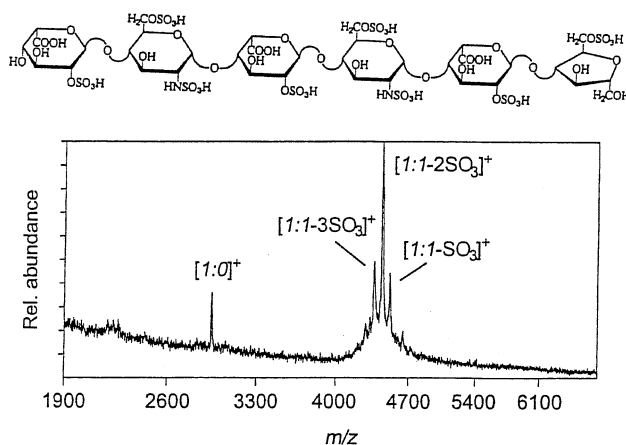
Because of its simplicity, high sensitivity, and ease of

over-all operation MALDI-MS became the sine qua non of protein chemistry. The enormous amount of highly specific and accurate data that can be produced in a single experiment, requiring only minute amounts of protein (such as that present in a spot on a 2-D gel) opened up new dimensions in biological research. The coincidence of these developments with two others—the explosion in computer technology that permits the use of large data bases on a desktop computer and the unraveling of the sequence of the human genome—led to a new field, “proteomics”. The rest of it is now history. This could not have been imagined over forty years ago, when we first demonstrated the feasibility of sequencing small peptides by mass spectrometry!

## Heparin

To minimize ambiguities when using MALDI-derived molecular weights of peptides generated by enzymatic or chemical digestion as well as those of proteins, it was important to make these measurements as accurate as possible. In an effort to optimize the experimental parameters using a series of well defined peptides as internal standards, Peter Juhasz in our laboratory one day picked the oxidized A-chain of bovine insulin ( $A_{ox}$ ) as the standard and bovine insulin as the “unknown”. Surprisingly, there was no signal at all at  $m/z$  2552.7 for the  $(M + H)^+$  ion of the standard, but a very large peak at  $m/z$  8264.9 in addition to the expected one at  $m/z$  5734.5 for the  $(M + H)^+$  ion of insulin. Puzzled at first we soon realized that the signal at high mass represented the protonated complex of the two components of the mixture,  $(M_{A_{ox}} + M_{ins} + H)^+$ , calculated 8266.3, caused by the high acidity of  $A_{ox}$ , which contains four sulfonic acid groups in the form of cysteic acid. This complexing phenomenon, observed by chance, turned out to be rather general. When a number of highly acidic compounds we had at hand were mixed with peptides, they also ionized as protonated complexes. Not unexpectedly, more basic peptides worked best.

The most highly acidic biologically significant materials are heparin and other related sulfated glycosaminoglycans (GAG), like heparan sulfate, chondroitin, etc. From earlier, unsuccessful efforts to obtain mass spectra of short heparin fragments we still had left over a sample of a hexasaccharide with eight sulfuric acid

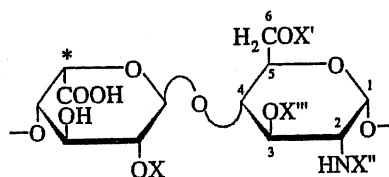


**Figure 15.** MALDI mass spectrum of the complex of the hexasaccharide shown and a peptide containing seven arginines and two lysines (reproduced from Reference [54], copyright 1994, National Academy of Sciences, U.S.A.).

half-ester groups. An equimolar mixture of 3 pmol thereof with a tetradecapeptide of mol. wt. 2942.41 containing seven arginines and two lysines gave an excellent MALDI spectrum (Figure 15) [54].

The implications of this experiment were very exciting as it opened a possibility for developing a strategy for the determination of the sequence of fragments of heparin—and ultimately perhaps of heparin itself—analogous to that we had developed for peptide and protein sequencing outlined in the previous section. The fact that there are no efficient, sensitive, and fast methods existing for the analysis of these highly acidic substances made such an effort even more important.

From a procedural point of view, the analogies with protein sequencing are striking: (1) While proteins are linear “polymers” of a set of  $\alpha$ -amino acids, heparin and its congeners are linear polymers of disaccharides, consisting of one uronic (glucuronic or iduronic) acid and one glucosamine (Structure 13); (2) while there are 20 well defined amino acids commonly found in proteins, there are 32 different known modifications of the repeating disaccharide 13 (depending on the number and position of sulfate groups or N-acetylation), all of them differing in mass by at least 4 Da [55]; and (3) like proteins, which can be degraded specifically by proteolytic enzymes or chemical reactions, heparin can be



13

Repeating unit of GAG heparin. X, X', and X''' –H or SO<sub>3</sub>H, X'' –SO<sub>3</sub>H or COCH<sub>3</sub>. Both epimers can occur at \* (L-iduronic or D-glucuronic acid).

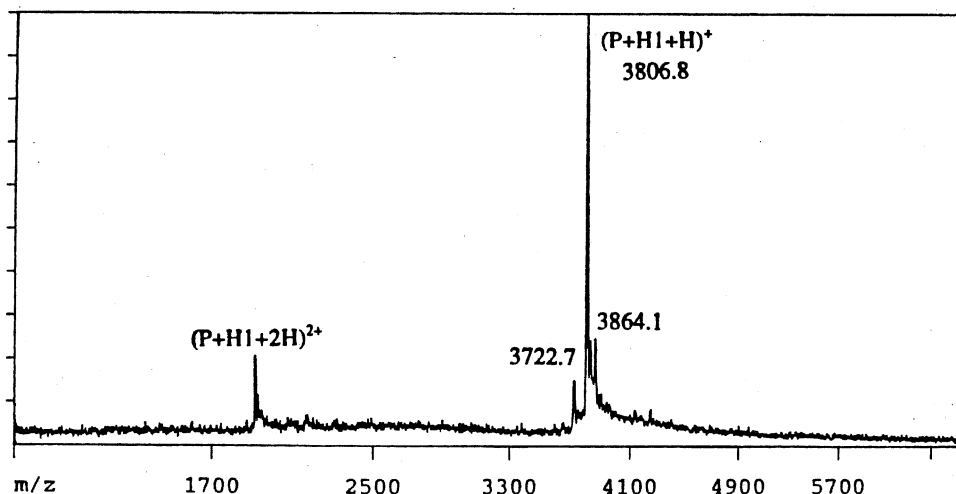


Figure 16. MALDI mass spectrum of the same hexasaccharide as in Figure 15, but complexed with  $(\text{Arg-Gly})_{10}$  (reproduced from Reference [56], copyright 1960, with permission of Elsevier Science).

cleaved enzymatically with heparinases, at specific glycosidic bonds; there are also some oxidative reactions which split the carbohydrate backbone at certain bonds.

The first step in the design of such a sequencing strategy was to develop an experimental procedure for the reliable determination of the true molecular weights of heparin fragments of the size expected from such enzymatic or chemical cleavage reactions. Clearly, the elimination of one to three  $\text{SO}_3$  groups from  $-\text{OSO}_3\text{H}$  to  $-\text{OH}$  (see Figure 15) must be avoided, because it simulates the presence of less sulfated molecules. Because of the quite regular spacing of the acidic groups along the linear disaccharide backbone it was expected that complexes with peptides of similar length and equally spaced arginines (the most basic amino acid) would be most stable and might thus not fragment. This turned out to be correct when we began using a properly designed synthetic peptide  $(\text{Arg-Gly})_n$ , where  $n = 10$ . It was extended to  $n = 15$  when it was found

that the number of strongly basic groups in the peptide must exceed the number of sulfates in the oligosaccharide. The result of these optimization experiments, which also included the evaluation of various matrices and wavelengths, is shown in Figure 16. As expected from these data, the molecular weights of compounds of different sizes present in a mixture could be measured reliably. It was also observed that the relative signal decreases with decreasing number of sulfate groups. Thus the method is more sensitive for the larger molecules, the direction desirable for its practical application [56].

Thus, the stage was set to explore the sequencing of heparin fragments by their enzymatic cleavage followed by identification of the products by MALDI. Heparinase I was chosen because the specificity of this enzyme was best known: It preferentially cleaves at the glycosidic bond between glucosamine sulfated at the amino group and the 4-hydroxyl of an iduronic acid

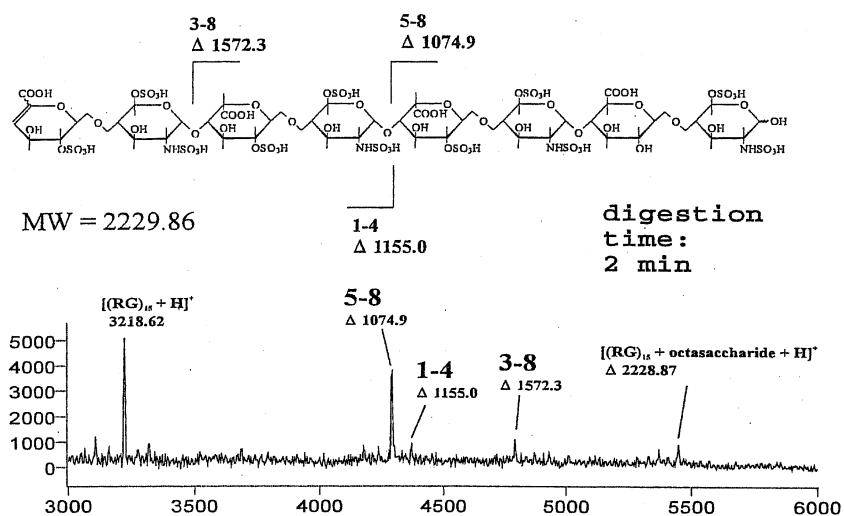


Figure 17. MALDI mass spectrum of the complex of  $(\text{Arg-Gly})_{15}$  with the heparinase I digest of the octasaccharide shown (for details see text).



moiety sulfated at the 2-hydroxyl. This is an elimination reaction, leaving two products, one with the glucosamine at the "reducing end", the other with a  $\Delta$ -4,5-dehydro uronic acid at the "non-reducing end".

An octasaccharide was digested with heparinase I and aliquots analyzed at various time points by MALDI after addition of the peptide (Arg-Gly)<sub>15</sub>. The spectrum obtained after only 2 min of digestion (Figure 17) shows that the octasaccharide has been almost completely degraded. The tetrasaccharide 5–8 which can be formed from the octasaccharide as well as from the first product (3–8) is most abundant because heparinase I cleaves tetrasaccharides only slowly. At this time point the signals for fragments 3–8 and 1–4 are small but clearly observable [57].

These results demonstrated the feasibility of sequencing heparin and heparin like sulfated polysaccharides by mass spectrometry. In the course of an extensive study [55, 58] the experimental conditions were optimized using heparinases I and II. To simplify the sometimes complex mixture of products, capillary electrophoresis was explored successfully. Because sometimes identical subunits can be formed during digestion, it was found useful to mass-label the reducing end by conversion to a semicarbazone.

During that time we began an extensive collaboration with Professor Ram Sasisekharan and his research group at MIT, who were interested in exploring the detailed mechanism of the cleavage of heparin by heparinases. For that purpose, our method of characterization and sequencing of sulfated GAGs was extremely well suited because it was fast, required little material, and allowed the following of the time-course of cleavage [59–61]. With my official retirement in 1996, graduation of my last student (Andrew Rhomberg) in 1998, and closing of my research laboratory, I was happy to transfer the technology, including the Vestec 2000 mass spectrometer, to that group. It is gratifying to see our new methodology being applied so quickly to the study of these important biological processes.

## Conclusions

The successful determination of the structure of a number of indole alkaloids starting in 1958 brought mass spectrometry, until that time confined to physics, the petroleum industry, and then analytical chemistry, to the attention of organic chemists. Academic and pharmaceutical research laboratories quickly embraced this methodology and added it to the arsenal of physical methods now so widely accepted.

At the same time we began to develop a strategy and procedure to determine the sequences of small peptides in the complex mixtures produced by the chemical or enzymatic hydrolysis of proteins. Over about two decades, mass spectrometry emerged as an alternative to the automated Edman degradation, which it complemented by its ability to determine the end-group and sequence of N-terminally blocked peptides, long

stretches of hydrophobic amino acid sequences, as well as the characterization and location of post-translational modifications. The explosion of new and efficient ionization techniques and developments in instrumentation, beginning in the late 1980's, made mass spectrometry the preeminent technology in protein research.

The observation that highly acidic compounds, normally not amenable to mass spectrometry, ionized efficiently as protonated complexes with well-defined basic peptides made it possible to determine the molecular weight of heparin fragments and similar highly sulfated glycosaminoglycans with very high sensitivity. Based on these experiments, a strategy for the sequencing of such compounds by their specific cleavage with enzymes (heparinases) was developed, in analogy to that which we had so successfully employed for peptides and proteins. This work opens the way to study the structure, chemistry, and biological function of heparin and heparin-like glycosaminoglycans, a field so hampered in the past by the lack of suitable methodologies.

## Acknowledgments

It is not possible to give proper credit to all those who contributed to the work discussed in this article, but all my graduate students and postdoctoral associates are named in Reference [3]. The National Institutes of Health funded the work described over the entire period, beginning with grant GM 05472 awarded in 1958 and running continuously through 1998, and later through grant RR00317, 1966, which ran from through 1995. The early alkaloid work and high resolution mass spectrometer was also supported by a grant from the National Science Foundation.

## References

1. van Itallie, L.; Steenhauer, A. J. *Rauwolfia serpentina* Benth. *Arch. Pharm.* **1932**, *270*, 313–322.
2. Dorfman, L.; Furlenmeier, A.; Huebner, C. F.; Lucas, R.; MacPhillamy, H. B.; Mueller, J. M.; Schlittler, E.; Schwyzer, R.; St. Andre, A. F. *Rauwolfia* Alkaloids. VIII. The Constitution of Reserpin. *Helv. Chim. Acta.* **1954**, *37*, 59–75.
3. Biemann, K. The Massachusetts Institute of Technology Mass Spectrometry School. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 332–338.
4. Arnold, W.; von Philipsborn, W.; Schmid, H.; Karrer, P. Calabash Alkaloids. XXIII. C-Alkaloid T [sarpagine methyl ether, lochnerine]. *Helv. Chim. Acta.* **1957**, *40*, 705–716.
5. Woodward, R. B. New Development in the Chemistry of Natural Substances. *Angew. Chem.* **1956**, *68*, 13–20.
6. Biemann, K. The Determination of the Carbom Skeleton of Sarpagine by Mass Spectrometry. *Tetrahedron. Lett.* **1960**, *15*, 9–13.
7. Bartlett, M. F.; Dickel, D. F.; Taylor, W. I. The Alkaloids of *Tabernanthe iboga*. IV. The Structures of Ibogamine, Ibogaine, Tabernanthe, and Voacangine. *J. Am. Chem. Soc.* **1958**, *80*, 126–136.
8. Hesse, O. Studien Über Argentinische Quebrachodrogen. *Liebigs Ann.* **1882**, *211*, 249–282.
9. Witkop, B. Quebrachamine I. *J. Am. Chem. Soc.* **1957**, *79*, 3193–3200.
10. Cohen, L. A.; Daly, J. W.; Kny, H.; Witkop, B. Nuclear Magnetic Resonance (NMR) Spectra of Indoles. *J. Am. Chem. Soc.* **1960**, *82*, 2184–2187.

11. Kny, H.; Witkop, B. Quebrachamine. II. *J. Org. Chem.* **1960**, *25*, 635-637.
12. Mills, J. F. D.; Nyburg, S. C. Molecular Structure of Aspidospermine. *Tetrahedron Lett.* **1959**, *11*, 1-3.
13. Biemann, K.; Spiteller, G. The Structure of Quebrachamine. *Tetrahedron Lett.* **1961**, *9*, 299-304; Applications of Mass Spectrometry to Structure Problems. VIII. Quebrachamine. *J. Am. Chem. Soc.* **1962**, *84*, 4578-4586.
14. Biemann, K. Applications of Mass Spectrometry to Structure Problems. IV. Sarpagine. *J. Am. Chem. Soc.* **1961**, *83*, 4801-4805.
15. Biemann, K.; Friedmann-Spiteller, M.; Spiteller, G. An Investigation by Mass Spectrometry of the Alkaloids of *Aspidosperma quebracho blanco*. *Tetrahedron. Lett.* **1961**, *14*, 485-492.
16. Djerassi, C. Steroids Made It Possible. Organic Mass Spectrometry. *Org. Mass Spectrom.* **1992**, *27*, 1341-1347.
17. Biemann, K. *Mass Spectrometry: Organic Chemical Applications*. McGraw-Hill Book Company, Inc.: New York, 1962 (Reprinted 1998 as Vol I of ASMS Publications "Classical Works in Mass Spectrometry").
18. Biemann, K.; Friedmann-Spiteller, M.; Spiteller, G. Applications of Mass Spectrometry to Structure Problems. X. Alkaloids of the Bark of *Aspidosperma quebracho blanco*. *J. Am. Chem. Soc.* **1963**, *85*, 631-638.
19. Beynon, J. H. High-Resolution Mass Spectrometry of Organic Materials. *Proceedings of the Advances in Mass Spectrometry Conference*; Cambridge, UK, 1959; Volume Date 1958; pp 328-354.
20. Bommer, P.; McMurray, W. J.; Biemann, K. High Resolution Mass Spectrometry of Natural Products. Vinblastine and Derivatives. *J. Am. Chem. Soc.* **1964**, *86*, 1439-1440.
21. Neuss, N.; Gorman, M.; Hargrove, W.; Cone, N. J.; Biemann, K.; Büchi, G.; Manning, R. E. Vinca Alkaloids. XXI. The Structure of the Oncolytic Alkaloids Vinblastine (VLB) and Vincristine (VCR). *J. Am. Chem. Soc.* **1964**, *86*, 1440-1442.
22. Biemann, K. Mass Spectrometry of Selected Natural Products. *In Progress in the Chemistry of Organic Natural Products*; Zechmeister, L., Ed.; Springer Verlag: Vienna, 1966; pp 1-98.
23. Biemann, K.; McCloskey, J. A. Applications of Mass Spectrometry to Structure Problems. VI. Nucleosides. *J. Am. Chem. Soc.* **1962**, *84*, 2005-2006; Mass Spectra of Organic Molecules. II. Amino Acids. *Ibid.*, 3192-3193.
24. Hites, R. A.; Biemann, K. A Mass Spectrometer-Computer System Particularly Suited for Gas Chromatography of Complex Mixtures. *Anal. Chem.* **1968**, *40*, 1217-1221 and earlier references therein.
25. James, A. T.; Martin, A. J. P. Gas-Liquid Partition Chromatography. A Technique of the Analysis of Volatile Minerals. *Analyst* **1952**, *77*, 915-932.
26. Ryhage, R. Use of a Mass Spectrometer as a Detector and Analyzer for Effluents Emerging from High Temperature Gas Liquid Chromatography Columns. *Anal. Chem.* **1964**, *36*, 759-764.
27. Watson, J. T.; Biemann, K. High Resolution Mass Spectra of Compounds Emerging from a Gas Chromatograph. *Anal. Chem.* **1964**, *36*, 1135-1137.
28. Biemann, K.; Gapp, F.; Seibl, J. Applications of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides. *J. Am. Chem. Soc.* **1959**, *81*, 2274-2275.
29. Sanger, F.; Thompson, E. O. P. Amino Acid Sequence in the Glycyl Chain of Insulin. II. Peptides from Enzymic Hydrolyzates. *Biochem. J.* **1953**, *53*, 366-374.
30. Edman, Pehr. Stepwise Degradation of Peptides via Phenylthiohydantoins. *Acta Chem. Scand.* **1953**, *7*, 700-701.
31. Biemann, K.; Vetter, W. Separation of Peptide Derivatives by Gas Chromatography Combined with the Mass Spectrometric Determination of the Amino Acid Sequence. *Biochem. Biophys. Res. Commun.* **1960**, *3*, 578-584.
32. Biemann, K. The Coming of Age of Mass Spectrometry in Peptide and Protein Chemistry. *Protein Sci.* **1995**, *4*, 1920-1927.
33. Hudson, G.; Biemann, K. Mass Spectrometric Sequencing of Proteins. The Structure of Subunit I of Monellin. *Biochem. Biophys. Res. Commun.* **1976**, *71*, 212-220.
34. Nau, H.; Biemann, K. Computer-Assisted Assignment of Retention Indices in Gas Chromatography-Mass Spectrometry and its Application to Mixtures of Biological Origin. *Anal. Chem.* **1974**, *46*, 426-434.
35. Nau, H.; Forster, H.-J.; Kelly, J. A.; Biemann, K. Polypeptide Sequencing by a Gas Chromatograph-Mass Spectrometer-Computer System. Characterization of Complex Mixtures of Oligopeptidases Trimethylsilylated Polyamino Alcohols. *Biomed. Mass Spectrom.* **1975**, *2*, 326-339.
36. Carr, S. A.; Hauschka, P. V.; Biemann, K. Gas Chromatographic Mass Spectrometric Sequence Determination of Osteocalcin: A  $\gamma$ -Carboxyglutamic Acid Containing Protein from Chicken Bone. *J. Biol. Chem.* **1981**, *256*, 9944-9950.
37. Bohak, Zvi; Li, Shoen-Lung. The Structure of Monellin and its Relation to the Sweetness of the Protein. *Biochim. Biophys. Acta.* **1976**, *427*, 153-170.
38. (a) Gerber, G. E.; Anderegg, R. G.; Herlihy, W. C.; Gray, C. P.; Niemann, K.; Khorana, H. G. Partial Primary Structure of Bacteriorhodopsin: Sequencing Methods for Membrane Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 227-231. (b) Khorana, H. G.; Gerber, G. E.; Herlihy, W. C.; Gray, C. P.; Anderegg, R. G.; Nihei, K.; Biemann, K. The Amino Acid Sequence of Bacteriorhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 5046-5050.
39. Putney, S. D.; Royal, N. H. Neumann; de Vegvar, H.; Herlihy, W. C.; Biemann, K.; Schimmel, P. R. Primary Structure of a Large Aminoacyl-tRNA Synthetase. *Science* **1981**, *213*, 1497-1501.
40. Morris, H. R. Studies Towards the Complete Sequence Determination of Proteins by Mass Spectrometry; a Rapid Procedure for the Successful Permethylolation of Histidine Containing Peptides. *FEBS Lett.* **1972**, *22*, 257-260.
41. Hunt, D. F.; Buko, A. M.; Ballard, J. M.; Shabanowitz, J.; Giordani, A. Sequence Analysis of Polypeptides by Collision Activated Dissociation on a Triple Quadrupole Mass Spectrometer. *Biomed. Mass Spectrom.* **1981**, *8*, 397-408.
42. Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. Fast Atom Bombardment of Solids (FAB): A New Ion Source for Mass Spectrometry. *J. Chem. Soc. Chem. Commun.* **1981**, *7*, 325-327.
43. Biemann, K. Fast Atom Bombardment Mass Spectrometry in Protein Sequencing. *Proc. Japanese Soc. Med. Mass Spectrom.* **1981**, *6*, 21-32.
44. Biemann, K.; Martin, S. A. Mass Spectrometric Determination of the Amino Acid Sequence of Peptides and Proteins. *Mass Spectrom. Rev.* **1987**, *6*, 1-76.
45. Biemann, K. Mass Spectrometric Methods for Protein Sequencing. *Anal. Chem.* **1986**, *58*, 1289A-1300A.
46. Papyannopoulos, I. A.; Gan, Z.-R.; Wells, W. W.; Biemann, K. A Revised Sequence of Calf Thymus Glutaredoxin. *Biochem. Biophys. Res. Commun.* **1989**, *159*, 1448-1454.
47. Johnson, R. S.; Biemann, K. The Primary Structure of Thioredoxin from *Chromatium vinosum* Determined by High-Performance Tandem Mass Spectrometry. *Biochemistry* **1987**, *26*, 1209-1214.
48. Johnson, R. S.; Biemann, K. Computer Program (SEQPEP) to Aid in the Interpretation of High-Energy Collision Tandem Mass Spectra of Peptides. *Biomed. Environ. Mass Spectrom.* **1989**, *18*, 945-957.
49. Meng, C. K.; Mann, M.; Fenn, J. B. Electrospray Ionization of Some Polypeptides and Small Proteins. *Proceedings of the 36th*

- ASMS Conference on Mass Spectrometry and Allied Topics*; San Francisco, CA, June 5–10, 1998; pp 771–772.
50. Karas, Michael; Hillenkamp, Franz. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60*, 2299–2301.
  51. Zia, J.; Annan, R. S.; Biemann, K. The Correct Molecular Weight of Myoglobin, a Common Calibrant for Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 32–36.
  52. Zeng, Z.; Biemann, K. Determination of N-linked Glycosylation of Yeast External Invertase by MALDI TOF Mass Spectrometry. *J. Mass Spectrom.* **1999**, *34*, 311–329.
  53. Henzel W. J. Stults J. T. Watanabe C. A Novel Approach for Identifying Proteins: Molecular Ion Searching of Protein Data Bases. *Proceedings of the 3rd Symposium of The Protein Society*; Seattle, WA, 1989; p M179.
  54. Juhasz, P.; Biemann, K. Mass Spectrometric Molecular Weight Determination of Highly Acidic Compounds of Biological Significance via their Complexes with Basic Polypeptides. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4333–4337.
  55. Rhomberg, A.; Biemann, K. Mass Spectrometric Analysis of Highly Acidic Polysaccharides. In *A Laboratory Guide to Glycoconjugate Analysis*; Jackson, B.; Gallagher, J. T., Eds.; Birkhauser: Basel, 1997; pp 77–89.
  56. Juhasz, P.; Biemann, K. Utility of Non-covalent Complexes in the MALDI Mass Spectrometry of Heparin-derived Oligosaccharides. *Carbohydr. Res.* **1995**, *270*, 131–147.
  57. Rhomberg, A. J.; Wolf, S.; Biemann, K. Mass Spectrometric Sequencing of Heparin and Heparan Sulfate Using Partial Digestion with Heparinases. *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*; Palm Springs, CA, June, 1997, pp 1026–1027.
  58. Rhomberg, A. J. Mass Spectrometric and Capillary Electrophoretic Investigation of Heparin, Heparinases, and Related Compounds. Ph.D. Thesis, Massachusetts Institute of Technology, 1998.
  59. Rhomberg, A. J.; Ernst, S.; Sasisekharan, R.; Biemann, K. Mass Spectrometric and Capillary Electrophoretic Investigation of the Enzymatic Degradation of Heparin-like Glycosaminoglycans. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4176–4181.
  60. Ernst, S.; Rhomberg, A. J.; Biemann, K.; Sasisekharan, R. Direct Evidence for a Predominantly Exolytic Processive Mechanism for Depolymerization of Heparin-like Glycosaminoglycans by Heparinase I. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4182–4187.
  61. Rhomberg, A. J.; Shriver, Z.; Biemann, K.; Sasisekharan, R. Mass Spectrometric Evidence for the Enzymatic Mechanism of the Depolymerization of Heparin-like Glycosaminoglycans by Heparinase II. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12232–12237.