Mass Spectrometry of Large Molecules

Lectures of a course held at the Joint Research Centre, Ispra (Italy)
5–9 September 1983

Edited by

S. Facchetti

Mass Spectrometry of Large Molecules

Lectures of a course held at the Joint Research Centre, Ispra (Italy)
5–9 September 1983

Edited by

S. Facchetti

Joint Research Centre, Ispra, Italy

Published for the Commission of the European Communities

ELSEVIER [illegible] Tokyo 1985

Commission of the European Communities
Joint Research Centre, Ispra (Varese), Italy

Publication arranged by:
Directorate-General Information Market and Innovation
Luxembourg

Published under licence by:
Elsevier Science Publishers B.V.
Molenwerf 1,
P.O. Box 211, 1000 AE Amsterdam,
The Netherlands

Distributors for the United States and Canada:
Elsevier Science Publishing Company Inc.
52, Vanderbilt Avenue
New York, N.Y. 10017

EUR 9421 EN

Library of Congress Cataloging in Publication Data
Main entry under title:

Mass spectrometry of large molecules.

Includes bibliographies and index.
1. Macromolecules--Analysis--Congresses. 2. Mass spectrometry--Congresses. I. Facchetti, S. (Sergio), 1931- . II. Commission of the European Communities. Joint Research Centre. Ispra Establishment.
QD380.M37 1985 574.19'285 84-28666
ISBN 0-444-42456-3 (U.S.)

Legal Notice

Neither the Commission of the European Communities nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

ISBN 0-444-42456-3

Printed in The Netherlands

Mass Spectrometry of Large Molecules

ISPRA — COURSES ON CHEMICAL AND ENVIRONMENTAL SCIENCE

A series devoted to the publication of courses and educational seminars given at the Joint Research Centre, Ispra Establishment, as part of its education and training programme.

Published for the Commission of the European Communities, Directorate-General Information Market and Innovation.

Volumes already published

- APPLICATIONS OF MASS SPECTROMETRY TO TRACE ANALYSIS
  Edited by S. Facchetti

- ANALYTICAL TECHNIQUES FOR HEAVY METALS IN BIOLOGICAL FLUIDS
  Edited by S. Facchetti

- OPTICAL REMOTE SENSING OF AIR POLLUTION
  Edited by P. Camagni and S. Sandroni

The publisher will accept continuation orders for this series which may be cancelled at any time and which provide for automatic billing and shipping of each title in the series upon publication. Please write for details.

PREFACE

In recent years considerable attention has been focused on new methods for the study of the large class of biologically important non-volatile and thermally unstable molecules. When these compounds are available in trace amounts, mass spectrometry provides the only means for investigation of their structure and for their correct identification through the determination of molecular weight. However, as many of these compounds are unstable and are readily transformed or decomposed, new ionization techniques for mass spectrometry have been developed to overcome these difficulties. The present volume reflects this situation. The most important ionization techniques in current use have been considered and the latest developments in the instrumentation in high molecular weight mass spectrometry have been illustrated. Most of the book, however, has been dedicated to the identification of molecules of biological and biomedical interest, and of polymers, as well as to the use of mass spectrometry in the study of structural problems and in research for applications. The book should therefore provide food for thought for biochemists, chemists and analytical mass spectrometrists.

This volume includes the lectures given at a course on mass spectrometry of large molecules. The course, organized within the framework of the Training and Education programme of the Joint Research Centre of the European Communities, was held at the Ispra Establishment from 5th to 9th September 1983. Most of the lectures, however, were updated by the authors during 1984.

The important role played by the lecturers, whose names and affiliations are given in a separate list, is gratefully acknowledged here. Further thanks go to those attending the course for their contributions to the discussions. Finally the editor is indebted to Mrs. A. Confortini Pinolini for her help in the organization of the course and for her skilled editorial assistance, to Mrs. M. Van Andel for the typing of the volume and to Dr. R. Self for having kindly revised the English texts.

Sergio Facchetti
Ispra, July 27th, 1984

Alphabetical list of Authors

G. AUDISIO	CNR, Area di ricerca, 35100 Padova, Italy
A.J.H. BOERBOOM	FOM-Institute for Atomic and Molecular Physics, Kruislaan 407, 1089 SJ Amsterdam, The Netherlands
A.L. BURLINGAME	Mass Spectrometry Resource and Dept. of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, U.S.A.
F. CAMPAGNARI	Laboratory of Biochemistry, Biology Group D.G. XII, Commission of the European Communities, Joint Research Centre-Ispra Establishment, 21020 Ispra, Varese, Italy
S. DAOLIO	CNR, Area di ricerca, 35100 Padova, Italy
A. DELL	Department of Biochemistry, Imperial College of Science and Technology, London SW7, U.K.
J. EAGLES	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
S. FACCHIN	CNR, Area di ricerca, 35100 Padova, Italy
G.R. FENWICK	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
S. FOTI	Istituto Dipartimentale di Chimica e Chimica Industriale, University of Catania, 95195 Catania, Italy

M. GLERIA	CNR, Area di ricerca, 35100 Padova, Italy
R.A. McDONELL	Department of Biochemistry, Imperial College Of Science and Technology, London SW7, U.K.
F.A. MELLON	Food Research Institute, Colney Lane Norwich NR4 7UA, U.K.
G. MONTAUDO	Istituto Dipartimentale di Chimica e Chimica Industriale, University of Catania, 95195 Catania, Italy
H.R. MORRIS	Department of Biochemistry, Imperial College of Science and Technology, London SW7, U.K.
M. OEHME	Norwegian Institute for Air Research, P.O. Box 130, 2001 Lillestrom, Norway
C. PAGURA	CNR, Area di ricerca, 35100 Padova, Italy
M. PANICO	Department of Biochemistry, Imperial College of Science and Technology, London SW7, U.K.
A. PARENTE	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
R. SELF	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
M.N. SHORT	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.

G.W. TAYLOR	Department of Clinical Pharmacology, Royal Postgraduate Medical School, Ducane Road, London W12 0HS, U.K.
P. TRALDI	CNR, Area di ricerca, 35100 Padova, Italy
E. VECCHI	CNR, Area di ricerca, 35100 Padova, Italy
U. VETTORI	CNR, Area di Ricerca, 35100 Padova, Italy
J.O. WHITNEY	Liver Center and Dept. of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, U.S.A.
D.H. WILLIAMS	University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

M. GLERIA	CNR, Area di ricerca, 35100 Padova, Italy
R.A. McDOWELL	Department of Biochemistry, Imperial College Of Science and Technology, London SW7, U.K.
F.A. MELLON	Food Research Institute, Colney Lane Norwich NR4 7UA, U.K.
G. MONTALDO	Istituto Dipartimentale di Chimica e Chimica Industriale, University of Catania, 95195 Catania, Italy
H.R. MORRIS	Department of Biochemistry, Imperial College of Science and Technology, London SW7, U.K.
M. OEHME	Norwegian Institute for Air Research, P.O. Box 130, 2001 Lillestrom, Norway
C. PAGURA	CNR, Area di ricerca, 35100 Padova, Italy
M. PANICO	Department of Biochemistry, Imperial College of Science and Technology, London SW7, U.K.
A. PARENTE	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
R. SELF	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.
M.N. SHORT	Food Research Institute, Colney Lane, Norwich NR4 7UA, U.K.

Preface

List of Authors

Table of contents

Classification of biological macromolecules. A survey for macromolecular analysis in mass spectrometry by F. Campagnari 1

Instrumentation in high molecular weight mass spectrometry by A.J.H. Boerboom 35

Simultaneous ion detection, by A.J.H. Boerboom 53

Direct inlet systems for identification of labile and/or large molecules, by P. Traldi, S. Daolio, S. Facchin, C. Pagura and U. Vettori 75

Structure and elucidation of "large" biological molecules by D.H. Williams 93

Mass spectrometry of intermediate molecular weight (1000 < m/z < 10,000 d) labile biological substances by A.L. Burlingame 105

Perspective on problems in biomolecular mass spectrometry by H.R. Morris, A. Dell, R.A. McDowell, M. Panico and G.W. Taylor 111

New biomedical applications of mass spectrometry by H.R. Morris, A. Dell, R.A. McDowell, M. Panico and G.W. Taylor 127

Modification of peptides for mass spectrometry by R. Self, M.N. Short and A. Parente 151

Structural studies on peptides by mass spectrometry 167
by D.H. Williams

Mass spectrometric analysis of bile acids in neonatal liver diseases, by J.O. Whitney and A.L. Burlingame 185

The potential of fast atom bombardment mass spectrometry in food and agricultural science, by R. Self, J. Eagles, F.A. Mellon and G.R. Fenwick 209

Negative ion chemical ionization mass spectrometry 233
by M. Oehme

Application of negative ion mass spectrometry to medium-size and large molecules, by M. Oehme 249

Pyrolysis mass spectrometry, by A.J.H. Boerboom 265

Direct mass spectrometry of polymers, by S. Foti and G. Montaudo 283

The mass spectrometer as synthetizer-analyser device in polymer chemistry : the polymerization of cyclophosphazenes by P. Traldi, M. Gleria, G. Audisio, S. Daolio and E. Vecchi 303

Subject Index 319

Mass Spectrometry of Large Molecules, Lectures of a course held at the Joint Research Centre, Ispra, Italy, 5—9 September 1983, S. Facchetti (Ed.). 1—34

CLASSIFICATION OF BIOLOGICAL MACROMOLECULES
A Survey for Macromolecular Analysis in Mass Spectrometry

F. CAMPAGNARI

1. PREMISE

The Lord Kelvin's statement that "science is phycics or stamp collecting" rests on a background of truth, but it can be challenged. In fact, the exercise of classification and the related theory of taxonomy have many good uses in modern sciences. Ordered collections of experimental notions in a given field can stimulate further studies and are of easy approach also to scholars of foreign disciplines. They are needed to open new areas of research to non-specialists.

Similar feelings led to select a general survey of biological macromolecules as an introduction to this course. Conceivably, mass spectrometry analysts would be helped to become more interested in biopolymers.

At first, I underestimated the responsibility of delivering this lecture. Due to the wide range of chemical forms under consideration, I had to take options. Therefore, the topic will not be covered in detail nor adequately. Important types of molecules will be recalled only with the aid of indicative examples. Moreover, I will attempt to tailor the classification to the standpoint of mass spectrometry by taking into account elemental composition, chemical structure and physical size of the biological macromolecules and largely neglecting structural conformation. This is a relevant omission that is at variance with respect to ordinary lectures on the same subject.

2. LIVING MATTER HAS CELLULAR ORGANIZATION

Chemically, living matter consists of organic molecules dispersed or dissolved in aqueous systems. It is produced and still contained in the microscopic cells (with typical diameters of about 10 μ) that represent the ubiquitous units of life organization. All forms of life on the earth have either a unicellular or multicellular composition.

A cell can be conceived essentially as a hydrated collection of biological materials enclosed by a peripheral plasmic membrane. Besides water, the substances of life include inorganic salts, simple metabolites, active organic compounds and, more notably, various types of macromolecules. The plasmic membrane is a fluid lipid bilayer embedding globular proteins and acts as a hydro-

phobic barrier to separate the biological materials from the environment, thus forming an isolated aqueous compartment. Any cell contains one or more macromolecules of deoxyribonucleic acid, DNA, i.e. of the genetic substance which stores the inherited specific information necessary to organize biological structures and functions.

Most primitive prokaryotic cells possess a single system of peripheral membranes surrounding a main metabolic compartment with or without a minor periplasmic space.

The more evoluted and complex eukaryotic cells, including those of plants and animals, are provided also of internal membranes and have multiple biochemical compartments. One intracellular membrane encompasses the nucleus, a site that contains DNA coiled around small proteins named histones to form thin fibers of cromatin. The information coded in the DNA is transcribed as specific messages in the molecules of ribonucleic acid, RNA.

Although originated from a same common ancestor called progenote, the nucleated eukaryotic cells diverged early in evolution from the two prokaryotic cell types of archaebacteria and eubacteria. In the course of their independent development, the cells of the three primary kingdoms acquired a few distinct features in the basic biochemical mechanisms for processing of genetic information. These differences persist in contemporary organisms derived from the three cell lineages and are reflected to a certain degree in the biosynthesis of macromolecular compounds.

3. CONSTITUENTS OF LIVING MATTER

In normal cells water amounts to about 70 or 80% of the mass, inorganic salts about 1% and organic molecules make up the remainder. Almost the same gross composition is found, with few exceptions, in the extracellular substance that fills the interstitial spaces of animal tissues. Water is abundant also in plants, although to variable extents in different species and vegetal formations.

The dry biomass of cells is almost totally (up to 95% in many cases) contributed by four elements of the first two periods: hydrogen, oxygen, carbon and nitrogen. They have the common property to form covalent bonds by electron-pair sharing. Hydrogen can share one electron, oxygen two, nitrogen three and carbon four. By combining with each other, these atoms fill up to completion their outer electron shells and thus they can give rise to many different molecules of great stability.

The wide variety of biomolecules synthesized in the cells is mainly due to the ability of carbon atoms to join each other in serial -C-C- bonds bridged occasionally by nitrogen, oxygen or phosphoesters in -C-N-C-, -C-O-C- and -C-O-P-O-C- linkage combinations, respectively. We all learned from organic

chemistry that carbon holds four covalent bonds and may bind four electron-pair sharing atoms as to produce a variety of linear, branched and cyclic chemical structures. The -C-C- bond systems with or without nitrogen, oxygen and phosphoester bridges are the skeletons of biological substances in the cells.

A conspicuous feature of living matter is the property to polymerize linearly simple organic compounds in extended gigantic macromolecules. These biopolymers are assembled stepwise by terminal addition of single building-blocks to growing chainlike structures. The reactions are carried out by enzymes that use the activated monomers (coupled by high-energy bonds to specific carriers) as substrates. The synthesized macromolecules form the largest fraction of organic cell constituents and contribute to more than 80% of the cell dry weight, much in excess of non-polymeric compounds.

In average eukaryotic cells from phylogenetically distant organisms, the relative proportions of the main chemical constituents are more or less similar. However, differentiated cells of certain tissues in animals and plants synthesize large amounts of materials that are needed for specialized functions. These products are macromolecular polymers in most cases and are either kept inside the cells or secreted as extracellular substances. Thus, chemical variations are introduced in living matter over a rather uniform fundamental pattern.

The gross chemical composition of rat liver (data from our laboratory) is given in Table 1 as a general example of the relative distribution of different materials in animal cells. Liver is a soft tissue with high cellularity and differs from other parenchymatous organs for its metabolic accumulation of glycogen, a storage polysaccharide.

The distribution of biological substances according to mass reflects their hierarchical complexity, is relevant to our definition of macromolecules and is reported in Table 2.

The group of light molecules with masses below 350 from Table 2 includes fuels, intermediary and biosynthetic metabolites up to the "chemical building-blocks" that are precursors of the macromolecules, the steroids, several hormones, neurotransmitters, prostaglandins and most vitamins.

The large compounds falling in the mass range from 350 to about 1000 can be divided into two types: lipids and complex molecules that have functional roles in biochemical processes and can be conveniently named as active compounds. The latter ones comprise: phosphorylated sugars, certain activated "building-blocks" (forms of nucleotides and sugars that react as immediate substrates of the enzymatic polymerizations), important coenzymes and some of the most complex antibiotics.

Although not defined as macromolecules, these large chemical structures either reach or exceed the limit of detection by ordinary mass spectrometers and will be shortly considered in our discussion.

TABLE 1

Percent gross composition of rat liver

	Wet tissue	Dry residue
Inorganic molecules		
Water	71.1	
Mineral salts	1.1	3.8
Small organic molecules		
Carbohydrates	0.1	0.4
Aminoacids	0.3	1.0
Nucleotides	0.1	0.4
Lipids	3.1	10.7
Macromolecules		
Polysaccharides	3.7	12.8
DNA	0.2	0.7
RNA	0.6	2.1
Proteins	19.7	68.1

TABLE 2

Order of biological materials according to mass

Materials	Mass range
Light molecules (fuels, building blocks, ...)	below 350
Large compounds (lipids, coenzymes, ...)	350 1000
Macromolecules (biopolymers)	10^3 10^{12}
Macromolecular complexes (complex proteins, ...)	10^4 10^8
Supramolecular structures (ribosomes, ...)	10^6 10^9

The terms "macromolecules" is reserved to compounds with masses above 1000 daltons, i.e. to all polymeric substances down to short oligomers. The fixed minimal size for macromolecules is already attained by chains of 4-5 nucleotides, 7-10 sugars or aminoacids and about 15 isoprene units.

Biopolymers are classified according to the chemical nature of their fundamental subunits and are assigned to four primary types: polysaccharides or glycans, polypeptides or simple proteins, nucleic acids and polyisoprenoids. They can be further distinguished in homopolymers and heteropolymers: the former ones are combinations of many identical subunits, the latter ones are made up of non-identical chemical residues.

In most chain-like heteropolymers, the alignment of different subunits is not random but it conforms to well established patterns. The distinct monomers either alternate in fixed repeated sequences giving rise to regular polymers (certain polysaccharides) or are ordered in variable sequences specified by the genetic messages "written" in the informational polymers (DNA, RNA, proteins).

Certain biological macromolecules occur in nature as mixed polymers that generally contain linear or branched chains of monosaccharides covalently linked to lipids or to peptides.

The range of the molecular weights of biopolymers extends over 9 orders of magnitude from the masses of compounds with a dozen of monomers to those of the polynucleotide strands of animal and plant DNAs containing many millions of subunit residues (each residue with a weight of about 330 daltons).

The macromolecular complexes are integral aggregates of polymers with a definite proportion of their individual constituents. They result from the cooperative association of single macromolecules that are held together by non-covalent bonds and fit precisely a given spatial arrangement.

The term "macromolecular complex" rather than the plain name "macromolecule" applies to any composite protein formed by multiple polypeptides. The various aminoacids chains are serially indicated by letters of the greek alphabet. Thus the normal human hemoglobin of adult life is a macromolecular tetramer with two pairs of non-identical polypeptides and it is denoted $Hb\alpha_2\beta_2$. The set of macromolecular complexes includes the antibodies, many enzymes, specific transport carriers, cell surface and membrane receptors and many functional or structural proteins.

Macromolecular systems such as the lipoproteins (polypeptides intimately associated with lipids), the contractile and motile elements and the multienzyme complexes possess a higher order of organization than the composite proteins. They resemble the next-in-hierarchy supramolecular structures from which they differ for the degree of compactness.

The supramolecular structures are dense macromolecular aggregates with constant composition and high stability. Due to their mechanical resistance, they are readily isolated as distinct morphological entities or true particles.

Among these supramolecular systems we should list: elementary fibers, microtubules, microfilaments, integrated sets of contractile elements and the different combinations of hyaluronate, link proteins and proteoglycans that form the ground substances of various connective tissues.

The most compact aggregates are those made up of protein and DNA or RNA. Such are the nucleosomes (repeating structures of DNA and histones in the chromatin, as mentioned above) and the nucleoprotein particles containing RNA molecules.

The largest and most prominent ribonucleoprotein particles are the ribosomes (machineries for protein synthesis) and the nucleoli (sites of production of ribosomal RNA, rRNA, in cell nuclei). These particles differ from the intracellular organelles (mitochondria, chloroplasts, lysosomes, nuclei, endoplasmic reticulum) that are encircled by membranes and form independent metabolic compartments.