MASS SPECTROMETRY of NUCLEOSIDES and NUCLEIC ACIDS

Edited by
Joseph H. Banoub
Patrick A. Limbach

CRC Press
Taylor & Francis Group

MASS SPECTROMETRY of NUCLEOSIDES and NUCLEIC ACIDS

Edited by
Joseph H. Banoub
Patrick A. Limbach

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-4402-7 (Hardback)

Library of Congress Cataloging-in-Publication Data

Mass spectrometry of nucleosides and nucleic acids / editors, Joseph H. Banoub, Patrick A. Limbach.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4200-4402-7 (hardcover : alk. paper)
1. Nucleic acids--Spectra. 2. Nucleosides--Spectra. 3. Mass spectrometry. I. Banoub, Joseph H. II. Limbach, Patrick A. III. Title.

QD433.M37 2010
572.80284--dc22 2009040943

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

MASS SPECTROMETRY of NUCLEOSIDES and NUCLEIC ACIDS

Preface

Nucleic acids and their components play an important role in a variety of fundamental biological processes. The elucidation of the molecular structure of deoxyribonucleic acid (DNA) in 1953 by Watson and Crick led to the so-called *central dogma of molecular biology*. This dogma proclaims that DNA is transcribed to form messenger ribonucleic acid (mRNA) of a complementary sequence that directs its own replication and the sequence of nucleobases in the RNA is then translated to the corresponding sequence of amino acids to form a protein. In addition to mRNA, two other RNA species, transfer RNA (tRNA) and ribosomal RNA (rRNA), are required for this translation process.

The field of genomics has evolved over the last few years, culminating in the sequencing and characterization of the entire human genome. Furthermore, genetic theory contributed to the concept of coding by genes. A gene is defined, biochemically, as a segment of DNA (or, in a few cases, RNA) that encodes the information required to produce a functional biological product—a protein. Current challenges include the unraveling of the relationship between the genome and the proteome. The genome can be used to predict the total potential proteome, including all of the modifications that can be carried out after the initial translation of mRNA to protein by tRNAs and rRNAs.

Consequently, nucleic acid research has had a most profound impact on molecular biology. The study of noncovalent interactions between nucleic acids and peptides or proteins is a highly active research field. Such interactions are involved in many cellular processes, including replication, regulation of gene expression, and DNA packaging. Nearly all the functions of nucleic acids are accomplished by interacting with proteins. Recognition and selectivity are achieved through noncovalent contacts between virtually all biopolymers (proteins, DNA, RNA, polysaccharides, and membrane lipids) that are present in living organisms. The formation and dissociation of these weak interactions are crucial in a vast number of biochemical events. In addition, weak interactions between nucleases and DNA or RNA are central to the recombination, repair, and replication of these molecules in cells.

It is not possible to overestimate the role of metal ions in determining the three-dimensional architecture of nucleic acids, their helices, and higher-order molecular assemblies. Protein–protein, protein–nucleic acid, protein–ligand, and antibody–antigen interactions are other examples that involve noncovalent interactions, which are also of fundamental importance for the pharmaceutical industry in the evaluation of potential drug candidates.

The ability of DNA adducts to induce mutagenesis and carcinogenesis is dependent on their chemical structure, stability, and ability to be recognized by specialized DNA repair proteins. The local DNA sequence context plays a major role in mediating the rate of nucleobase adducts repair and their mispairing potency. Furthermore, the exact location of a DNA adduct within the gene determines whether the resulting mutation affects the structure and function of the gene product.

Among the biophysical techniques employed to study the large biomolecular assemblies involved in these tasks, mass spectrometry presents the greatest potential based on its intrinsic characteristics and accessible information. Therefore, mass spectrometry has had a major role in elucidating the regulatory components and mechanisms involved in the progression from gene to functional protein. Furthermore, analyses of nucleic acids is required to establish their size, purity, and sequence as a prerequisite to their use as molecular probes or therapeutics in biomedical science.

The development of soft-ionization techniques for mass spectrometry has, without question, transformed the field of nucleic acid chemistry. Significant progress in the area of accurate mass

determination, sequencing, and the study of noncovalent interactions has been made possible by the use of ionization techniques such as matrix-assisted laser desorption ionization coupled to time-of-flight (MALDI-TOF) mass analyzers and electrospray ionization (ESI) coupled to conventional analyzers and to tandem mass spectrometric (MS/MS) instruments. Mass spectrometry is now one of the definitive methods for the characterization of isolated and synthetic DNA, RNA, and related products. In addition to accurately determining an intact mass, one can obtain information on the molecular structure and biopolymer sequence using a number of different mass spectrometry approaches. These approaches have provided valuable information on structurally specific biomolecular DNA and RNA interactions. They can also enable the direct determination of stoichiometry for post-translational modifications of specific DNA–drug, DNA–protein, RNA–protein, and DNA–RNA noncovalent and covalent associations. Furthermore, these mass spectrometry approaches can be applied to DNA and RNA sequencing analyses, single-nucleotide polymorphisms, genotyping, mutation detection, genetic diagnosis, and probing of viral structures.

The editors acknowledge the special efforts of the contributors to this CRC/Taylor & Francis publication: Diethard K. Bohme (Canada), William Buchmann (France), Ed Dudley (UK), Daniele Fabris (USA), Valérie Gabelica (Belgium), Kristina Håkansson (USA), Filip Lemière (Belgium), Russell P. Newton (UK), Natalia Y. Tretyakova (USA), Antonio Triolo (Italy), Jean-Jacques Vasseur (France), and Yinsheng Wang (USA). Each contributor, an expert in their particular discipline, has shared their knowledge and understanding of the field of nucleic acid mass spectrometry.

Our aim is to expose you, the reader, to the latest developments in the field of nucleic acid mass spectrometry. The past 15 years have seen dramatic progress in this field. Thus, in addition to contributions that review and summarize these developments and advances, this book also contains chapters describing the next generation of mass spectrometry analyses of nucleic acids and their complexes that have been enabled by such past accomplishments.

This book should be equally accessible to the nucleic acids expert who is a mass spectrometry novice as well as to those with expertise in mass spectrometry but a minimal appreciation of nucleic acids. The exciting developments in mass spectrometry technology have fueled incredible advances in our understanding of nucleic acids and their complexes. We believe these contributions have captured these advances in outstanding fashion, serving to inspire new findings and developments of interest to all in this field.

Finally, no work of this scope is possible without the guidance and support of the publisher. The editors wish to especially acknowledge Ms. Jill J. Jurgensen for her expert help in preparing and producing this book.

Joseph H. Banoub (Canada)
Patrick A. Limbach (USA)

Editors

Joseph H. Banoub was born in 1947 and obtained his BS degree from the University of Alexandria (Egypt) in 1969. In 1977 he obtained his PhD degree in organic chemistry from the University de Montreal (Quebec) under the supervision of Stephen Hanessian. He spent three years as a research associate with the MDS Health Group at the National Research Council of Canada, Ottawa, synthesizing human neoglycoconjugate vaccines. In 1979 he moved to St. John's, Newfoundland, as a research scientist with the Federal Government of Canada, Department of Fisheries and Oceans. During this period he worked on the structural elucidation of various bacterial antigens and the formation of artificial vaccines for aquatic Gram-negative bacteria. He is currently principal research scientist and head of the special project program of the Department of Fisheries and Oceans and his research focus has evolved into the mass spectrometry and tandem mass spectrometry study of biomolecules (bacterial lipopolysaccharides, nucleic acids, and proteomics). He has been an adjunct professor of biochemistry at Memorial University of Newfoundland since 1984 and an adjunct professor of chemistry since 2006. His present research interests include the uses of tandem mass spectrometry for the structural elucidation of biologically active molecules.

Patrick A. Limbach was born in 1966 and obtained his BS degree from Centre College (Kentucky, USA) in 1988. In 1992, he obtained his PhD degree in analytical chemistry from The Ohio State University under the supervision of Alan G. Marshall. He spent two years as a postdoctoral researcher in the laboratory of James A. McCloskey at the University of Utah. In 1995, he accepted his first faculty position in the Department of Chemistry at Louisiana State University in Baton Rouge, Louisiana. In 2001 he moved to his current location in the Department of Chemistry at the University of Cincinnati. He is a professor of chemistry, serves as department head, and is an Ohio Eminent Scholar. His research interests include the development of new mass spectrometry methods for analyzing RNAs, the identification of unknown modified nucleosides, the characterization of RNA–protein complexes, and investigating the role of modified nucleosides in biological systems.

Contributors

Janna Anichina
York University
Toronto, ON, Canada

Joseph H. Banoub
Fisheries and Oceans Canada
Science Branch
Newfoundland, Canada

and

Department of Chemistry
Memorial University of Newfoundland
Newfoundland, Canada

Diethard K. Bohme
York University
Toronto, ON, Canada

William Buchmann
Analysis and Modeling Laboratory for Biology and Environment
University of Evry-Valley of the Essonne
Evry, France

C. Callie Coombs
Department of Chemistry
University of Cincinnati
Cincinnati, Ohio

Ed Dudley
Department of Biological Sciences
Swansea University
Singleton Park, Swansea,
United Kingdom

Daniele Fabris
Department of Chemistry and Biochemistry
University of Maryland Baltimore County
Baltimore, Maryland

Valérie Gabelica
Mass Spectrometry Laboratory
University of Liège and FNRS
Liège, Belgium

Chunang Gu
Environmental Toxicology Graduate Program
University of California
Riverside, California

Nathan A. Hagan
Department of Chemistry and Biochemistry
University of Maryland Baltimore County
Baltimore, Maryland

Kristina Håkansson
Department of Chemistry
University of Michigan
Ann Arbor, Michigan

Haizheng Hong
Environmental Toxicology Graduate Program
University of California
Riverside, California

Mahmud Hossain
Department of Chemistry
University of Cincinnati
Cincinnati, Ohio

Farid Jahouh
Fisheries and Oceans Canada
Science Branch
Newfoundland, Canada

and

Department of Chemistry
Memorial University of Newfoundland
Newfoundland, Canada

Nicolas Joly
Department of Chemistry
Rue de l'Université
Béthune, France

Filip Lemière
Department of Chemistry
University of Antwerp
Antwerp, Belgium

Patrick A. Limbach
Department of Chemistry
University of Cincinnati
Cincinnati, Ohio

Patrick Martin
Department of Chemistry
Rue de l'Université
Béthune, France

Judith Miller-Banoub
Fisheries and Oceans Canada
Science Branch
Newfoundland, Canada

and

Department of Chemistry
Memorial University of Newfoundland
Newfoundland, Canada

Russell P. Newton
Biological Sciences, School of the Environment and Society
Swansea University
Singleton Park, Swansea, United Kingdom

Keiichiro Ohara
Max Mousseron Institute of Biomolecules
University of Montpellier
Montpellier, France

Michael Smietana
Max Mousseron Institute of Biomolecules
University of Montpellier
Montpellier, France

Natalia Y. Tretyakova
Department of Medicinal Chemistry
University of Minnesota
Minneapolis, Minnesota

Antonio Triolo
Menarini Ricerche Spa
Florence, Italy

Kevin B. Turner
Department of Chemistry and Biochemistry
University of Maryland Baltimore County
Baltimore, Maryland

Jean-Jacques Vasseur
Max Mousseron Institute of Biomolecules
University of Montpellier
Montpellier, France

Yinsheng Wang
Department of Chemistry
University of California
Riverside, California

Jiong Yang
Department of Chemistry
University of Michigan
Ann Arbor, Michigan

Contents

1 Overview of Recent Developments in the Mass Spectrometry of Nucleic Acid and Constituents

Joseph H. Banoub, Judith Miller-Banoub, Farid Jahouh, Nicolas Joly, and Patrick Martin

CONTENTS

1.1 INTRODUCTION

Nucleic acids and their components play an important role in a variety of fundamental biological processes.[1] The elucidation of the molecular structure of DNA in 1953 by Watson and Crick led to the so-called central dogma of molecular biology.[2] This dogma proclaims that DNA is transcribed to form a messenger RNA (mRNA) of a complementary sequence that directs its own replication, and the sequence of nucleobases in the RNA is then translated to the corresponding sequence of amino acids to form a protein.[2] In addition to mRNA, two other RNA species, transfer RNA (tRNA) and ribosomal RNA (rRNA), are required for this translation process. Nucleic acid research has consequently had a most profound impact on molecular biology, initiating an avalanche of new disciplines. The field of genomics has evolved over the last few years, culminating in the sequencing and characterization of the entire human genome.[3] Current challenges include the unraveling of the relationship between the genome and the proteome. The genome can be used to predict the total potential proteome, including all of the modifications that can be carried out after the initial translation from mRNA to protein. Mass spectrometry (MS) has had a major role in elucidating the regulatory mechanism involved in the progression from gene to functional protein.[4]

Nucleotides represent the primary level for the storage and transmission of genetic information. They are not only crucial to the organization and execution of protein synthesis but also function as secondary messengers, metabolic regulators, and components of vitamins.[5] They also act as high-energy intermediates driving thermodynamically unfavorable enzyme-catalyzed reactions. Two classes of heterocyclic nitrogenous bases are found in nucleic acids; the two purine derivatives, adenine (A) and guanine (G), and the three pyrimidine derivatives, uracil (U), thymine (T), and cytosine (C). There have been many studies devoted to the separation, isolation, and purification of nucleic acids from biological samples as well as the separation and identification of constituents of hydrolyzed RNA and DNA at the nucleotide, nucleoside, and nucleobase levels.[6] Elucidation of the MS fragmentation patterns for simple nucleosides and nucleotides is a key step in developing an understanding of the MS fragmentation patterns of more complex oligonucleotides and nucleic acids. Seminal reviews by McCloskey and collaborators have established the fundamental importance of MS techniques and applications to nucleic acid research and other related biotechnological fields.[7–9]

The MS analysis of oligonucleotides is of primary importance since establishing the size, purity, and sequence of nucleic acids is a prerequisite to their use as molecular probes in biomedical science. Significant progress in the area of accurate mass determination, sequencing, and study of noncovalent interactions has been made possible by the use of MS ionization techniques such as matrix-assisted laser desorption/ionization coupled to time-of-flight (MALDI-TOF) and electrospray ionization (ESI) coupled to conventional analyzers and to tandem mass spectrometric (MS/MS) instruments.

Nucleic acids and their constituents, due to their polarity and thermal lability, have presented a considerable challenge for analysis by MS. Mixtures of a variety of nucleobases and nucleosides can be analyzed by combined gas chromatography-mass spectrometry (GC-MS), although it is usually

necessary to convert them to more volatile derivatives. The GC-MS analysis of more complex nucleosides, with polar modifications in either the glycone (glycosyl portion) or the nucleobase (aglycone), had not been very successful.[10] However, with the development of newer ionization methods, it has become feasible to perform, without any derivatization, MS analysis on nucleosides, nucleotides, and other large related oligonucleotides (DNA and RNA).

Since the 1960s, it has been known that the mutagenic activity of many substances is linked to the degree with which they covalently bind to DNA. The analysis and measurement of these carcinogen–DNA adducts have been proposed as a means of evaluating the extent of human exposure to specific carcinogens and of monitoring the presence and effects of these specific carcinogens in the environment.[6,11a–e,12] Highly sensitive and structurally informative analytical techniques are thus needed for the general characterization of DNA modifications. MS has the potential to provide structural information and, as such, has played an important role in the structural elucidation of covalently modified nucleic acids.

Many ionization modes, such as electron ionization (EI) and chemical ionization (CI), have been reviewed and employed for the structural elucidation of nucleic acids, their constituents, and their covalently modified nucleic acids.[7–10,12] All of these techniques provide, when coupled with MS/MS, structural information for either single nucleobase-adducts or derivatized single nucleobase adducts. In general, these methods suffer from either a lack of sensitivity necessary for the analysis of trace levels of components derived from small amounts of DNA, or the need to use chemical derivatization techniques to increase volatility and decrease thermal lability. The new technologies developed in the last century, such as fast atom bombardment-mass spectrometry (FAB-MS), MALDI-TOF-MS, ESI-TQ-MS (triple quadrupole, TQ), quadrupole-ion-trap (QIT), and Fourier transform ion cyclotron resonance (FT-ICR), hybrid instruments such as the quadrupole orthogonal time-of-flight (Q-TOF), quadrupole-ion-trap-time-of-flight (QIT-TOF), and quadrupole-ion-trap-Fourier transform-ion cyclotron resonance-tandem mass spectrometry (QIT-FT-ICR-MS/MS) instruments have allowed the analysis of nonderivatized adducts of nucleosides and nucleotides. In addition, capillary electrophoresis (CE) coupled to MS/MS has been used for adduct detection.[6,11a–e,12]

Thus, the intense interest in the MS analysis of nucleic acids and their components has produced a wealth of published work. The major objectives of this chapter are to put into perspective the progress in current biotechnological research in this area, highlight the most popular ionization methods, and illustrate the diversity of strategies employed in the characterization and sequencing of DNA and RNA oligomers, nucleosides, nucleotides, and adducts of biomedical importance.

1.2 GAS-PHASE TECHNIQUES

The analysis of nucleic acids and derived products isolated from complex biological matrixes is synonymous with analysis at very low concentrations. Thus, tremendous efforts have been made to develop and implement MS methodologies for the study of the monomeric constituents of nucleic acids.[4,13–15a–c] Biemann and McColskey elucidated detailed fragmentation pathways for the major ions present in both electron ionization (EI) and chemical ionization (CI) mass spectra of simple purine and pyrimidine nucleosides.[16–18] Derivatization, which is of value in many cases, requires an additional sample manipulation step, and may produce unwanted and unexpected side products.[19–22]

1.2.1 Electron Impact Ionization-MS for the Detection of DNA Adducts

So far the biological significance of cytosine methylation has not been fully understood, but there is growing evidence suggesting a link between the development of human cancer and the perturbation of the methylation patterns due to the lack of DNA cytosine methyltransferase (Mtase) substrate specificity. A novel *in vitro* assay called mass tagging was developed, which permitted quantification of the DNA substrate preferences of cytosine Mtases.[23] This approach involved the labeling of target cytosine residues in the synthetic DNA duplexes with stable isotopes such as ^{15}N. Methylation

was then measured by GC-EI-MS to determine the formation of 5-methylcytosine (5 mC) and DNA. Substrate selectivity was then determined through MS by the absence or presence of the label in 5 mC. This mass-tagging approach has proven to be a powerful tool for examining the substrate selectivity of cytosine DNA Mtases and could easily be adapted to the study of other kinds of selectivity.

1.2.2 Electron-Capture Ionization-MS for the Detection of DNA Adducts

Electron capture (EC) is a sensitive ionization technique for MS analysis, providing selectivity toward electrophoric compounds. Advances in instrumentation have led to a more widespread application of this method in biomedical and pharmaceutical analysis.

Fedtke and Swenberg contributed a short review of MS techniques for DNA adducts, with emphasis on electron-capture negative ion chemical ionization-mass spectrometry (EC-MS) detection, especially with regard to their method for N^2,3-ethenoguanine.[24] They included in their work, a summary of earlier and more general reviews on the detection of DNA adducts as well as the normal constituents of DNA. Similarly, Giese et al. emphasized their own work in a short review on the same general subject, but with more attention to method development and instrumental aspects of EC-MS.[25] Detection of DNA adducts by means of different types of ionization techniques has also been reviewed.[26–29]

Electrophores are compounds that can be readily detected by the two closely related gas-phase detection techniques of EC-MS and electron-capture detection (ECD). These two analyses require that the electrophore be both thermally stable and volatile.[30,31] Usually, the compound is delivered to one of these detectors as a peak eluting from a GC column, although a direct insertion probe and a liquid chromatography (LC) belt interface can also be linked to the EC-MS instrument.[30,31]

In ECD, a gas-phase current of electrons is monitored, and the electrophore reduces the current by capturing one of the electrons. Similar EC takes place in the EC-MS, but under partial vacuum conditions, and the ion(s) formed from the analyte are measured instead. Once an electrophore captures an electron, it becomes an anion radical and it can either exist long enough to reach the detector or immediately fragment.

For standards, the two detectors are similarly sensitive, but the greater specificity of the EC-MS makes it considerably more powerful for biological samples. Aside from special analytes that are unusually volatile or nonvolatile, it is difficult to reach the low picogram level by GC-ECD when real samples requiring electrophore derivatization are tested.[32,33] Largely, this is because electrophoric contaminants are present or formed in the reagents and solvents during sample preparation.[34]

DNA adducts, due to their lack of volatility and thermal stability, are not good electrophores for detection by GC-EC-MS. Therefore, as part of sample preparation, one or more chemical reactions are necessary to make the DNA adducts available for detection by GC-EC-MS.

It must be kept in mind that the overall electrophore response in the GC-EC-MS is a combination of the recovery of the compound from the GC part into the EC ion source; the inherent EC properties of the compound; other events (e.g., wall effects) in the ion source; and the transmission efficiency (including the lifetime) of the electrophore anion to the detector.[35] A strong electrophore can be detected as a pure standard at the low attomole level on a modern instrument.[36]

Typically, the DNA adduct is converted into an electrophore, which is detected as either a parent anion radical (no dissociative EC) or as an exclusive anion fragment (dissociative EC), which is formed along with the corresponding neutral. Either way, the electrophore that is typically selected forms a single ion (aside from the associated isotope peak) maximizing the sensitivity. Ideally, the electrophore-labeled DNA adduct would give two or three anion fragments in similar amounts, and their ratio would be checked for each sample as an extra guard against interferences. This would be worth the slightly higher detection limit.

Although some standards of derivatized DNA adducts can be detected by GC-ECD (Table 1.1), DNA adducts that are derived from biological samples are usually only detectable by GC-EC-MS.

TABLE 1.1
DNA Adducts and Monomers Detected by Electrophore-Labeling GC-MS or GC-ECD (Detection is by GC-EC-MS Unless Indicated Otherwise)

		Sample Preparation		Reported Detection Limit		
Type of Adduct	**Target on DNA**	**Procedure**	**Yield (%)**	**Standard**	**DNA**	**Reference**
4-Aminobiphenyl	*C*8-G (mostly)	DNA/NaOH/hexane/PFP	—	—	0.32 in 10^8 (100 g of DNA)	38,39
2-Aminofluorene	*C*8-G	NH_2NH_2/PFBzal	—	25 amol (S/N = 10)	—	40,41
Benzo[*a*]pyrene	*N*2-G (mostly)	DNA/HCl/heat/KO_2/PFBz/Si-SPE	22	14 amol (S/N = 20)	5 in 10^7 (100 g of DNA)	36,42
	—	DNA/HCl/heat/CH_3I/C_6H_6	—	1 fmol	—	43
Cytosine	—	PFB/Me/GC-ECD	35	7 fg	—	44–46
Etheno	N^2,3-G	DNA/HCl/IEC/C18-Si-SPE/PFBz/Si-SPE	—	190 amol (S/N = 10)	60 fmol/mol of G for 1 mg of DNA	24,47
Ethyl	O^2-T	PFBz/TLC/Si-FC	56	25 pg (S/N > 100)	—	48
Ethyl	O^4-T	PFBz/TLC	57	—	—	48,49
Ethyl	O^4-Tdn	PFBz/TLC/GC-ECD	55	0.44 fmol	—	30,50
Hydroxy	*C*8-dG	DNA/nucleases/TFA-acid/hydrazine/C18-SPE/ acetyl/hydrolysis/ PFBz/Si-SPE	—	1.8 fmol	4 in 10^6 for 30 g of DNA	51,52
Hydroxy	*C*8-dG	Urine/C18(OH)-Si-SPE/acetyl/PFBz/HPLC/ EtOAc	—	1.8 fmol	1.8 pmol in 50 L of urine	51,52
Hydroxy	CH_3 on T	PFBz/TLC/Si-FC/PFBz/TLC	—	200 zmol (S/N = 3)		53
4′-Hydroxybutyl	O^6-G	DNA/HCl/Ab/PFBz/TMS	—	—	0.12–3 mol/mol of G	54
2′-Hydroxyethyl	*N*7-G	DNA/heat/HCl/HNO_2/PFBz/Si-SPE	9.7	1.3 amol (S/N = 10)	100 pg in 100 g of DNA	36,55,56

Source: Adapted from Kellersberger, K.A. et al., *J. Am. Soc. Mass Spectrom.* 16, 199, 2005. With permission.