DERIVING STRUCTURAL INFORMATION FROM NON-CODING RIBONUCLEIC ACIDS BY MASS SPECTROMETRY

Jari Aho
Master's thesis
29.4.2008
University of Helsinki
Laboratory of Analytical Chemistry

HELSINGIN YLIOPISTO) HELSINGFORS UNIVERSITET) UNIVERSITY OF HELSINKI

Tiedekunta/Osasto) Fakultet/Sektion) Faculty	Laitos) Institution) Department
Faculty Of Science	Department of Chemistry

Tekijä) Författare) Author
Jari Aho

Työn nimi) Arbetets titel) Title
Deriving structural information from non-coding ribonucleic acids by mass spectrometry

Oppiaine) Läroämne) Subject
Analytical chemistry

Työn laji) Arbetets art) Level	Aika) Datum) Month and year	Sivumäärä) Sidoantal) Number of pages
Master´s thesis	29.4.2008	98

Tiivistelmä) Referat) Abstract

The purpose of this study is to describe the development of application of mass spectrometry for the structural analyses of non-coding ribonucleic acids during past decade. Mass spectrometric methods are compared of traditional gel electrophoretic methods, the characteristics of performance of mass spectrometric, analyses are studied and the future trends of mass spectrometry of ribonucleic acids are discussed.

Non-coding ribonucleic acids are short polymeric biomolecules which are not translated to proteins, but which may affect the gene expression in all organisms. Regulatory ribonucleic acids act through transient interactions with key molecules in signal transduction pathways. Interactions are mediated through specific secondary and tertiary structures. Posttranscriptional modifications in the structures of molecules may introduce new properties to the organism, such as adaptation to environmental changes or development of resistance to antibiotics.

In the scope of this study, the structural studies include i) determination of the sequence of nucleobases in the polymer chain, ii) characterisation and localisation of posttranscriptional modifications in nucleobases and in the backbone structure, iii) identification of ribonucleic acid-binding molecules and iv) probing of higher order structures in the ribonucleic acid molecule. Bacteria, archaea, viruses and HeLa cancer cells have been used as target organisms. Synthesised ribonucleic acids consisting of structural regions of interest have been frequently used.

Electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI) have been used for ionisation of ribonucleic analytes. Ammonium acetate and 2-propanol are common solvents for ESI. Trihydroxyacetophenone is the optimal MALDI matrix for ionisation of ribonucleic acids and peptides. Ammonium salts are used in ESI buffers and MALDI matrices as additives to remove cation adducts. Reverse phase high performance liquid chromatography has been used for desalting and fractionation of analytes either off-line of on-line, coupled with ESI source. Triethylamine and triethylammonium bicarbonate are used as ion pair reagents almost exclusively.

Fourier transform ion cyclotron resonance analyser using ESI coupled with liquid chromatography is the platform of choice for all forms of structural analyses. Time-of-flight (TOF) analyser using MALDI may offer sensitive, easy-to-use and economical solution for simple sequencing of longer oligonucleotides and analyses of analyte mixtures without prior fractionation. Special analysis software is used for computer-aided interpretation of mass spectra.

With mass spectrometry, sequences of 20-30 nucleotides of length may be determined unambiguously. Sequencing may be applied to quality control of short synthetic oligomers for analytical purposes. Sequencing in conjunction with other structural studies enables accurate localisation and characterisation of posttranscriptional modifications and identification of nucleobases and amino acids at the sites of interaction. High throughput screening methods for RNA-binding ligands have been developed. Probing of the higher order structures has provided supportive data for computer-generated three dimensional models of viral pseudoknots.

In conclusion. mass spectrometric methods are well suited for structural analyses of small species of ribonucleic acids, such as short non-coding ribonucleic acids in the molecular size region of 20-30 nucleotides. Structural information not attainable with other methods of analyses, such as nuclear magnetic resonance and X-ray crystallography, may be obtained with the use of mass spectrometry. Sequencing may be applied to quality control of short synthetic oligomers for analytical purposes. Ligand screening may be used in the search of possible new therapeutic agents.

Demanding assay design and challenging interpretation of data requires multidisclipinary knowledge. The implement of mass spectrometry to structural studies of ribonucleic acids is probably most efficiently conducted in specialist groups consisting of researchers from various fields of science.

Avainsanat) Nyckelord) Keywords
Non-coding ribonucleic acids, mass spectrometry, sequencing, posttranscriptional modifications, higher order structures, non-covalent interactions

Säilytyspaikka) Förvaringsställe) Where deposited
University of Helsinki, Laboratory of Analytical Chemistry

Muita tietoja) Övriga uppgifter) Additional information
Rakenteellisen tiedon johtaminen koodaamattomista ribonukleiinihapoista massaspektrometrialla

TABLE OF CONTENTS

ABBREVIATIONS 3
SYMBOLS 6
1 INTRODUCTION 7
2 RIBONUCLEIC ACIDS 9
2.1 Structure and chemical properties of ribonucleic acids 10
2.2 Non-coding ribonucleic acids 17
2.3 Ribonucleic acids from biological sources 18
2.4 Synthetic ribonucleic acids 18
3 STRUCTURAL ANALYSES OF RIBONUCLEIC ACIDS 20
3.1 Fields of application 21
3.2 General strategic approaches 22
4 TRADITIONAL METHODS OF ANALYSIS 24
4.1 Gel electrophoresis 24
4.2 Sequencing 26
4.2.1 Sanger dideoxy method 26
4.2.2 Enzymatic sequencing 29
4.2.3 Alkaline hydrolysis 31
4.3 Interactions 32
4.4 Higher order structures 34
5 MASS SPECTROMETRIC METHODS OF ANALYSIS 39
5.1 Principle of mass spectrometry 40
5.2 Mass spectrometry of ribonucleic acids 41
5.2.1 Sample preparation 41
5.2.2 Sample pretreatment 42
5.2.3 Sample introduction and ionization 42
5.2.4 Mass analysis 48
5.2.5 Detection 50
5.2.6 Tandem mass spectrometry 50
5.3 Application of mass spectrometry to structural studies of ribonucleic acids 52
5.3.1 Sequencing 52
5.3.2 Posttranscriptional modifications 63
5.3.3 Higher order structures 69
5.3.4 Noncovalent interactions 76
5.4 Overview of the instrumentation 85
5.5 Analytical performance 88

6 COMPARISON OF TRADITIONAL AND MASS SPECTROMETRIC METHODS 89
7 OUTLOOK 91
8 SUMMARY AND CONCLUSIONS 92
9 REFERENCES 93

ABBREVIATIONS

ACN	acetonitrile
ACTH 11-24	adrenocorticotropin hormone analog
AGE	agarose gel electrophoresis
BDMAB	butyldimethylammonium bicarbonate
BE	arrangement of sector mass spectrometer, where magnetic analyzer (B) is in the front of the electrostatic analyzer (E)
Bek	bekanomycin
BKT	ethylene glycol bis[3-(2-ketobutyraldehyde) ether] (bikethoxal)
BSP	bovine spleen phosphodiesterase
cDNA	complementary DNA
CDTA	1,2-cyclohexanediaminetetraacetic acid
CHB	4-[bis(2-chlorethyl)amino]benzenebutanoic acid (chlorambucil)
CHCA	α-cyano-4-hydroxycinnamic acid
CID	collisionally-induced dissociation
CMCT	1-cyclohexyl-3--(2-morpholinoethyl)-carbodiimide metho-p-toluene sulfonate
CPT	*cis*-diamminedichloroplatinum (II) (cisplatin)
CSP	calf spleen phosphodiesterase
Da	Dalton, a non-SI mass unit equal to unified mass unit (u)
DEPC	diethylpyrocarbonate
DHB	dihydroxybenzoic acid
DMS	methylsulfinylmethane (dimethyl sulfoxide)
DNA	deoxyribonucleic acid
E. coli	bacterium *Escherichia coli*
EB	arrangement of sector mass spectrometer, where electrostatic analyzer (E) is in the front of the magnetic analyzer
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
ESI-FT-ICRMS	electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry
FIV-PK	feline immunodefiency virus pseudoknot
fmol	femtomole, 10^{-15} mol
FT-ICRMS	Fourier-transform ion cyclotron resonance mass spectrometry
HCV	hepatitis C virus
HeLa	immortal human breast cancer cell line
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
HIV-1	human immunodefiency virus 1
HPA	hydroxypicolinic acid
HPLC	high performance liquid chromatography
IMAC	immobilized-metal ion affinity chromatography
IP-RPLC	ion-pair reverse phase liquid chromatography
IR	infrared
IRMPD	infrared multiphoton dissociation

ITMS	ion trap mass spectrometry
IUPAC	International Union of Pure and Applied Chemistry
K	kilo-, 10^3
KT	β-ethoxy-α-ketobutyraldehyde.
kV	kilovolt
LC	liquid chromatography
LC/MS	liquid chromatography/ mass spectrometry
LV	lividomycin
m/z	mass-to-charge ratio
MALDI	matrix-assisted laser desorption ionization
MH^+	protonated molecular ion
miRNA	microribonucleic acid
MRM	multiple reaction monitoring
mRNA	messenger ribonucleic acid
MS/MS	tandem mass spectrometry
MS^n	multiple stage mass spectrometry
MW	molecular weight
$NaBH_4$	sodium tetrahydroborate
NC	nucleocapsid protein
Nd:YAG	neodymium-doped yttrium aluminium garnet
NH_4OAc	ammonium acetate
nm	nanometer, 10-9 m
NM	2-chloro-N-(2-chloroethyl)-N-methyl-ethanamine (nitrogen mustard)
NS	nozzle-skimmer
NS-CID	nozzle-skimmer collisionally induced dissociation
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDG	1, 4-diacetalbenzene (phenyl diglyoxal)
pH	measure of the acidity or alkalinity of a solution: negative logarithm of hydrogen ion concentration
PM	paromomycin
pmol	picomole, 10^{-12} mol
ppm	parts per million
PSD	post-source decay
Q	quadrupole mass spectrometer
QIT	quadrupole ion trap
QITMS	quadrupole ion trap mass spectrometryr
Q-LIT	quadrupole linear ion trap
QqQ	triple quadrupole mass spectrometer
RNA	ribonucleic acid
RNAse	ribonuclease
RPLC	reverse-phase liquid chromatography
rRNA	ribosomal ribonucleic acid

SDS	sodium dodecyl sulphate
SIM	selected ion monitoring
siRNA	small interfering ribonucleic acid
Sis	sisomycin
SL1A	human immunodefiency virus stemloop 1A
snRNA	small nucleolar ribonucleic acid
snRNP	small nucleolar riboprotein
SORI	sustained off-resonance irradiation
SORI-CID	sustained off-resonance irradiation collisionally induced dissociation
SRM	selected ion monitoring
sTT	2,4,6-trichloro-1,3,5-triazine (cyanuric acid)
SVP	snake venom phosphodiesterase
TEAA	tetraethyl ammonium acetate
TEAB	tetraethyl ammonium bicarbonate
TFAA	trifluoroacetic acid
THAP	trihydroxyacetophenone
Tob	tobramycin
TOF	time-of-flight
TOFMS	time-of-flight mass spectrometry
tRNA	transfer ribonucleic acid
UV	ultraviolet
W_n/w_n	*w*-series ion at position n counted from the 3´ terminus
VPK	viral pseudoknot
Y_n/y_n	*y*-series ion at position n counted from the 3´ terminus
Z_n/z_n	*z*-series ion at position n counted from the 3´ terminus

SYMBOLS

A	adenine nucleobase
A>p	adenosine 3´ monophosphate (cyclic phosphate)
A_3-B_3	oligonucleotide ion belonging to ion series *a* which has lost a nucleobase (B). Subscripted 3 indicates the position of the nucleotide
a_n-B	oligonucleotide ion belonging to ion series *a* with a lost nucleobase (B). Subscripted n indicates the position of the nucleotide
Ap	adenosine monophosphate
AU	dinucleotide consisting of adenosine and guanosine nucleotide
C	cytosine nucleobase
C_3-B_3	oligonucleotide ion belonging to ion series *c* which has lost a nucleobase (B). Subscripted 3 indicates the position of the nucleotide
Cm	2'-O-methylcytidine
C_{Me}p	methylated cytidine with phosphorylated 2´ carbon
Cp	cytosine monophosphate
D	dihydrouridine
dA	deoxyadenine
dG	deoxyguanine
G	guanine nucleobase
G>p	guanine monophosphate (cyclic phosphate)
Gp	guanosine monophosphate
$m^2{}_2G$	N^2,N^2-dimethylguanosine
m^3U	3-methyluridine
m^4Cm	4,2'-*O*-dimethylcytidine
m^5C	5-methylcytosine
$m^6{}_2A$	N^6,N^6-dimethyladenosine
m^7G	7-methylguanosine
N	atomic nitrogen
p7	nucleocapsid protein
R	oligonucleotide of any length and sequence
S	Svedberg, a non-SI unit of coefficient of sedimentation
T	tesla, a derived SI unit of the density of magnetic field , Vs/m^2
U	uracil nucleobase
Up	uridine monophosphate
Å	ångström, a special unit of length, 10^{-10} m
Ψ	pseudouridine
ψ-RNA	human immunodefiency virus packaging signal ribonucleic acid
µg	microgramme, 10^{-6} g
µL	microlitre, 10^{-6} L
16S	part of the small 30 S subunit of the bacterial ribosomal ribonucleic acid
23S	part of the large 70S subunit of the bacterial ribosomal ribonucleic
5S	part of the large 70S subunit of the bacterial ribosomal ribonucleic acid

1 INTRODUCTION

Ribonucleic acids (RNA:s) are polymeric compounds, which form a heterogeneous group of macromolecules abundantly present in the cells of all living organisms. Together with structurally closely resembling deoxyribonucleic acids (DNA:s), they form a larger group of biomolecules generally referred as nucleic acids, possessing important functions in the mediation of flow of genetic information in the process of protein synthesis.

The discoveries of catalytic RNA[1,2] and small nucleolar RNA:s (snRNA)[3] have demonstrated the presence of a new class of RNA molecules, which are not translated to protein, but do participate on the regulation of gene expression at many stages of cell signalling through transient interactions with key molecules. RNA transcripts associated to this class are generally referred as "non-coding RNA:s".

RNA molecules are processed further after the transcription phase inside the cell. Originally linear molecules may fold into various secondary and tertiary structures and undergo different chemical derivatisations. Regions of different higher order structures and posttranscriptional modifications have been found to play important roles in the functionality of RNA molecules and interactions with other molecules. For example, conformational changes and modification of bases have been found to assist organisms in adaptation to extreme environmental conditions and development of resistance towards antibiotics.

Transfer-RNA:s (tRNA) and ribosomal RNA:s (rRNA) are generally grouped with the class of non-coding RNA:s. By studying the form and function of these molecules, the structure-dependent functionality and different forms of RNA-ligand interactions are demonstrated. Both tRNA and rRNA have also been recognised as targets of therapeutic agents in the treatment of bacterial and viral diseases.

In the study of non-coding RNA, the knowledge of primary and higher order structures of polymer chain, localisation and identification of posttranscriptional modifications and characterization of interactions with other molecules is needed. Established traditional methods for deriving structural data employ various strategies of chemical modification and controlled fragmentation of target RNA, separation of produced fragments in gel electrophoresis and image analysis of separated products visualised in the gel.

Mass spectrometric methods are emerging into the field of structural analyses of ribonucleic acids offering advantages over traditional gel-based methods in the terms of

speed of analysis, automation, sample consumption, mass accuracy and high throughput of detection. Mass spectrometric techniques are based on vaporization of analytes into charged gas-phase ions, which may be accurately separated and weighed. Utilising traditional strategies of analyte pretreatments in combination with tandem mass spectrometric instrumentation and multiple-stage analyses, detailed structural information may be gathered. Furthermore, mass spectrometric methods provide only plausible way to study posttranscriptional modifications.

In this study, the application of mass spectrometry to structural analyses of non-coding RNA is reviewed. In the following chapters the properties of ribonucleic acids are studied more in detail, traditional and mass spectrometric methods of analysis are compared and the characteristics of performance of mass spectrometric analyses are studied. Finally, the future trends of mass spectrometry of RNA are discussed.

2 RIBONUCLEIC ACIDS

Ribonucleic acids can be found in large amounts inside the cells of all living organisms, including viruses. Three major classes of RNA can be distinguished: messenger-RNA (mRNA), transfer-RNA and ribosomal RNA, of which ribosomal RNA is the most abundant. Ribosomal RNA consists of subunits, which are classified by their sedimentation coefficients. The nomenclature follows the classification: for example, the three ribosomal subunits of bacterium *Escherichia coli* are called 5S, 16S and 23S, where the numeral refers to the sedimentation coefficient and S to the Svedberg unit. The relative distribution of major species of RNA is presented in Table 1 using *E. coli* as a model organism. There is no data available for the relative abundances of the heterogeneous group of small non-coding RNA:s.

Table 1. RNA molecules in a bacterial cell.[4]

Type	Relative abundance (%)	The coefficient of sedimentation (S)	Molar mass (kDa)	Number of nucleotides in a molecule
rRNA (23S)	80	23	12000	3700
rRNA (16S)		16	550	1700
rRNA (5S)		5	360	120
tRNA	15	4	250	75
mRNA	5		heterogeneous	

Molecules of non-coding RNA are usually short, single- or doublestranded stretches of RNA originating from the transcription of regions of chromosomes, which have been formerly considered transscriptionally silent.[5] Species of this class of RNA have been found to be involved in gene silencing,[6, 7] co-activation of steroid receptors,[8] X-chromosome inactivation[9] and repressing the activity of a transcription factor important for T-cell-receptor -mediated immune response system,[10] to mention a few.

Target RNA used in structural studies can be obtained by extracting native RNA from actual cells of living organisms, synthesizing it in the laboratory or purchasing synthesised oligonucleotides from commercial suppliers. Using automated solid-phase synthesis

methods,[11] the base sequence of a synthesised RNA molecule can be engineered to reconstitute naturally occurring molecules either as whole, native molecules or as segments of regions of interest. In addition, the sequence can be designed completely artificial for different analytical or functional purposes.

In this chapter, the characteristics of ribonucleic acids as analytes are described and features of non-coding RNA are more closely observed.

2.1 Structure and chemical properties of ribonucleic acids

A ribonucleic acid molecule is a linear polymer molecule where the monomer unit consists of an acidic phosphate group, a basic purine or pyrimidine compound and a five-carbon sugar compound ribose.[4] This monomer unit is also called a nucleotide and the basic compounds are referred as nucleobases; a nucleotide lacking the phosphate group is nucleoside. Nucleobases are given alphabetical symbols to facilitate presentations of molecular structure in written form. The nomenclature, applied symbols and other data concerning the components of ribonucleic acid monomer is presented in the Table 2.

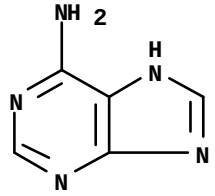
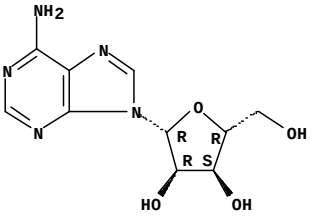
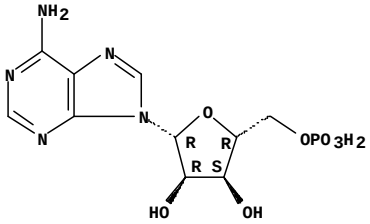
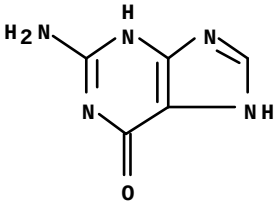
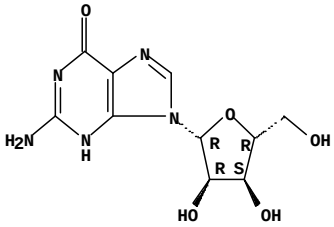
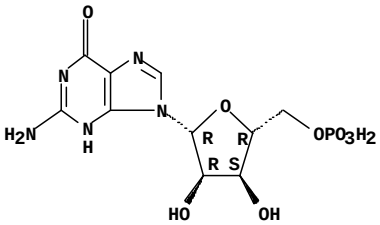
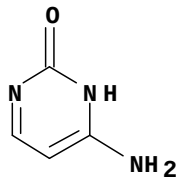
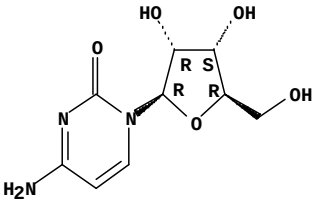
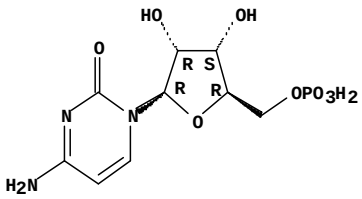
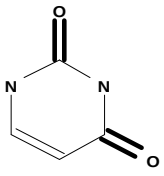
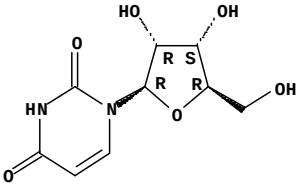
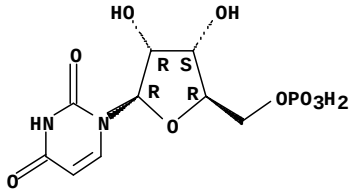
The polymer chain is formed of ribose sugar units connected by phosphodiester bonds between the 3´ and 5´ carbons of adjacent ribose rings. Ionisable phosphate groups render the RNA molecule polar, water-soluble and acidic and thus create a large negative net charge to the molecule. Any of the two purine compounds (adenine or guanine) or pyrimidine compounds (uracil or cytosine) may be connected to the 1´ carbon of the ribose. The end of the chain where a 5´ carbon of the ribose is found unconnected to any other nucleotide is called as 5´ end of the chain. Similarly, the other end is called as the 3´ end. Instead of nucleotides, the carbons of these positions may harbour hydroxyl or phosphate groups and may be used as sites of different labelling and derivatisation. The general structure of oligoribonucleic acid polymer is presented in the Figure 1.

The order of nucleotides in the chain is generally expressed through the order of the nucleobases as the "sequence of nucleobases", or "base sequence". Common practice is to read the sequence to 5´→3´ direction. In naturally occurring RNA molecules the nucleobases can appear in any order in the oligonucleotide chain, dictated by the sequence of nucleobases of the corresponding gene.

The sequence of an oligonucleotide chain may be presented in written form using alphabetic symbols. For example, a tetraribonucleotide having a base sequence of adenine,

uridine, cytosine and guanidine may be presented as 5´-AUCG-3´. Occasionally also phosphate groups and terminal hydroxyl groups are denoted where appropriate, using (HO-) for hydroxyl groups and (p) for phosphate groups. For example, a dinucleotide fragment (AU) of aforementioned tetranucleotide having hydroxyl group at the 5´ terminus and a phosphorylated 3´ terminus, may be presented as 5´-HO-AUp-3´.

Table 2. Ribonucleic acid monomers.

Symbol	Nucleobase	Nucleoside	Nucleotide
A			
	adenine $C_5 H_5 N_5$ MW: 135.13 Da	adenosine $C_{10} H_{13} N_5 O_4$ MW: 267.24 Da	adenosine 5´-monophosphate $C_{10} H_{14} N_5 O_7 P$ MW: 347.22 Da
G			
	guanine $C_5 H_5 N_5 O$ MW: 151.13 Da	guanosine $C_{10} H_{13} N_5 O_5$ MW: 283.24 Da	guanosine 5´-monophosphate $C_{10} H_{14} N_5 O_8 P$ MW: 363.22 Da
C			
	cytosine $C_4 H_5 N_3 O$ MW: 111.10 Da	cytidine $C_9 H_{13} N_3 O_5$ MW: 243.22 Da	cytosine 5´-monophosphate $C_9 H_{14} N_3 O_8 P$ MW: 323.20 Da
U			
	uracil $C_4 H_4 N_2 O_2$ MW: 112.09 Da	uridine $C_9 H_{12} N_2 O_6$ MW: 244.20 Da	uridine 5-monophosphate $C_9 H_{13} N_2 O_9 P$ MW : 324.18 Da

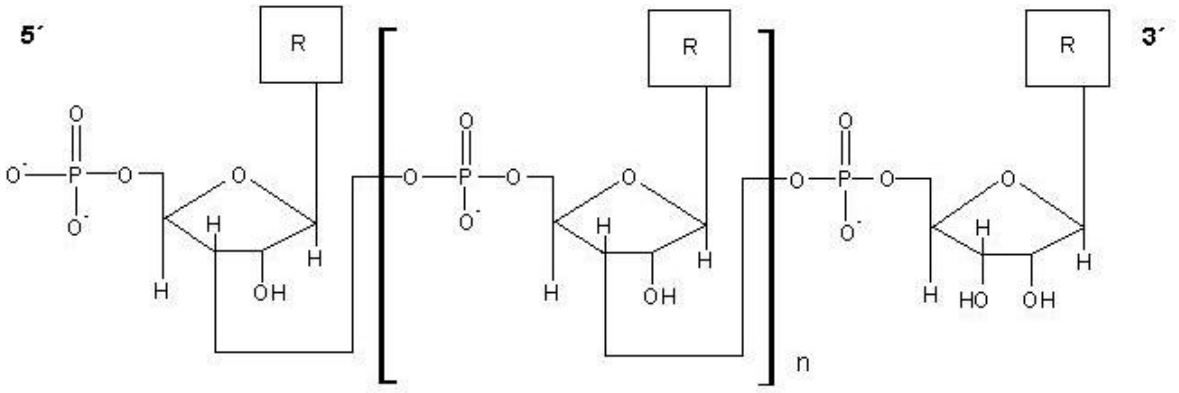


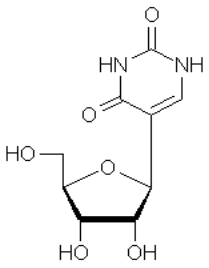
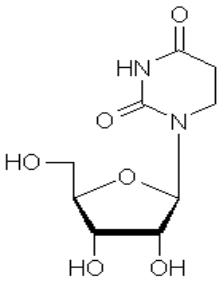
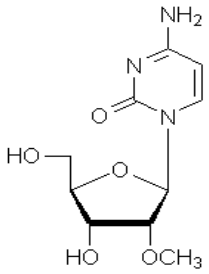
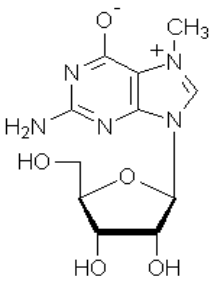
Figure 1. General form of a ribonucleic acid polymer. R denotes any of the nucleobases adenine, guanine, cytosine or uracil.

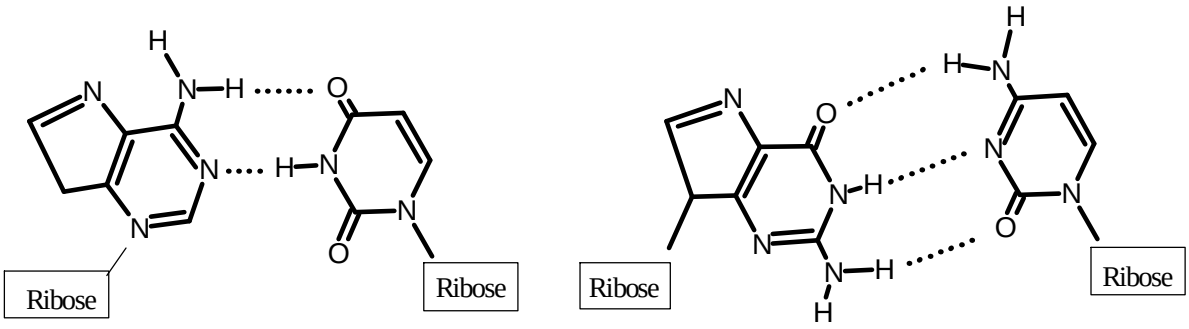
After transcription many nucleotides undergo chemical modifications which fine-tune properties and functionality of different RNA species. Most common modifications involve various alkylations and methylations of nucleobases and ribose moiety. In addition, uridine has one structural isomer, pseudouridine, which has exactly the same molar mass. Characterized modifications are given alphanumeric symbols to facilitate structure presentations. Some examples of modified nucleotides are presented in Table 3.

DNA differs from RNA by having a hydrogen atom instead of hydroxyl group attached to the 2´ carbon of ribose and a thymine pyrimidine base instead of uracil. The reactivity of 2´-hydroxyl group contributes to many differences in the chemical stability and reactivity of these different types of nucleic acids.

An adenine base can form two stable hydrogen bonds with uracil while guanine can bind more strongly to cytosine with three hydrogen bonds. This type of bonding of nucleobases is called complementary base-pairing and is demonstrated in Figure 2.

Table 3. Selected modified nucleotides of ribonucleic acids.[12]

Nucleotide	Symbol	Molecular formula	Molecular structure	Molar mass (Da)
pseudouridine	Ψ	$C_9 H_{12} N_2 O_6$		244.20
dihydrouridine	D	$C_9 H_{14} N_2 O_6$		246.22
2'-*O*-methylcytidine	Cm	$C_{10} H_{15} N_3 O_5$		257.25
7-methylguanosine	m^7G	$C_{11} H_{16} N_5 O_5$		297.27



Adenine-Uridine base-pairing Guanine-Cytosine base- pairing

Figure 2. Complementary base-pairing in RNA. Hydrogen bonds are denoted with dotted lines.

Due to complementary base pairing, RNA molecules are capable to bind to itself forming complementary double-stranded regions, hairpin loops and intermolecular bridges between strands in close proximity and hairpin-strand bridges called pseudoknots. These secondary structures have been found to provide sites of interaction with other molecules[13] or involve in different catalytic activities creating different classes of RNA with diverse functionality. RNA may form double-stranded hybrid structures also with DNA due to structural similarity, common nucleobases and the complementary binding capability of uracil of RNA to adenine in the DNA. This sequence-specific base pairing can be exploited in many ways in analytical applications. Schematic figures illustrating different secondary structures present in RNA are shown in Figures 3 and 4.

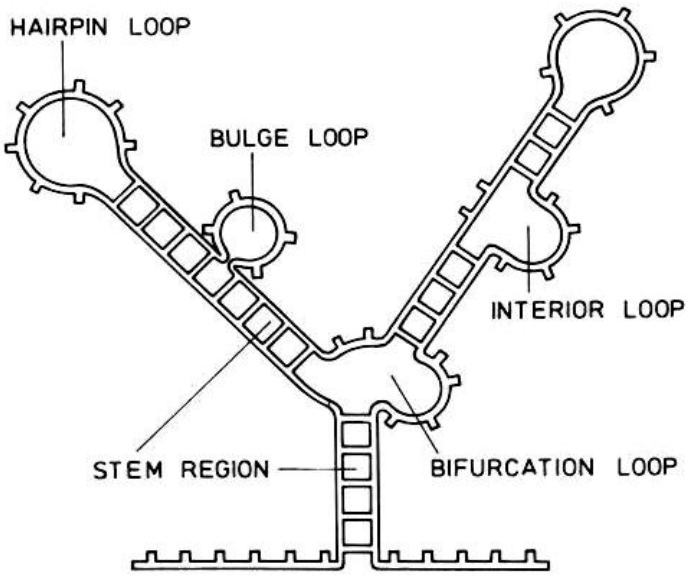


Figure 3. Different secondary structures of RNA.[14]

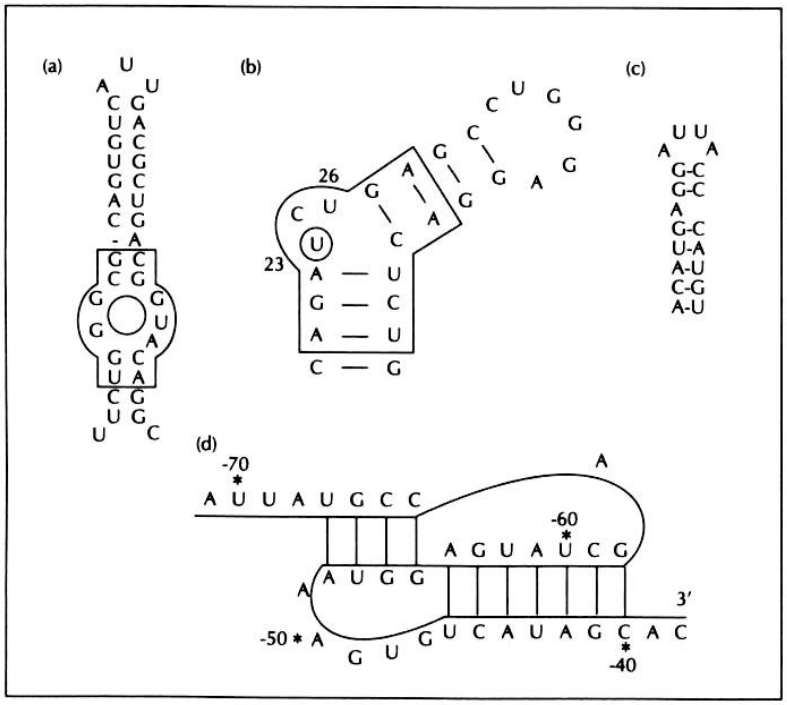


Figure 4. More detailed examples of secondary structures**.** (a) RNA bubble; (b) RNA bulge; (c) RNA hairpin and (d) RNA pseudoknot.[13]

Some base modifications have an impact on binding of different ligands to RNA molecules, such as preventing binding of aminoglycoside drug molecules to viral or bacterial tRNA molecules resulting in resistance of the particular drugs in question.[15, 16] In archaeal organisms, 2´-*O* -methylations of ribose sugar moieties are shown to assist adaptation to live at extremely high temperatures by stabilizing the secondary structures of the oligonucleotide strands.[17] A large proportion of dihydrouridine bases in tRNA of some psychrotropic bacteria increases the flexibility of tRNA at the low temperatures.[18]

With the exception of dihydrouridine, the double bonds in the ring structures of nucleobases provide chromophores for ultraviolet (UV) spectroscopy by having an absorption maximum at 260 nm.

2.2 Non-coding ribonucleic acids

Systematic analysis of human chromosomes 21 and 22 discovered ten times more transcribed RNA products of RNA than expected by the predicted protein-coding genes.[5] Further analyses of this activity showed that 36 % of the binding sites of the three transcription factors studied were located within or outside of well-known genes and are correlated significantly with non-coding RNA:s implicating, that these regions are regulated.[19] Analyses of the pool of transcribed RNA (transcriptome) products in mouse[20] predict thousands of non-coding RNA:s of yet unknown function. In addition, results of a comparative computational study of genomes of human, mouse, rat and dog suggest that at least 20% of human genes are regulated by non-coding RNA:s called microRNA:s (miRNA:s). Some known species of non-coding RNA species are presented in the Table 4.

Table 4. Selected species of non-coding RNA.

Name	Type	Length (nucleotides)	Known function	Reference
siRNA (small interfering RNA)	double-stranded	20 - 22	gene silencing	4, 6
miRNA (microRNA)	single-stranded	21 - 25	gene silencing	4
snRNA (small nuclear RNA)	structured	~100-300	component in spliceosomes, which perform posttranscriptional splicing of mRNA in eukaryotic cells	3, 4

On the basis of research on the functionality of non-coding RNA:s, synthetic short oligoribonucleotides are manufactured for diagnostic and therapeutic agents for artificial regulation of gene expression.[6]

2.3 Ribonucleic acids from biological sources

Ribonucleic acids in living cells of target organisms have to be released from the cells, concentrated and purified from other cell matrix material. Traditional methods base on cell wall lysis and purification of ribonucleic acids using organic extraction and alcohol precipitation.[21] Commercial methods often employ solid phase extraction, where the use of toxic organic solvents is avoided. In these methods, the solid phase extraction tubes are fitted into centrifuge tubes facilitating rapid binding, washing and eluting procedures.

The quality and integrity of RNA molecules are compromised by the presence of cytosolic and environmental oligoribonucleotide-degrading enzymes (*ribonucleases*) at all stages of purification.[22] Ribonucleases cleave oligoribonucleotides effectively even in small amounts reducing the amount of intact target RNA molecules available for analyses. Ribonucleases can be inactivated only reversibly using chaotropic guanidium salts or alkylating agents such as diethylpyrocarbonate (DEPC). Ribonucleases can be introduced to samples from environmental sources such as dust particles from air, surfaces of equipment or working facilities, reagent solutions and most importantly, from the fingers, hair or skin of a person working with the samples. Safe working with RNA often calls for facilities, instruments and other material specifically devoted to working with RNA. Commercial, ribonuclease-free reagents and laboratory consumables are recommended, although there are laborious methods for cleaning of equipment and preparing reagent solutions for RNA work.

2.4 Synthetic ribonucleic acids

Many problems involved in successful extraction of native, intact RNA from actual biological sources can be circumvented by using synthetic RNA, which has been constructed to replicate the base sequence and native structure of its biological counterpart. Oligoribonucleotides of any sequence may be constructed from activated monomers using automated solid-phase technologies.[23] The correct nucleobase sequence is built according to the sequence in the corresponding gene. Copies of large, native ribonucleic molecules may also be produced *in vitro* using standard vector-mediated cloning procedures.[24]

Radioactive labels, photoreactive base analogs and other modifications can be introduced to RNA during synthesis to facilitate various analytical applications.[25] Designed oligonucleotides are often used in the initial stages of method development and actual RNA from biological sources is then used for method validation.

Number of RNA analyses call for the use of DNA which is synthetized to replicate the sequence of RNA using reverse transcriptase polymerase and deoxynucleotide monomers. The reverse transcriptase is an enzyme capable of building a complementary strand to ribonucleic acid strands by adding complementary deoxynucleotides to the template RNA in a process of polymerase chain reaction (PCR). The DNA product constructed from RNA template is called complementary DNA (cDNA).

Small amounts of unspecific sequences and by-products remain in the synthesis mixture so purification steps are needed also here. However, the purification of synthetic RNA is straightforward comparing to isolation of intact RNA from biological matrices, comprising often just liquid chromatography or polyacrylamide gel electrophoresis. Commercial synthetic products are often used without prior purification.

3 STRUCTURAL ANALYSES OF RIBONUCLEIC ACIDS

In the scope of this study, the structural analyses of non-coding ribonucleic acids consist primarily of determination of the base sequence (sequencing) of the target molecule, identifying posttranscriptional modifications in nucleobases, studying higher-order structures of native molecule and interrogating noncovalent interactions with other molecules. Bacteria, archaea, viruses and human cancer cells from the immortal HeLa cancer cell line have been used as target organisms.

All approaches of structural analyses of ribonucleic acids are, in certain extent, based on the controlled derivatisation or fragmentation of the target molecule prior the analysis. Often both treatments are used in combination, and this phase of sample manipulation is referred in the future as "sample pretreatment" in the context of this text.

With the exception of sequencing, these pretreatment of target RNA cause either increments in the analyte mass or alterations in the distribution of products of fragmentation. After the analysis of the products, the structural features of the original target molecule may be deduced. Results are interpreted by comparing analysis data of pretreated analytes to partially or completely untreated reference samples of the species of original target molecule. The nucleobase sequence of the target molecule must be known beforehand to enable comparison. Sequence may be determined using different sequencing methods, or it may be deduced from the base sequence of the gene corresponding to the ribonucleic acid analyte in question.

Application-specific strategies of analyses have been designed for each type of structural analysis. Strategies comprise of different methods for the sample pretreatment, analysis and the interpretation of the results of the analysis.

In this chapter, the fields of application and general strategic approaches of structural analyses of ribonucleic acids are described.

3.1 Fields of application

Sequencing is performed when information of the base composition and order in the oligonucleotide chain is needed. Sequencing data may be used to determine the base sequence in previously unknown species of RNA, to verify the base sequences of known target molecules or, in conjunction with other structural analyses, to locate and identify posttranscriptionally modified bases or sites of interaction with other molecules. The verification of target sequences is used especially in the quality control of synthetized oligonucleotides.[26]

The studies of posttranscriptional modifications have focused on detecting and identifying modified nucleotides in tRNA:s and rRNA:s of bacteria, archaea and viruses. Studies have confirmed the presence of suspected modifications or revealed new types of modifications. Comparative studies between different organisms have showed similarities in modification patterns in same conserved regions between different organisms, implying important biological properties. Altogether 96 base modifications have been identified in different species of RNA, tRNA harboring the largest number of modifications.[12]

The study of noncovalent interactions include 1) the study of RNA-protein interactions in spliceosomal complexes,[27] 2) binding of nucleocapsid protein and the packaging signal RNA of human immunodeficiency virus HIV-1[28] and 3) screening RNA-binding ligands for therapeutic use. The major topics of research include i) identifcation of the complex-forming ligands and ii) the sites of interaction and iii) determination of binding affinities and dissociation constants.[28]

Study of higher-order structures of RNA molecule deals with mapping of single-stranded regions and hairpin-shaped (also known as *stemloop*) structures which may act as sites for noncovalent interactions with other molecules as well as identifying intramolecular contact sites.

3.2 General strategic approaches

Traditional methods of sequencing are based on production of sets of fragments with differing lengths and arrangement of the products in order of increasing size. Fragments are produced by chemical or enzymatic degradation of the original target molecule. The sequence is then read by recognizing the identities of terminal bases in each fragment. Mass spectrometric applications may also use further fragmentation of the products in the gas phase. The most probable base sequences are then deduced with the use of computer sequencing software using the exact measured masses of fragments.

Simple RNA-ligand interactions are studied by detecting experimentally formed complexes of RNA molecule and the bound ligand. Small complexes are differentiated from the unbound species by their mass or decreased electrophoretic mobility. The complex formation between RNA and RNA-binding proteins may be strengthened by covalently binding the counterparts with the use of special crosslinking reagents or irradiation with ultraviolet light.

The regions of ligand binding in the large molecules are mapped by analysing the differences in the fragmentation patterns between complexed analytes and unbound reference samples. The determination of exact sites of interaction of RNA-protein complexes involve 1) chemical crosslinking of nucleobases and peptide side chains in the contact areas, 2) isolation of fragments consisting of crosslinked, short fragments of peptide and oligonucleotide chain and 3) identifying the crosslinked amino acid and nucleobase by sequencing techniques. The analysis of peptides and proteins is beyond the scope of this study and is not discussed further. Literature over the subject is abundant.

Higher order structures are differentiated from unstructured or unbound regions of oligonucleotide chain by probing the target molecule with or special reagent molecules. Single-stranded regions are susceptible for derivatisation and fragmentation. Intramolecular spatial distances between the oligonucleotide strands in a three-dimensional molecule may be probed with the use of bridge-forming reagent molecules. Regions derivatised with probe molecules may be mapped by comparing fragmentation patterns of probed samples to those of untreated reference samples.

Posttranscriptional modifications generally increase the mass of a modified nucleotide, so the strategies of studying modifications are based on mass spectrometric detection of increments in expected masses in molecular and fragment ions. The location of modified nucleobases can be determined comparing masses of enzymatically digested molecules of

reference and analysed RNA, identifying fragments with altered masses and sequencing of the fragment in question. The resolution of traditional gel-based method is insufficient to separate analytes with minor mass shifts. Generally, LC/MS-based methods are employed.

4 TRADITIONAL METHODS OF ANALYSIS

Established methods for structural studies of ribonucleic acids employ a variety of vertical or horizontal slab gel electrophoretic methods, where charged analytes are separated according their electrophoretic mobility in a sieving gel matrix under the influence of an electric field.[29] Separated analytes are detected on the basis of irradiation of radioactive labels or emission of fluorescent dyes attached to the analyte molecules. The equipment and materials are cheap and easily available. Analyses are, in principle, easy to perform and analytes may be recovered from the gel after the analysis. However, methods often are composed of many, laborious and time-consuming steps and require handling of toxic chemicals.

Gel electrophoretic methods are well applicable for sequencing, studies of interactions and higher order structures but fail to provide accurate information about posttranscriptional modifications. In this chapter, principle of gel electrophoresis and strategies of gel-based structural analyses are described.

4.1 Gel electrophoresis

Two main types of gel electrophoretic methods exist, classified by the most commonly used gel matrices, agarose (*agarose gel electrophoresis*, AGE) and polyacrylamide (*polyacrylamide gel electrophoresis*, PAGE). Matrices having different pore sizes for the separation of analytes with a wide range of molar masses are available. A model of a vertical electrophoresis apparatus is illustrated in Figure 5.

Denaturating agents, such as urea, formaldehyde and methylmercuric hydroxide are added to gels to disrupt possible double-stranded secondary structures in ribonucleic acids. Denaturating conditions ensure that oligonucleotides appear in linear form so that the separation may occur solely on the basis of the fragment length. Fluorescent dyes may be added to the gel matrix to be attached to the analytes during the electrophoresis or the separated analytes may be stained afterwards in the gel.

Ribonucleic acid samples are prepared by dissolving purified RNA in mildly basic aqueous electrophoresis buffer consisting of a biological buffer**,** chelating agent (ethylenediaminetetraacetic acid; EDTA) and detergent (sodium dodecyl sulphate, SDS). If autoradiography is used for detection of analytes, radioactive ^{32}P-phosphate label is incorporated to the 5´end of analyte molecules in a separate labelling reaction prior to

sample loading.

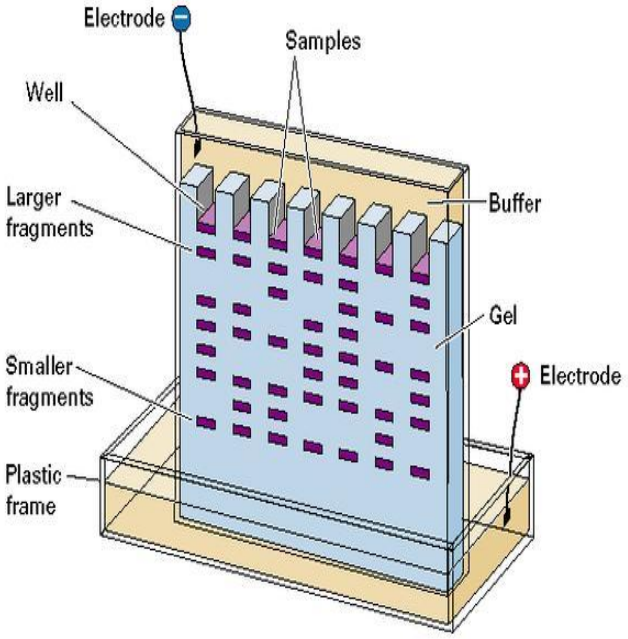


Figure 5. Vertical gel electrophoresis apparatus.[30]

Samples are introduced to gel by loading small amounts of sample solution into wells prepared into the cathode end of gel. When the electric field is applied, negatively charged RNA oligonucleotides start to migrate towards the anode through the pores in the gel matrix and along the lane defined by the boundaries of the sample well. Larger fragments advance at slower speed than smaller ones, so fragments of different sizes eventually concentrate to own distinct bands at different distances from the sample well level. The electrophoresis run is terminated after appropriate time and the concentrated bands are visualized by recording the radioactivity of the bands to X-ray film (autoradiography) or photographing the fluorescence upon UV irradiation of the gel.

Results of analyses are obtained by studying the gel image. Depending on the type of application, different information can be deduced from the migration patterns of the analyte bands. Fragments of interest can be recovered from the gel for downstream applications by cutting out the corresponding bands with razor blade and proceeding with organic extraction or elution and alcohol precipitation of the sample.

Assigning molar masses, or more commonly, lengths of oligonucleotide fragments, can be achieved by comparing the distances of migration in the gel between the analytes and

reference samples (*ladders*) consisting of sets of oligonucleotide fragments of known length and molar masses. Constructing a calibration curve, where the relation between the log value of the individual fragment mass and its distance of migration from the starting point in the gel is formed, the mass or the length of a sample fragment can be calculated by measuring the distance of migration of the band of interest.[29] Computer-aided systems for image capturing[31] and analysis[32] have been developed.

4.2 Sequencing

The traditional sequencing methods are based on controlled production of sets of fragments of the molecule to be sequenced, separation of fragments in gel electrophoresis and subsequent analysis of the gel image. Fragments are produced either using the method of controlled termination of replication) of the target molecule (*Sanger dideoxy method*) or by chemical or enzymatic hydrolysis. From the image of size-ordered, overlapping and aligned fragments, the sequence of bases can be read by recognising the identity of the terminal base in each fragment in the order of increasing length.

4.2.1 Sanger dideoxy method

The method of controlled termination of replication is adapted from the Sanger dideoxy method of sequencing DNA.[33, 34] In this method, complementary DNA copies of various lengths are produced from the template RNA in four separate reactions using reverse transcriptase polymerase, four regular deoxynucleotide triphosphate molecules as building blocks and a synthetic 20-30 nucleotides long DNA molecule (*primer*). The reaction starts at the edge of the bound primer molecule, whose nucleobase sequence is designed for complementary base pairing with the region of template RNA outside the region to be sequenced.

Each reaction mixture contains also a small concentration of a dideoxy analog of one of the normal deoxynucleotide as a random terminator of the synthesis reaction, a different analog in each reaction. The dideoxy analog of a nucleotide lack a hydroxyl group at the 3´ position of the ribose, thus preventing the polymerase enzyme to incorporate any more nucleotides after the addition of a dideoxy analog. Due to the low concentration of the terminator molecule, the incorporation of the dideoxy analog occurs only sporadically allowing the synthesis to proceed to subsequent sites of possible termination. Radioactive label is added either in the primer or in one of the deoxynucleotides, so that the products

can be visualized from the gel. The result of each reaction is a distribution of DNA products of different lengths, each ending to the dideoxy analog used in that particular reaction.

When all four sets of products are subjected to electrophoresis in adjacent lanes, the sequence of the oligonucleotide can be read from the gel image. The base calling is done by recognising the order of the bands and their respective terminator bases starting from the bottom of the gel image and proceeding to the top. The resulting sequence is read in the 5′→3′ direction, transcribing each read nucleobase of cDNA to a complementary nucleobase of RNA. Principle of sequencing using the Sanger dideoxy method is described in Figure 6.

Target RNA to be sequenced

↓

3´-UGACGCUUGUCA-5´

| | |

5´-ACT-3´

↑

DNA primer bound to complementary starting site

Products of sequencing reactions:

	Dideoxy analog in sequencing reaction			
	dideoxy-G	dideoxy-A	dideoxy-C	dideoxy-T
Species produced in reactions	5´-ACTGCGAACAG-3´	5´-ACTGCGAACA-3´	5´- ACTGCGAAC-3´	5´- ACTGCGAACAGT -3´
	5´-ACTGCG-3´	5´-ACTGCGAA-3´	5´-ACTGC-3´	
	5´-ACTG-3´	5´-ACTGCGA-3´	5´-ACT-3´	

The gel image after electrophoretic separation and visualization:

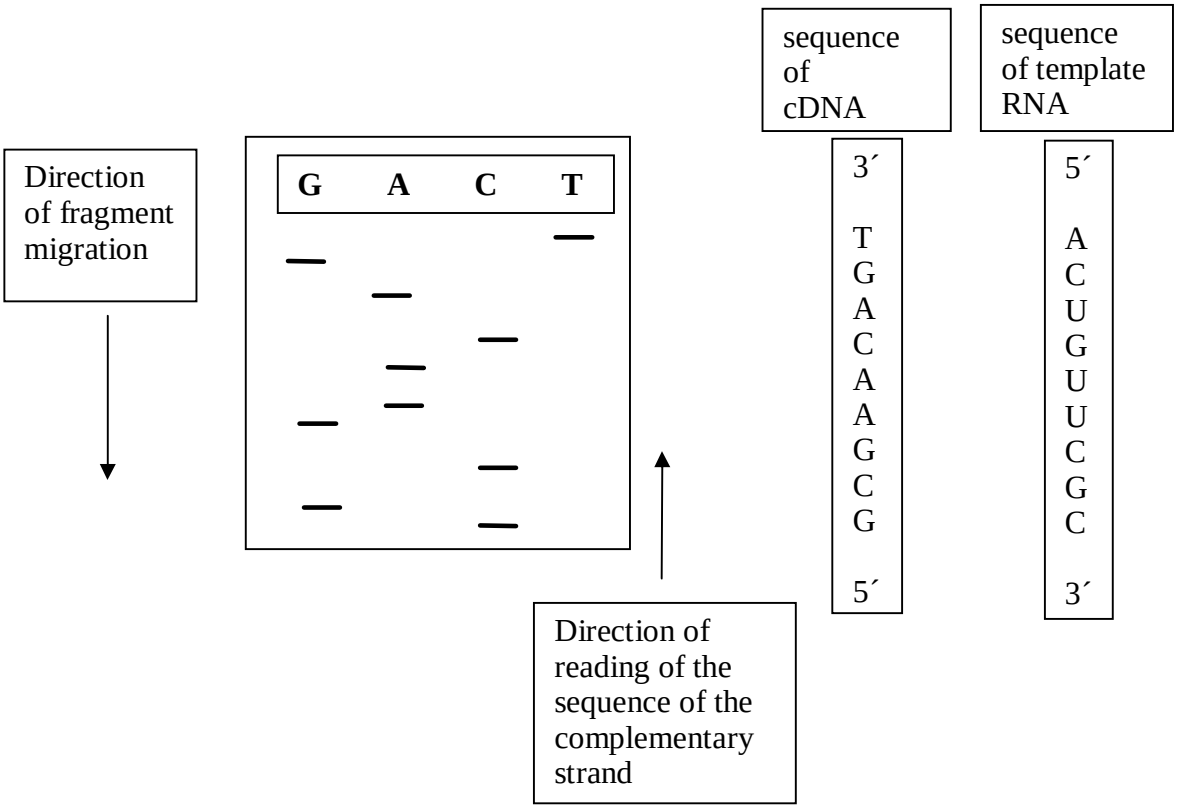


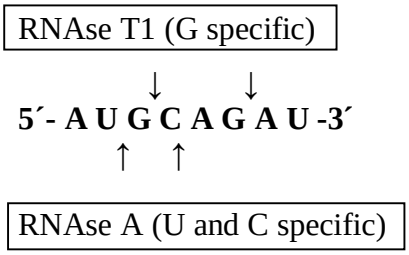
Figure 6. The principle of Sanger dideoxy sequencing of RNA. Figure adapted from the book by Berg, Tymoczko and Stryer.[11]

4.2.2 Enzymatic sequencing

Sequencing strategies using enzymatic hydrolysis are based on the production of overlapping sets of fragments of different lengths derived from the target oligonucleotide. This is accomplished either by sequential removal of individual nucleotides from the either end of the chain or inducing cleavages of the chain using site-specific enzymes.

The classical enzymatic method for sequencing RNA employs sets of enzymes, which recognise specific nucleobases in the oligonucleotide chain and cleave phosphodiester bonds only at these sites.[35] This class of enzymes is also known as base-specific endonucleases. The reaction conditions are adjusted so that strand scission may occur only once per molecule but at different sites of recognition in different molecules. Digestion provides sets of fragments, where original termini are present and the identity of the base at the site of cleavage is known. Figure 7 illustrates the production of sets of fragments from a single oligomer with the use of two different base-specific enzymes.

A.



B.

5´terminal digestion products	
RNAse T1	**RNAse A**
5´-AUGp-3´ 5´-AUGCAGp-3´	5´-AUp-3´ 5´-AUGCp-3´

Figure 7 .A) predicted cleavage sites of a hypothetical oligonucleotide upon digestion with nucleases T1 and A and (B) products of digestion.

Target oligonucleotide is digested in several separate reactions, each using an enzyme with different base specificity. Enzymes are chosen so that series ending to all four normal nucleobases are obtained. However, there are no strictly uridine or cytosine-specific enzymes, so these bases have to be resolved using a combination of enzymes to produce exclusive information.

In order to function efficiently, the acidity and concentration of divalent cations of the reaction solution must be adjusted to meet the requirements of the enzyme used. Commonly used endonucleases and other specifications are presented in the Table 5.

Table 5.. Enzymes used in structural studies of RNA.[35-37]

Enzyme	**Base specifity**	**Products**[1)]	**Cofactor**	**Optimal pH**
RNAse T1	3´ end of G	5´-**R**Gp-3´		7.5
RNAse T2	3´ end of all 4 residues, preferably A	5´-**R**p-3		4.5
RNAse U2	3´ end of A	5´-**R**A-3		4.5
RNAse PhyM	3´ end of A and U	5´-**R**Ap-3 5´-**RU**p-3	with urea	4.5
		5´-**R**Ap-3 5´-**R**Up-3 5´-**R**Gp-3	without urea	4.5
RNAse CL3	3´ end of C	5´-**R**Cp-3		6.5
RNAse A	3´ end of C or U	5´-**R**Cp-3 5´-**R**Up-3		7.5
RNAse V1	double-stranded regions without sequence preferences	5´-**R**p-3	divalent cations	7.5

[1)] **R** denotes the remaining chain of the oligonucleotide of any sequence

When the digestion products of each reaction are separated in the adjacent lanes of gel during electrophoresis, the base sequence can be read from the gel image much in the same fashion as when using the Sanger chain termination method.

The hydrolysis proceeds through a cyclic phosphate intermediate, which occasionally remains in the cyclic form. The mechanism of hydrolysis of a phosphodiester bond is presented in the Figure 8.

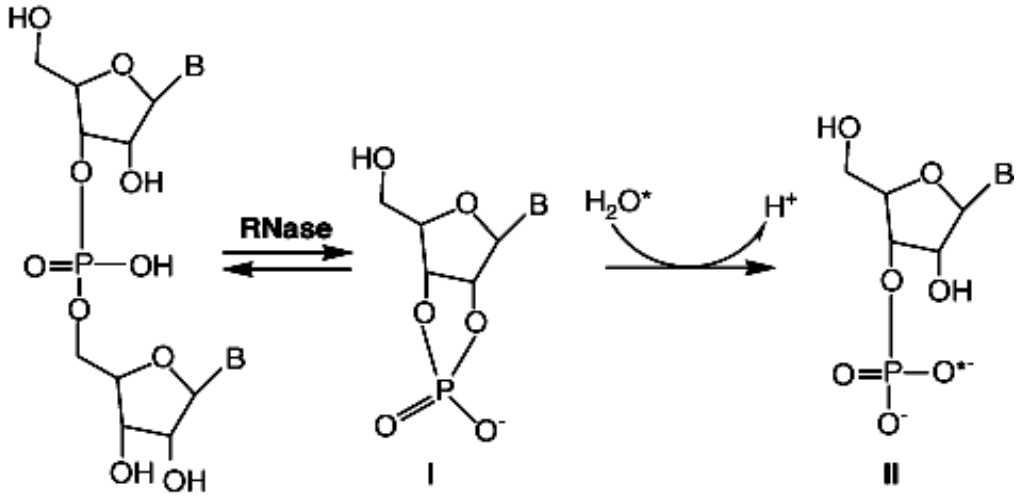


Figure 8. General mechanism of hydrolysis of nucleic acids by ribonucleases.[38]

4.2.3 Alkaline hydrolysis

Fragment ladders may be produced also chemically using a method of limited alkaline hydrolysis, where oligonucleotides are incubated in a solution of ammonium hydroxide at pH ~10 at high temperature (60-90 ˚C).[39-41] Incubation produces strand scissions at the phosphodiester bonds without any site specifity through nucleophilic attack of the hydroxyl ion to the 3´ hydroxyl group of the ribose. Taking aliquots of the reaction solution at specified time intervals, sets of fragments of all possible lengths of the target oligonucleotide may be produced. Prolonged incubation will lead to complete hydrolysis of the oligonucleotide to monomers. Ladders of alkaline hydrolysis are run in parallel with other samples to facilitate determination of lengths of fragments in the lanes of other samples. The mechanism of alkaline hydrolysis is illustrated in Figure 9.

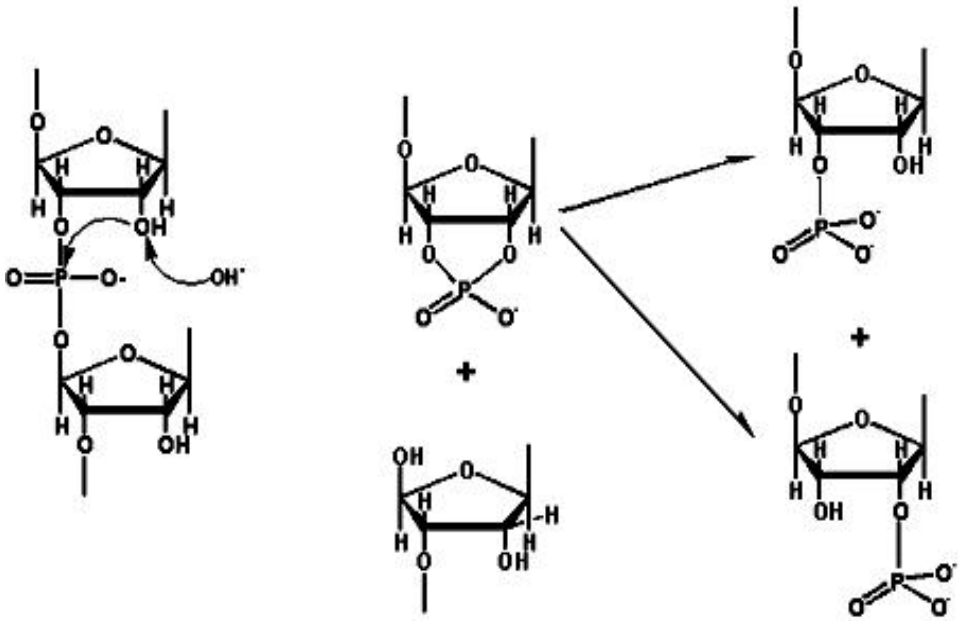


Figure 9. The mechanism of alkaline hydrolysis.[42]

4.3 Interactions

The noncovalent RNA-ligand –interactions can be detected using *electrophoretic mobility shift assay* (or also known as *gel retardation assay*),[43] where the complex can be detected from the decreased length of migration compared to unbound reference RNA molecule. Analyses of RNA-protein interactions usually employ different crosslinking methods, where covalent bonds between nucleobases and peptide sidechains are formed at the sites of contact by chemical crosslinking agents[44, 45] or irradiation with ultraviolet light.[25] UV irradiation enables the study of complexes of RNA also in the living cells; RNA-ribonucleoprotein crosslinks have been studied extensively in this manner.[27, 46] Crosslinking with UV irradiation can be enhanced using synthetic RNA, where photoreactive base analogs have been inserted.[47]

UV irradiation conditions must be carefully optimized, because the longer the time the complex is irradiated, more crosslinks are formed but modifications in the protein part or nicking of the oligonucleotide strands may be induced.[27]. Generally, irradiation at 254 nm for few minutes has been used. Because of the choice of mild irradiation conditions, the occurrence of crosslinks is low and makes the detection of crosslinked sites challenging.

The sites of interaction can be approximated using controlled enzymatic fragmentation studies called RNA footprinting.[28] In the footprinting experiment, complexed trial RNA and uncomplexed reference RNA and are both cleaved with base-specific enzymes or

chemical reagents. Bound ligands may protect some cleavage sites when the digestion step is performed, resulting absence of corresponding fragments in the gel image, when the band patterns of the complexed RNA and uncomplexed reference RNA are compared. The gaps in the band pattern of trial RNA form the “footprint” of the bound ligand or ligands. The principle of the method is illustrated in Figure 10.

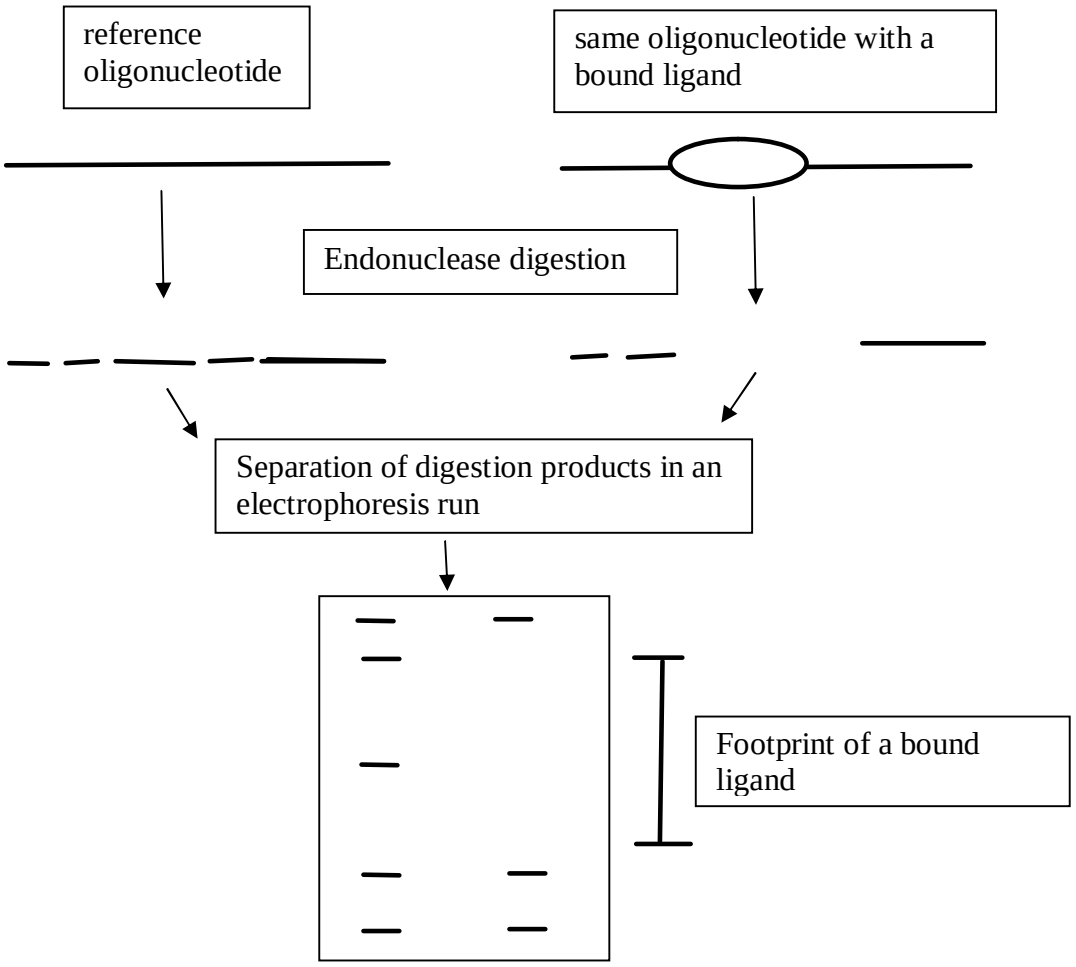


Figure 10. A schematic diagram of a footprint experiment.

More accurate localisation can be achieved by premature termination of synthesis cycles in a process called *primer extension*, where complementary single-stranded DNA copies of the regions of interest in the RNA are produced with the use of reverse transcriptase polymerase.[48, 49] A suitable primer is attached to a desired complementary single-stranded region. The polymerase recognises the boundary of single- and doublestranded regions, set by the primer, and starts extending the primer by adding complementary deoxynucleotides to the RNA strand beginning from the 5´ end of the primer. Unless inhibited, the synthesis

continues towards the 5´ end of the target RNA until the end of the molecule is reached. Polymerases cannot proceed through crosslinks, modified bases, secondary structures or bound ligands in the template RNA, resulting in premature stops of transcription and shortened products. When the reaction mixture is analysed in PAGE together with sequencing reactions, the resulting bands in the primer extension experiment indicate locations of crosslinks in the sequence. For larger molecules, several primers for different specific sites of RNA are used for mapping of the molecule in its complete length.

4.4 Higher order structures

Secondary structures have traditionally being mapped by means of three basic approaches: 1) probing the available cleavage sites in the single-stranded regions of RNA with base-specific endonucleases,[36] 2) derivatising the nucleobases of the single-stranded regions with base-specific chemical probes followed by chemical scission of strands[37] and 3) producing series of complementary DNA (cDNA) copies of the single-stranded regions of RNA with a primer extension method.[50] Single-stranded regions are first delimited by enzymatic digestion or tagging with structure probes. Products of digestion or primer extension are then separated by gel electrophoresis and the image analysed.

In the enzymatic probing of higher order structures, the double-stranded and tertiary structures inhibit single-strand specific enzymes cleaving at the sites predicted from the sequence information resulting in lack of expected fragmentation products in the image, much in the same fashion as in the footprinting studies of interactions. The regions of cleavage protection can be pointed out and identified accordingly. To verify the presence of secondary regions, nucleases specific for only double-stranded regions may be used. Figure 11 presents results of a hypothetic enzymatic structure probing of a simulated ribonucleic acid molecule.

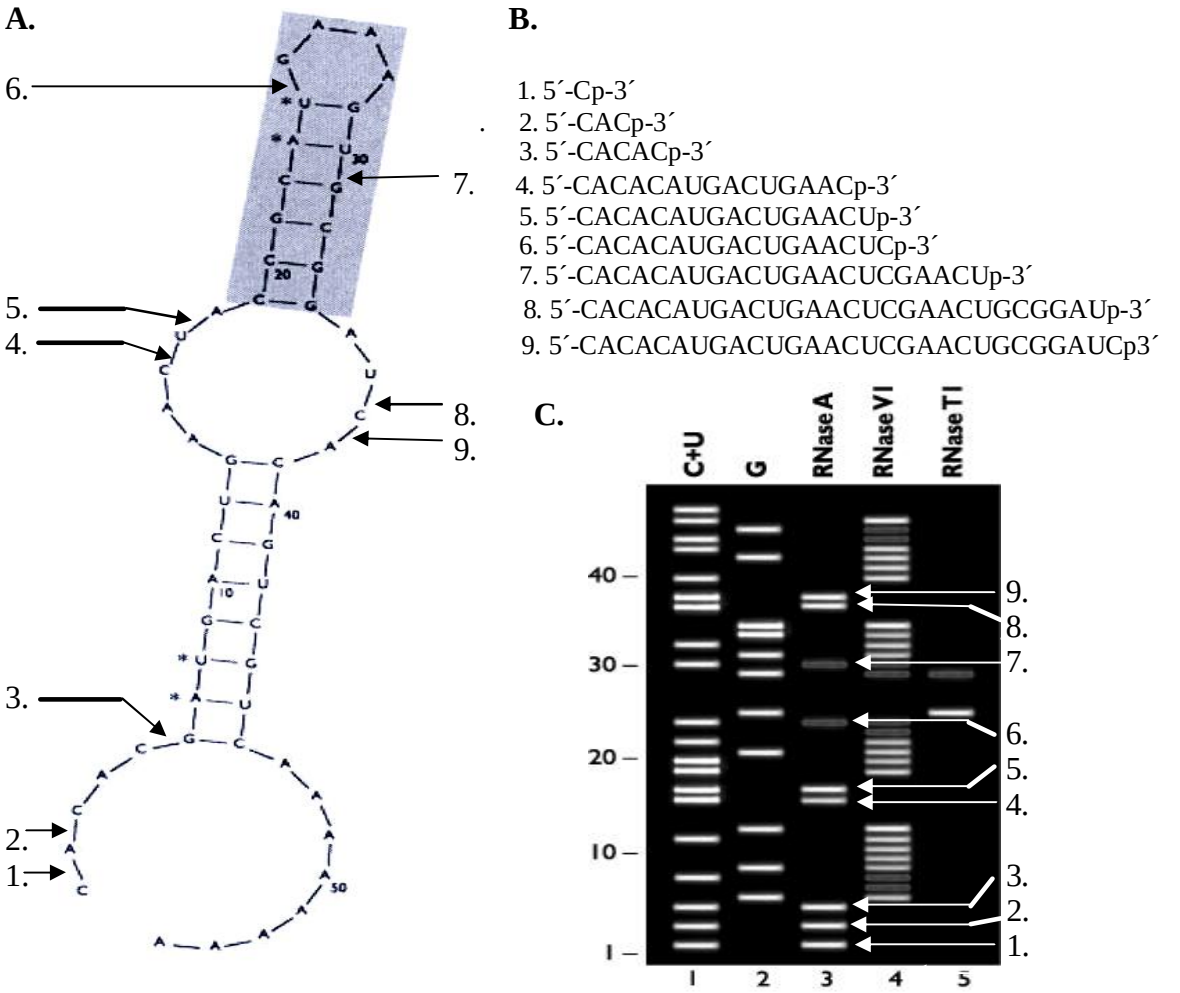
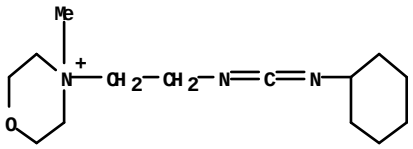
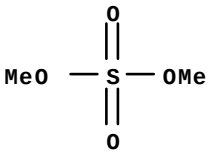
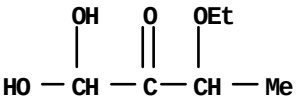


Figure 11. Schematic presentation of a simulated enzymatic structure probing experiment with a hypothetical structured RNA target molecule digested with selected enzymes.[51] For clarity, the cleavage sites and products of digestion are presented in detail only for the endonuclease RNase A in panels A and B. **Panel A:** illustration of the secondary structure of a target RNA molecule. The numbered arrows point to sites available to the cleavage by endonuclease RNase A. In this context, the shaded area of stemloop bears no special significance. **Panel B:** a list of produced oligonucleotides from the enzymatic digestion of the target RNA with the use of RNase A. Numbers correspond to the numbered sites of single cleavage in the RNA molecule shown in the panel A. **Panel C:** Simulated gel image of electrophoresis run of the products of enzymatic probing experiment. Lanes 1 and 2: products of enzymatic sequencing reactions with the use of endonucleases RNase V1 and T1. Lane 3: products of digestion with the use of RNase A. Lanes 4 and 5: products of digesion by using endonucleases RNase V1 and T1. Numbered white arrows on the right indicate the level of migration of bands corresponding to digestion products listed in panel B.

Single-stranded regions may be mapped more accurately with specific chemical structure probe reagents[52] to derivatise all nucleobases in the places, where the bases are not protected by base-pairing or ligand binding. Probes are also frequently referred in the literature as *solvent accessibility* probes, implying that the probed areas are susceptible to interactions with solvent molecules.

Probe reagents attack specific nitrogens of nucleobases depending on their specifity. Attachment of the probe molecules renders the RNA strand susceptible to cleavage at the binding site with other chemical reagents, such as aniline and hydrazine. Probing may be performed by incubating target RNA with chosen reagent for an appropriate time in aqueous solution of suitable pH. Prior to probing, the RNA samples are renaturated with a short heating and slow cooling treatment to ensure the biologically native and correct folding of the target molecule. The probe:RNA target concentration ratio must also be experimentally determined to avoid induction of conformational changes to the molecule which would lead to erroneous interpretation. Most commonly used structure probes and some properties of them are presented in the Table 6.

Table 6. Common structure probes.[52]

IUPAC name	Acronym	Molecular structure	Molar mass (g/mol)	Specifity	Optimal pH
1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluene sulfonate	CMCT		424	N at 3´ position of uracil N at 1´ position of guanine	8.0
methylsulfinylmethane (dimethyl sulfoxide)	DMS		126	N at 3´ position of cytosine N at 1´ position of adenine N at 7´ position of guanine	~7
β-ethoxy-α-ketobutyraldehyde (kethoxal)	KT		148	N at 1´ position of guaninel N at 2´ position of guanine	<7

DMS reacts primarily with nitrogens at positions 7´ of guanine, 1´ of adenine and 3´ of cytosine at neutral pH. Subsequent treatment with aniline induces strand scission at cytosine bases; prior reduction of the 7, 8 double bond with sodium borohydride is needed for guanine. No strand scission occurs at adenine, so this modification must be detected by other means, such as primer extension. CMCT reacts primarily with nitrogens at positions of 3´ of uridine and 1´ of guanine, preferably with uridine. Kethoxal reacts specifically with unpaired guanosines forming a new ring structure connecting nitrogens at positions 1´ and 2´ of guanine ring. Reaction mechanisms are presented in Figure 12.

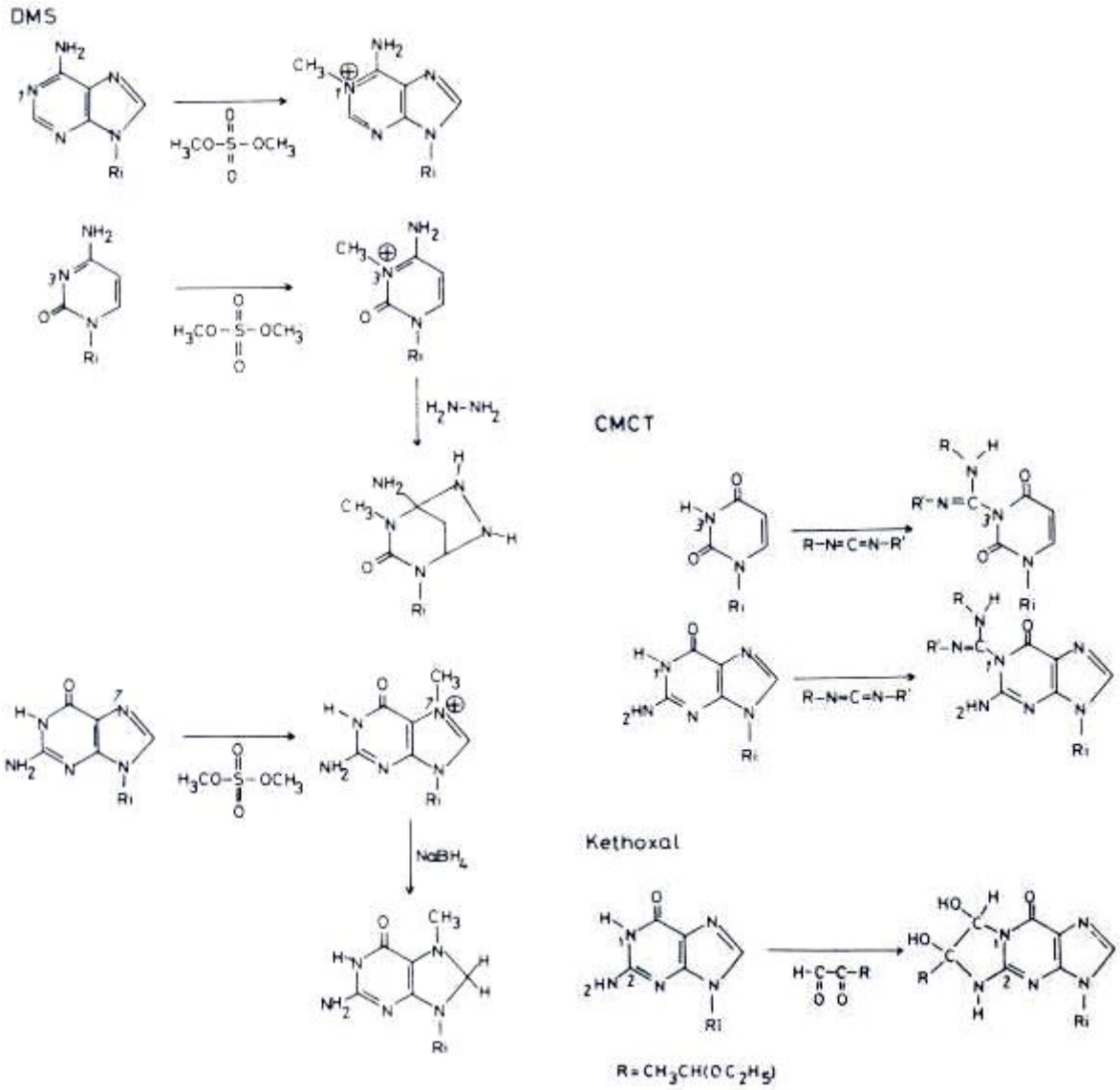


Figure 12. Mechanisms of the derivatisation of RNA with structure probes. **A:** DMS alkylation of 1´-N of adenine, DMS alkylation of 3´-N of cytosine followed by hydrazine treatment and alkylation of 7´-N of guanine followed by sodium borohydride ($NaBH_4$) reduction. **B:** CMCT modification at 3´-N of uridine and 1´-N of guanine and kethoxal modification at 1´-N of guanine. Figure adapted with modifications from a paper of Ehresmann *et al.*[52]

5 MASS SPECTROMETRIC METHODS OF ANALYSIS

In contrast to traditional methods, mass spectrometry enables detailed studies of posttranscriptional modifications and tools for studying three-dimensional structures. In this chapter, the fundamentals of mass spectrometric analyses of ribonucleic acids are described and the application of mass spectrometry to structural studies is declared. An overview of instrumental arrangements is made and some details of analytical performance of methods are presented.

Rapid mass spectrometric methods for sequencing short ribonucleic acids have been developed.[26, 38, 39, 53-67] Most commonly, techniques are practiced in conjunction with other structural studies of RNA and in the verification of base sequences of short oligoribonucleotides. Few attempts have been made to establish completely unknown sequences.[58]

Mass spectrometric approaches have been developed for fast and efficient detection and characterization of interactions of RNA with different ligand molecules.[41, 46, 68-77] RNA-protein interactions have been studied by investigating binding of small nuclear ribonucleic acids (snRNA:s) to small nuclear ribonucleoproteins in spliceosome assemblies[74, 76, 77] and binding of nucleocapsid protein to the human immunodeficiency virus (HIV-1) packaging signal RNA (ψ-RNA).[75] Spliceosomes remove introns from messenger RNA molecules in the RNA processing phase. Binding of nucleocapsid protein to the packaging signal is essential for the replication of virus and thus is a site of interest in the search for possible antiviral drugs. Also effective screening methods for RNA-binding ligands in the investigation of therapeutic agents have been developed.[46]

Bacterial, archaeal and eukaryotic tRNA and rRNA have been investigated to discover novel posttranscriptional modifications. Some presumptive findings obtained with traditional methods have been confirmed. Also the role of modifications in adjustment to environmental changes and development of antibiotic resistance has been examined.[17, 18, 78-86]

The methods for determination of higher order structures in RNA have been successfully transferred to mass spectrometric environment.[87-91] Structured ribonucleic acid molecules may exceed in size or be too difficult to crystallize properly to be examined by traditional methods of structural analyses, such as X-ray crystallography and nuclear magnetic resonance (NMR).[87, 89] Studies of secondary and tertiary structures of viral pseudoknots

and stemloops have produced data to support computer-aided designing of three-dimensional models of these structures.[90]

5.1 Principle of mass spectrometry

Mass spectrometry consists of various methods based on accurate measurement of mass and number of analyte molecules, which are released in the gas phase by different ionization methods.[92] Depending on the type of ionization method, intact or fragmented and singly- or multiply charged ions are extracted from the sample, released ions are directed towards mass analyser and separated by their mass-to-charge ratio (*m/z*) and subsequently detected. The separation of ions is conducted using adjustment of electromagnetic fields or kinetic energy provided. During separation, ions are passed to the detector in groups of one *m/z* ratio at a time and the number of respective ions counted. Commonly used detectors measure electric currents originating from cascades of electron emissions, which are initiated by an impact of analyte ion to the first dynode in the series of dynodes in an electron or photon multiplier tube. Internal or external calibration of the analyser by using reference substances of known molar masses is required for accurate molar mass determinations. General overview of mass spectrometry is presented in the Figure 13.

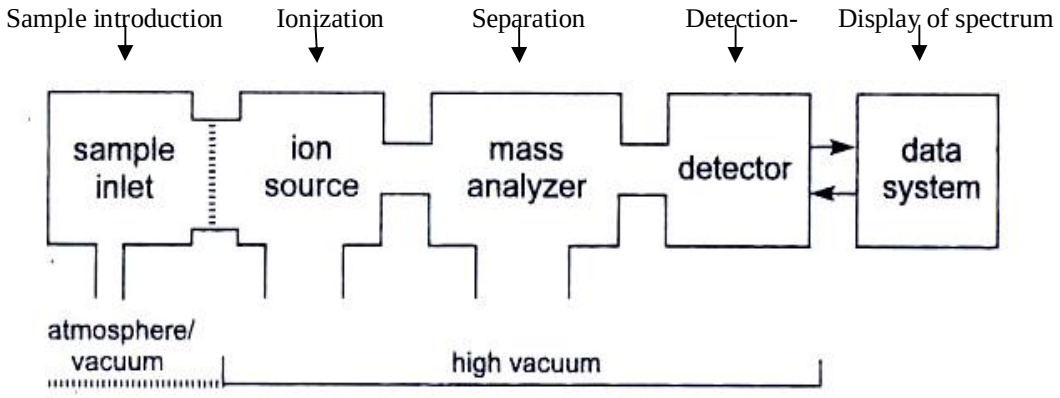


Figure 13. General overview of mass spectrometric analysis.[93]

The information from a mass spectrum can be used for assigning exact molar masses for analytes and deriving structural information of the molecule on the basis of identification of characteristic losses of neutral fragments.

5.2 Mass spectrometry of ribonucleic acids

The sizes of the RNA molecules limit the applicability of ionization methods available. Furthermore, to obtain meaningful structural information, excessive fragmentation of molecules is not desired. Intensive fragmentations of large molecules result in complex peak patterns which are not possible to interpret unambiguously. The research of oligonucleotides was facilitated greatly by the development of electrospray ionization (ESI) and matrix-assisted laser desorption (MALDI) methods, which are capable of ionizing large biomolecules while keeping the ions intact.[92]

ESI enables exact determination of masses of large molecules and allows coupling of liquid chromatography for fractionation of mixtures of ribonucleic analytes. Mixtures of ribonucleic acids may be analysed directly using MALDI. Various mass analysers may be used with ESI and MALDI. Tandem techniques permit gaining of more detailed structural features from the analytes.

In the following sections, the principles and instrumentation and of mass spectrometry of ribonucleic acids is described at every stage of analysis.

5.2.1 Sample preparation

So far, ribonucleic acids extracted from original biological sources have been used almost exclusively only in studies of spliceosomal protein-RNA crosslinks and naturally occurring posttranscriptional modifications in bacteria and archaea. Otherwise, commercially available and *in vitro* produced partial or whole molecules of tRNAs, rRNAs, stemloops, pseudoknots and other regions of interest have been widely used. Inserting modifications to oligoribonucleotide backbone may give protection against ribonucleases. [71]

Commercial products have often been used without prior purification whereas the biological, in-house synthesised or *in vitro* produced ribonucleic acid analytes require purification steps such as ion exchange chromatography, reverse phase liquid chromatography (RPLC), solid phase extraction, PAGE, dialysis and alcohol precipitation, to mention few.

5.2.2 Sample pretreatment

Depending on the field of application and the goal of analysis, the methods of sample pretreatment are chosen to produce species of analytes, which are informative for the particular application in question. Combinations of different mass-shifting derivatisations, enzymatic or chemical structure probing or isotopic labelling may be used. In most cases, the ultimate steps of pretreatment involve enzymatic or chemical hydrolysis of samples either to just reduce the sizes of large analytes to more manageable sizes[58] or to finalize the analytical fragment production.

CMCT, DMS and KT have been used for structure probing. Combination of endonucleases RNase A and T1 has been most commonly used along with different phosphodiesterases.. There are still no specific digestion enzymes available for hydrolysis of U or C nucleobases: a fact which contributes to the limitations of mass spectrometric sequence determinations.

5.2.3 Sample introduction and ionization

In electrospray ionization, solvent-phase ionic analytes are released to gas phase by spraying the analyte solution through a narrow capillary needle surrounded by cylindrical electrode. A potential difference of 3-6 kV is applied between the capillary needle tip and the cylindrical electrode producing a fine mist of highly charged solvent droplets emerging from the tip. The solvent evaporates rapidly leaving intact analyte ions in the gas phase. Ions are directed and focused to the mass analyser through two stages of vacuum and ion optics. A schematic diagram of electrospray ionization is presented in Figure 14.

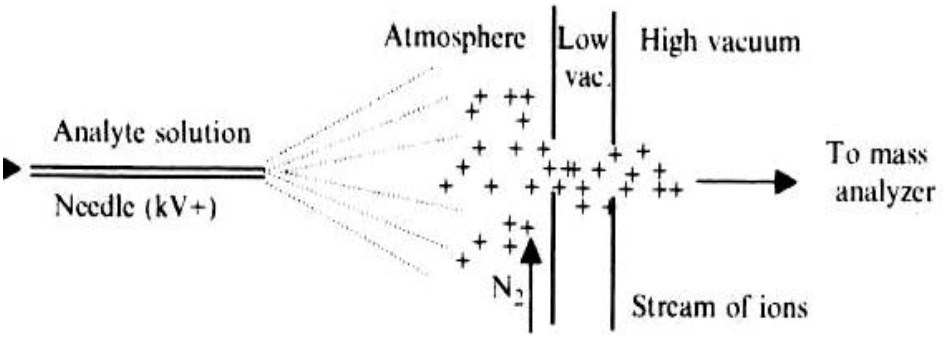


Figure 14. Schematic presentation of electrospray ionization.[94]

Droplets may be positively or negatively charged depending on the sign of the voltage applied. Ribonucleic acids are most commonly analysed in negative mode. Desolvation process can be assisted by means of auxiliary gas flow (*nebulising gas*) outside the spray needle in parallel direction. .Drying gas (*sheath gas*, *curtain gas*) flow, usually nitrogen, is passed through the stream of droplets assisting the evaporation of the solvent. To prevent formation of corona discharges at the spray needle tip, especially in the negative ion mode, analyses, a flow of air or sulphur(VI) fluoride ($SF_{6)}$ is often applied to pass near the needle tip.[95]

Ribonucleic acids are ionized as intact, multiply charged ions appearing in several charge states due to partial ionization of the phosphate groups.[96] Alkali metal ions form easily adducts with phosphate groups leading to distribution of ion species with different charge states. Furthermore, high resolution spectrometers render the single peaks to a cluster of isotope peaks due to presence of naturally occurring isotopes of C, N, O, and P. The isotopic pattern may be used in the definition of the charge state of the ion.

A general formula for observed ions is

$$(M-[n+m]H+mNa/K)^n \qquad (1)$$

where M denotes the molecular ion, [n+m] is the total number of lost protons (H), m is the number of adducted Na or K ions (Na/K) and n is the overall charge state of the ion.[97]

Patterns of adducted molecular peaks may not be resolved which will lead to formation of single large peaks, reducing the signal-to-noise ratio and accuracy of determination of masses.[96] Therefore, addition of metal cations to analyte solutions has to be avoided at all stages of analyses where possible. For example, sodium-containing buffer salts must be replaced by more volatile ammonium equivalents. Complexing agents, such as EDTA or cyclohexane-1,2-diaminetetra-acetic acid (CDTA) may be added to remove divalent cations. Various liquid chromatographic methods may be used for removal of residual salts prior to ionization.

Solutions containing purified oligonucleotides of single species can be introduced to the ESI ion source using direct infusion with syringe pumps. Complex mixtures of oligonucleotides, such as pools of products from enzymatic hydrolysis, have to be fractionated by using chromatographic methods before ionization, otherwise overlapping ion series of multiple charge states, originating from different analyte molecules, may lead tounresolved spectra. High performance liquid chromatographic (HPLC) separation

systems are commonly used, because the ESI sources allow on-line coupling of effluents from chromatographic separations. Furthermore, HPLC separations enable simultaneous cation removal and desalting of the samples.[26]

For direct infusion, volatile organic modifiers are added to aqueous RNA solutions to facilitate nebulisation through lowering of surface tension. Acetonitrile, methanol, 2-propanol are commonly used in concentrations of 50%. Volatile organic bases, such as ammonium acetate, piperidine or imidazole may be added to replace metal cation adducts with ammonium counterions. High overall charge states of large oligonucleotide analytes may be reduced adding volatile organic acids, such as trifluoroacetic acid (TFAA) or 1, 1, 1, 3, 3, 3-hexafluoro-2-propanol (HFIP), to reduce spectral complexity and enhance the sensitivity of detection.

A widely practiced mode of HPLC of ribonucleic acids is known as *ion pair* RPLC (IP-RPLC),[98] where the aqueous eluent contains amphiphilic reagent ions, which form ion pairs with the ionized phosphate groups of single-stranded nucleic acid backbone. The formations of ion pairs render the analyte molecules hydrophobic, resulting in interaction of analytes with the stationary phase. Non-polar nucleobases contribute to the additional solvophobic interactions. The interactions grow stronger as the length of the adsorbed molecule increases. The analytes are desorbed from the stationary phase according to the increasing length of the molecules by increasing gradually the concentration of organic modifier in the eluent flow. Most commonly used ion pair reagents are triethylamine (TEA) and triethylammonium bicarbonate (TEAB). For efficient ion pair formation, the temperature of the eluent is held at 60-70 ˚C to maintain ribonucleic acids free from secondary structures.[98]

The detection of analytes is based on the UV absorption of nucleobases at 260 nm. Separated fractions can be collected on-line during analysis with the use of a post-column splitter, which allows small amount of analyte enter the ion source while directing most of the flow to fraction collector.[35, 99]

The flow rate of the effluent to the ESI source and the composition of the effluent has to be adjusted to be compatible with the ion source.[95] Both adjustments can be performed by post-column splitting of the flow and addition of modifying agents by using tee splitters and a syringe pump. Using post-column techniques, solvent composition optimal for the chromatographic separation may be further modified for most efficient ionization. Using nanobore chromatographic columns, flow speed adjustment is not necessary. All modifiers,

along with buffers and ion pair reagents must be volatile. The ion strength must be maintained low, in millimolar range. Also the pH is best kept around 7 to avoid acid- or base-catalysed hydrolysis of RNA analytes.

Figure 15 illustrates a schematic model of coupling liquid chromatography to an ESI ion source.

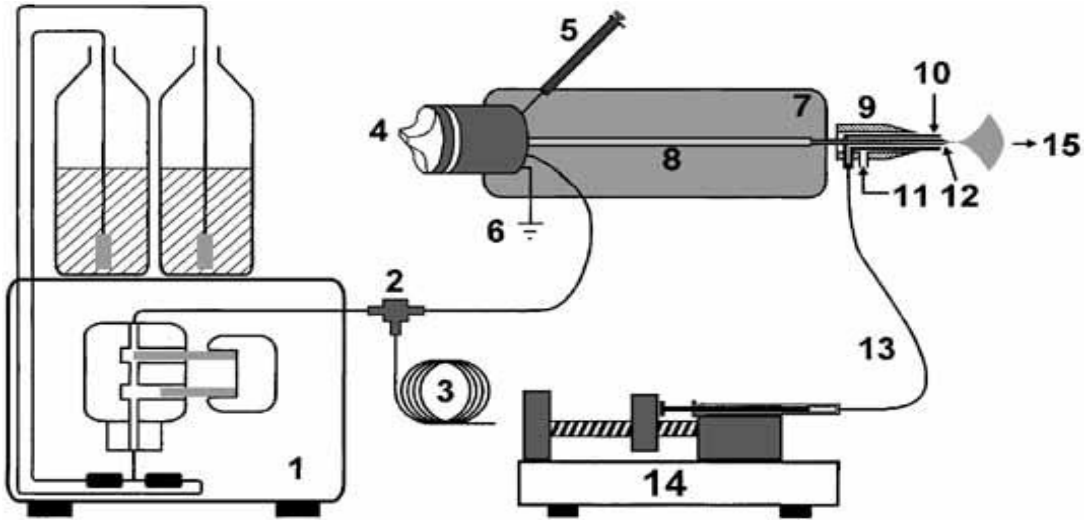
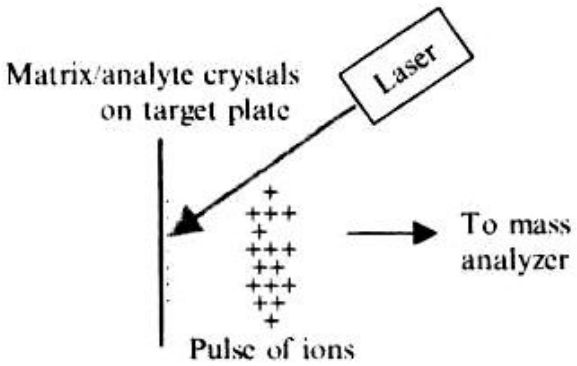


Figure 15. Schematic illustration of instrumentation for nucleic acid analysis by mass spectrometry coupled with HPLC. 1) gradient micropump; 2) splitting tee; 3) restriction capillary; 4) microinjector; 5) syringe for sample injection; 6) ground; 7) column thermostat; 8) capillary column; 9) triaxial electrospray ion source for connections for sheath gas and liquid; 10) metal needle at high voltage; 11) sheath gas line; 12) fused silica spray capillary; 13) sheath liquid line; 14) syringe pump; 15) mass analyser.[100]

In MALDI, analytes are desorbed to gas phase from the surface of crystallized sample spots using the energy of laser irradiation.[101] Sample spots are made by first mixing a low amount of analyte to a matrix medium capable of absorbing laser light. Then the mixture is allowed to dry onto a stainless steel target plate. Upon irradiation, the matrix transfers the energy of absorbed laser light to the analyte molecules leading to analyte ionization and ejection from the sample spot.

Analyte molecules are transferred to gas phase as intact, singly charged ions. Ions having molar masses between 3 kDa- 300 kDa can be ionized. Method allows analyses from the mixtures of analytes without prior fractionation, because singly-charged ion species are

resolved easily. However, mixtures may be fractionated using LC and the fractions of interest concentrated and lyophilized for further analysis using MALDI. A schematic illustration of the ionization method is presented in Figure 16.



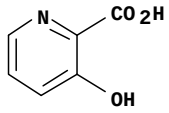
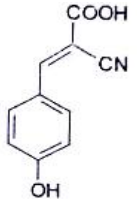
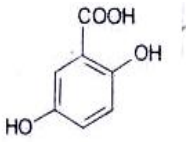
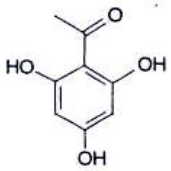
c

Figure 16. Schematic presentation of matrix-assisted laser desorption ionization.[94]

With a few exceptions,[39, 102] nitrogen-type lasers with wavelengths in ultraviolet region have been used for MALDI ionization. The pulse widths of few nanoseconds of irradiation have generally used for ribonucleic acids. The spectrum is usually formed by averaging spectra of several tens of successive shots.

The matrix is chosen to have similar solubility to the analyte to ensure even dispersion of analyte throughout the sample spot. The absorption maxima may differ between various matrices, and it should coincide with the operating wavelength of the laser used. 2,4,6-trihydroxyacetophenone (2,4,6 -THAP) has proven to be most suitable matrix for detection of both oligoribonucleotides and peptides.[69, 74] 2,4,6-THAP has been used also in combination with its structural isomer 2,3,4-THAP.[64, 79]. Table 7 presents some commonly used matrices in mass spectrometric analyses of ribonucleic acids.

Table 7. Matrices commonly used in mass spectrometry of ribonucleic acids.

Matrix name	Molecular structure	Laser type	Wavelength (nm)	Reference #
3- hydroxypicolinic acid (3-HPA)		Nd:YAG[a]	355	102, 39
α-cyano-4-hydroxycinnamic acid (CHCA)		N_2	337	68, 41, 74
2,5-dihydroxybenzoic acid (2,5-DHB)		N_2	337	41, 74
2,4,6-trihydroxyacetophenone (2,4,6-THAP)		N_2	337	31, 54, 64,73

[a] Neodymium-doped yttrium aluminium garnet

The co-crystallization of the matrix and the analyte is a critical step affecting the level of ionisation and the overall quality of spectra. The carryover of impurities from isolation, synthesis, purification and enzymatic and chemical pretreatments of ribonucleic acids may interfere the formation of crystals and cause peak broadening through adduct formation. Shaler *et al.* [103]studied the effects of impurities such as Tris buffer salts, urea, dithiotreitol, complexing agents and glycerol to the quality of MALDI spectra, although only with DNA analytes and 3-HPA as matrix . The overall tolerable concentration level of these impurities was in the range of 0.25-0.5 M.

MALDI may not be that sensitive to the impurities in the sample as the ESI, but the cation adduction remains still as a problem. Ammonium citrate has been added frequently to the matrix to remove metal cations from the mixture.[38, 54, 57, 64, 79]The suitability of various other ammonium salts as MALDI matrix additives have been evaluated, [104, 105]although only with using DNA as analyte. Alternatively, the matrix and sample may be treated with

beads coated with ion exchange material to remove metal ions from the sample mixture.

5.2.4 Mass analysis

The first generation of mass spectrometers, called magnetic sector instruments, separated analyte ions using the effect of homogenous magnetic field to charged analyte ions.[106] Due to the energy of magnetic field applied perpendicular to the direction of entering ions, the flight paths are curved at different radii depending on the linear momentum of the ion, which in turn is dependent on the *m/z* ratio of the particular ion. By sequential adjustment the magnetic field, scattered ions of selected *m/z* ratio are focused and allowed to reach the detector while other ions collide the walls of the magnetic sector device.

Modern double-focusing instruments use an additional sector (electrostatic analyser) to produce radial electric field in combination with magnetic sector. In the electrostatic analyser, ions are separated by their kinetic energies. In the "forward" or "EB" geometry, electrostatic analyser (E) is in the front of the magnetic analyser (B) and vice versa in the "reverse" or "BE" geometry.

Time-of-flight (TOF) analysers separate ions of different *m/z* ratios by their different velocities in a field-free flight tube of known length. Following ionization, the analyte ions are accelerated to flight in an electric field. Equal kinetic energy is provided for all ions in the acceleration phase, but due to different ion masses, lighter ions travel faster. Ions are detected at the end of the tube. The flight tube is kept in high vacuum to ensure collision-free flight path for the analytes.

Ionization, especially MALDI, may introduce distribution of spatial arrangement and times of generation to ions of same *m/z* ratio due to sample spot geometry and duration of pulses. Reflector instruments employ a retarding electric field at the end of the tube to focus and decrease the spatial scatter of groups of ions of same *m/z* ratio. The reflector field slows down the speed of entering ions and eventually ejects them out at the opposite direction. Ions with larger kinetic energy penetrate deeper than those with smaller energy, thus spending more time within the reflector area. When ions are ejected, the distribution of ions first entering the reflector area is eliminated and the front of emerging ions concentrated.

Ions can be focused also using a delay between generation and acceleration of ions, when ions with larger initial speeds receive weaker acceleration than ions closer to the acceleration field (*delayed extraction* or *time lag focusing*).

Original TOFMS instruments use linear structure comprising straight flight tubes with detector located at the end of the tube. Modern instruments extract sequential packets of ions from the ion current to flight tube attached orthogonally to the ion flow coming from the ion source. Reflector fields and delays in acceleration are employed in both configurations.

In quadrupole devices, ions are directed between a group of four parallel metal poles, which have been electrically coupled together pairwise. A positive direct current is applied to one pair of poles and a negative to the other. In addition, radio frequency alternating current is applied so that a phase difference of 180º is formed between the pairs of poles. The flow of analyte ions is gradually scanned using adjustment of electric field or radio frequency alternative current, enabling ions of different *m/z* ratios to pass through the analyser in a sequential manner.

Quadrupole ion trap (QIT) analysers capture ions in stable orbital trajectories between a space confined by two hyperbolic electrodes and a ring electrode. Direct current potential is applied between hyperbolic electrodes and radio frequency potential to ring electrode. Ions are passed to detector one *m/z* ratio at a time.

Fourier transform ion cyclotron resonance mass spectrometer (FT-ICRMS) measures electric currents induced by the circular motion of ions trapped inside a box-shaped chamber. Ions are trapped inside an electromagnetic field and forced into circular trajectories. Ions of different *m/z* ratios circulate with individual cyclotron frequencies. Detection is based on measurement of electric current, which is induced to the receiving plates by the ion circulating in the close proximity. The resulting complex wave is converted to a mass spectrum using Fourier transformation.

5.2.5 Detection

Three modes of data acquisition exist: 1) scanning, 2) selected ion monitoring (SIM) and 3) selected reaction monitoring (SRM).

Scanning implies rapid repetitive measurement of a range of masses at a certain speed. Ions of all masses within that mass range are counted during each cycle of scan. Decreasing mass range or increasing the scanning time may enhance the sensitivity of the detection because more ions per *m/z* ratio are counted. This may also lead to loss of data in failure to detect ions outside the mass range or inability to resolve two different analyte species eluting in the same fraction, when mass spectrometry is coupled to chromatography.

If there is only a few masses of analytical interest to be detected, such as compound-specific fragmentation products, the mode of *selected ion monitoring* may be used. In this mode, scanning can be restricted to these values leading to increased relative scanning times per *m/z* value and thus enhanced sensitivity.

In *selected reaction monitoring,* a characteristic product ion of decomposition of a chosen reaction is selected as a precursor ion for further fragmentation and a certain characteristic product of that precursor ion is monitored in another analyser, in a tandem mass spectrometric environment. This technique is used to selectively monitor whether expected precursor ion-product ion pairs are formed. If multiple pairs of precursor-product ions are monitored, the mode of analysis is called as *multiple reaction monitoring* (MRM), although the usage of the term is not recommended by IUPAC.

5.2.6 Tandem mass spectrometry

Tandem mass spectrometry is a general expression to various analysis techniques, where ions of interest are selected on the basis of *m/z* ratio and subjected to a second mass spectrometric analysis.[106] The selected ion, also called the precursor ion, is dissociated either by using spontaneous metastable fragmentation or by providing additional external activation using different techniques. When several successive series of precursor ion selection and dissociation originating from the same original primary precursor ion is performed, the technique is known as *multiple stage mass spectrometry.* Tandem mass spectrometry is also known as MS/MS or MS^2 and multiple stage mass spectrometry as MS^n, where "n" denotes the number of stages of precursor ion selection and dissociation.

The precursor ion activation and dissociation is most commonly conducted using *collisionally induced dissociation* (CID) where the precursor ions collide with neutral gas molecules (helium, nitrogen, argon) in a separate collision chamber. The chamber is located between the devices for selecting the precursor ion and analyzing the product ions.

Ions trapped inside a FT-ICRMS cell can be activated using laser irradiation of infrared wavelength (*infrared multiphoton dissociation*, IRMPD), but the s*ustained off-resonance irradiation* (SORI) has established as a standard CID method in FT-ICRMS. In this method, a rf-excitation pulse slightly off the resonance frequency is applied to a selected ion to be dissociated. Irradiation increases the radial amplitude of the circulating ion, increasing the number of collisions with the gas inside the cell. During prolonged irradiation period, the internal energy of the ion increases gradually leading eventually to dissociation.

CID may be induced also in the vacuum area between the nozzle of ESI ion source and skimmer cone, when using elevated potential applied bet the nozzle and the skimmer cone (*nozzle-skimmer CID, NS*).[107] In the MALDI-TOFMS instruments, ion fragmentation can be achieved through metastable ion dissociation of activated ions during the flight. Excessive internal energy is provided to ions using higher laser irradiation energy. Metastable dissociation is also called as *post source decay* (PSD).

Tandem mass spectrometric analyses may be performed either *tandem-in-space*, where two separate analysers operating with continuous extraction of ion beams are coupled together, or *tandem-in-time*, where steps of tandem mass analyses are performed in the same instrument capable of processing stored ions, such as ion trap and FT-ICRMS analysers. When individual instruments are coupled together, the precursor ion selection, dissociation and product ion analysis are performed in separate instruments. Most common MS/MS instrumentation combines three quadrupoles (QqQ), where one quadrupole, indicated as lowercase q, is kept as a site of precursor ion dissociation. Two reflector TOFMS instruments may be combined so that one operates at linear mode and the other in a reflector mode. A collision cell may be placed between the instruments.

Ion trap and FT-ICRMS instruments can perform MS/MS and MS^n experiments by ejecting out all ions except the chosen precursor ion and subjecting it to further fragmentation. Fragmentation is conducted in the ion trap instruments by CID with a collision gas and in FT-ICRMS instruments by giving excess translational energy to the precursor ions.

5.3 Application of mass spectrometry to structural studies of ribonucleic acids

Modern mass spectrometric analyses of RNA make use of traditional strategies and employ same methods in many stages of sample pretreatment. Some traditional gel-based analyses have been implemented to mass spectrometric platforms in quite straightforward manner: main differences being the method and accuracy of detection and molar mass determination. One advantage of mass spectrometric methods is the possibility to couple different chromatographic separation systems on-line to mass analysers, which enables additional preparative and analytical procedures, reduced analysis times and enhanced throughput of analyses. Another advantage is the possibility to study individual analyte fragments in more detail through multiple successive steps of fragment selection and further fragmentation using tandem mass spectrometry.

In the following sections the development of the past decade in mass spectrometric analyses of different fields of structural analyses is described.

5.3.1 Sequencing

The sequencing of RNA with MS is based on analysis of fragments of the original molecule produced by 1) enzymatic digestion, 2) fragmentation upon ionization or 3) collisionally induced dissociation of a selected precursor ion using tandem MS techniques.

Foundations were laid in the studies of McLuckey *et al.*[53] where the gas phase dissociation mechanisms of oligodeoxynucleotides were investigated by ESI-QITMS. Under low energy CID the backbone cleavage of an oligonucleotide chain occurred at four possible sites across the phosphodiester bond. The dissociation is usually initiated by a loss of a nucleobase, leading to the cleavage of the 3´ phosphodiester bond of the nucleotide, where the nucleobase was lost. Single cleavage produces two complementary ions containing the original termini. These findings led to characterization of four sets of complementary ion series and proposal of nomenclature for these series, as illustrated in Figure 17.

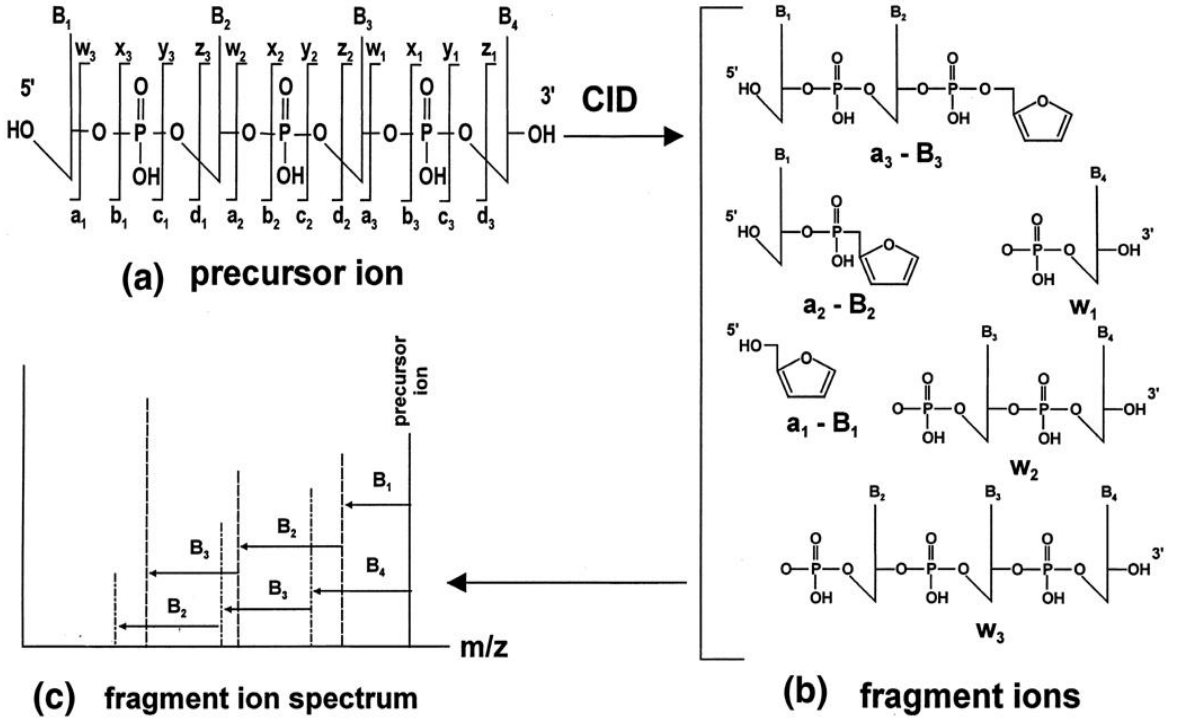


Figure 17. A schematic presentation of ion series produced by collisionally-induced fragmentation of oligonucleotides.[100](a) Notation and characterization of ion series of *a-d* starting from the 5´-end and *w-z* starting from the 3´-end of the oligonucleotide chain. (b) Products of complementary ion series *a* and *w*. Term "a_n-B" denotes the loss of a nucleobase from an "*a*"-type ion. (c) Spectrum showing peaks formed from cleavages of the precursor ion.

The fragments containing 5´ terminus are designated as ion types of *a-d* and the fragments containing the 3´terminus are designated as ion types of *w-z*. Thus, the complementary ion pairs are *a/w*, *b/x*, *c/y* and *d/z*. The subscripted numbers indicate the number of bases belonging to that particular ion, numbers being counted from the corresponding terminus. The nucleobases are designated as B with the numerical subscript, referring to the position of the base in the sequence counted from the 5´ terminus. The proposed nomenclature is widely practiced in current reports of mass spectrometric studies.

A few studies of the fragmentation mechanisms of RNA show that the nucleobase loss is not a prerequisite for backbone dissociation[66, 67, 100] and the production of c-ions and their complementary y-ions is preferred. It has been shown that oligoribonucleotides are more stable in gas phase than oligoribonucleotides because the 2´-OH bond has a stabilizing

effect on the N-glycosidic bond.[66]

The general strategy of the determination of nucleobase sequences based on CID spectra comprises following steps: 1) identification of fragment ions belonging to the complementary ion series; 2) assignment of general base compositions to these fragments on the basis of measured masses; 3) deduction of the order of the assigned nucleobases in individual fragments and finally 4) alignment of fragments of complementary ion series in correct order. Assigning base compositions and base sequences within separate fragments is based on incremental fitting of individual nucleotide masses to match the measured masses of the fragments in question, taking into account the different ion types and number of other constraints applicable. For example, prior knowledge of the identities of terminal bases arising from the treatment of target oligonucleotide with base-specific enzymes may be used.

Iterative algorithms are used in sequence determination. Calculations tend to be too labourious and time-consuming to be performed manually: computer software for assigning base sequences from spectra produced with CID has been designed. On the basis of provided data of measured spectrum and other input parameters, programs assign best-fitting base-sequences to fragments displayed in the spectrum and assemble the total sequence from the fragments. Laboratories may write their own software suitable for their needs[58, 88] and may be available upon request[62] ,but programs are also freely available to be used via internet [108] or to bedownloaded from internet sites[109].

Sequencing software operate with data files of the measured spectrum containing mass and abundance values of the peaks found in the spectrum. Additional basic input data needed is the type of nucleic acid, the charge state, molar mass and nature of the 3´ and 5´ termini (whether a phosphate or hydroxyl group is present) of the precursor ion and allowable error tolerance of mass determinations in *m/z* units. Other prior knowledge such as enzymes used in digestion and identity and location of posttranscriptional modifications (in sugar, phosphate or nucleobase moiety) may be provided to assist narrowing down the number of candidate sequences. Accumulation of sequence candidates with isomeric nucleobase composition restricts the unambiguous sequence determinations to oligomers of 15-20 nucleotides length. A schematic representation of an algorithm of a certain sequencing program is displayed in Figure 18.

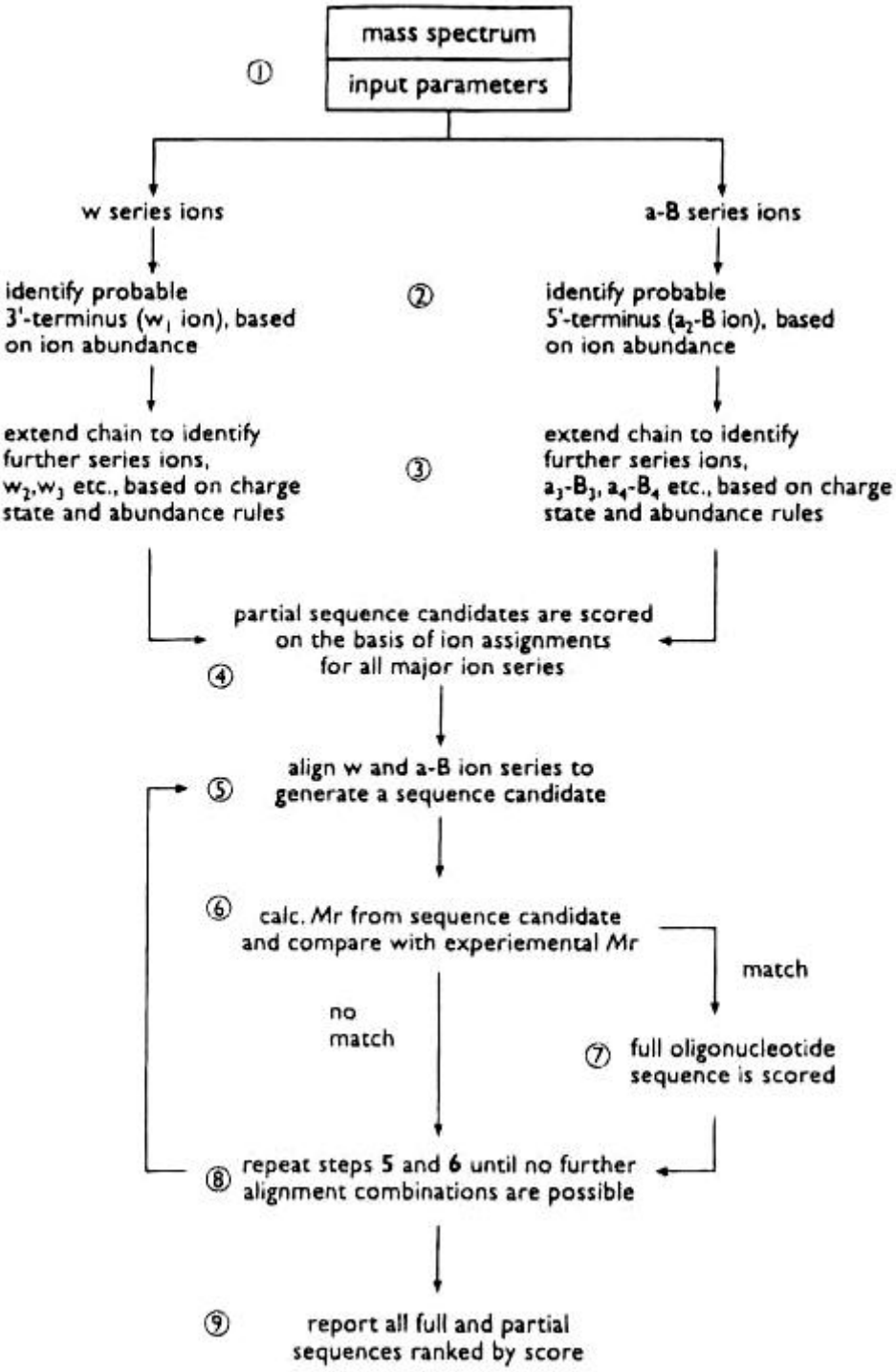


Figure 18. A flow chart presentation of an example of an sequencing algorithm.[58]

Few analyses of RNA have been conducted using post source decay or nozzle-skimmer dissociation as a means of fragmentation. Berhane and Limbach[38] used PSD fragmentation in conjunction of MALDI-TOFMS analysis of RNAse T1 digestion product of *E. coli* tRNA molecule. Kirpekar *et al.*[79] used the method to determine the base sequence and a location of a modified nucleotide in a fragment of 5S rRNA. Meng[65] studied nozzle-

skimmer dissociation as high-throughput alternative for the conventional collision-gas-induced dissociation methods. Nozzle-skimmer fragmentation proceeds without precursor ion selection, which potentially leads to higher sensitivity of analysis. Method was easily implemented for sequencing short (5-6 -mer) oligonucleotides of single species. A mixture of two 5-6 nucleotides long oligonucleotides was also resolved successfully, but more complex mixtures resulted unresolved, overlapping spectra of different oligonucleotide species. An example of a spectrum produced by nozzle-skimmer fragmentation in ESI-MS is presented in Figure 19.

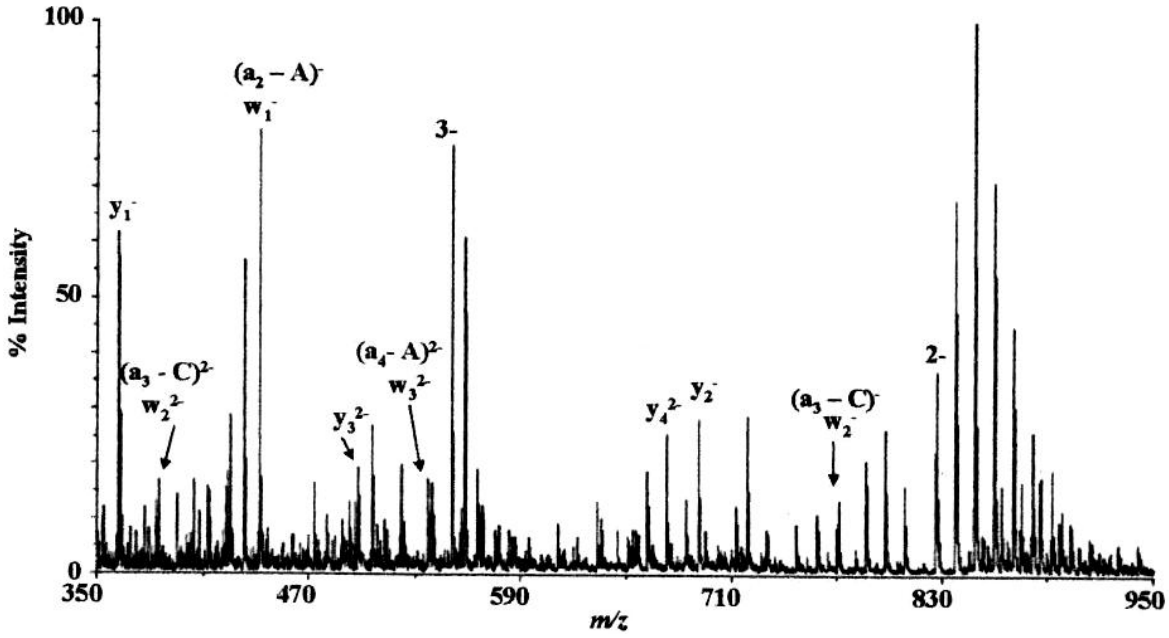


Figure 19. ESI -TOFMS spectrum of nozzle-skimmer fragmentation products of synthetic oligoribonucleotide 5´-AAGACp-3´. Analytes were dissolved in water and infused using syringe pump through in-house pulled fused silica capillary microemitters (360 µm o.d. by 50 µm i.d., tip i.d.15 µm) Nozzle potential was 140 V. No further information concerning experimental conditions of analysis with TOF analyser is given.[65]

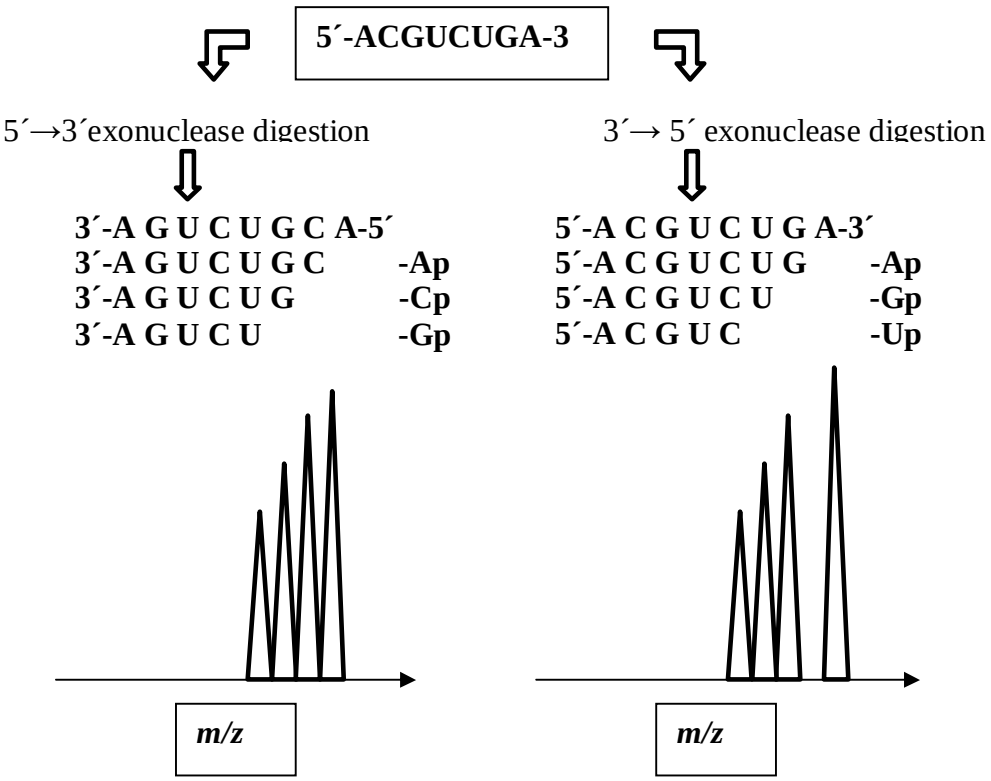
More easily readable sequence-determining spectra became available when a method called *ladder sequencing* using MALDI-TOFMS was introduced. The principle of method was first described by Pieles *et al.*[54] The method was developed for the sequencing of DNA, but same protocols can be applied to RNA as well. In this method, the oligonucleotide to be sequenced is degraded stepwise using phospodiesterase enzymes, which hydrolyse phosphodiester bonds sequentially starting from the either end of the oligonucleotide chain, depending on the specifity of the enzyme in question. Such enzymes

are also referred as exonucleases in distinction to base-specific enzymes, which cleave phosphodiester bonds from the specific sites across the oligonucleotide chain. However, the mechanism of hydrolysis of phosphodiester bond is the same. Table 8 presents commonly used exonucleases in enzymatic ladder sequencing.

Table 8. Common exonucleases in RNA analyses.

Enzyme	Action	Products	Reference
Bovine spleen phosphodiesterase (BSP)	hydrolyses phosphodiester bonds without base specificity, starting from the 5´end	3´ nucleotides	57, 59
Calf spleen phosphdiesterase (CSP)	hydrolyses phosphodiester bonds without base specificity, starting from the 5´end	3´ nucleotides	41, 54
Snake venom phosphdiesterase (SVP)	hydrolyses phosphodiester bonds without base specificity, starting from the 3´ end	5´ nucleotides	54, 57, 59, 86

During hydrolysis, the chain is truncated one nucleotide at a time, proceeding until total hydrolysis of the oligonucleotide to monomers has occurred. A set of fragments, where each product is one nucleotide shorter than the previous, is obtained by taking aliquots of the reaction mixture at a specific time intervals and quenching the reaction. Hydrolysis is usually performed to both directions along the oligonucleotide chain in two separate reactions, where an enzyme with different specification for direction is used. After the mass analysis of the fragments, the base sequence can be determined from the mass differences between adjacent fragment peaks. The principle of exonuclease sequencing is presented in Figure 20.



	Nucleotide leaving group			
	Gp	**Ap**	**Cp**	**Up**
Mass (Da)	345.20	329.20	305.18	306.17

Figure 20. Principle of exonuclease sequencing. An oligonucleotide is subjected to digestion from the 3´ end with snake venom phosphodiesterase and separately from the 5´end with bovine spleen phosphodiesterase. The identity of each base residue is determined from the mass difference between adjacent peaks. Adapted and modified from a paper by Smirnov *et al.*[110]

Resulting two overlapping sequences can be aligned to complete and confirm the total sequence. The unambiguous determination of the sequence may not always be possible due to the small mass difference between cytosine and uridine (1 Da). The resolution capacity of TOFMS instruments is not sufficient to resolve digestion products, thus sequence ambiguities are encountered. An example of a spectrum from an exonuclease digestion

experiment is given in Figure 21.

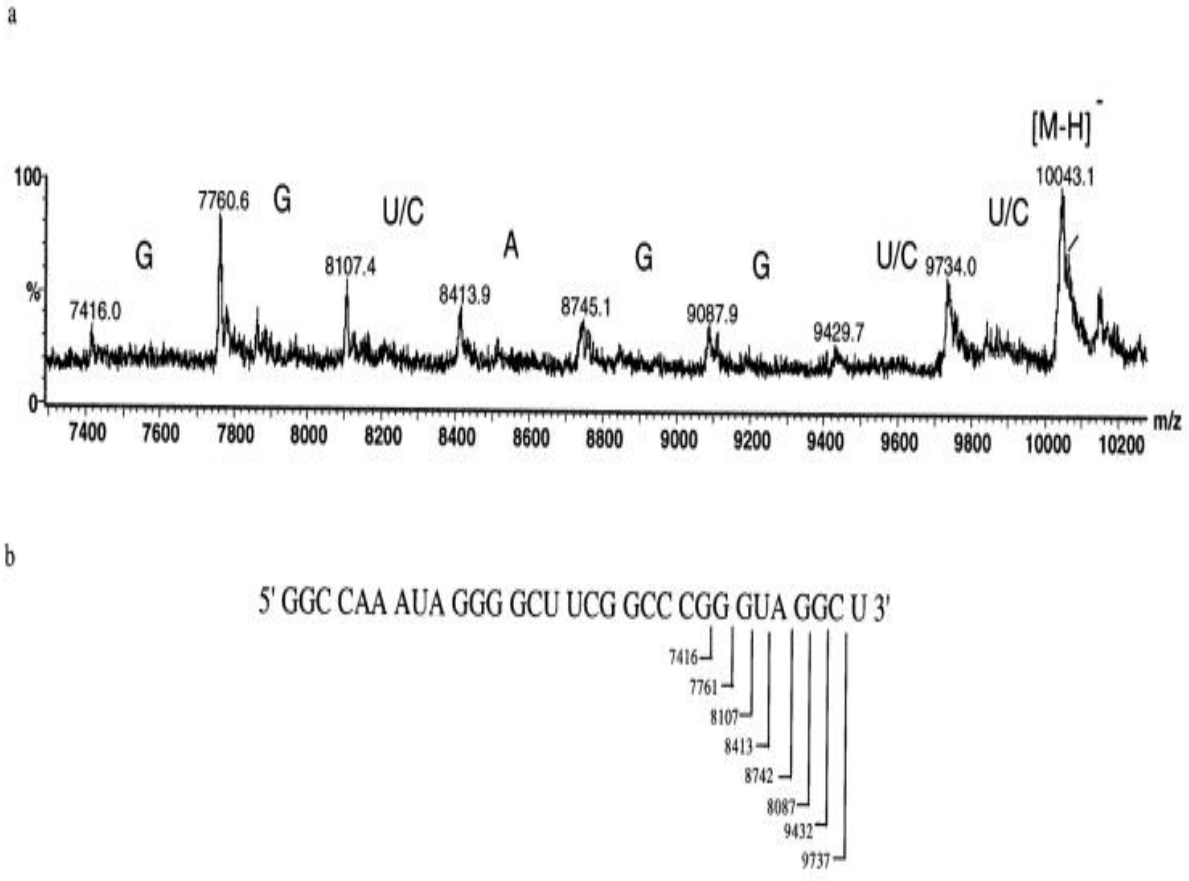


Figure 21. (a) Spectrum of a synthetic 31-mer oligoribonucleotide after two minutes of digestion with a snake venom phosphatase, analysed using UV-MALDI-TOFMS with delayed extraction and in the negative ion mode. MALDI matrix: 75 mg/ml 3-HPA in ACN: water (1:1, v/v). Acceleration voltage: 22 kV. Mass differences arising from the cleavages of U and C cannot be differentiated and are indicated in the peak notation as U/C. (b) Schematic representation of fragments observed in the spectrum.[59]

Mass ladders may be constructed also in traditional fashion using limited alkaline hydrolysis.[39, 41, 68]

Fragment ladders produced by traditional method of combination of different base-specific endonucleases may be analysed with MALDI-TOFMS as well. An example of a spectrum of a sequencing experiment using endonuclease digestion is presented in Figure 22.

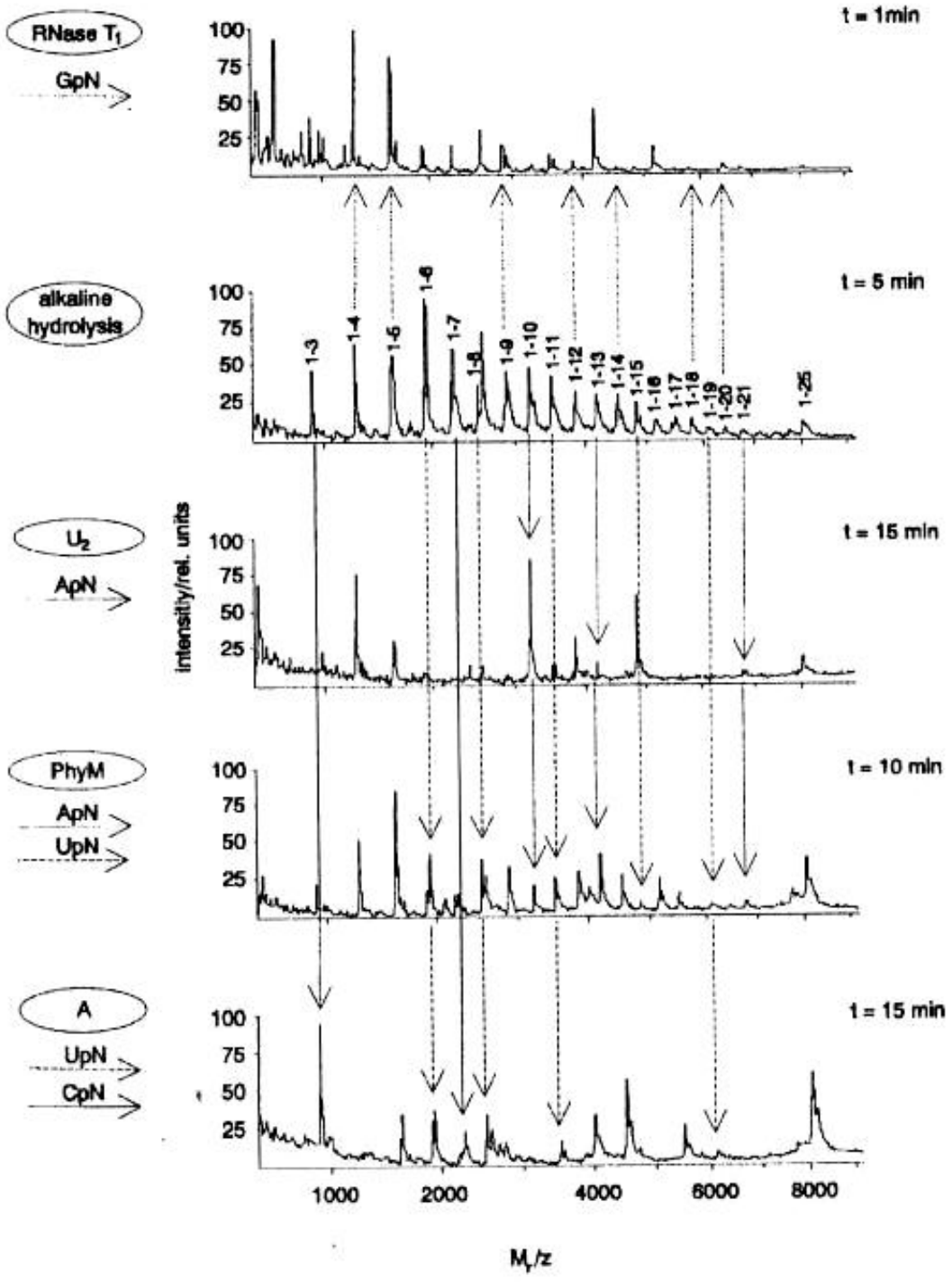


Figure 22. UV-MALDI-TOFMS mass spectra of a 25-mer synthetic RNA digested with a combination of endonucleases.[39] Minute values on the right refer to duration of digestion. Analyser was used in reflector mode and ions were detected in positive mode. Laser type: Nd:YAG operating at 335 nm. Pulse width: 5 ns. Acceleration voltage: 22 kV. MALDI matrix: 3-HPA.

As mentioned earlier, the enzymatic digestion with T1 proceeds through an intermediate product of cyclic phosphate. The reaction does not always proceed to completion, resulting accumulation of intermediate products and increased spectral complexity**.** To improve differentiation of actual digestion products from unspecific products and artefacts, Berhane and Limbach[38] developed a method for isotopic labelling of T1 digestion products during enzymatic hydrolysis. The hydrolysis is performed in water, where 50 % of the oxygen is replaced with a radioactive isotope ^{18}O. During hydrolysis, the radioactive oxygen is incorporated into the phosphate group forming to the 3´ end of the digestion product. The labelled digestion products can be detected from the mass spectra from the presence of doublet peaks separated by 2 *m/z* units. In addition, doublet peaks indicate fragments containing original 3´ termini, providing aid to sequence assignments**.**

Several approaches to overcome the difficulty to distinguish between uracil- and cytosine-containing nucleotides has been suggested, but the problem has remained to this day Faultisch and his co-workers[57] differentiated fragments containing 5´ terminal C from those containing 5´ terminal U by their different intensities of spectrum peaks. The greater relative abundance of C-terminal fragments is a result of different rate of phosphodiester bond hydrolysis. Tolson and Nicholson[59] used the traditional method of derivatisation of uracil bases with hydrazine, followed by a strand scission at the site of modification with a reaction with aniline.

Spottke and her co-workers circumvented the problem of small difference between uridine and cytidine by developing a "reversed version" of Sanger sequencing method, where sets of fragments, all ending to one of the four common nucleobases, were produced in separate reactions and analysed individually.[64] The sequence could be read from the combined and overlaid spectra of the separate analyses. This approach enabled the assignment of identities to each peak in the spectrum without ambiguities between uracil- and cytidine-containing fragments, because they were produced in separate reactions.

Aforementioned method employed an *in vitro* produced copy of a region of RNA of interest, where a binding site of a T7 polymerase was ligated. The enzyme was used to produce full-length complementary RNA copies in four separate enzymatic transcription reactions. Each reaction mixture contained a phosphothioate analog of one of the normal nucleotides, which was incorporated in random fashion into the product being synthesised instead of the normal counterpart. After the replication reactions, each set was digested using an exonuclease, which cleaved the chain in 3´→5´ direction until was inhibited by the modified analog. Sets of products after enzymatic digestion were then subjected to

MALDI-MS and resulting spectra overlayed. The sequence of the oligonucleotide could then be read from the aligned spectrum image. An example of a sequence determination using this method is displayed in Figure 23.

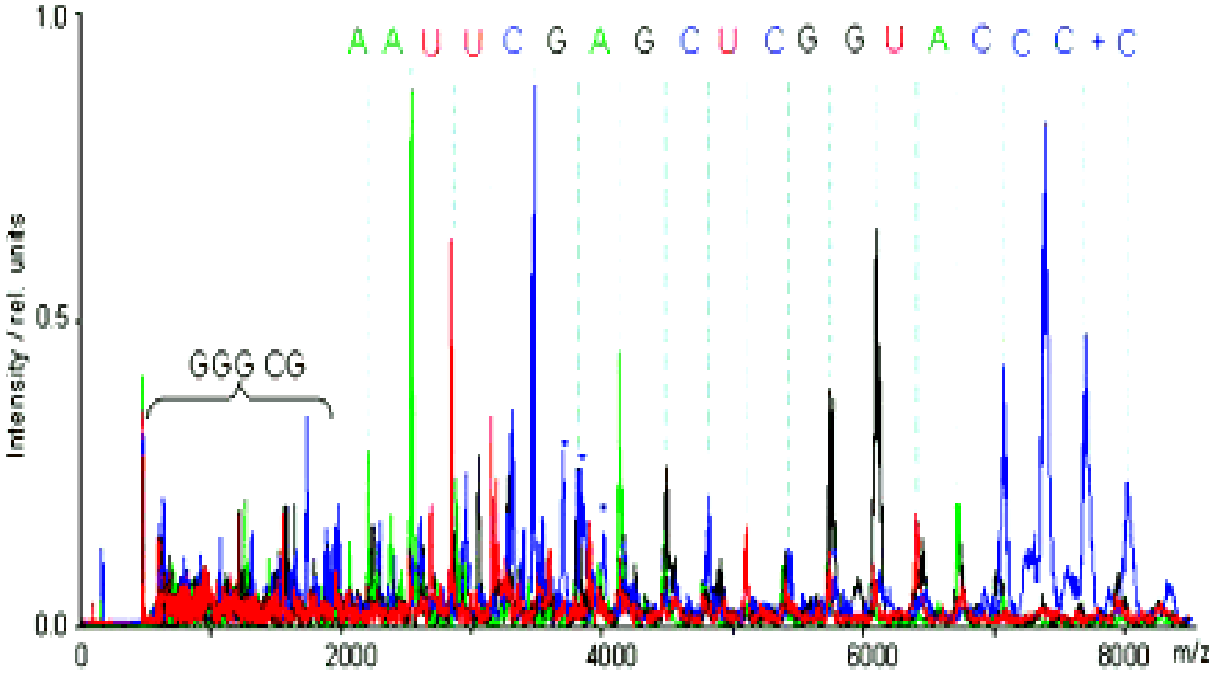


Figure 23. A composite image of MALDI-TOFMS spectra from four separate transcription and digestion experiments using a method of "reverse Sanger sequencing".[64] Nitrogen laser operating at 337 nm was used. Spectra were acquired in positive ion mode. Ions were accelerated to 25 kV by delayed extraction. MALDI matrix: ternary mixture of 0.2 M 2, 3, 4-THAP in 50% ACN:0.2 M 2, 4, 6-THAP in 50% ACN:0.3 M diammonium citrate (1:2:1, v/v/v).

Method suffers from the generation of by-products during the *in vitro* transcription of RNA template. To overcome this, the group extracted the target products from the reaction mixture by capturing them to magnetic beads coated with oligonucleotides with regions of complementary sequence.

5.3.2 Posttranscriptional modifications

Kowalak *et al.*[78] developed a two-phased LC/MS approach, which is based on mass spectrometric analysis of fragmentation products of total and partial enzymatic digestions. In the first step, RNA is digested completely to individual nucleosides using a combination of phosphodiesterase and alkaline phosphatase treatments. The pool of nucleosides is fractionated by RPLC and the identities of nucleosides are determined using catalogued values of retention times[111] and information of mass spectra.[78] Both modified and unmodified nucleosides can be seen in the chromatogram and relative molecular ratios can be calculated from the peak areas.[17] An example of an chromatographic separation of nucleosides is presented in the Figure 24.

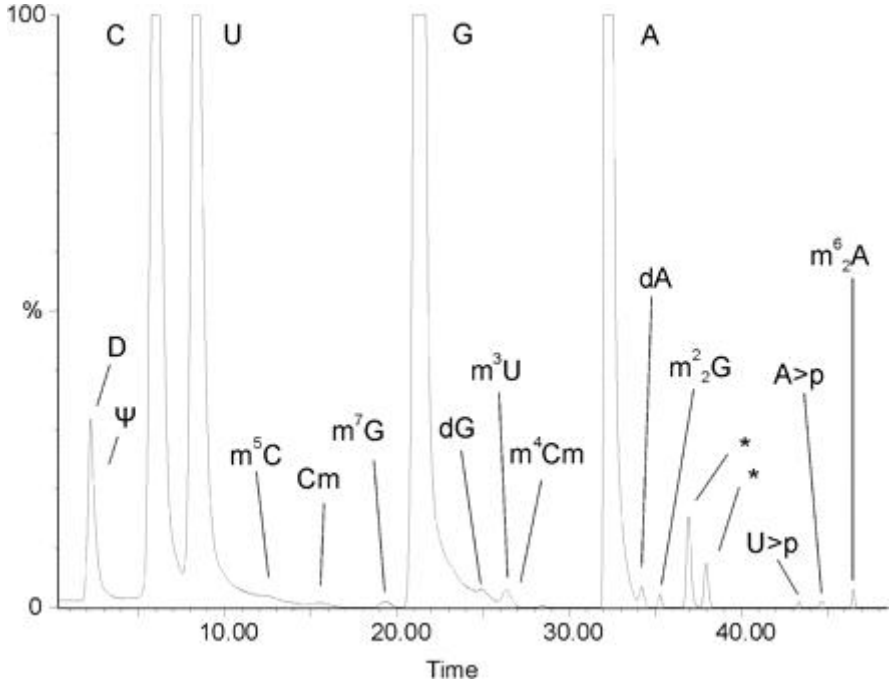


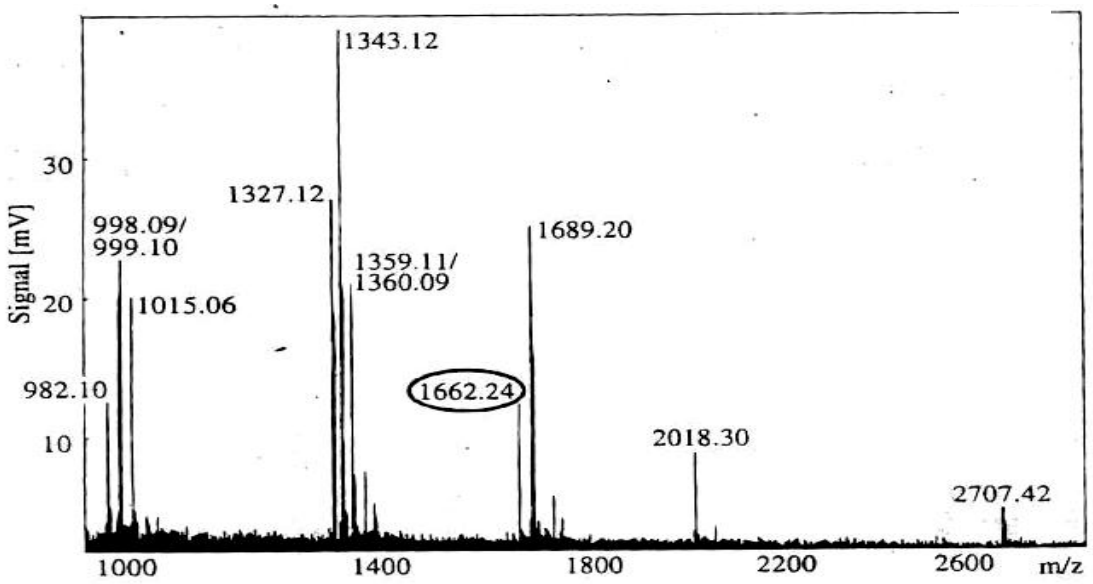
Figure 24. HPLC chromatogram of pool of nucleosides after total enzymatic hydrolysis of *Clostridium acetobutylicum* 16S rRNA.[86] Annotation of modified nucleosides according Crain, Roszenski and McCloskey[12], except A>p and U>p denoting cyclic phosphate. Deoxyadenine (dA) and deoxyguanine (dG) nucleotides are present probably as artefacts of reverse transcription of the *in vitro* synthesis of target RNA. Chromatographic conditions: Atlantis dC18 column (0.32 x 150 mm, 3 µm particle size), gradient elution with 50 mM NH_4OAc (pH 6.0)/ACN buffer, at flow rate of 5 µL/min. Detection: absorbance at 240-300 nm using diode array detector.

In the second phase RNA is partially digested using site-specific endonuclease, most commonly T1. Fragments are separated in RPLC and fragment masses determined using mass spectrometry. By comparing the determined fragment masses of the sample to the predicted fragment masses calculated on the basis of the sequence of unmodified RNA**,** fragments harboring modifications can be detected from the increase of fragment mass. Finally, the exact location of the modified bases can be determined by tandem mass spectrometric sequencing of target fragments.

Dihydrouridine lacks suitable chromophore for UV detection, but it can be detected in mass analyser using selected ion monitoring of *m/z* 247, representing MH^+ form of the molecule.[18] Pseudouridine can be detected after incorporation of specific mass labels prior to digestion and mass analysis or using selected ion monitoring of pseudouridine-specific fragmentation products in tandem mass spectrometry. Kirpekar *et al.* used cyanoethylation of pseudouridines to introduce 53 Da increases to pseudouridine bases.[83] Pomerantz *et al.*[84] used nozzle-skimmer dissociation and monitoring of production of characteristic ions.

Technically more straightforward, although somewhat more limited analyses can be performed using techniques based on MALDI ionization, where fragment separation step can be avoided. Kirpekar *et al.* .[79]developed a method often referred as "*RNAse mapping*" where the RNA sample is digested with a combination of two endonucleases T1 and A and is analysed in MALDI-TOFMS using post-source decay fragmentation. Cation adducts are removed prior the analysis in a treatment with beads coated with cation exchange material. Mass silent modifications, such as pseudouridines, cannot be detected readily. An example spectra of "RNAse mapping" experiment is shown in Figure 25.

A.



B.

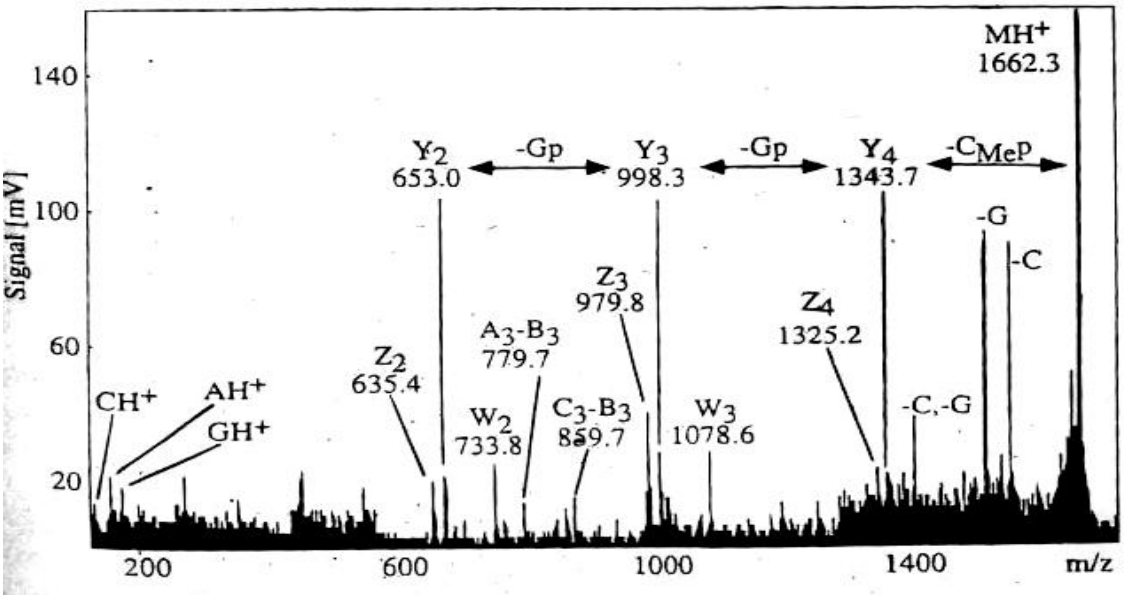


Figure 25. Analysis of posttranscriptional modification using MALDI-TOFMS[79]. The spectra in panels A and B were both recorded in positive ion mode, using reflector and delayed extraction but different MALDI matrices. **A**: Spectrum of fragments of *Sulfolobus* 5S rRNA produced with enzymatic digestion with ribonuclease A. The *m/z* value inside a circle denotes a peak of a fragment not expected from the gene sequence. MALDI matrix: 5 g/L 3-HPA in 50% ACN. **B**: MALDI post source decay spectrum of the unexpected fragmentation product. Nomenclature of identified ion series and nucleotide cleavages is according to McLuckey *et al.*[53] Losses of the nucleobases from the fragments are indicated by $-B_n$, where n denotes the position of the nucleobase read from the 5´end. Spectrum indicates the presence of a methylated cytidine ($C_{Me}p$) at the 5´ terminus. MALDI matrix: 0.2 M 2 ,4, 6-THAP in 50 % ACN:2, 3, 4-THAP in 50% MeCN: 0.3 M diammonium citrate (2:1:1, v/v/v).

Meng and Limbach used ^{18}O labelling of endonuclease digests in development of ESI-FT-ICRMS method for enzymatic mapping of modifications in RNA.[82] According to their findings, the use of isotopic labelling in conjunction with FT-ICRMS analysis at the 10 ppm mass accuracy level enables unambiguous sequence determination up to the length of 20 nucleotides, when there is no more than four methylated nucleosides per digestion product and the number of any nucleotides is constrained. Figure 26 illustrates the effect of ^{18}O labelling of the endonuclease digests.

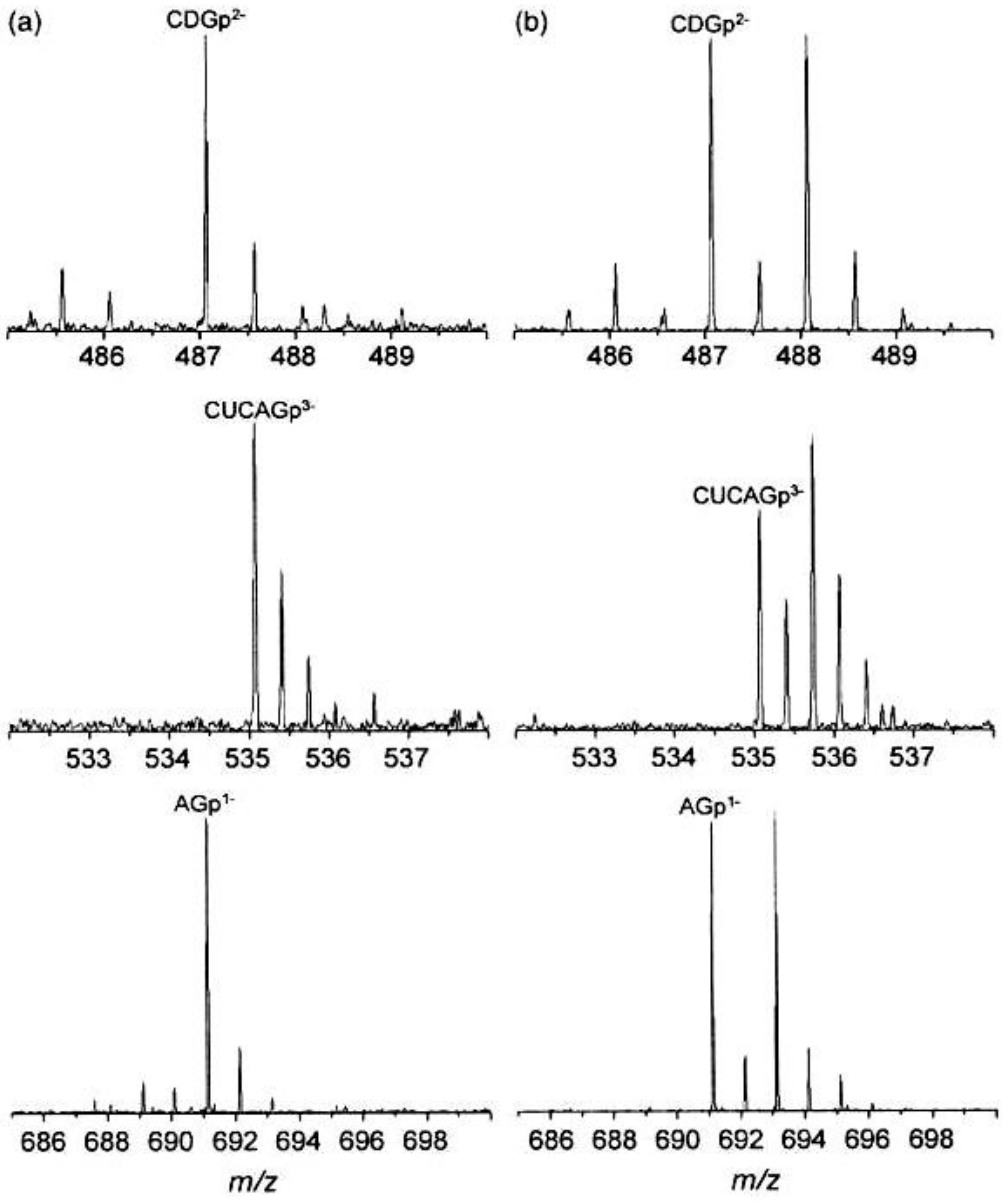


Figure 26. ESI-FT-ICRMS mass spectra of products of digestion with RNAse T1. (a) Hydrolysis without labelling. (b) Hydrolysis performed with ^{18}O labelling.[82] Analytes were dissolved in 50% 2-propanol and infused directly through microemitter (needle i.d. 30 µm) at a flow rate of 200-500 nl/min. Spectra were acquired in positive ion mode. ESI capillary voltage: 1.5-3.0 kV. Ion trapping voltage: -6 V. Quadrupole ion guide operated at 2.1 MHz. Strength of the magnetic field: 7.0 T. Ion excitation: broadband RF sweep for 4 ms. Data acquisition rate: 2 MHz, 512 K data points.

Specific regions of interest of target molecule can be isolated for the further studies by a combination of shielding of the area of interest and total hydrolysis of the remaining molecule.[81, 83] The shielding can be accomplished by hybridizing the target region of RNA with a synthetic complementary DNA oligonucleotide. The isolation of fragment of interest is then made either by total hydrolysis of the rest of the molecule using exonucleases, which cannot cleave double-stranded regions or which cleave specifically from the edges of double-stranded regions. The DNA-RNA hybrid is then separated from the remaining material using gel electrophoresis.

5.3.3 Higher order structures

Yu *et al.* developed mass spectrometric methods for investigation of secondary structures based on straight implementation of traditional methods employing structure probes.[87] Here, the RNA hairpins of the HIV-1 packaging signal were derivatised with structural probes CMCT, DMS and KT, digested using RNA:se T1 and the resulting fragments were analysed using ESI-FT-ICRMS. RNA was probed with each reagent separately producing set of individual probing experiments. Using the knowledge of base sequences of the whole RNA molecule and expected digestion products, the detection of fragments containing probes were accomplished by comparing the mass values of the digestion products measured before and after structure probing. The identity of probe and the modified base were determined using computer analysis of fragment masses.[108] If ambiguities were encountered, the identification of modified bases could be done by sequencing using SORI-CID fragmentation. The complete map of probed bases in the sequence was deduced by combining the data from separate probing experiments. The data confirmed that the regions of accessible sites of hairpin loops were completely derivatised by the probes. When the packaging signal was allowed to form complexes with the nucleocapsid protein p7, the binding of the probes was entirely inhibited.

This approach/method was further developed by Kellersberg *et al.*[88] so that the digestion phase was completely left out and the fragmentation of probed RNA was produced directly by SORI-CID and the sequence ambiguities resolved using MS^n techniques. Also new algorithm software called MS2Links was developed for spectrum analysis.

Yu and Fabris attempted to develop a method where the probing could be performed using all three aforementioned probes simultaneously in the same reaction mixture.[89] These mixtures were digested with ribonucleases T1 and A and the products were analysed using ESI-FT-ICRMS. Successful methods were accomplished for the binary combinations of CMCT+DMS and KT+DMS. In the combination of CMCT +KT, more efficiently binding CMCT reagent displaced adducted KT resulting absence of KT adducts in the spectrum. An example spectrum of the experiment is displayed in Figure 27.

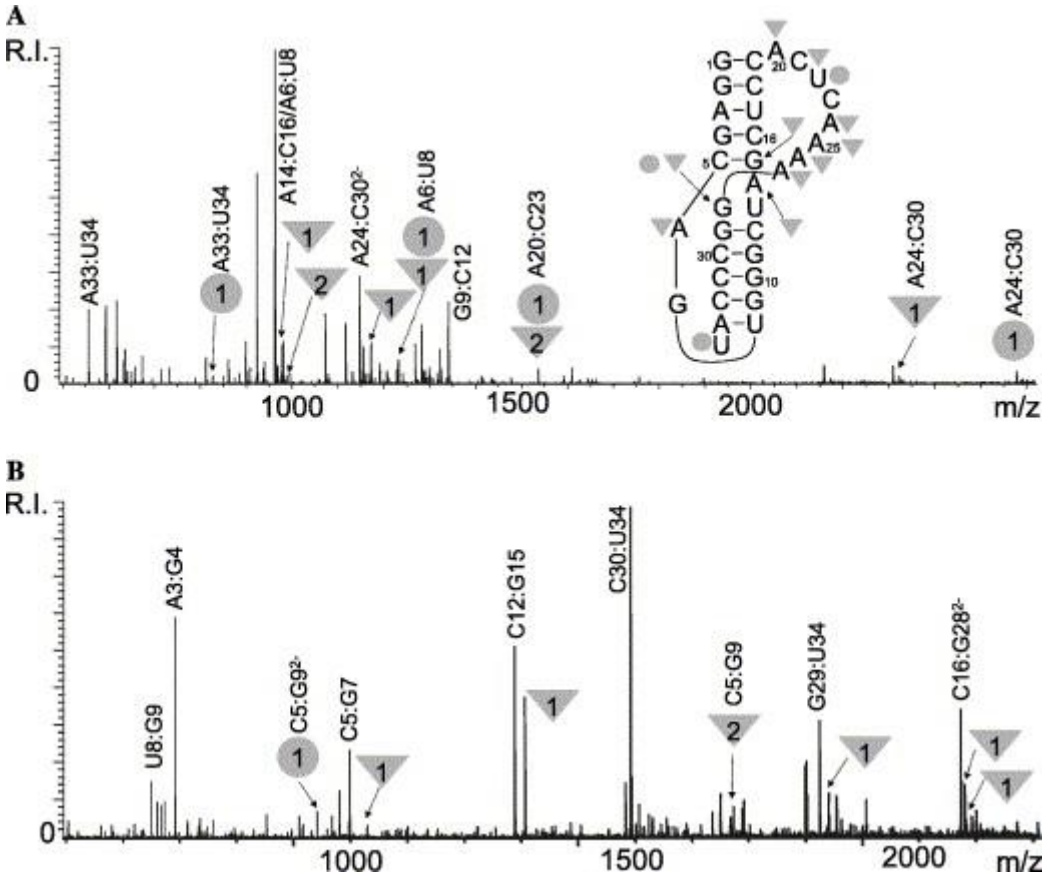


Figure 27. ESI-FTMS spectra of higher order structure probing experiment using combination of CMCT and DMS on mouse mammary tumour virus pseudoknot (VPK). Probes are labelled with (●) for CMCT and (▼) for DMS. Products of digestion are expressed by a region between indicated 3´- and 5´ -terminal bases separated by a colon (":"). The structure of the VPK is displayed in the inset. **A:** Products of digestion with RNase A. **B:** Products of digestion with RNase T1.[89] Analytes dissolved in 10 % 2-propanol were infused directly by a syringe pump at a flow rate of 2µL/min. Spectra were acquired in negative ion mode. The strength of the magnetic field of the analyser was 7.0 T. No further information concerning experimental conditions of analysis with FT-ICR is given.

Yu and her group developed also a method for studying three-dimensional structures of feline immunodeficiency virus ribosomal frameshifting pseudoknot (FIV-PK) using structural probes to map single-stranded regions and bis-(2-chloroethyl-)methylamine (*nitrogen mustard,* NM) as a bifunctional crosslinking reagent to investigate intramolecular spatial relationships between strands within the molecule.[90] NM forms a covalent bridge between guanine or adenine nucleobases located at average distances of 8 ± 3 Å. After

digestion with nucleases, crosslinked and probed fragments were analysed using ESI-FT-ICRMS mass spectrometry and spectrum interpretation software. Ions containing crosslinked fragments were recognised from the measured fragment masses equalling the sum of two crosslinked fragments and the crosslinker molecule. Crosslinkers may also form incomplete bridges when there is no other suitable nucleobase at suitable distance and thus have to be differentiated from actual complete crosslinked species.

Probing data was combined with information such as bond distances, torsion and dihedral angles obtained from a database of RNA structures to build a three-dimensional computer-aided molecular model of the pseudoknot structure. Figure 28 presents a schematic model of the FIV-PK and sites of adduction by structure probes. Products of the ESI-FT-ICRMS spectrum of probing analysis of FIV-PK are shown in Figure 29.

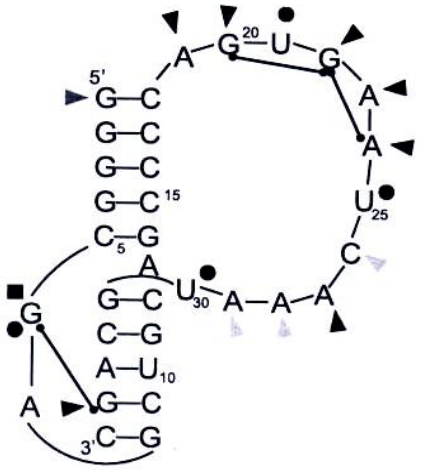
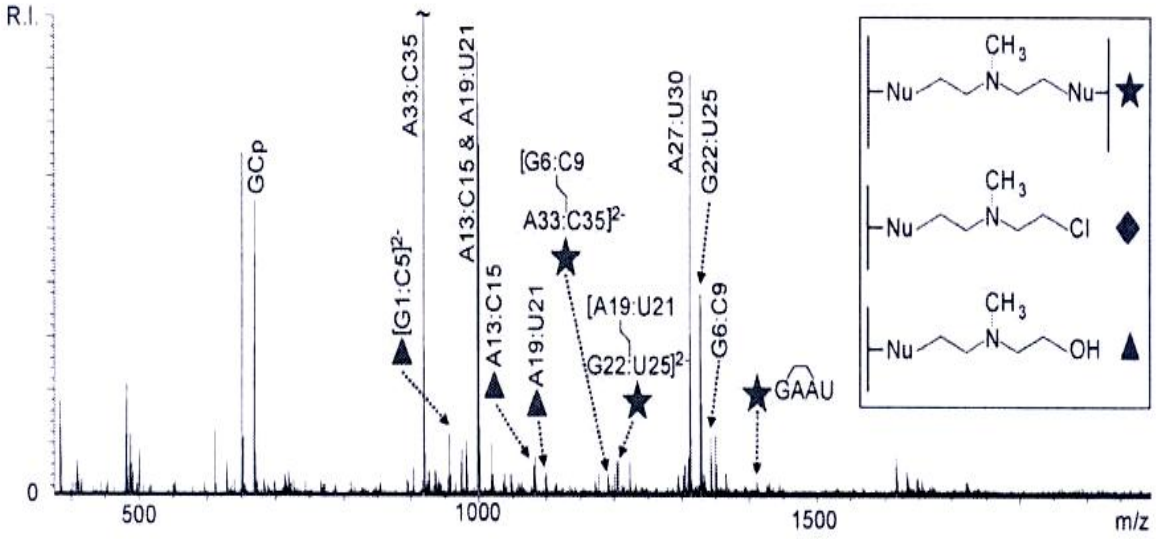


Figure 28. Secondary structure of FIV-PK and sites of modification by DMS (◄), CMCT (●), KT (■) and NM crosslinks (●—●). Modified image from studies of Yu and colleagues.[90]



Crosslinked nucleobases	Sequences of the crosslinked species[a]	Expected Mass[b]	Observed Mass[c]
G6-G34	**G**AGC-A**G**C	2342.46	2342.81
G20-G22	A**G**Up-**G**AAUp	2408.40	2408.63
G22-A24	**GAA**Up	1410.26	1410.32

[a] Boldfaced nucleobases indicate actual crosslinked nucleobases.

[b] Mass calculated on the basis of the composition of nucleobases in digested fragments and the mass of the NM crosslinker.

[c] Mass of a matching fragment observed in the mass spectrum.

Figure 29. ESI-FT-ICRMS spectrum of products of RNAse A digestion of FIV-PK after NM crosslinking. The colon symbol (:) denotes a range of nucleotides. The structures of mono-adducts and bifunctional adducts are shown in the inset. Observed crosslinked species are listed below the spectrum. Analytes dissolved either in 10 mM NH_4OAc:IP (1:1, v/v) or IP:CDTA (2:1, v/v) were infused directly into the ESI source by a syringe pump at a flow rate of 2 µL/min. Spectra were recorded in negative mode. The strength of the magnetic field of the analyser was 7.0 T. No further data of settings of FT-ICR is given. Modified image from studies of Yu and colleagues.[90]

Yu and her team investigated several candidate crosslinkers using di- and trinucleotide RNA targets [91] in order to expand the selection of crosslinkers with different properties,

such as binding distances and specifities towards different nucleobases. The adduct stability was also studied. Finally, HIV-1 virus stemloop SL1A was used as a target in the actual experiments. Results showed that a few crosslinkers are available for adenine, guanine and cytosine, while there are no convenient crosslinkers for uridine and thymine. Table 9 lists some characteristics of the crosslinks used in the experiments. Molecular structures of the crosslinkers are displayed in Figure 30.

Table 9. Characteristics of selected bifunctional crosslinking agents. Adapted with modifications from the paper by Zhang *et al.*[91]

Name[a]	Base specificity[b]	Δm[c] (Da)	Crosslinking distance (Å)[d]	Probe:substrate[e]	Reaction conditions[f]	Adduct stability[g]
BKT	G	230.079	4.4–6.4	5:1	20mM ammonium borate (pH 8.0), 37°C, 2h	+
PDG	G	190.027	7.0–8.0	5:1	20mM ammonium borate (pH 8.0), 37°C, 2h	+
CPT	G>A>C	225.995	3.2	2:1	10mM ammonium cacodylate pH 6, 37°C, 24h	+++
NM	G>A>C	83.073	8.0–11.0	9:1	10mM ammonium acetate/cacodylate (pH 8.0), 37°C, 2h	++
CHB	G>A>C	231.126	8.0–11.0	9:1	10mM ammonium acetate/cacodylate (pH 8.0), 37°C, 8h	++
sTT	G	110.962	2.4	60:1	50mM ammonium borate, (pH 6.0), 37°C, 2h	++

[a] Abbreviated name:
BKT: Ethylene glycol bis[3-(2-ketobutyraldehyde) ether] (bikethoxal)
PDG: 1, 4-diacetalbenzene (phenyl diglyoxal)
CPT: *cis*-diamminedichloroplatinum (II) (cisplatin)
NM*:* 2-chloro-N-(2-chloroethyl)-N-methyl-ethanamine (nitrogen mustard)
CHB: 4-[bis(2-chlorethyl)amino]benzenebutanoic acid (chlorambucil)
sTT: 2,4,6-trichloro-1,3,5-triazine (cyanuric acid)
[b] Target nucleobases in decreasing order of affinity
[c] Increment of mass upon adduction of the crosslinker
[d] Distance or range of distances between crosslinked regions
[e] Experimental probe:substrate ratio
[f] Experimental reaction conditions
[g] Estimated stability at the experimental conditions: +++ very stable; ++ stable; + partially reversible

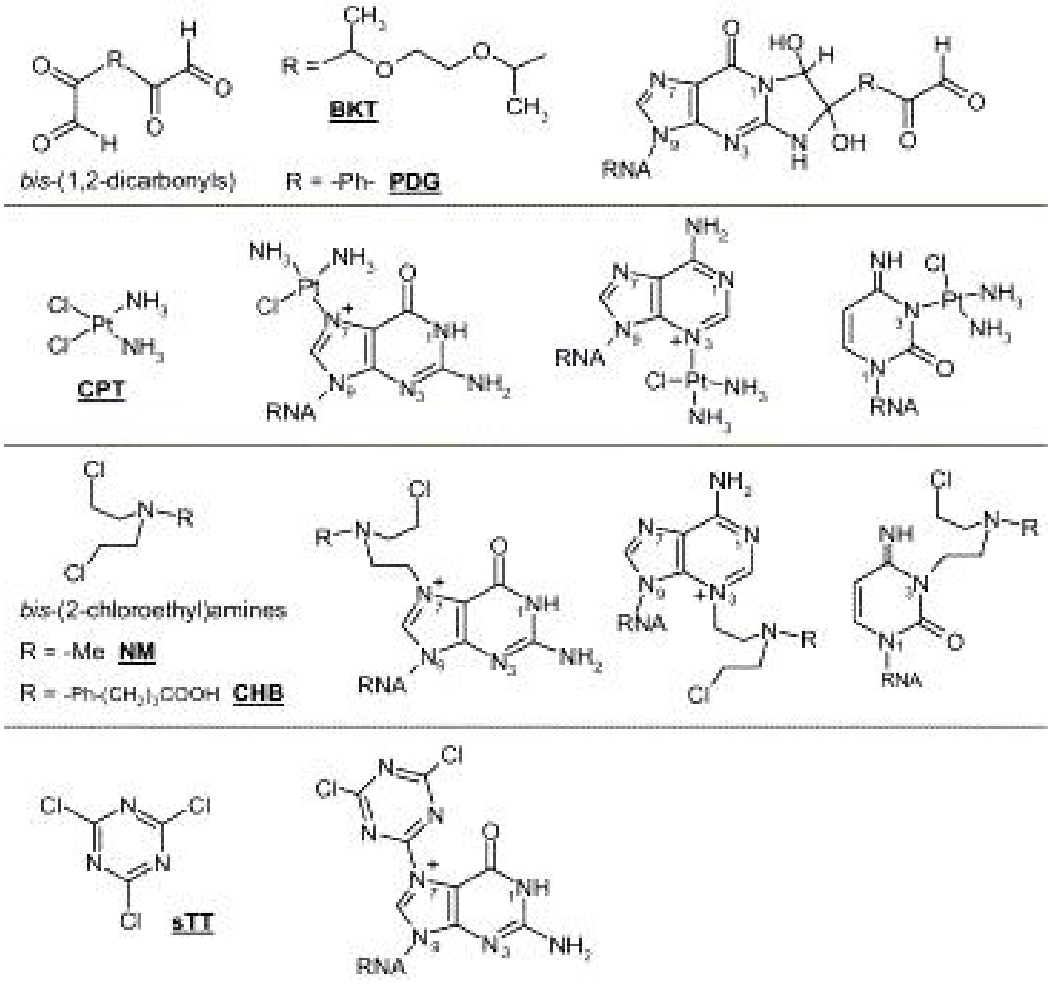


Figure 30. Molecular structures of selected intermolecular crosslinkers and sites of adduction at target nucleobases. **BKT:** bikethoxal; **PDG:** phenyl-diglyoxal; **CPT**: cisplatin; **NM**: nitrogen mustard; **CHB**: chlorambucil; and **sTT**: sym-triazine trichloride.[91]

5.4 Noncovalent interactions

Simple detection of RNA-RNA, RNA-protein or other RNA-ligand complexation[46, 69-73] may be accomplished in straightforward manner by ionizing and measuring masses of single components and formed complexes. Complexes are allowed to form simply bringing trial counterparts together in a solution and incubating the solution sufficient period of time to ensure complex formation. Masses of peaks in the spectrum equalling sum of counterpart analyte masses verify the complex formation. Multiple ligand binding may be detected in similar manner. The affinity of binding molecule to a target molecule can be calculated from the relative intensities of the peaks of the complexes and individual components. Figure 31 presents a spectrum of a complexing experiment with one polypeptide molecule and two species of RNA molecules.

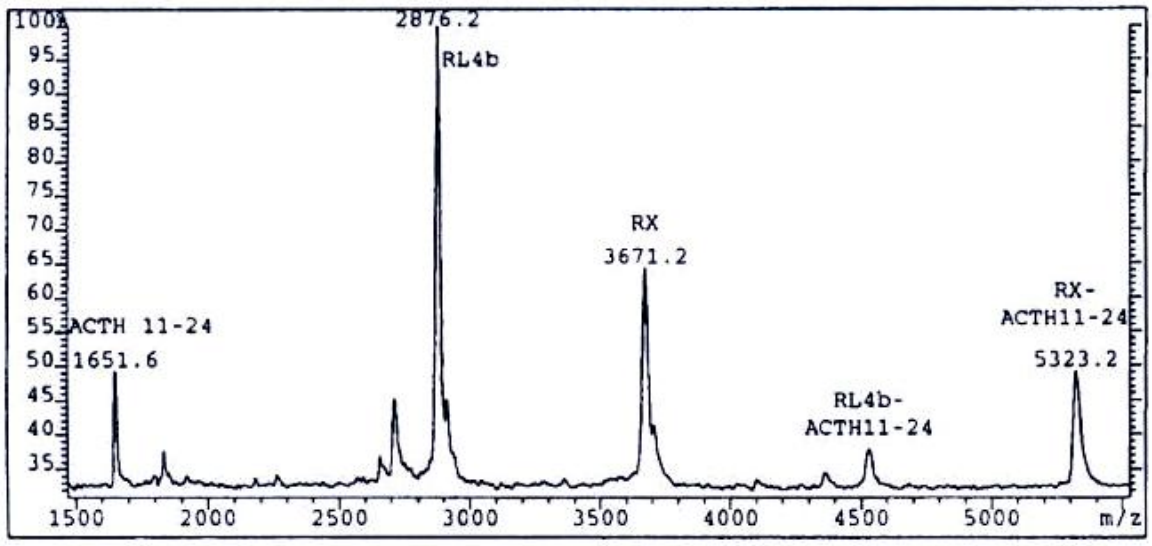


Figure 31. MALDI mass spectrum of a complexation of human adrenocorticotropin peptide (ACTH 11-24) with two species of synthetic structured RNA:s (RX and RL4b). The individual peaks of the peptide and the two RNA:s as well as their complex molecules are indicated.[69] The spectrum was recorded in negative ion mode with nitrogen-type laser. MALDI matrix: mixture of 20 mg/ml 2, 4, 6-THAP:0.2 M ammonium citrate (2:1, v/v). Acceleration voltage: 22 kV.

Complex formation may be detected with MALDI-TOFMS or ESI-FT-ICRMS analyses directly from suitable solvent solutions without any prior purification[69, 71] or after off- or in-line chromatographic removal of salts and metal cations with the use of size-exclusion chromatography or ion exchange columns. [72, 73]

Hofstadler *et al.*[71] laid the basis for a multiplexed screening of complexes in the binding studies of five aminoglycoside antibiotics to synthetic prokaryotic 16 S rRNA in a simultaneous assay using ESI-FT-ICRMS. The identities of bound ligands were deduced from the calculated mass difference between the peak of the target RNA and the peak of complexes, when the masses of individual ligands were known. Figure 32 shows a spectrum of the results of the aforementioned binding experiment.

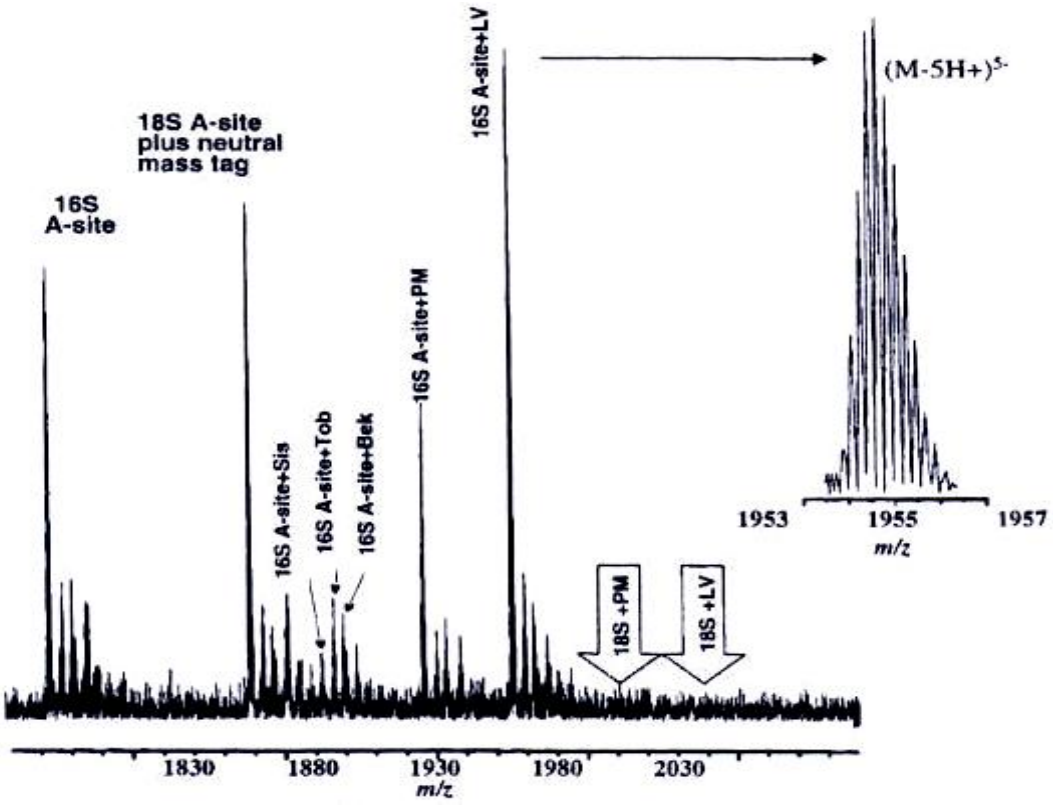


Figure 32. ESI-FT-ICRMS spectrum of a multiplexed screening of aminoglycoside components capable of binding to 16S rRNA, especially to a location known as the A-site. Polyethylene glycol (PEG) -tagged synthetic eukaryotic 18S rRNA construct was used as an indicator of toxicity of a trial substance in case formed complexes of 18S rRNA and the substance in question were detected. The PEG tag was attached to the 18S rRNA to increase the mass of the molecule and thus facilitate the identification of the control peak. Free 16S rRNA peak as well as the aminoglycoside-rRNA complex molecule peaks are indicated. Abbreviations used: Sis = sisomicin; Tob = tobramycin; Bek = bekanomycin; PM = paromomycin; LV = livodomycin. In the inset, the isotopic resolution of the complexes is demonstrated. Analytes dissolved in electrospray buffer composed of 50 mM NH_4OAc (pH 7,0):IP (1:1, v/v) were directly infused at a rate of 1.5 µL/min. Capillary voltage: 5 kV. Image adapted with modifications from the paper by Hofstadler *et al.* [71]

Gooding *et al.* [46] were able to screen approximately 300000 compounds against hepatitis viral RNA using a high throughput flow injection ESI-QITMS. By sequential complexation of target RNA with up to 10 pooled ligand molecules in same reaction mixture, screening of about 10000 compounds per day was possible. Using increasing concentrations of ligands, determination of the dissociation constants of the complexes was possible. Figure 33 presents spectra from a series of titrations of RNA of viral origin with increasing amounts of paromomycin, an antibiotic compound belonging to the group of aminoglycosides. The data was used to determine the dissociation constant of the binding reaction.

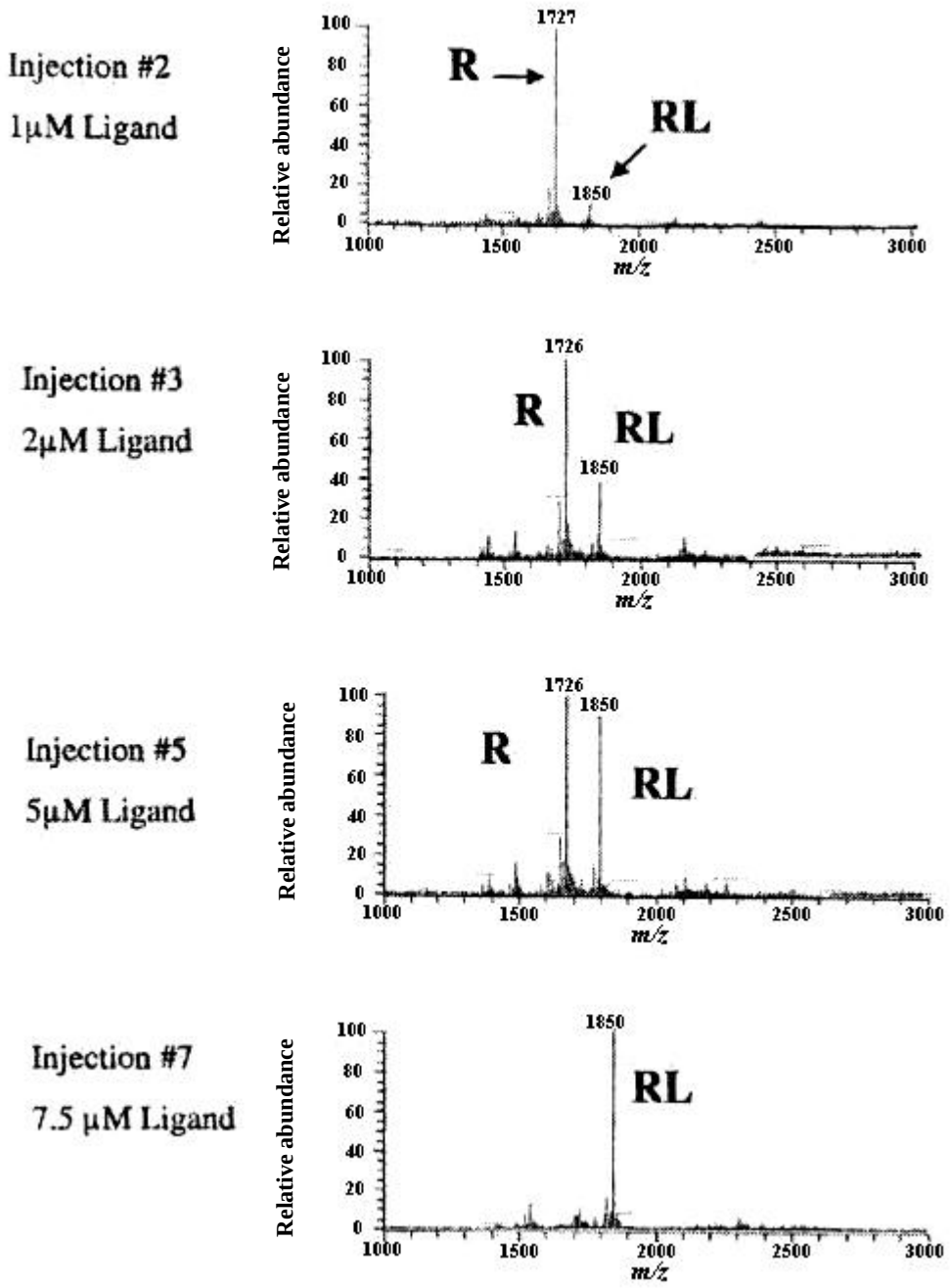


Figure 33. ESI-QITMS spectra of a series of complexation studies with a subdomain of a hepatitis C virus (HCV) RNA and increasing amounts of paromomycin. R= free target HCV RNA; RL= HCV/paromomycin complex. The signal of the complex intensifies and the target RNA diminishes as the concentration of the paromomycin increases. Analytes dissolved in buffer solution containing 50 mM NH4OAc in 50% 2-propanol were introduced to the ESI source automatically with flow injection method with the use of a Tecan robotic autosampler. 5 mM NH4OAc in 50% 2-propanol was used as a mobile phase, with a flow rate of 200 μL/min. Capillary voltage: 5 kV. No information regarding the instrument settings of the ion trap analyser is given. Image adapted with modifications from the paper by Gooding *et al.* [46]

Turner and his colleagues [75] used multiplexed screening assays with ESI-FTMS in a study, where dissociation of preformed complexes of HIV-1 nucleocapsid protein (NC) and the packaging signal RNA (Ψ-RNA) was investigated. Figure 34 displays spectra of a dissociation experiment with a complex of nucleocapsid protein, one RNA stemloop and two experimental compounds.

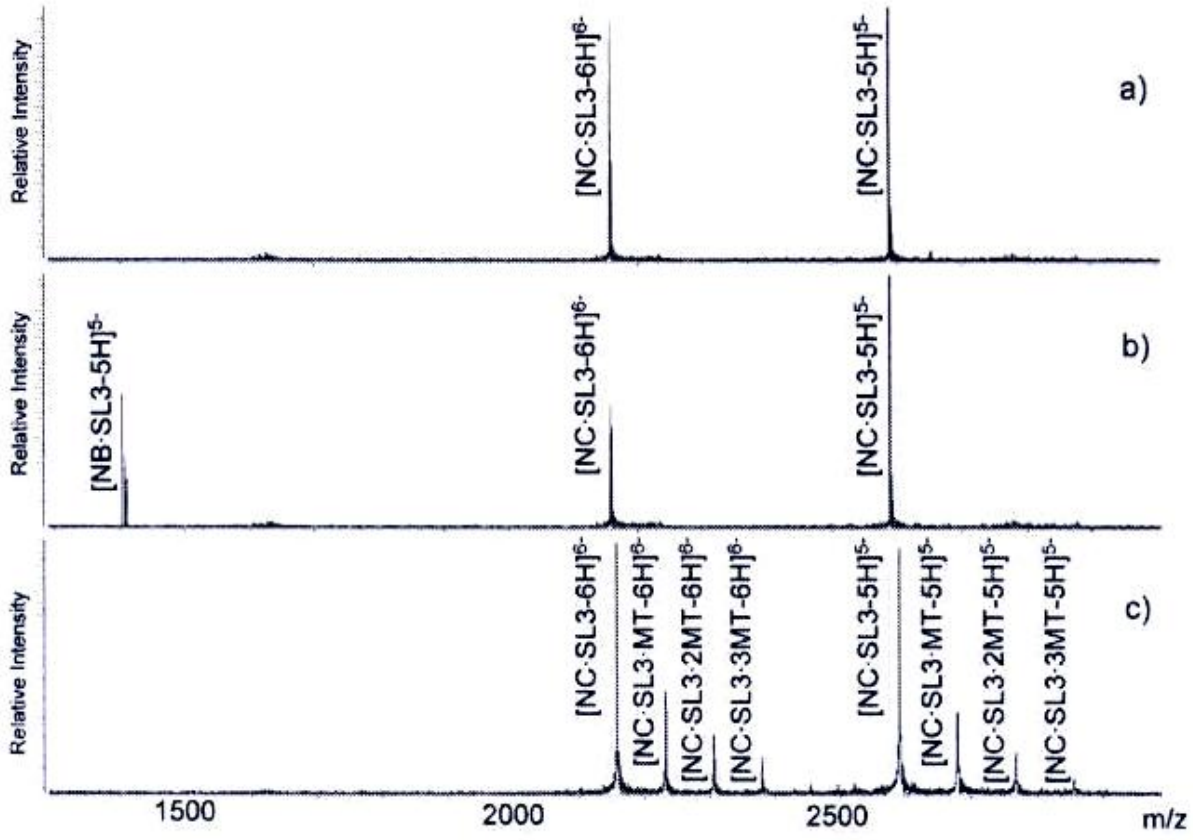


Figure 34. ESI-FT-ICRMS spectra of an experiment, where the complex formed by human immunodeficiency virus nucleocapsid protein (NC) with a stemloop structure (SL3) of a packaging signal RNA is probed with neomycin B (NB) and mitoxantrone (MT). [75]**A**: Free complex ion appearing in two charge states. **B**: Addition of 5-fold excess concentration of neomycin B induced dissociation of the NC/SL3 complex resulting in a peak of the complex formed by NB and SL3 at charge state of -5. **C**: No complex dissociation upon addition of mitoxantrone. Instead, series of ternary complexes (NC:SL3:MT) with one to three mitoxantrone adducts were found. Analytes dissolved in 10 mM IP were infused directly into the ESI source through uncoated borosilicate nanospray needle. Approximately 5 µL of each sample was loaded into the needle. Spectra were acquired in negative ion mode. The strength of the magnetic field of the analyser was 7.0 T. No further data of experimental conditions of analysis with FT-ICR is given.

In the characterisation of exact points of contact,[40, 41, 74, 76, 77, 112] RNA and protein are crosslinked chemically together from the contact sites and both moieties are digested enzymatically so that only short chains of nucleotides and amino acids remain bound together at the crosslinked sites. Traditional crosslinking techniques involving chemical crosslinkers and UV irradiation are employed. The crosslinked fragments are separated from the non-crosslinked material using chromatographic methods with detection of UV absorbance at 220 (aromatic amino acid side chains) and 260 nm (nucleobase ring structures). Fractions containing putative crosslinks are directed to mass spectrometric analyses. The approach is used in different modifications; the general outline of the method may be presented as in the Figure 35. Figure 36 presents spectra of a detected RNA-protein complex and the sequence.

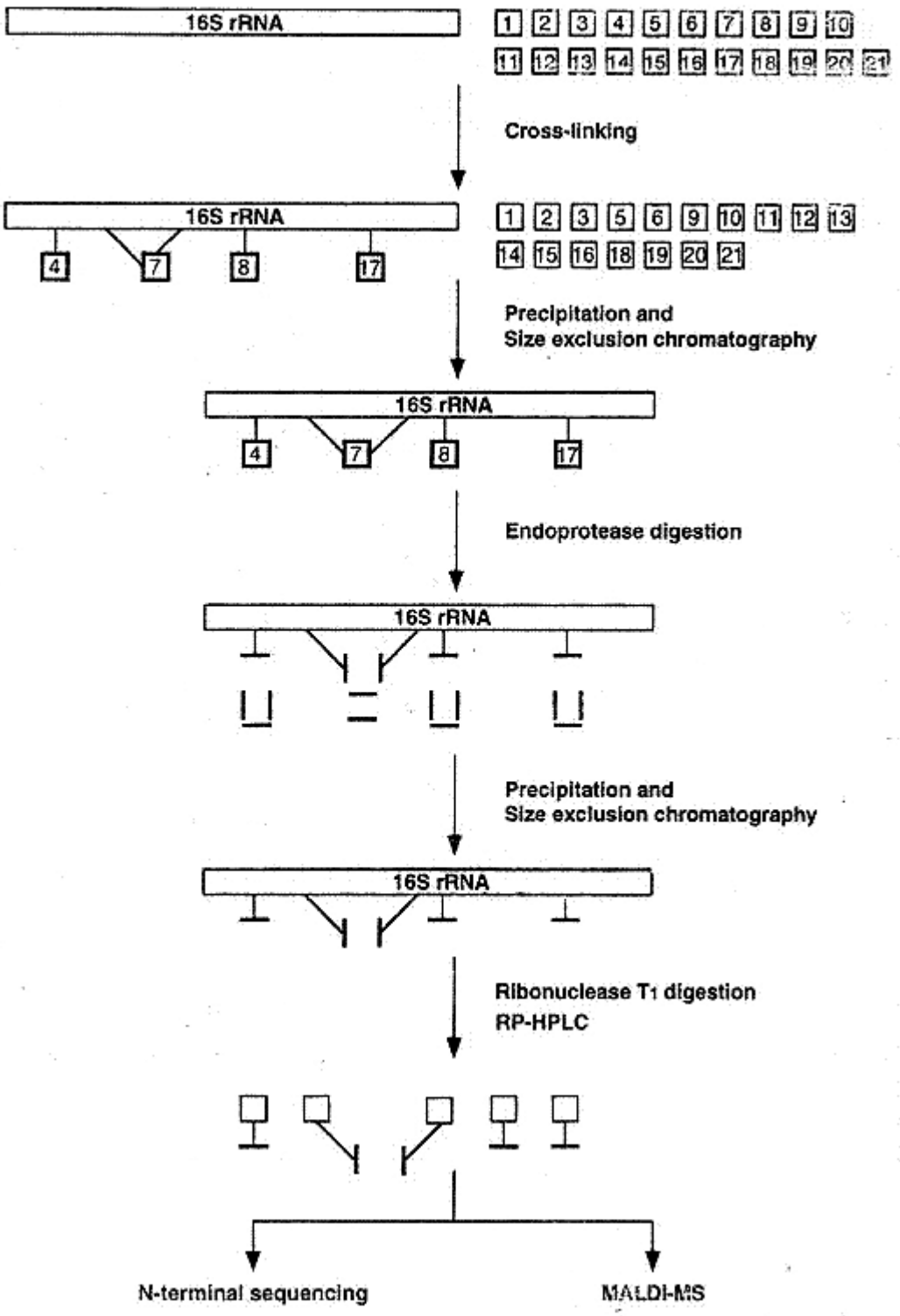


Figure 35. Schematic presentation of general features of determination of RNA-protein crosslinks between *E. coli* 16S rRNA and ribonucleoprotein particle.[41] Numbered boxes represent ribosomal proteins. Empty boxes illustrate short fragments of rRNA after enzymatic digestion.

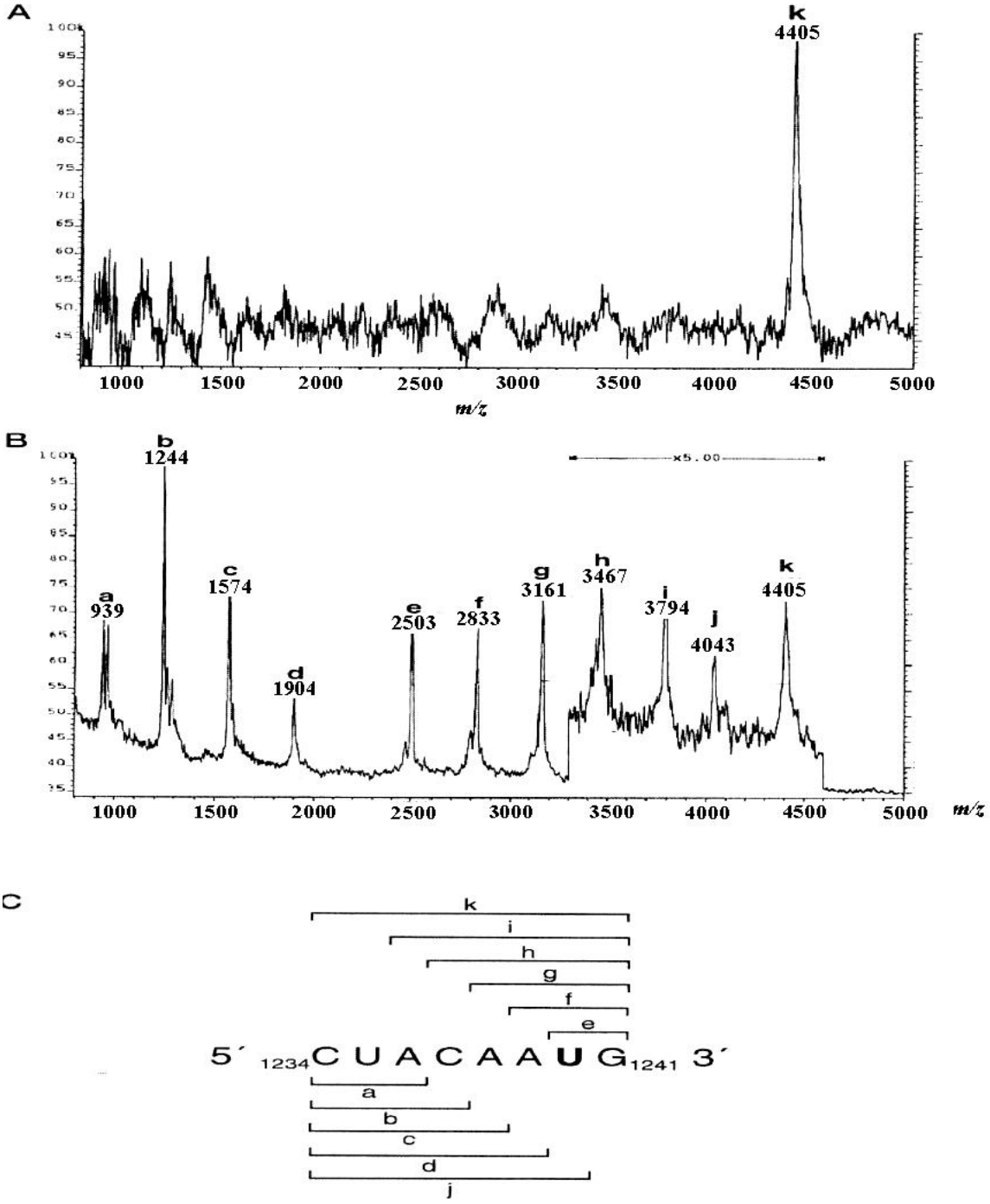


Figure 36. MALDI-TOFMS spectra of a crosslinked complex of *E. coli* 16S rRNA and ribonucleic protein S7. **A**. Mass spectrum of a complex from a isolated fraction after RPLC separation. **B.** Mass spectra of the complex after alkaline hydrolysis sequencing treatments. Figures above peaks are fragment masses in Daltons. Peaks between the masses 3300-4600 *m/z* are magnified 5-fold. Same conditions of mass analysis were applied for both analyses in A and B: negative ion mode, nitrogen-type laser operating at 337 nm, pulse width 4 ns, acceleration voltage 22 kV. MALDI matrix: CHCA in water:acetonitrile (3:2, v/v) + 0.1% TFAA. **C:** Determined sequence of the fragment of 16S rRNA crosslinked to ribonucleoprotein. Crosslinked nucleobase is denoted with bold letter. Lettered bars present the sequences of the fragments in the corresponding peaks in the spectrum (B). [68]

Lentz and his team[76] improved the sensitivity of the method using microcolumn in liquid chromatographic separation of crosslinks and precursor ion scan for *m/z* 79 to selectively detect phosphate-containing crosslinked species. Ions supposedly containing crosslinks were subjected to collisionally induced dissociation in Q-LIT instrument and the sequence of ions determined through multiple reaction monitoring of decomposition of nucleotides to separate nucleobases. This approach allowed the reduction of sample amount needed down to 5 fmol. In addition, the exact location of crosslinks and the identities of cross-linked nucleobase and peptide could be determined.

Hölsken[77] developed an enrichment method for RNA-protein crosslinks based on selective binding of phosphate-containing crosslinks to beads coated with Fe(III)-ions (*Immobilized-Metal Ion Affinity Chromatography*, IMAC).[113] Samples treated with IMAC beads were analysed with MALDI-TOFMS. Although they managed to reduce the amount of starting material to ≤50 pmol, that was still higher than in the method by Lentz. Furthermore, the unambigous determination of the crosslinking positions was not achieved.

5.5 Overview of the instrumentation

Instrumental arrangements and chosen methologies used in the papers reviewed in this study are presented in the Table 10, grouped by fields of application. The summary indicates that the LC-ESI-FT-ICRMS is the platform of choice for all forms of structural studies, although MALDI-TOF instrumentations have shown high versatility. MALDI-TOFMS may offer a sensitive, easy-to-use and economical solution for simple sequencing of longer oligonucleotides and analyses of analyte mixtures without prior fractionation.

Table 10. Instrumental setups in structural studies of ribonucleic acids by mass spectrometry.

Field of application	Method of ionization		Analyser								HPLC coupling	Reference #
	ESI	MALDI	Sector	TOF	Q	FT-ICR	QqQ	QTOF	QIT	TOF/TOF		
Sequencing	●			●							on-line IP-RPLC	65
	●					●					-	56, 67
	●						●				-	58, 60
	●							●			-	60, 61, 66
	●								●		on-line IP-RPLC -	26 53, 60
		●		●							-	38, 102, 57, 39, 59, 64
Interactions	●		●								-	72, 73
	●					●					-	70, 71, 75
	●						●				on-line RPLC	76
	●								●		-	46
		●		●							- on-line RPLC off-line RPLC	68, 69 41 74,77
		●								●	off-line RPLC	74

Table 10 continues.

Field of application	Method of ionization		Analyser								HPLC coupling	Reference #
	ESI	MALDI	Sector	TOF	Q	FT-ICR	QqQ	QTOF	QIT	TOF/TOF		
Post-transcriptional modifications	●					●						82
	●						●				on-line IP-RPLC	80, 84
	●							●			on-line IP-RPLC	84, 85, 86
	●								●		on-line IP-RPLC	85
		●		●							-	79, 81, 83
		●						●			-	81, 83
Higher order structures	●					●					-	87-91

5.6 Analytical performance

Values describing method performance, such as limits of detection or quantisation, mass accuracy, reproducibility, repeatability etc. have been given scarcely. In most cases, studies do not pursue to determine these features accurately. Furthermore, the descriptions of experimental conditions, such as instrumental parameters, buffer compositions applied are often insufficient or ambiguous making evaluation and comparison of methods difficult.

Highest mass accuracies of 1-5 ppm were reached with ESI-FTICR instruments[63, 67, 87] External calibration with synthetic oligodeoxynucleotides [26, 38, 54, 55, 59, 65] or peptide standards[74, 77] have been the most commonly used methods for instrument tuning.

Sequences of oligomers of 5-20 nucleotides are commonly determined. A sequence of 55-mer oligoribonucleotide has been confirmed using MALDI-TOF[55]. The resolving power of MALDI-TOF does not allow unambiguous determination of completely unknown sequences due to small mass difference between uridine and cytosine nucleobases. Spectral complexity and accumulation of isobaric base compositions generally prevent sequence determination above 15 nucleotides using ESI. The largest molecule reported to being ionized was 461 nucleotides (150 kDa).[55] No sequence was attempted to derive.

The sample amounts used with MALDI ionization have been around picomoles (1-10 pmol). The sample consumption in ESI is dependent on the concentration of the sample solution, flow rate and length of analysis time. Generally, concentrations of 5 - 20 pmol/µL are used with flow rates of 1 - 3 µL/min, 5-60 pmol/min sample consumptions are reached. Sample consumptions may be reduced considerably, when nanospray emitters are used and flow rates of 100-500 nL/min may be used.

Coupling of LC is essential for the studies of posttranscriptional modifications and crosslinked RNA-protein complexes. According to findings of Dickman and Hornby[98], the acetonitrile elution during IP-RPLC might also serve as a general method for removing ribonucleases from the samples.

6 COMPARISON OF TRADITIONAL AND MASS SPECTROMETRIC METHODS

Both traditional gel-based and modern mass spectrometric methods of structural studies of ribonucleic acids employ same basic approaches, where ribonucleic acid samples are manipulated according the strategy of the method in question, separated according the size, results of separation visualized and analyses made by interpreting the images of separation. Derivatisation of analytes with chemical probes, crosslinking the counterparts in complexes and production of analytically informative fragments with the use of enzymes are common for both traditional and mass spectrometric analyses.

Traditional gel-based methods separate analyses according their electrophoretic mobility in a polymer matrix, where electric field is applied. Analyte molecules differing one nucleotide in length may be separated in suitable conditions, leading to ability to estimate molar masses with the accuracy of average molar mass of a nucleotide, about 300 Da.[114] This level of resolution does not allow exact determination of small mass increases due to posttranscriptional modifications. Mass differences ranging from mass unit level down to ppm levels may be resolved with mass spectrometers. In addition, tandem mass spectrometric techniques allow further fragmentation of analytes to extract more detailed structural data.

Traditional methods are time-consuming, labour-intensive and demand usage of hazardous chemicals. Numerous stages of analysis involving preparation of gels, radioactive labelling, electrophoretic separation, gel image visualization and possible recovery of analytes from the gel matrix may extend the analysis times to several days, whereas mass spectrometric analyses may be performed in matter of few hours. Results may be compromised by many sources of error: samples may be lost during loading or they may get stuck in the wells; band broadening or smearing may occur and impurities, dirt, dust or air bubbles in the gel may induce interfering background fluorescence upon scanning of the gel.

Some automation may be employed in traditional methods for electrophoresis, detection and image production. Also image analysis may be computer assisted. Mass spectrometric applications may be fully automated in sample introduction, fragmentation, ionization, separation, detection, spectrum production and fragment mass analysis. Also the use of radioactive labels and toxic dyes for analyte visualization may be avoided.

Sample consumption in traditional methods is many orders of magnitude larger than in mass spectrometric methods. The actual comparison of sample amounts used is not straightforward, as the amounts used to load gels is usually declared in micrograms while the amounts of sample introduced to mass spectrometers is presented in picomoles. However, when using relation where 1 pmol of RNA of 1000 nucleotides length equals approximately 0.66 μg[114] and noting that the lengths of RNA analytes used in mass spectrometric analyses lie between 2-150 nucleotides, the sample masses dwell on the nanogram levels.

The traditional gel-based sequencing methods are still superior to mass spectrometric applications in the terms of sequence length and ease of interpretation. Automated sequencers are capable of sequencing molecules with lengths of several hundred nucleotides,[23]whereas using mass spectrometric methods the lengths of sequences are generally limited to about 20 nucleotides.

The drawbacks of mass spectrometry of ribonucleic acids is the high cost of instrumentation, need for meticulous controlling of sample purity at all stages of analysis and requirement for specialized computer software for interpretation of complex spectra. Successful on-line coupling of liquid chromatographic separation techniques calls for optimization of many parameters of both chromatography and ionization.

7 OUTLOOK

While the supply of analytical services and the number of published studies for mass spectrometric studies of DNA, protein and other biomolecules is high, the activity of mass spectrometric studies of RNA remains relatively low. When conducting a search of laboratories on the basis of published papers and by using a collection of links provided by a online resource site of mass spectrometry,[115] it appears that the research work using mass spectrometric techniques has concentrated to specialized groups in few universities in Europe[116-119] and USA.[120-124]

The emphasis of structural studies of ribonucleic acids is apparently moving towards studies of RNA-ligand interactions and posttranscriptional modifications, which provide supporting data for medical applications or studies of genomics.

In the search of potential drug molecules, mass spectrometric methods are applied in screening of natural and tailored RNA-binding ligands as well as studying structures and dynamics of formation of complexes.[119] Work is conducted in collaboration with groups of chemists, biologists, pharmacists and researchers of experimental medicine. For example, the inhibition of replication of viruses by ligand-mediated interference is of general interest.[119, 124, 125] Furthermore, assays are developed for identification and quantification of small modified RNA species related to or suspected to be of importance in cancer development or hereditary diseases.[122]

Development of techniques for sequencing,[116, 117]solid-phase syntheses[120] and studies of gas-phase dissociation of oligonucleotides[117, 123] provide support for the mass spectrometric studies of ribonucleic acids in general. There is still lot of work available to be done in the level of basic techniques, for example in the development of standardized reference materials for calibration of instruments and validation of methods. Moreover, the problem of differentiating masses of uracil and cytosine persists. Strictly uracil- or cytosine-specific enzymes are yet to be developed.

8 SUMMARY AND CONCLUSIONS

In this literature review, the development of mass spectrometric methods and approaches for studying structures of non-coding species of ribonucleic acids was described. Short introduction to non-coding RNA and the significance of this type of analytes was presented. The traditional gel-based methods were taken into comparison with the mass spectrometric methods. The general strategies of structural analyses in common for both traditional and mass spectrometric methods were declared. Virtues and shortcomings of both approaches were discussed. A brief outlook to the present research activity of mass spectrometric structural studies was taken.

Deriving structural information from ribonucleic acids with the use of mass spectrometry is based on indirect interpretation of data from exact mass measurements of specially pretreated analytes. Depending on the field of analysis, analytes may be derivatised with chemical probes, complexed with different ligands, transcribed to complementary DNA, fragmented with nucleases or in the gas-phase using different dissociation techniques. Partial sequences may be deduced from the exact measured oligonucleotide masses and other types of structural characterization may be performed by analyzing mass shifts of pretreated analytes in comparison to untreated reference analytes.

In conclusion, mass spectrometry provides tools for accurate analyses of short ribonucleic acids, which suits well for the analyses of small species of non-coding RNA, such as siRNA and miRNA and other at the molecular size region of 20-30 nucleotides. With the use of mass spectrometry, structural information not attainable with other methods of analyses may be obtained from the studies of ribonucleic acids.

However, the assay design, or which pretreatment methods to use, may prove to be demanding. Furthermore, the successful interpretation and application of gathered data requires multidisciplinary knowledge including fields of chemistry, biology, pharmacy, physics and bioinformatics. Therefore, the implementation of mass spectrometry to structural studies of ribonucleic acids is probably most efficient in co-operation with other specialist groups of scientists. Most likely these factors have an impact on that the mass spectrometric applications of structural studies of RNA have not found widespread use, but are practiced only among few specialist groups in Europe and USA.

9 REFERENCES

1. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. and Cech, T. R. *Cell **31*** (1982) 147-157.

2. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. *Cell **35*** (1983) 849-857.

3. Lerner, M. R. and Steitz, J. A. *Proc. Natl. Acad. Sci. USA **79*** (1979) 5495-5499.

4. Berg, J. M., Tymoczko, J. L. and Stryer, L. *Biochemistry*, 6th ed.,W.H.Freeman and Company, New York, 2007, p. 108-120.

5. Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. A. and Gingeras, T. R. *Science **296*** (2002) 916-919.

6. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. *Nature **391*** (1998) 806-811.

7. Elbashir, S. M., Lendeckel, W. and Tuschl, T. *Genes Dev. **15*** (2001) 188-200.

8. Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M. and O'Malley, B. W. *Cell **97*** (1999/4/2) 17-27.

9. Cooper, G. M. and Hausman, R. E. *The Cell - A Molecular Approach,* 3rd ed.,ASM Press, USA, 2004, p. 260.

10. Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., Hogenesch, J. B. and Schultz, P. G. *Science **309*** (2005) 1570-1573.

11. Reference 4, p.139.

12. http://library.med.utah.edu/RNAmods/ (21.1.2008).

13. Nagai, K. RNA-protein interactions. *Curr. Opin. Struct. Biol. **2*** (1992) 131.

14. Pleij, C. W. A. and Bosch, L. in *Methods Enzymol., 180,* Dahlberg, J. E.and Abelson, J. N , Academic Press, Inc, USA, 1989, p. 289.

15. Kehrenberg, C., Schwarz, S., Jacobsen, L., Hansen, L. H. and Vester, B. *Mol. Microbiol. **57*** (2005) 1064-1073.

16. Skinner, R., Cundliffe, E. and Schmidt, F. *J. Biol. Chem. **258*** (1983) 12702-12706.

17. Kowalak, J. A., Dalluge, J. J., McCloskey, J. A. and Stetter, K. O. *Biochemistry **33*** (1994) 7869-7876.

18. Dalluge, J., Hamamoto, T., Horikoshi, K., Morita, R., Stetter, K. and McCloskey, J. *J. Bacteriol. **179*** (1997) 1918-1923.

19. Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K. and Gingeras, T. R. *Cell **116*** (2004) 499-509.

20. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. *Nature* ***420*** (2002) 563-573.

21. Maniatis, T., Fritsch, E.F. Sambrook, J. *Molecular Cloning: A Laboratory Manual,* 3. ed., Cold Spring Harbor Laboratory, USA, 2001, p. A8.9-A8.17.

22. Reference 21, p. A7.82.

23. Reference 4, p. 138-139.

24. Milligan, J. F. and Uhlenbeck, O. C. in *Methods Enzymol. 180,* Dahlberg, J. E.and Abelson, J. N , 1989, p. 51.

25. Hanna, M. in *Methods Enzymol. 180,* Dahlberg, J. E. and Abelson, J. N Academic Press, Inc., USA, 1989, p. 383.

26. Holzl, G., Oberacher, H., Pitsch, S., Stutz, A. and Huber, C. G. *Anal. Chem.* ***77*** (2005) 673-680.

27. Piñol-Roma, S., Adam, S. A., Choi, Y. D. and Dreyfuss, G. in *Methods Enzymol. 180,* Dahlberg, J. E.and Abelson, J. N, Academic Press, Inc., USA, 1989, p. 410.

28. McPike, M. P., Goodisman, J. and Dabrowiak, J. C. in *Methods Enzymol. 340,* Chaires, J.B and Waring, M.J., 2001, p. 431-449.

29. Rickwood, D. and Hames, B. D. in *Gel Electrophoresis of nucleic acids: a practical approach,* 2nd ed.,IRL Press, UK, 1990,

30. http://www.cbs.dtu.dk/staff/dave/roanoke/fig5_33.jpg (22.1. 2008).

31. http://www.labtrade.com/alpha/alphaimager.htm (21.1. 2008).

32. http://www.applied-maths.com/bionumerics/bionumerics.htm (21.1 2008).

33. Zimmern, D. and Kaesberg, P. *Proc. Natl. Acad. Sci.* ***75*** (1978) 4257.

34. Hahn, G. S., Strauss, E. G. and Strauss, J. H. in *Methods Enzymol. 180,* Abelson, J. N and Simon, M.I. Academic Press, Inc., USA, 1989, p. 121.

35. Kuchino, Y. and Nishimura, S. in *Methods Enzymol. 180,* Abelson, J. N. and Simon, M.I., Academic Press, Inc., USA, 1989, p. 154.

36. Knapp, G. in *Methods Enzymol. 180,* Dahlberg, J. E.and Abelson, J. N., Academic Press, Inc., USA, 1989, p. 192.

37. Krol, A. and Carbon, P. in *Methods Enzymol. 180,* Dahlberg, J. E.and Abelson, J. N., Academic Press, Inc., USA, 1989, p. 212.

38. Berhane, B. T. and Limbach, P. A. *J. Mass Spectrom.* ***38*** (2003) 872-878.

39. Hahner, S., Ludemann, H., Kirpekar, F., Nordhoff, E., Roepstorff, P., Galla, H. and Hillenkamp, F. *Nucl. Acids Res.* ***25*** (1997) 1957-1964.

40. Thiede, B. *Methods Mol. Bio.* ***118*** (1999) 63.

41. Thiede, B. and Wittmann-Liebold, B. in *Methods Enzymol. 318*, Celander, D.W., Academic Press, Inc., USA, 2000, p. 438-446.

42. http://digilander.libero.it/itisaltamura/arizona/acidinucleici/rna_hydrolysis.gif (24.1. 2008).

43. Reference 29, p. 202.

44. O'Gorman, W., Thomas, B., Kwek, K. Y., Furger, A. and Akoulitchev, A. *J. Biol. Chem.* ***280*** (2005) 36920-36925.

45. Hanna, M. M. in *Methods Enzymol. 318,* Celander, D.W., Academic Press, Inc., USA, 2000, p. 22.

46. Gooding, K.B., Higgs, R., Hodge, B., Stauffer, E., Heinz, B., McKnight, K., Phipps, K., Shapiro, M., Winkler, M., Ng, W.L. and Julian, R.K. *J Am Soc Mass Spectrom.* ***15*** (2004) 884-892.

47. Meisenheimer, K. M., Meisenheimer, P. L. and Koch, T. H. in *Methods Enzymol. 318,* Celander, D.W., Academic Press, Inc., USA, 2000, p. 88.

48. Boorstein, W. R. and Craig, E. in *Methods Enzymol. 180,* Dahlberg, J. E.and Abelson, J. N , Academic Press, Inc., USA, 1989, p. 346.

49. Wilms, C., Noah, J. W., Zhong, D. and Wollenzien, P. *RNA* ***3*** (1997) 602.

50. Brunel, C. and Romby, P. in *Methods Enzymol. 318,* Celander, D.W., Academic Press, Inc., USA, 2000, p. 3.

51. http://www.ambion.com/techlib/tn/92/9214.html (21.1. 2008).

52. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. and Ehresmann, B. *Nucl. Acids Res.* ***15*** (1987) 9109-9128.

53. McLuckey, S. A., Van Berker, G. J. and Glish, G. L. *J Am Soc Mass Spectrom.* ***3*** (1992/1) 60-70.

54. Pieles, U., Zürcher, W., Schär, M. and Moser, H. E. *Nucleic Acids Res.* ***21*** (1993) 3191.

55. Kirpekar, F. *Nucleic Acids Research* ***22*** (1994) 3866.

56. Little, D., Thannhauer, T. and McLafferty, F. *Proc. Natl. Acad. Sci.* ***92*** (1995) 2318-2322.

57. Faulstich, K., Worner, K., Brill, H. and Engels, J. W. *Anal. Chem.* ***69*** (1997) 4349-4353.

58. Ni, J., Pomerantz, S. C., Rozenski, J., Zhang, Y. and McCloskey, J. A. *Anal. Chem.* ***68*** (1996) 1989.

59. Tolson, D. A. and Nicholson, N. H. *Nucleic Acids Res.* ***26*** (1997) 446.

60. Ni, J. and Chan, K. *Rapid Commun. Mass Spectrom.****15*** (2001) 1600-1608.

61. Schürch, S., Bernal-Méndez, E. and Leumann, C. J. *J Am Soc Mass Spectrom.* ***13*** (2002) 936-945.

62. Rozenski, J. and McCloskey, J. A. SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry. *J Am Soc Mass Spectrom.****13*** (2002) 200-203.

63. Sannes-Lowery, K. A. and Hofstadler, S. A. *J.Am. Soc. Mass Spectrom.* ***14*** (2003) 825.

64. Spottke, B., Gross, J., Galla, H. and Hillenkamp, F. *Nucl. Acids Res.* ***32*** (2004).

65. Z. Meng.*Towards High Through-put Analysis of RNA Using Mass Spectrometry,* academic dissertation,University of Cincinnati,USA,2004.

66. Tromp, J. M. and Schürch, S. *J Am Soc Mass Spectrom.* ***16*** (2005) 1262-1268.

67. Yang, J. and Håkansson, K. *J Am Soc Mass Spectrom .****17*** (2006) 1369-1375.

68. Urlaub, H., Thiede, B., Müller, E., Brimacombe, R. and Wittmann-Liebold, B. *J. Biol. Chem.* ***272*** (1997) 14547-14555.

69. Thiede, B. and Janta-Lipinski, M. v. *Rapid Commun. Mass Spectrom.* ***12*** (1998) 1889-1894.

70. Liu, C., Tolic, L. P., Hofstadler, S. A., Harms, A. C., Smith, R. D., Kang, C. and Sinha, N. *Anal. Biochem.****262*** (1998) 67-76.

71. Hofstadler, S. A., Sannes-Lowery, K. A., Crooke, S. T., Ecker, D. J., Sasmor, H., Manalili, S. and Griffey, R. H. *Anal. Chem.* ***71*** (1999) 3436-3440.

72. Hoyne, P. R., Benson, L. M., Veenstra, T. D., Maher, L. J.,III and Naylor, S. *Rapid Commun. Mass Spectrom.****15*** (2001) 1539-1547.

73. Cavanagh, J. *Anal. Chem.****75*** (2003) 3281.

74. Kühn-Hölsken, E., Lenz, C., Sander, B., Lührmann, R. and Urlaub, H. *RNA* ***11*** (2005) 1915-1930.

75. Turner, K. B., Hagan, N. A. and Fabris, D. *Nucl. Acids Res.* ***34*** (2006) 1305.

76. Lenz, C., Kühn-Hölsken, E. and Urlaub, H. *J Am Soc Mass Spectrom.****18*** (2007) 869-881.

77. Kuhn-Holsken, E., Dybkov, O., Sander, B., Luhrmann, R. and Urlaub, H. *Nucl. Acids Res.* ***35*** (2007) e95.

78. Kowalak, J. A., Pomenrantz, S. C., Crain, P. F. and McCloskey, J. A. *Nucl. Acids Res.* ***21*** (1993) 4577-4585.

79. Kirpekar, F., Douthwaite, S. and Roepstorff, P. *RNA* ***6*** (2000) 296-306.

80. Noon, K. R., Guymon, R., Crain, P. F., McCloskey, J. A., Thomm, M., Lim, J. and Cavicchioli, R. *J. Bacteriol.* ***185*** (2003) 5483-5490.

81. Andersen, T. E., Porse, B. T. and Kirpekar, F. *RNA* ***10*** (2004) 907-913.

82. Meng, Z. and Limbach, P. A. *Int. J. Mass Spectrom.* ***234*** (2004) 37-44.

83. Kirpekar, F., Hansen, L. H., Rasmussen, A., Poehlsgaard, J. and Vester, B. *J. Mol. Biol* ***348*** (2005) 563.

84. Pomerantz, S. C. and McCloskey, J. A. *Anal. Chem.* ***77*** (2005) 4687-4697.

85. Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyuauchi, K. and Suzuki, T. *Nat Struct Mol Biol* ***14*** (2007) 349.

86. Emmerechts, G., Barbe, S., Herdewijn, P., Anne, J. and Rozenski, J. *Nucl. Acids Res.* ***35*** (2007) 3494-3503.

87. Yu, E.T. and Fabris, D. *J. Mol. Biol.* ***330*** (2003) 211-223.

88. Kellersberger, K. A., Yu, E.T., Kruppa, G. H., Young, M. M. and Fabris, D. *Anal. Chem.* ***76*** (2004) 2438-2445.

89. Yu, E.T. and Fabris, D. *Anal. Biochem.* ***334*** (2004) 356-366.

90. Yu, E.T.,Zhang, Q. and Fabris, D. *J. Mol. Biol.* ***345*** (2005).

91. Zhang, Q., Yu, E. T., Kellersberger, K. A., Crosland, E. and Fabris, D. *J Am Soc Mass Spectrom.* ***17*** (2006) 1570-1581.

92. Williams, D.H. and Fleming, I. *Spectroscopic Methods In Organic Chemistry*, 5th ed., McGraw-Hill Publishing Company, England, 1995, p. 170-179.

93. Gross, J. H. *Mass spectrometry: a textbook*, Springer, Berlin, 2004, p. 4.

94. Baldwin, M. A. in *Methods Enzymol. 420*, Burlingame, A.L., Academic Press, Inc., USA, 2005, p. 3.

95. Cole, R. B. *Electrospray Ionization Mass Spectrometry: Fundamentals, instrumentation & applications,*Wiley-Interscience, New York, 1997, p. 323-339.

96. Reference 95, p. 424-431.

97. Hoffmann, E. D. and Stroobant, V. *Mass Spectrometry: Principles And Applications* ,2nd ed.,John Wiley & Sons, Chichester, 1999, p. 276.

98. Dickman, M. J. and Hornby, D. P. *RNA* ***12*** (2006) 691.

99. Azarani, A. and Hecker, K. H. *Nucl. Acids Res.* ***29*** (2001) e7.

100. Huber, C. G. and Oberacher, H. *Mass Spectrom. Rev.* ***20*** (2001) 310-343.

101. Reference 97, p. 28-44.

102. Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Lezius, A., Hahner, S., Karas, M. and Hillenkamp, F. *Nucl. Acids Res.* ***22*** (1994) 3866-3870.

103. Shaler, T. A., Wickham, J. N., Sannes, K. A., Wu, K. J. and Becker, C. H. *Anal.*

Chem. ***68*** (1996) 576-579.

104. Zhu, Y. F., Taranenko, N. I., Allman, S. L., Martin, S. A., Haff, L. and Chen, C. H. *Rapid Commun. Mass Spectrom.* ***68*** (1996) 1591.

105. Li, Y. C. L., Cheng, S. W. and Chan, T. -. D. *Rapid Commun. Mass Spectrom.* ***12*** (1998) 993.

106. Reference 97 p. 65-154.

107. Reference 93, p. 451.

108. http://library.med.utah.edu/masspec/mongo.htm (06/13 2007).

109. http://arch.debian.org./ (27.1. 2008).

110. Smirnov, I. P., Roskey, M. T., Juhasz, P., Takach, E. J., Martin, S. A. and Haff, L. A. *Anal. Biochem.* ***238*** (1996) 19-25.

111. Pomerantz, S. C. and McCloskey, J. A. in *Methods Enzymol. 193,* McCloskey, J.A., Academic Press, Inc., USA, 1990, p. 796.

112. Sannes-Lowery, K. A., Hu, P., Mack, D. P., Mei, H. and Loo, J. A.. *Anal. Chem.* ***69*** (1997) 5130-5135.

113. Li, S. and Dass, C. *Anal. Biochem.* ***270*** (1999) 9-14.

114. http://www.informaxinc.com/downloads/Conversion_Factors_Y14479_Conversion.pdf (17.2. 2008).

115. http://www.spectroscopynow.com/coi/cda/landing.cda?chId=4&type=Link (17.2. 2008).

116. http://medchem.rega.kuleuven.ac.be/mass.htm (17.2. 2008).

117. http://www.dcb.unibe.ch/groups/schuerch/research.htm (17.2. 2008).

118. http://www.sdu.dk/Om_SDU/Institutter_centre/Bmb_biokemi_og_molekylaer_biologi/Forskning/Forskningsenheder/PR.aspx (17.2. 2008).

119. http://www.sfb579.uni-frankfurt.de/en/index.htm (17.2. 2008).

120. http://www.chem.cornell.edu/faculty/index.asp?fac=44 (17.2. 2008).

121. http://www.chem.tamu.edu/services/ms/index.html (17.2. 2008).

122. http://www.che.uc.edu/faculty_staff/limbach_pat.html (17.2. 2008).

123. http://www.pharmacy.utah.edu/medChem/faculty/mccloskey/ (17.2. 2008).

124. http://www.umbc.edu/chem/general/user/fabris (17.2. 2008).

125. http://roswell.ca.sandia.gov/etyu.html (17.2. 2008).