QUANTITATIVE MASS SPECTROMETRIC ANALYSIS OF RNA-PROTEIN CROSS-LINKS

Dissertation

for the award of the degree

“Doctor of Philosophy” (Ph.D.)

Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

within the doctoral program Biology
of the Georg-August University School of Science (GAUSS)

submitted by

SAADIA QAMAR

from Lahore, Pakistan

Göttingen, 2019

Members of the Thesis Committee

Prof. Dr. Henning Urlaub (Supervisor & Reviewer)	Department of Cellular Biochemistry, Bioanalytical Mass Spectrometry Group, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany. Bioanalytics, Institute for Clinical Chemistry, University Medical Center, Göttingen, Germany.
Prof. Dr. Heike Krebber (Second Reviewer)	Department of Molecular Genetics, Institute for Microbiology and Genetics, Georg-August-University, Göttingen, Germany.
Prof. Dr. Claudia Höbartner	Institute for Organic Chemistry, Julius-Maximilians-University, Würzburg, Germany.

Further members of the Examination Board:

Prof. Dr. Markus T. Bohnsack	Institute for Molecular Biology, University Medical Center, Göttingen, Germany.
Prof. Dr. Reinhard Lührmann	Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.
Prof. Dr. Jörg Stülke	Department of General Microbiology, Institute for Microbiology and Genetics, Georg-August-University, Göttingen, Germany.
Prof. Dr. Peter Schu	Department of Cellular Biochemistry, Centre for Biochemistry and Molecular Cell Biology, Georg-August-University, Göttingen, Germany.

Date of the oral examination: 04.03.2019

Dedication

This thesis is dedicated to:

Almighty Allah, my Creator and my Master,

My great teacher and messenger, Prophet Mohammed (PBUH),

who taught us the purpose of life,

My beloved homeland Pakistan, the warmest womb;

My dear parents, who never stop giving of themselves unconditionally in countless ways,

My sweet sisters, who support and stand by me,

when things look bleak,

My friends who encourage and support me,

All the people in my life who touch my heart.

Acknowledgement

Prima facie, I am grateful to ALMIGHTY GOD who granted me strength and courage to complete this task.

I wish to extend my sincere thanks to Prof. Dr. Henning Urlaub, for providing valuable guidance and sharing expertise.

Besides my supervisor, I am indebted to the rest of the thesis committee members: Prof. Dr. Heike Krebber and Prof. Dr. Claudia Höbartner for their insightful comments which incented me to broaden my research from various perspectives and the members of the examination committee for their time.

I am gratefully indebted to Fanni Laura Bazsó & Kuan-Ting Pan, who were involved in the validation survey for this research. Their passionate participation and suggestions helped in overcoming the hurdles faced during the work.

With a special mention to Dr. Klaus Hartmuth, Dr. Olexandr Dybkov, Dr. Sergey Bessonov, Dr. Jana Schmitzová, Juliane Moses, and Lab staff Uwe Pleßmann, Monika Raabe, Thomas Conrad, Irene Öchsner, Hossein Kohansal, Gabi Heyne, Marion Killian, Ulrich Steuerwald, it was nice to have an opportunity to work with you in your facilities.

Last but by no means least, also to everyone in the impact hub Dr. Christof Lenz, Dr. Carla Schmidt, Dr. He-Hsuan Hsiao and all the lab fellows, it was good experience sharing the laboratory with you.

I intend to express gratitude to all the friends, Dr. Aneela Javed, Rabia, Naila, Sunit Mandad, Sara, Dr. M. Qasim, Altaf Sahab, Shahida aapi and Jusra who supported me throughout this venture.

Finally, I must express my sincere and profound appreciation and gratitude for my parents especially my dearest Mother, whom I missed a lot throughout this venture but her invisible constant prayers and countless love has always been and will remain with me forever. My Father, his concern and patience were remarkable and I am thankful to him, also for giving me confidence and

unceasing encouragement throughout this period of study. My sisters for their everlasting love and moral support, without them this accomplishment would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two dimensional
2DE	Two dimensional gel electrophoresis
3D	Three dimensional
ACN	Acetonitrile
ADC	Analog to digital converter
AGC	Automatic gain control
APS	Ammonium peroxodisulfate
AUC	Area under the curve
BRAT	Brain tumor
BSA	Bovine serum albumin
CID	Collision-induced dissociation
CLIP	Cross-linking and immunoprecipitation
CRM	Charge residue model
Cryo-EM	Cryo-electron microscopy
CSD	Cold shock domain
CTD	C-terminal domain
DDA	Data dependent acquisition
DHB	2,5-dihydroxybenzoic acid
DIA	Data independent acquisition
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
dsRBDs	Double-stranded RNA-binding domain
DTT	Dithiothreitol
ECD	Electron capture dissociation
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry
ETD	Electron transfer dissociation
FA	Formic acid
FDR	False discovery rate
FT-ICR	Fourier transform ion cyclotron resonance
Hb	Hunchback
HCD	Higher energy collision dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hnRNP	Heterogeneous nuclear ribonucleoprotein
HPLC	High performance liquid chromatography
HRAM	High resolution and accurate mass
IEM	Ion evaporation model
IM	Immonium ion
kDa	Kilo daltons
LB	Lysogeny broth

LC	Liquid chromatography
LIT	Linear ion trap
LTQ	Linear trap quadrupole
m/z	Mass-to-charge (ratio)
MALDI	Matrix assisted laser desorption ionization
MBP	Maltose binding protein
MCP	Microchannel plate
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
Nano-ESI	Nano-electrospray ionization
Nano-LC	Nano-liquid chromatography
NMR	Nuclear magnetic resonance
Nos	Nanos
NREs	Nanos response elements
NTC	Nineteen complex
OMSSA	Open mass spectrometry search algorithm
PAGE	Polyacrylamide gel electrophoresis
PCI	Phenol-chloroform-isoamyl alcohol
PCV	Packed cell volume
PMSF	Phenylmethylsulfonyl fluoride
ppm	Parts per million
PRM	Parallel reaction monitoring
PTM	Post-translational modification
Pum	Pumilio
QIT	Quadrupole ion trap
RBD	RNA binding domain
RBP	RNA binding protein
rf	Radio frequency
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RP	Reversed phase
RP-LC	Reverse phase liquid chromatography
rpm	Rounds per minute
RRM	RNA recognition motif
SDS	Sodium dodecyl sulphate
SE	Size exclusion
SEC	Size exclusion chromatography
sno	Small nucleolar
snRNA	Small nuclear RNA
snRNP	Small nuclear RNP
SR	Serine and arginine rich
SRM	Selected reaction monitoring
SRPs	Signal recognition particles
TDC	Time to digital converter
TEAB	Triethylammonium bicarbonate
TEMED	Tetramethylethylenediamine

TFA	Trifluoroacetic acid
TOF	Time of flight
TRIM	Tripartite motif
tRNA	Transfer RNA
UTP	Uridine-5'-triphosphate
UTR	Untranslated region
UV	Ultraviolet
XIC	Extracted ion chromatogram
ZnF	Zinc finger

SUMMARY

Discoveries made in recent years have revealed that a copious number of RNAs get associated with miscellaneous number of RNA-binding proteins throughout their life cycle in a eukaryotic cell. Together they constitute the Ribonucleoprotein (RNP) complexes. The structural studies of these RNP complexes provide a valuable insight into the binding modes and functional implications of their interactions within the complexes. Most of the RNA-binding proteins have more than one binding domain and various kinds of RNA binding motifs (RBM) that recognize the specific RNA sequence. The identification and characterization of these RNA binding domains/motifs is utmost essential for the better understanding of the function of RNP complexes.

To investigate the interactions of the RNA-binding proteins within RNP complexes, UV-induced cross-linking followed by mass spectrometry (MS) has proved to be a promising and straightforward technique. During the course of this study, the protocol is modified and optimized for the interaction analysis of large RNP complex assemblies which led to the identification of predicted as well as unknown RBMs. However, by using the conventional protocol, relatively less complex RNP assemblies have also been analyzed.

The first part of the research work mainly emphasized on the qualitative analysis of the protein-RNA cross-links. One of the protein-RNA complex studied is *in vitro* assembled Brat-NHL-hb RNA complex. Six hb RNA contact sites have been found in Brat-NHL protein by UV-induced cross-linking followed by mass spectrometric analysis which provided the basis for mutation studies to confirm the binding activity of Brat protein with hb RNA during embryogenesis in Drosophila.

Another protein-RNA complex studied by the aforementioned technique is comprised of yeast splicing factor CWC2 along with U6 and U4 snRNAs. Seventeen probable contact sites within CWC2 protein have been identified that have undergone cross-linking reaction with U6 and U4 snRNAs. Later on this

CWC2-U6/U4 snRNAs complex is also used as a model system for the quantitative studies of the cross-links.

Based on the qualitative analysis of CWC2-U6/U4 cross-links, the quantitative analysis of protein-RNA cross-links has been established. For this, relative quantitative approach has been adopted by using isotope labeled RNAs. The results have been used to determine the stoichiometry of the cross-links.

In the last part of the research work, the UV-induced cross-linking approach has also been utilized to investigate the interaction sites within large RNP complexes (H/E complex) isolated from the *HeLa* nuclear extract by assembling on MS2-tagged pre-mRNAs. It has led to the identification of several contact sites within the known RNA binding regions as well as in the regions which have not been reported before as RNA binding regions.

The studies conducted during the research work have contributed in the identification and characterization of protein-RNA interactions within the aforementioned complexes and also provided the quantitative insight into the protein-RNA interactions. The results will also contribute in improving the data analysis approach for protein-RNA cross-links and will serve as an outlook to future directions of this project.

1 INTRODUCTION

1.1 Proteomics

The major challenge in cell biology is to reveal the mechanisms occurring within the cells. In addition to the knowledge provided by the genome in the form of genetic information, the detail comprehension of the proteins and metabolites is extremely necessary for the understanding of the biological processes (Angel et al., 2012; Hsu & Chen, 2016). The term proteome first emerged in 1994. It was designated to the whole set of proteins that is been encoded by the organism's genome (Wilkins et al., 1996). Proteomics refers to the study of proteome explicitly involving the protein identification, characterization and quantification as well as interactions of proteins with each other and with other biomolecules such as nucleic acids or lipids etc. present in their surrounding environment (Lane, 2005). These interactions lead to the formation of large macromolecular assemblies that play a vital role in many cellular processes (Alberts, 1998). Therefore it is indispensable to understand the structure of protein complexes in order to have detailed insight in the mechanism of the cellular processes occurring in living organisms (Sali et al., 2003; Robinson et al., 2007; Walzthoeni et al., 2013).

Over the past few decades the conventional structural techniques like nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and cryogenic electron microscopy (cryo-EM) have been developed to generate the structural information of the large macromolecular complexes (Grassucci et al., 2007; Faini et al., 2016). These techniques however have their limitations. Both NMR and X-ray crystallography require relatively large quantity of materials usually in milligrams (mg) (Svergun et al., 1997). Many times the high concentration of protein results in its precipitation in NMR. Also NMR is limited to complexes of molecular weight approximately 40-50 kDa. On the other hand the X-ray crystallography provides high resolution protein structures but all the proteins do

not easily get crystallized (Loo, 1997). As compared to the X-ray crystallography, cryo-EM requires less material (~10 µg/ml) for crystallization (Sjöberg et al., 2005) however the resolution obtained is relatively lower than that obtained by the NMR spectroscopy and X-ray crystallography (Grassucci et al., 2007). Recently, with the development of mass spectrometry (MS) based methods along with advanced computational approach, can generate low resolution structural information of macromolecular assemblies (Chen et al., 2010; Jennebach et al., 2012). Generally, the mass spectrometry technique is relatively fast and not limited by the size of the protein complex to be analyzed requiring nanogram (ng) to microgram (µg) of sample amount (Walzthoeni et al., 2013). Although the information obtained by the mass spectrometry based methods about the three dimensional (3D) structure of the protein complex is not complete but in combination with other structural techniques and computational approaches, it can be used to determine the structure of these large macromolecular assemblies.

1.2 Mass Spectrometry

Much credit has been given to the studies carried out by Sir J. J. Thomson (1906 physics Nobel Laureate for the discovery of electron) and his student F. W. Aston (1922 chemistry Nobel laureate for establishing mass spectrograph and element's isotope measurements) which lead to the development of mass spectrometry (Thomson, 1921; Aston, 1933). Mass spectrometry that is regarded as a microanalytical technique is applied for the detection and determination of the composition, molecular structure and the amount of an analyte (Watson & Sparkman, 2007). It is an accurate way of measuring the mass to charge ratios (*m/z*) of ions. During last 30 years, it has proved to be significant for the structural characterization and analysis of biomolecules (Griffiths et al., 2001).

Every mass spectrometer is comprised of three components: An ion source ─ to convert the molecules into gas phase ions; Mass analyzer ─ to separate the ions according to their *m/z* ratios through electric or magnetic fields; A detector ─ to

detect the number of emerging ions. These detected signals are proportional to the abundance of each species. In addition to these components the mass spectrometer also has an inlet device e.g. liquid chromatography or a direct injecting probe for introducing the analyte into an ion source and a computer for controlling the instrument, processing the data and producing the mass spectrum in a suitable format (Figure 1.1) (Lane, 2005; Girolamo et al., 2013).

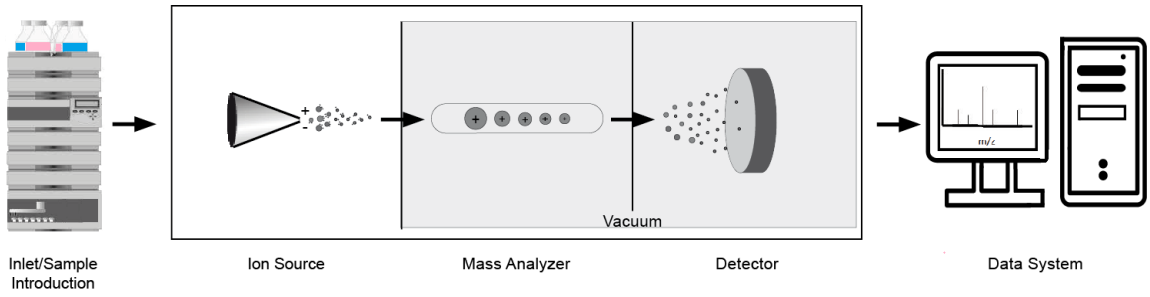


Figure 1.1: Components of a mass spectrometer. The mass spectrometer is comprised of three basic components: An ion source ─ to generate ions, Mass analyzer ─ to separate ions on the basis of *m/z* ratios and a detector ─ to detect ions and their abundance. The mass analyzer and the detector are always under vacuum. The mass spectrometer is usually coupled to an inlet device to introduce sample into an ion source and a computer for processing of the data

1.2.1 Ion Source

It is a first component of a mass spectrometer where the sample is first vaporized and then ionized. For ionization, a low amount of internal energy is transmitted to the molecule. There are two widely used soft ionization techniques for the analysis of proteins and peptides, electrospray ionization (ESI) (Yamashita & Fenn, 1984) and matrix-assisted laser desorption ionization (MALDI) (Karas & Hillenkamp, 1988; Tanaka et al., 1988; Hillenkamp & Karas, 1990) which has revolutionized the field of biology. This pioneering work on ESI and MALDI was recognized and 2002 Nobel prize was awarded in chemistry to John Fenn and Koichi Tanaka.

1.2.1.1 Electrospray Ionization (ESI)

In electrospray ionization, the ions are transferred from solution into gaseous phase at atmospheric pressure. The sample is sprayed at low flow rates (nl/min to µl/min) via thin needle to which high voltage (1-6 kV) is applied. This electric field is built up between needle and a counter electrode. The positive potential results in the accumulation of positive ions at the tip surface and drawn out attaining a cone-like shape known as "Taylor cone" (Lane, 2005). This cone turns into a filament as it holds more electric charge. When the surface tension exceeds the applied electrostatic force, fine positively charged droplets are produced. These droplets fly towards the counter electrode and pass through either a heated capillary or film of heated nitrogen resulting in the evaporation of solvent. The electric charge density increases at the surface of the droplets and when the electrostatic repulsion becomes higher than the surface tension they split into even more smaller droplets (Lane, 2005). The process continues until the solvent is completely removed or an ion desorbs from a droplet (Griffiths et al., 2001). The ions are produced in the gas phase by either of the two known mechanisms: the ion evaporation model (IEM) (Iribarne & Thomson, 1976; Thomson & Iribarne, 1979; Nguyen & Fenn, 2007) and the charge residue model (CRM) (Dole et al., 1968; Willoughby et al., 1998). Generally, the ions produced are in the form $[M+H]^+$ or $[M+nH]^{n+}$ (or $[M-H]^-$/$[M-nH]^{n-}$). Since large macromolecules have several ionizable sites, multiply charged ions will also be produced.

The advent of nano-electrospray ionization (nano-ESI) has extended the electrospray ionization mass spectrometry (ESI-MS) applications. It has considerably increased the sensitivity for high concentration low volume samples. In nano-ESI-MS low flow rate 10-300 nl/min is used which improves the ionization efficiency by generating smaller droplets (Griffiths et al., 2001). Now-a-days mass spectrometers are online coupled to separation techniques like liquid chromatography (LC) etc. bringing together the advantages of separation techniques and nano-ESI to biological mass spectrometry (Karas et al., 2000).

1.2.1.2 Matrix-assisted Laser Desorption Ionization (MALDI)

Since MALDI technique has not been used in the current study so it has been briefly described here. The sample to be analyzed is mixed and co-crystallized with an excess of matrix material and is irradiated with a nitrogen laser at a wavelength of 337 nm. It sublimates and ionizes the analyte without fragmentation. Highly accepted ionization mechanism includes proton transfer in gas phase to the expanding plume from photoionized matrix molecules (Lane, 2005; De Hoffmann & Stroobant, 2007). The ions are then accelerated towards an analyzer by an electrostatic field.

1.2.2 Mass Analyzer

It is a second component of mass spectrometer which is specified for the separation and detection of ions. The mass analyzer performance and its utility can be assessed based on several parameters including mass resolution, mass accuracy, mass range and MS/MS acquisition and precision. Presently, four main types of mass analyzer are widely used in proteomics: time of flight (TOF), quadrupole, ion trap, and Fourier transform ion cyclotron resonance (FT-ICR) analyzers. These mass analyzers can be used both as stand-alone analyzers or in combination in tandem mass spectrometer (Girolamo et al., 2013).

1.2.2.1 Time of Flight (TOF)

It separates ions on the basis of their velocity. The *m/z* ratios are based on the measurement of the time required by the ions to travel from the source to the detector. All ions are produced at the same time and are accelerated by applying a fixed potential into TOF drift tube. As all ions with similar charge carry identical kinetic energy, therefore the ions with lower *m/z* value achieve higher velocities and reach the detector earlier than the ones carrying higher *m/z* value, after travelling through a fixed distance of 0.5-2 meters (Guilhaus, 1995; Mamyrin, 2001; Glish & Vachet, 2003). The performance of TOF instrument is improved by using two techniques. Firstly, by creating a time delay between an ion formation

and its extraction from the source, also known as "delayed extraction" (Kovtoun, 1997; Kovtoun & Cotter, 2000). Secondly, by including ion mirrors, or reflectrons that create a retarding field and deflect the ions, sending them back to the detector through the flight tube. Thereby compensating for the minor differences in the velocities of ions with the same *m/z* ratios (Mamyrin, 1994; Scherer et al., 2006).

1.2.2.2 Quadrupole

The quadrupole mass analyzer consists of four equally spaced parallel rods arranged around a central axis. Direct voltage is applied to two of the rods and an alternating current or radio frequency (rf) potential to the other two rods. The flight of the ions between these rods depends upon the voltages applied. The specific direct and alternating current voltages allow only ions defined with certain *m/z* ratio to travel through the analyzer to the detector (Glish & Vachet, 2003; Dawson, 2013).

1.2.2.3 Ion Trap

(A) Quadrupole Ion Trap (QIT)

In ion trap, the ions are first trapped and then detected on the basis of their *m/z* ratios. An ion trap may be conceptualized as a quadrupole that has been bent around on itself to configure a closed loop. The ions are trapped inside the three hyperbolic electrodes, the ring electrode, the entrance cap electrode and the exit cap electrode. Thus the ion trap is also referred as the "quadrupole ion trap". The ions are subjected to the oscillating electric field in an ion trap. This electric field is generated by a radio frequency (rf) voltage applied to the ring electrode. Inside ion trap, the Helium gas is responsible for removing excess energy from the ions by collision and "dampens" the ions into the center of trap. The oscillation frequency of the ions depends upon their *m/z* ratios therefore the ions with different *m/z* ratios will exit the ion trap at different voltages and time. The ions are then transferred to the detector (March, 2000; March & Todd, 2005).

(B) Linear Ion Trap (LIT)

The linear ion trap (LIT) is also referred as two-dimensional quadrupole ion trap (2D-QIT). It consists of two pairs of parallel mounted planar electrodes and a pair of z-electrodes that are used as the endcaps (Song et al., 2006). The radio frequency (rf) potential is applied for ion trapping in the radial and axial directions. The ions are confined radially in x and y directions by two dimensional rf field and axially in z direction by potentials applied to the endcaps, thus limiting the longitudinal flow of ions which are later detected by conventional means (Tabert et al., 2006). Linear ion traps have high injection and trapping efficiencies and high storage capacities that enhance the sensitivity, resolution and mass accuracy (Hager, 2002).

(C) Orbitrap

The orbitrap mass analyzer works by orbital trapping of ions. It is comprised of an outer barrel like electrode and a central spindle-like electrode along the axial axis. The outer electrode serves two purposes, ion excitation and detection. For the efficient ion introduction the electric field is modified at the injection port by using field compressor, a small portion of the outer electrode (Zubarev & Makarov, 2013). The electric field created by applying the voltage to outer and central electrodes allow the injected ions to cycle around the central electrode and at the same time oscillate along the horizontal axis (Hu et al., 2005; Eliuk & Makarov, 2015). The image current from the coherently oscillating ions is detected on the receiver plates as a time domain signal. This signal is Fourier transformed into frequency domain as in FT-ICR and is converted into a mass spectrum (Scigelova & Makarov, 2006).

1.2.2.4 Fourier Transform Ion Cyclotron Resonance (FT-ICR)

It uses the magnetic field to determine the *m/z* ratios of ions. It consists of an ESI ion source, ion optics to transfer ions into the magnetic field and an ICR cell or Penning trap. The ions are trapped, exposed to the magnetic field, oscillated with a cyclotron frequency that is inversely related to the *m/z* ratio, analyzed and

finally detected. The use of a Penning trap enhances the detection time and thus sensitivity and resolution (Marshall et al.,1998; Marshall & Chen, 2015).

1.2.3 Ion Detector

It is the last component of mass spectrometer. After passing through the mass analyzer, the ion beam strikes the detector. The ions separated on the basis of *m/z* ratio are electrically detected by the detector. There are various types of ion detection systems based on sensitivity, accuracy and response time.

Electron multiplier is an extensively used detection system. Whereby the secondary electrons are repeatedly emitted as a result of constant collisions of energized particles at suitable surface leading to the escalation of ions' signal intensity. The electron multiplier can be either of discrete dynode type or a continuous dynode type (Niessen & Falck, 2015). It can be used in combination with quadrupole and ion trap instruments.

Microchannel plate (MCP) is other type of ion detector that has its application in combination with time of flight instruments. The MCP is an array of miniature electron multipliers aligned parallel to one another, mostly along a small angle with the surface. It is backed up by the time to digital (TDC) or analog to digital (ADC) converters that enable the high speed acquisition of data (Dubois et al., 1999).

In FT-ICR and Orbitrap based mass spectrometers, the ion detection is based on the detection of high frequency image currents generated by the coherent movement of ions. The signals of all ions with different *m/z* ratios are detected simultaneously (Niessen & Falck, 2015).

1.2.4 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) involves minimum two stages of mass analysis coupled with either a chemical reaction or a dissociation process that brings the difference in ion mass or charge (De Hoffmann & Stroobant, 2007). In tandem mass spectrometry (MS/MS) experiment, the first analyzer is used to isolate the precursor or parent ion, which undergoes fragmentation to yield product or daughter ions and neutral fragments which are then analyzed by a second mass analyzer (Madeira & Florêncio, 2012). The widely used types of fragmentations in tandem mass spectrometry are collision induced dissociation (CID), high energy collision dissociation (HCD) and electron transfer dissociation (ETD).

Hybrid mass spectrometers utilize different types of analyzers for the first and second stages of mass analysis. The general purpose of designing these hybrid instruments is to combine various performance attributes exhibited by discrete types of analyzers into single mass spectrometer i.e. mass resolving power, the ion kinetic energy for dissociation and speed of analysis etc.(Glish & Burinsky, 2008).

The MS/MS instruments can be classified in two categories: "tandem in space" mass spectrometers by coupling of two physically distinct analyzers like TOF analyzers and quadrupole mass filters and "tandem in time" mass spectrometers through conducting an appropriate sequence of events within an ion storage device like QIT, FT-ICR and orbitraps (Glish & Vachet, 2003).

For the current studies, linear trap quadrupole (LTQ)-Orbitrap (Velos and XL) mass spectrometer has been used. It is a Fourier transform hybrid mass spectrometer equipped with ESI source that is-line coupled to nano-flow reverse phase liquid chromatography (RP-LC) system. The LTQ-Orbitrap instrument consists of dual cell linear ion trap coupled to rf-only C-shaped "C-trap", HCD collision cell and the orbitrap mass analyzer. The linear ion trap selectively fills the C-trap with ions of interest. The C-trap stores and collisionally cools the ions before injecting into the orbitrap. The orbitrap analyzer utilizes the image current

detection and Fourier transformation for mass analysis (Olsen et al., 2005; Makarov et al., 2006a).

The precursor ion scan (MS1) is carried out by orbitrap whereas product ion scan (MS2) is performed with CID fragmentation in LTQ simultaneously. However, if desired, the CID generated fragments can be transferred to the orbitrap for the measurement. Depending upon the experimental necessity, HCD fragmentation can also be performed in HCD collision cell of the selected ions which are then measured in orbitrap (Olsen et al., 2009). Thus the combination of the strengths of two different mass analyzers i.e. high sensitivity and high scan rates of LTQ mass analyzer and high mass resolution and high mass accuracy of orbitrap mass analyzer, has led to high throughput, high accuracy (≤ 5 ppm) and high resolution (> 100,000) of the LTQ-Orbitrap mass spectrometer (Makarov et al., 2006b).

The other Fourier transform based hybrid instrument used during the course of study is Quadrupole-Orbitrap (Q Exactive) mass spectrometer. The instrument is equipped with quadrupole which serves as a precursor ion selection device configured with high-efficiency C-trap, HCD collision cell and the orbitrap mass analyzer (Michalski et al., 2011). Both MS1 and MS2 measurements are carried out by orbitrap. Hence, the instrument combines the high performance quadrupole precursor selection with high resolution (up to 140,000 FWHM) and accurate mass (better than 1 ppm) (HRAM) orbitrap detection (Kelstrup et al., 2012).

1.2.5 Mass Spectrometric Analysis

The tandem mass spectrometry coupled with liquid chromatography and the availability of genome sequence information has revolutionized the field of mass spectrometry. There are two main strategies used in proteomics for protein identification by mass spectrometry: top-down proteomics and bottom-up proteomics. In top-down approach, the intact proteins, mostly selected multiple

charge ions generated by ESI as precursor ions, are subjected to gas phase fragmentation. Conversely the bottom-up approach involves the enzymatic digestion of proteins into peptides. The peptides can be separated using reverse phase liquid chromatography which is coupled online/offline to ESI-MS/MALDI. The data generated by mass spectrometer is used for the identification and quantification of specific peptides (Niessen & Falck, 2015; Gillet et al., 2016; Zhang et al., 2013).

The bottom-up proteomics is comprised of three main approaches: data-dependent acquisition (DDA) or shotgun approach; targeted data acquisition by selected reaction monitoring (SRM)/multiple reaction monitoring (MRM)/parallel reaction monitoring (PRM); and data-independent acquisition (DIA).

In DDA based methods, a complete and unbiased coverage of proteome is achieved by acquiring the mass spectra at MS1 level of the most abundant precursor ions, co-eluting at a specific point in gradient elution and MS2 level of the fragment ions, generated by the sequential isolation and fragmentation of the precursor ions (Aebersold & Mann, 2016). Instruments with quadrupole-orbitrap or TOF analyzers are most suitable for DDA approach.

In targeted data acquisition, the proteins of interest are preselected and known. In the first stage the precursor ion of particular mass is isolated which is then fragmented and product ion of certain mass is selected in the second stage for detection. The targeted analysis can be carried out by SRM/MRM/PRM approach. The SRM involves the monitoring of single fixed ion mass while MRM is the parallel acquisition of multiple SRM transitions (Bauer et al., 2014). In PRM, the full scan of each transition by a precursor ion is performed (Law & Lim, 2013). The SRM, MRM and PRM experiments are carried out on triple quadrupole and quadrupole-orbitrap mass spectrometers (Gallien et al., 2012).

In DIA, all the ions within selected *m/z* range of precursors are fragmented at the same time thus generating the MS/MS records of all the analytes in the biological sample (Doerr, 2015; Hu et al., 2016). Mass spectrometers with quadrupole, TOF

or orbitrap analyzers like Q-TOF and Q-Orbitrap are appropriate for the DIA approach.

1.2.6 Mass Spectrometry Based Fragmentation

Tandem mass spectrometry helps in providing detail information about the structure, composition and sequence of biomolecules. During this process, the selected ions of interest are fragmented on interaction with the collision gas such as nitrogen or helium or argon (Platner, 2013). There are several fragmentation techniques. For the current studies CID and HCD methods have been used for the analysis of samples containing RNA and proteins.

For the MS analysis of proteins, bottom up approach has been adopted. For this, trypsin is widely used protease to digest complex protein sample. Trypsin cleaves proteins C-terminally at arginine (R) or lysine (K) amino acid residues, except when either of it is followed by proline (P) amino acid (Wang, 2011). Usually the peptides from acidic solution are fragmented in positive ion mode (Seidler et al., 2010). The fragmentation occurs along the peptide backbone (C-N, N-C and C-C bonds) (Frese, 2013). To annotate peptide sequence ions, Roepstorff and Fohlman nomenclature (Roepstorff & Fohlman, 1984) is widely used which was later modified by Biemann (Biemann, 1990). As a result of fragmentation, if the charge is retained on N-terminus of the ion, it is designated as b-ion whereas if the charge is retained on C-terminus of the ion, it is designated as y-ion. Depending on type of fragmentation method used, relatively less common a, c, x and z ions can also be created (Newton et al., 2004) (Figure 1.2). In addition to these, the mass spectrum can also has the signals generated by the neutral loss of NH_3 or H_2O from ions, immonium ions and the a-type ions, formed by the loss of CO from b-ions (Griffiths et al., 2001; Kapp et al., 2003). The peptide mass spectrum is interpreted by calculating the difference in mass between the adjacent peaks that corresponds to the mass of an amino acid of either b or y ions series (Niessen & Falck, 2015).

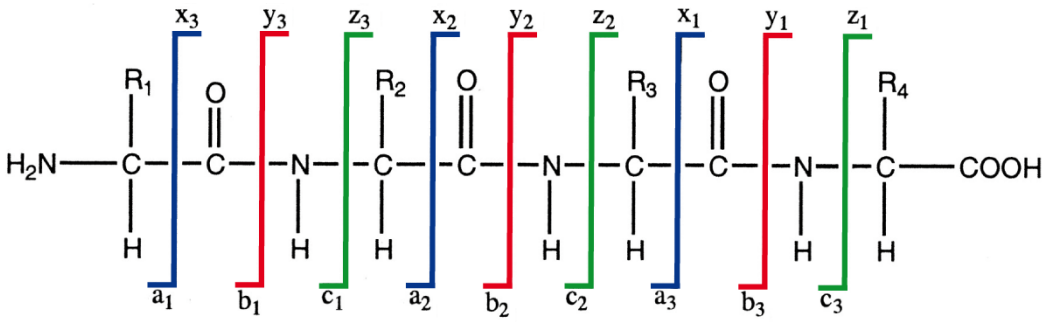


Figure 1.2: Schematic representation of peptide fragmentation and nomenclature of fragments (Biemann, 1990). The collision induced dissociation mainly generates b and y ions due to the fragmentation of the peptide bond.

Besides the tandem mass spectrometry has frequently been used for the elucidation of structure and sequences of proteins, it has proven its potential for the investigation of nucleic acids. The fragmentation of RNA has been relatively less studied than that of DNA. The studies using MALDI have revealed that the RNA is more stable than DNA (Kirpekar & Krogh, 2001). For mass spectrometric analysis, the RNA is digested with RNase. The nucleic acids from basic solutions are usually fragmented in negative ion mode. The oligonucleotide fragment ions are annotated according to the nomenclature, similar to the one used for peptides, proposed by McLuckey, Berkel and Glish (Mcluckey et al., 1992). The 5‘-terminus fragment ions a, b, c and d and their complementary 3‘-terminus fragment ions z, y, x and w respectively are generated upon cleavage of phosphodiester chain (Figure 1.3). Regardless of the type of fragmentation method used, there is nucleobase loss due to the cleavage of N-glycosidic bond between the nucleobase and the sugar moiety followed by the dissociation of phosphodiester bond of the nucleotide that undergoes base loss. The mass difference in the series of 5‘-terminus and 3‘-terminus fragment ions can be used to determine the nucleotide sequence (Huber & Oberacher, 2001).

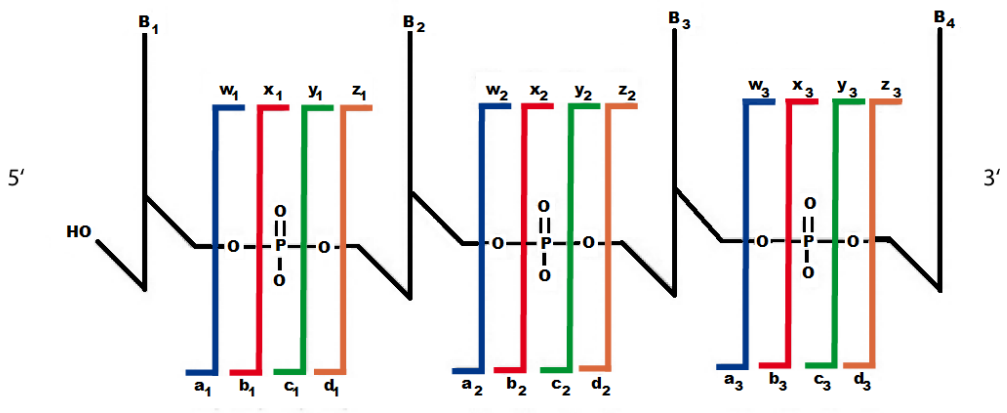


Figure 1.3: Schematic representation of RNA fragmentation and nomenclature of fragments (Mcluckey et al., 1992). The cleavage of phosphate backbone results in the generation of 5'-terminus fragment ions a, b, c and d and their complementary 3'-terminus fragment ions z, y, x and w.

1.2.7 Proteomic Data Analysis

Bottom-up approach leads to considerable increase in the sample complexity which in turn makes the data analysis complicated. Therefore, prior purification, separation, fractionation and enrichment of the protein sample is necessary to reduce the complexity. For this the protein samples are usually processed by gel electrophoresis and chromatography. The peptides obtained are then analyzed by liquid chromatography coupled to mass spectrometry (LC-MS) (Lane, 2005; Gillet et al., 2016).

The data generated is comprised of information regarding *m/z*, retention time and intensities of the observed peptides along with fragment ion spectra (Schmidt et al., 2014). The correct assignment of the spectrum to the peptide sequence is the primary and crucial step for proteomic data processing. Several computational softwares are available for the automatic assignment of the peptide sequence to the spectrum (Nesvizhskii et al., 2007).

The chromatographic peak of the peptide can be used to calculate the area under the curve (AUC) for quantification of a peptide (Schmidt et al., 2014). To identify the peptide sequence, the MS2 spectrum is scanned using the search algorithm such as Andromeda (Cox et al., 2011) or Open Mass Spectrometry Search

Algorithm (OMSSA) (Geer et al., 2004) etc. against protein sequences database. The acquired peptide spectrum is searched and correlated with *in silico* generated peptide spectrum from protein sequences database (Lane, 2005; Nesvizhskii et al., 2007; Gillet et al., 2016). These searches are made according to the user specified filtering criteria like protease and labels used, mass tolerance and types of post translational modifications expected etc. A target-decoy based false discovery rate (FDR) approach is used to limit the coincidental peak matches to a certain number. The score is calculated, based on the statistical significance of the match between the acquired and the theoretical peptide spectrum. It helps in differentiating the correct from incorrect identifications. Generally, the peptide hits with only best scores are further considered for the statistical analysis (Lane, 2005; Nesvizhskii et al., 2007).

1.3 Quantitative Proteomic Studies

In the past centuries mass spectrometry (MS) has been evolved as technique used not only for providing the qualitative information of proteins but also their abundance in the sample (Urban, 2016). Quantitative proteomics can be defined as a technique for determining the global protein levels (Yan & Chen, 2005). Quantitative protein profiling has proven to be essential for the investigation of biological processes. In addition to mass spectrometry, various kinds of techniques have been developed for quantitative analysis like two dimensional gel electrophoresis (2-DE), fluorescence microscopy and enzyme-linked immunosorbent assay (ELISA) etc. (Wilm, 2009). In certain cases, such as in the presence of multiple isoforms and modified forms of proteins, the proteins are undetectable and unmeasurable by immunological or other techniques. The mass spectrometry has the ability to distinguish between various isoforms and modified protein and also to quantify them (Hale, 2013).

In shotgun proteomics, the peptides are quantified by determining the signals of their precursors at both MS1 and MS2 levels (Gillet et al., 2016). Numerous

strategies have been developed for quantitative analysis by mass spectrometry (Frese, 2013). These strategies can be divided into two major categories: absolute and relative quantification. Absolute quantification is the determination of precise amount of a biological molecule under study e.g. copy number of protein per cell whereas the relative quantification is the determination of relative levels of biomolecules between two or more states/samples e.g. percentage or fold change increase or decrease in protein abundance in response to some treatment (Ong & Mann, 2005; Hale, 2013).

These quantitative measurements can be carried out either with or without labeling, referred as labeled quantitation and label-free quantitation respectively. In labeled quantitation, the labels (^{2}H, ^{13}C, ^{15}N or ^{18}O as heavy labels and ^{1}H, ^{12}C, ^{14}N and ^{16}O as light labels) are incorporated either metabolically *in vivo* or chemically or enzymatically *in vitro* in the samples to be analyzed. In label-free quantitation, peak intensities or identification frequency of peptides is used for the calculation of protein abundance (Kito & Ito, 2008; Bruce et al., 2013). For the comparative studies of assembled RNP complex, the relative quantification has been carried out by using technique explained below.

1.3.1 Quantitation Based on Extracted Ion Chromatogram (XIC)

After mass spectrometric analysis, the ion chromatograms of certain mass to charge ratio of the peptides are extracted and used for quantification. The area under the curve (AUC) of the peptide peak with a particular *m/z* at a given time is directly related to the amount (Ong & Mann, 2005). This process of peak area determination is called ion extraction and its resulting chromatogram is referred as extracted ion chromatogram (XIC). Different peptides have different ionization efficiency, therefore for relative quantification, the peak area of same ion species from different samples are compared. For protein correlation profiling, the alignment of total ion chromatograms belonging to different samples is carried out. The ion species, for which MS2 spectra have been generated and their respective peptide sequences have been identified are correlated on the basis of

their chromatographic retention times to the ones of other samples either with or without any MS2 spectra. This method provides relatively reliable estimation of protein ratios (Schulze & Usadel, 2010). It is also cost effective as no labeling reagents are used and is versatile and can be applied to any type of sample. However, there are some limitations of XIC-based relative quantification approach. It can be error-prone due to variations that can occur during sample processing and analysis as well as the ion suppression effect because of co-detected ions or interfering substances like abundant background proteins etc. Some of these limitations can be minimized by spiking in the calibrant or the identical amount of standard peptide into the samples to be compared and by developing an algorithm that can align multiple runs and compare the XIC of each peptide ion (Ong & Mann, 2005; Kito & Ito, 2008).

1.4 Quantitative Nucleic Acid Studies

Mass spectrometry has been recognized as an emerging powerful tool for the identification, characterization and quantification of nucleic acids. Recently, the electrophoretic separation of RNA followed by in-gel RNase digestion and LC-MS analysis has been successfully developed for the identification of RNA along with its post-transcriptional modifications. Like proteins, the nucleic acids can also be subjected to isotope labeling. The stable isotope labeling of RNA can be carried out chemically, enzymatically or metabolically by using heavy isotopes such as ^{2}H, ^{13}C and ^{15}N. Without altering the biochemical and biophysical properties of the oligoribonucleotide, it facilitates the accurate quantification. The known concentrations or certain ratio of light and heavy isotopically labeled RNAs are combined together and are digested with RNase and analyzed by LC-ESI-MS. In the MS spectrum generated, the light and heavy peptide pairs separated by a specific mass difference can be observed. The mass difference between the differentially labeled oligonucleotides correspond to the number of atoms of the elements present. Therefore, this method can be used for the deduction of the partial elemental composition as well as the quantification of the

oligoribonucleotides. The relative quantification is achieved by comparing the area under the curve (AUC) or signal intensities of light and heavy isotopically labeled oligoribonucleotide pair. Compared to some other polymers, the data generated by nucleic acids fragmentation is very complex leading to the limited success in the nucleic acid studies by mass spectrometry (Waghmare & Dickman, 2011; Borland & Limbach, 2017). Few strategies have been developed for the identification and quantification of nucleic acids but still there is a room for further advancements (Meng & Limbach, 2005; Popova & Williamson, 2014; Paulines & Limbach, 2017).

1.5 Qualitative Studies

Various classes of proteins like histones, helicases, ligases, polymerases, transcription factors and ribosomal proteins interact with nucleic acids to play a crucial role in gene expression, gene regulation as well as cell function. Some proteins recognize and bind to a specific sequence motif of single-stranded nucleic acids whereas others recognize and bind to that of double-stranded nucleic acids or they bind to nucleic acids with a particular three dimensional structure (Steen & Jensen, 2002). The post-genomic era has faced one of the major challenges regarding mapping of these biological interactions. For identification and characterization of these interactions a variety of techniques have been established including the direct binding methods like electrophoretic mobility shift (Hellman & Fried, 2007) and fluorescence anisotropy assays (Heyduk et al., 1996), RNA-mapping methods such as chemical probing (Duval et al., 2017) and cross-linking and immunoprecipitation (CLIP) (Ule et al., 2003), genetic methods such as yeast three-hybrid system (SenGupta et al., 1996), microarray-based methods (Tenenbaum et al., 2000) and mass spectrometry methods (Kvaratskhelia & Grice, 2008). Every method has its own strengths and limitations however these limitations can be compromised by using the combination of these techniques (Gordiyenko & Robinson, 2008; Hegarat et al., 2008; Nakamura et al., 2012).

1.5.1 Protein-Nucleic Acid Cross-linking

In order to study the protein-RNA interactions, the complexes assembled either *in vivo* or *in vitro* by using recombinant components or by incubating the cellular or nuclear extracts with *in vitro* transcribed tagged RNA are purified for further analysis (Schmidt et al., 2012; Faoro & Ataide, 2014). The assembled protein-nucleic acid interactions are oftenly held together by weak non-covalent interaction forces resulting in the dissociation of interaction partners during the ionization process. In order to stabilize the protein nucleic acid complex, covalent bonds are generated between the interaction partners with cross-linking technique. The cross-linking along with mass spectrometry furnishes information about interacting partners of the complex more quickly and with great sensitivity and can subsequently be used for predicting the three-dimensional molecular modeling of protein-nucleic acid interfaces (Steen & Jensen, 2002).

The protein-nucleic acid complex can be cross-linked either by using chemical cross-linking reagents such as formaldehyde and nitrogen mustard etc. that connects the reactive groups of the interacting partners present within a certain distance range (Tretyakova et al., 2015) or by photoactivation using UV light that generates a so called zero-length cross-link, connecting the reactive groups of the interacting partners lying in close proximity with each other. It makes use of the natural reactivity of nucleobases on excitation by UV light (Meisenheimer & Koch, 1997). Moreover, there are some photo-inducible cross-linkers such as 2-iminothiolane and DTT etc. that can get activated on UV light exposure and connects the reactive groups of the interacting partners present adjacent to each other (Wower et al., 1981; Zaman et al., 2015). After cross-linking, the cross-linking site up to amino acid residue and nucleotide level can be determined by mass spectrometric analysis and bioinformatics.

1.5.1.1 UV-Induced Protein-Nucleic Acid Cross-linking

During the course of current studies, the UV-induced photo cross-linking along with mass spectrometry techniques have been used to study the interactions of protein-RNA complexes. The proteins and nucleic acids can absorb the radiation

of wavelength in the 250-280 nm range to generate photochemical cross-links between the interacting partners without distorting the protein-nucleic acid complex conformation (Steen & Jensen, 2002). However, the low cross-linking yield (1-5%) is the major limitation of this approach (Kramer et al., 2011) which can be overcome by using photosensitive analogs such as halonucleotides e.g. 4-thiouridine, 6-thioguanosine and 5-bromouracil etc. can be used. These photoactivable nucleotide analogs can be incorporated during *in vivo* or *in vitro* synthesis of oligonucleotides. They can enhance the cross-linking yield and minimize the structural distortions of the ligands (Meisenheimer & Koch, 1997; Kramer et al., 2011). In principle, all nucleotides and amino acids can undergo cross-linking reaction, however they vary in their reactivity. According to cross-linking studies, lysine, tryptophan, tyrosine, phenylalanine and cysteine are the most reactive amino acids whereas uracil is the most reactive nucleotide followed by guanine, cytosine and adenine (Shetlar et al., 1984a; Shetlar et al., 1984b; Schmidt et al., 2012).

Sample preparation, cross-linking reaction and its resulting heterogeneous mixture of products and ionization are the major challenges in structural analysis of peptide-oligonucleotide heteroconjugates by using MS-based approach.

1.5.1.2 Purification of Peptide-Oligonucleotide Heteroconjugates

The purification of the assembled complex is one of the important steps for the analysis of protein-nucleic acid interactions. The purification can be carried out by using biochemical techniques such as gradient centrifugation etc. or by affinity-based purification methods using genetically epitope-tagged proteins or tagged nucleic acids as a bait for macromolecular complex or by using antibodies against one of the nucleic acid interacting proteins and nucleic acid (Schmidt et al., 2012).

Depending upon the complexity of the sample, the peptide-oligonucleotide heteroconjugates are purified from non-cross-linked components by using combination of various chromatographic techniques. The peptide-nucleic acid heteroconjugates derived by the hydrolysis of highly complex protein-nucleic acid

cross-linked sample with endoproteinase can be subjected to size exclusion chromatography under denaturing conditions to separate cross-linked peptide-nucleic acid heteroconjugates and non-cross-linked nucleic acid from the non-cross-linked peptides (Urlaub et al., 1995; Urlaub et al., 2002). Thereafter, the nucleic acid is hydrolyzed with nucleases and the peptide-oligonucleotide heteroconjugates are separated from the non-cross-linked oligonucleotides by C18 reversed phase liquid chromatography (RP-LC). The peptide-oligonucleotide heteroconjugates are more hydrophobic and bind to the C18 RP column whereas the non-cross-linked oligonucleotides do not bind and are easily removed (Urlaub et al., 2008). Owing to the nucleic acid phosphate backbone, the peptide-oligonucleotide heteroconjugates behave the same as phosphopeptides. Therefore, to enrich the heteroconjugates the same enrichment protocols can be adopted as for phosphopeptides for example TiO_2 solid-phase extraction (Larsen et al., 2005; Luo et al., 2008). The peptide-oligonucleotide heteroconjugates can be enriched by adsorbing to the titanium dioxide beads in the presence of DHB, providing the acidic environment for the competitive binding of the peptide-oligonucleotide heteroconjugate against unphosphorylated peptides (Richter et al., 2009) (Figure 1.4).

The relatively less complex samples composed of few proteins and short nucleic acid, after reconstitution and cross-linking, can be hydrolyzed by endoproteinases and nucleases and the peptide-oligonucleotide heteroconjugates are purified directly by using C18 RP-LC and TiO_2 solid-phase extraction (Kramer et al., 2011).

1.5.1.3 Mass Spectrometric Analysis of Protein-Nucleic Acid Cross-links

Since peptide-oligonucleotide heteroconjugate is composed of two different types of polymers, exhibiting different physicochemical properties therefore they require different conditions for ionization in MS. The peptides and oligonucleotides are ionized in positive and negative ion modes respectively. During the current studies the interest lies in the identification of the protein region interacting with RNA therefore the ionization is performed in positive ion mode (Schmidt et al.,

2012). However, the presence of excess of non-cross-linked components hampers the ionization of the cross-links. The increased hydrophilicity of peptide-oligonucleotide heteroconjugates due to presence of oligonucleotide moiety cross-linked to the peptide in comparison to the unmodified non-cross-linked peptides may cause lower ionization efficiency. For the relative ionization improvement, the oligonucleotide part of the cross-linked heteroconjugate should be made as small as possible, maximum up to four nucleotides by using nucleases (Steen & Jensen, 2002; Qamar et al., 2015).

For the mass spectrometric analysis of peptide-oligonucleotide heteroconjugates, the HCD fragmentation has proved to be better than the CID fragmentation methods. The mass spectrometers with orbitrap analyzers carry out the HCD fragmentation with high accuracy. It helps in differentiating the distinct signals generated by peptide and nucleotide fragmentation such as the signals of immonium ion of tyrosine (*m/z* 136.0762) and the RNA marker ion of adenine (*m/z* 136.0623). In addition, the peptide-oligonucleotide heteroconjugate spectrum generated by HCD fragmentation usually has long y-ion series, high intensity signals of a2 and b2 ions, signals of immonium and internal ions and nucleic acid marker ions that improve its identification.

1.5.1.4 Protein-Nucleic Acid Cross-links Data Analysis

The low signal intensity in MS/MS spectrum and the wide variety of potentially cross-linked nucleotide fragments has made the interpretation of the data obtained by the mass spectrometric analysis of protein-nucleic acid cross-linking, very challenging and laborious. The cross-linking is usually an additive reaction. The molecular weight of the peptide-oligonucleotide heteroconjugate is the sum of the molecular weight of the peptide and the oligonucleotide moiety cross-linked to it. The MS/MS spectra obtained are usually prevailed by the signals of peptide fragments. In this case, the cross-linked nucleotide moiety can only be deduced by calculating the mass difference between the experimental peptide-oligonucleotide heteroconjugate and the peptide (Kramer et al., 2011).

In recent years, a semi-automated data analysis approach has been developed for the unbiased analysis of peptide-oligonucleotide heteroconjugates. It is comprised of RNP^{xl} tool (Kramer et al., 2014) in OpenMS environment (Sturm et al., 2008; Bertsch et al., 2011) using OMSSA (Geer et al., 2004) as the search engine. The oligonucleotide fragment mass is dealt as variable modification while searching for PTM against database. After endonuclease digestion and TiO_2 enrichment, the heteroconjugates with maximum four nucleotides are possible. Taking this into consideration, 69 different mass combinations out of four different nucleotides and 829 different mass combinations out of four different nucleotides along with RNA/DNA modifications like loss of H_2O and HPO_3 etc. are possible. During the database search, these mass variants are used to generate the theoretical precursor fragment spectra for every original spectrum. By subtracting the molecular weight of the oligonucleotide from the experimentally determined molecular weight of heteroconjugate, the molecular weight of the peptide can be deduced. The sum of combination of oligonucleotide fragment mass along with peptide mass and spectrum which fits to the experimental precursor mass along with its candidate spectrum will yield a most probable hit. Additionally, different filters can be applied according to the experimental design such as for comparing the experimental sample with the control one and also for removing the pure peptide hits etc. (Kramer et al., 2014). The manual inspection of the MS/MS spectrum is immensely important to screen the exact amino acid and nucleotide, undergone cross-linking reaction. The presence of signals of marker ions of nucleic acid base resulting by the nucleic acid fragmentation and the shift of b or y ion series or signals of immonium and internal ions by the mass of cross-linked nucleotide fragment or adduct, indicate the cross-linked amino acid along with cross-linked nucleotide (Qamar et al., 2015). Oftenly, the cross-linking bond formed between peptide and oligonucleotide is labile to HCD fragmentation resulting in the identification of the cross-linked peptide but the identification of the single cross-linked amino acid residue is no longer possible. So far, the unavailability of a software that can handle all the aspects of peptide-oligonucleotide heteroconjugate fragment spectra makes the requirement of the

completely automated system for protein-nucleic acid cross-linking data indispensable.

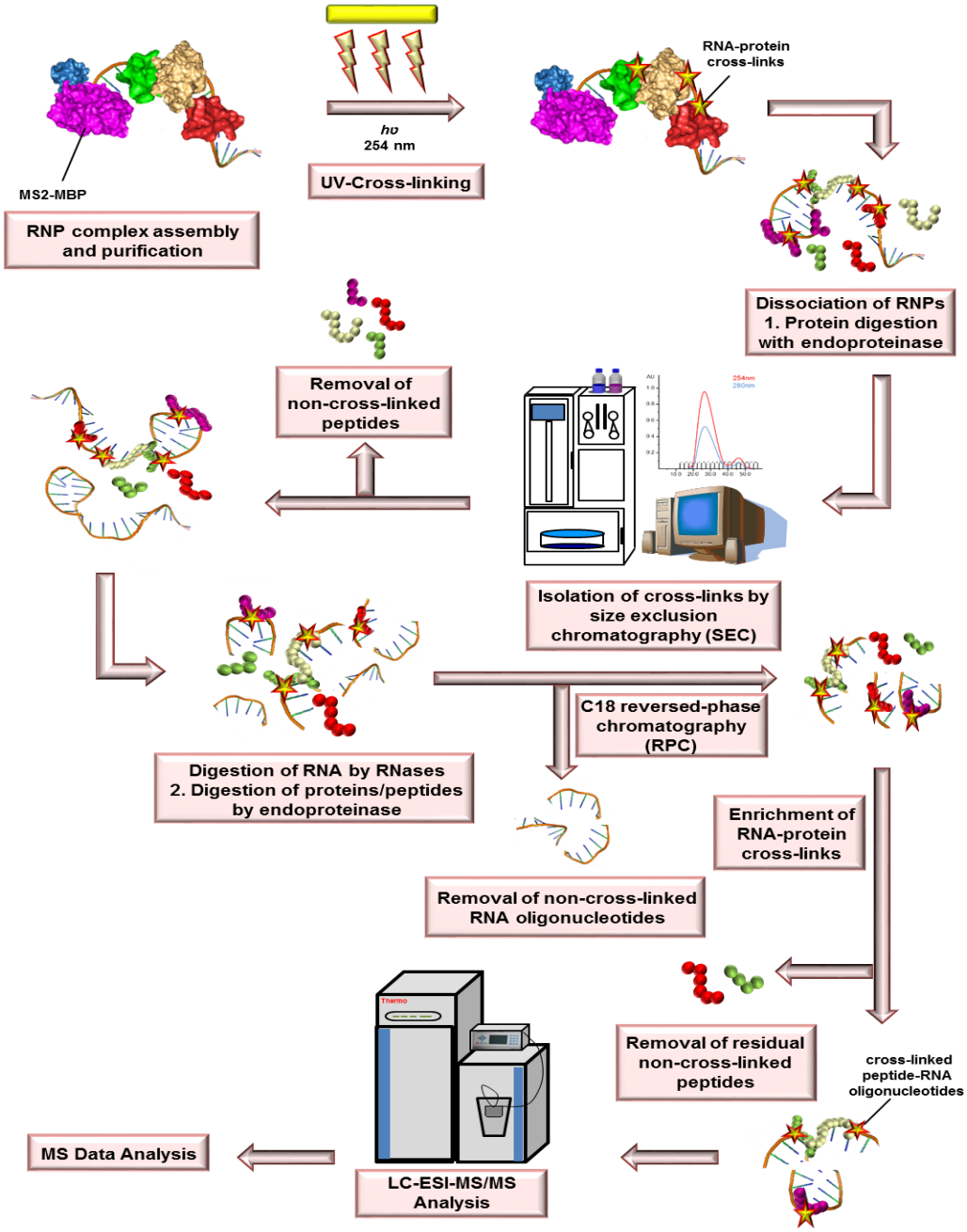


Figure 1.4: Workflow of cross-linking protocol (Figure adapted from Qamar et al., 2015). The *in vitro* transcribed MS2-tagged pre-mRNA is incubated with *HeLa* nuclear extract for the RNP complex assembly. The assembled RNP complex is purified. The purified RNP complex is then UV-cross-linked. The proteins in the sample are digested with trypsin. The cross-links are isolated and the non-cross-linked peptides are removed by administering the sample to the SEC. The RNA is hydrolyzed by the RNases. The non-cross-linked RNA oligonucleotides are removed by RPC. The RNA-protein cross-links are subsequently enriched by using TiO_2 solid phase extraction. The sample is then subjected to LC-ESI-MS/MS analysis. The RNP^{xl} pipeline in an OpenMS environment is used to analyze MS data.

1.6 Biological Complexes Studied Using UV-Induced Cross-linking and Mass Spectrometry

1.6.1 Brat-NHL-hb RNA Complex

The post transcriptional gene regulation plays a crucial role in the regulation of metabolism and development. Generally, the RNA binding proteins (RBPs) interact by recognizing the cis-acting sequences in the 3' UTR of target mRNAs and regulate their translation, stability and localization. Embryogenesis in Drosophila, is one of the context in which the translational regulation of maternally encoded mRNAs is indispensable. The translational repression of maternal hunchback (hb) mRNA at the posterior results in the development of the anterior–posterior axis (Irish et al., 1989; Struhl, 1989). Pumilio (Pum) along with two cofactors binds to the specific sequences (Nanos response elements, NREs) located in the hb mRNA's 3' untranslated region (UTR) to repress the translation of hb mRNA (Wharton et al., 1998; Sonoda & Wharton, 1999; Sonoda & Wharton, 2001; Wang et al., 2002). These cofactors include Nanos (Nos) protein that plays role in generation of abdominal segmentation (Wang & Lehmann, 1991) and Brain Tumor (Brat) protein, which is a NHL domain protein (named after NCL-1, HT2A, and LIN-41), and is recruited through interactions with Pum and Nos (Slack & Ruvkun, 1998; Adams et al., 2000; Arama et al., 2000). The NHL domain of TRIM-NHL proteins has been reported as RBD (Kwon et al., 2013). It is arranged in six-bladed β propeller structure, similar to the WD40 fold (Edwards et al., 2003) and has a positively charged top surface showing the ability to interact negatively charged molecules like nucleic acids. The two Nanos Response Elements (NREs) located in the hb mRNA's 3′UTR has found to be responsible for the regulation of hb mRNA translation by BRAT-PUM-NOS complex. Each NRE is comprised of two sequence motifs designated as BoxA and BoxB. The Pum protein binds to the BoxB motif, containing PUM consensus-binding site, UGUANAUA where N = A/C/G/U via conserved RNA-binding domain (RBD), PUM's C-terminal PUF-homology domain. The previous studies suggest

that Pum and Nos proteins directly and Brat protein indirectly interact with NREs (Wharton & Struhl, 1991; Murata & Wharton, 1995; Sonoda & Wharton, 2001). Recently it has been reported that Brat protein's C-terminal NHL domain directly binds to the sequences in and around hb's BoxA motifs in a PUM-independent manner (Loedige et al., 2015).

1.6.2 CWC2-U6/U4 snRNAs Complex

The eukaryotic mRNA processing involves the removal of introns from pre-mRNA by the spliceosomes. There are five main components of spliceosomes: U1, U2, U4, U5 and U6 snRNPs (small nuclear RNPs). In addition to this, the splicing reaction requires various protein complexes. The NTC (nineteen complex) which is comprised of scaffold protein Prp19 and a number of associated splicing factors is essential for the stable association of the U5 and U6 snRNPs with spliceosome (Chan et al., 2003). It specifies the interactions among RNAs and between RNA and protein (Villa & Guthrie, 2005; Hogg et al., 2010). The yeast NTC contains minimum ten components (Ajuh et al., 2000), out of which only CWC2/NTC40 (hereafter referred to as CWC2), has a tendency to bind to RNA. The CWC2 protein is known to contain two N-terminal RNA-binding motifs i.e. CCCH-type ZnF (zinc finger) and an RRM (RNA recognition motif) and a flexible C-terminus that can interact with Prp19 protein's WD40 domain (Vander et al., 2010). The CWC2 protein as a whole, exhibits the normal capacity to bind RNA with low sequence specificity, whereas this binding affinity has found to be reduced by RRM along with flexible C-terminus region of CWC2 (McGrail et al., 2009). However, the exact CWC2-RNA interaction mechanism is unclear. In yeast, the CWC2 has found to be interacting with U6 snRNA during splicing (McGrail et al, 2009). It has been observed that the CWC2 can cross-link to many catalytically significant RNA regions like U6-ISL, a region upstream to the ACAGAGA box, and the pre-mRNA intron close to the 5' SS, arranging the CWC2 at or near the catalytic center of the purified catalytically active spliceosome. *In vitro*, the absence of CWC2 results in the assembly of

catalytically inactive B^{act}-like complex (Rasche et al, 2012). Hence the CWC2 may involve in the generation of an active conformation of the spliceosome's catalytic center by interacting with various catalytic inter RNA interaction network sites.

1.6.3 RNP Complexes from *HeLa* Nuclear Extract

In eukaryotic cells mRNA, an intermediary between transaction of information and execution of information, is a highly regulated molecule subjected to RNA processing and regulation (Mitchell & Parker, 2014). It is an established fact that mRNA concentrations and protein concentrations are less than perfectly correlated and protein synthesis regulates the gene expression (Sonenberg & Hinnebusch, 2009; Schwanhausser et al., 2011). RNA biology is contrived by the interplay of RNA with RNA binding proteins in (RBPs) in ribonucleoproteins i.e. RNPs (Glisovic et al., 2008; Li, 2008). mRNA control modulation is driven through interactions between individual mRNA and proteins. The resultant assemblies are comprised of complex structures designated as mRNPs. Importance of ribonucleoproteins is accentuated by the fact that they perform major functions of expression and regulation including but not limited to protein synthesis by ribosomes, telomerase RNP targeted chromosome maintenance, editing of mRNA by spliceosome, RNase P directed leader sequence removal from tRNA, small nucleolar (sno)RNPs administered synthesis of ribosomal RNAs and small nuclear RNAs (snRNAs), translocation of protein by signal recognition particles (SRPs), editing complex involvement in RNA editing and finally gene silencing by siRNPs or miRNPs which contain small interfering RNAs and microRNAs respectively (Li, 2008). Disruption and dysregulation of proper expression and function of RBPs can contribute to major human diseases like cancer, metabolic disorders, neurodegenerative and myotonic dystrophy etc. (Castello et al., 2012).

The understanding of eukaryotic mRNAs is subjected to the understanding of constituents, assemblage, rebuilding processes and function of mRNPs.

Underlying molecular functions can be unrevealed by the precise knowledge about RBP binding sites.

Past decades present robust interplay between innovative method development, analysis and establishment regarding RNA biology (McHugh et al., 2014). Different traditional and state of the art technologies are adopted for both *in vitro* and *in vivo* methods, to study RNAs bound by individual RBPs, or *vice versa*. RNA-protein complexes are the valuable source of insight information about the plethora of binding modes and structural implications of interactions (Hall, 2002). In addition to the conventional way of identifying mRNA binding protein through analysis of RNA processing, translation and degradation, more dynamic way of identifying them is by means of RNA binding domains. RNA-interacting proteins can be categorized according to the RNA motifs with which they interact (Li, 2008).

It is quite appreciable that although most RNA binding proteins are comprised of few RNA associating modules, the functional repertoire of these proteins is broadened through copies of multitude structural variability of RNA binding domains (RBDs) to manage the great structural diversity of the substrate i.e. RNA. There are specific combinations of RNA binding modules that structurally and functionally arrange to facilitate myriad of different interactions and regulatory matters (Lunde et al., 2007). Multiple copies of modules confer proteins with attributes of enormous affinity and specificity in comparison to the individual domains that may offer weak affinity for short stretches of RNA. Consequently, through various modules option, interaction surfaces can be created presenting more affinity and specificity for a specific target. Interactions of such sort bear the privilege of controlling the formation of complexes through disassembling whenever is required. Not only specificity of individual domain is important but also the relative arrangement of domains is key to proper functionality.

The RNA recognition motif (RRM) or RNA binding domain/Ribonucleoprotein domain can modulate its fold in order to cater the task of identifying many RNA

and proteins so as necessary to perform biological functions relevant to post transcriptional gene regulation (Maris et al., 2005).

Examples of classical RBDs include RNA-recognition motif (RRM), K-homology (KH), double-stranded RNA-binding domain (dsRBDs), RGG box, cold shock domain (CSD), Zinc finger and the Pumilio (PUM) domain (Lunde et al., 2007; Castello et al., 2016) etc.

The studies conducted by using HEK298 (Baltz et al., 2012), *HeLa* (Castello et al., 2012) and mESC (Kwon et al., 2013) cells, suggest that in addition to classical RBDs, several RNA-interacting proteins possess non-classical RBDs.

1.7 Aim of the Study

In recent times much emphasis has been laid on revealing the composition and regulation of various RNP complexes. In order to have detailed insight into the RNA-protein interactions governing the gene regulation, various mass spectrometric approaches have been developed. But the limitation of most of the purification methods such as sample degradation, loss of interacting partners of the RNP complexes and low overall yields etc. as well as the intricate mass spectrometric data analysis have hampered the study of these RNP complexes.

During the course of study, much attention has been paid to overcome the aforementioned limitations. In addition to RNP complexes comprised of one or few proteins, highly complex RNP assemblies composed of several macromolecules have been studied by using the UV-induced cross-linking approach coupled with mass spectrometry.

Generally, the low cross-linking yield and the presence of non-cross-linked peptides and oligonucleotides after endoproteinase and nuclease digestion, pose hurdle in mass spectrometric analysis of RNA-protein heteroconjugates. One of the aims is to optimize the existing protocols described by Deckert et al., 2006 for isolation of RNP complex and by Urlaub et al., 2002 and Luo et al., 2008 for

purification and enrichment of cross-links, so that they can be used to study the interactions of the proteins of highly complex sample like RNP complex assembled from *HeLa* nuclear extract which can later be detected by mass spectrometric approach.

Various proteins may or may not interact differently with miscellaneous RNAs within RNP complexes. In order to have a detailed insight, the quantitative analysis of the RNA-protein cross-links has been established during the current study. The relative quantification approach can help in the comparative analysis of different domains and motifs of single or multiple proteins, acting differently with different RNAs.

2 MATERIALS AND METHODS

2.1 Materials

All media and buffers were prepared in deionized distilled water. The solutions and buffers used for cross-linking and MS analysis were prepared in LiChrosolv® water. The media were sterilized by autoclaving at 121 °C and 15 lbs/in^2 pressure for 15 min. All buffers and solutions were filter sterilized by using filter of 0.20 µm pore size. Cell culture grade plasticware was also used. All the glassware and plasticware were sterilized at 180 °C for 1 h by using hot air oven.

2.1.1 Chemicals and Reagents

Acetic acid	Merck, Darmstadt
Acetone	Merck, Darmstadt
Acetonitrile (LiChrosolv)	Merck, Darmstadt
Agarose, Ulra Pure	Invitrogen, Carlsbad
Ammonium bicarbonate	Fluka, Switzerland
Ammonium hydroxide solution (28–30% (v/v))	Merck, Darmstadt
Ammonium peroxydisulfate	Merck, Darmstadt
Boric acid	Merck, Darmstadt
Bovine serum albumin (BSA)	Sigma-Aldrich, Steinheim
Bromophenol blue (sodium salt)	Sigma-Aldrich, Steinheim
Calcium chloride dihydrate	Merck, Darmstadt
Chloroform	Merck, Darmstadt
Coomassie Brilliant Blue G-250	Fluka, Switzerland
2,5-Dihydroxybenzoic acid (DHB)	Sigma-Aldrich, Steinheim
di-Potassium hydrogen phosphate trihydrate	Merck, Darmstadt
di-Sodium hydrogen phosphate	Merck, Darmstadt
Dithiothreitol (DTT)	Roth, Karlsruhe
Ethanol	Merck, Darmstadt
Ethidium bromide	Roth, Karlsruhe
Ethylenediaminetetraacetic acid (EDTA-Na_2 salt)	Roth, Karlsruhe
Formaldehyde	Merck, Darmstadt
Formaldehyde-d_2 solution (D_2O)	Sigma-Aldrich, Steinheim
Formamide	Merck, Darmstadt
Formic acid	Fluka, Switzerland
Glycerol	Merck, Darmstadt

Glycogen	Roche, Mannheim
Hydrochloric acid	Merck, Darmstadt
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Merck, Darmstadt
Iodoacetamide	Sigma-Aldrich, Steinheim
Magnesium acetate	Merck, Darmstadt
Magnesium chloride	Sigma-Aldrich, Steinheim
Maltose	Merck, Darmstadt
Methanol (LiChrosolv)	Merck, Darmstadt
N,N,N',N'-Tetramethylethylendiamine (TEMED)	Sigma-Aldrich, Steinheim
NP-40	Sigma-Aldrich, Steinheim
Ortho-Phosphoric acid	Merck, Darmstadt
Phenylmethylsulfonyl fluoride (PMSF)	Roche, Mannheim
Potassium acetate	Merck, Darmstadt
Potassium chloride	Merck, Darmstadt
Potassium dihydrogen phosphate	Merck, Darmstadt
Potassium hydroxide	Merck, Darmstadt
Silver nitrate	Merck, Darmstadt
Sodium acetate	Merck, Darmstadt
Sodium carbonate	Merck, Darmstadt
Sodium chloride	Merck, Darmstadt
Sodium dihydrogen phosphate monohydrate	Merck, Darmstadt
Sodium dodecyl sulfate (SDS)	Merck, Darmstadt
Sodium hydroxide	Merck, Darmstadt
Trifluoroacetic acid (TFA)	Roth, Kalsruhe
Tris (hydroxymethyl) aminomethane (Tris base)	Roth, Kalsruhe
Urea	Sigma-Aldrich, Steinheim
Water (LiChrosolv)	Merck, Darmstadt
Xylene cyanol FF	Fluka, Switzerland

2.1.2 Commercial Buffers, Solutions and Kits

1 M Triethylammonium bicarbonate (TEAB) pH 8.5	Sigma-Aldrich, Steinheim
10 X CutSmart® buffer	New England Biolabs, USA
10 X NEBuffer 3.1	New England Biolabs, USA
100 bp DNA Ladder, extended	Roth, Kalsruhe
5 X Transcription buffer	Promega, Mannheim
6 X DNA gel loading dye	New England Biolabs, USA
Bradford solution (Bio-Rad Protein Assay)	Bio-Rad, Munich
NuPAGE® Antioxidant	Invitrogen, USA
NuPAGE® LDS sample buffer (4 X)	Invitrogen, USA
NuPAGE® MOPS SDS Running buffer (20 X)	Invitrogen, USA
NuPAGE® Sample reducing agent (10 X)	Invitrogen, USA
Phenol:Chloroform:Isoamyl alcohol (PCI) (25:24:1)	Roth, Karlsruhe
Pre-cast 4-12% Bis-Tris Gel	Invitrogen, USA
Precision Plus Protein™ Standards	Bio-Rad, Munich

PureLink® HiPure Filter Plasmid Maxiprep Kit	Invitrogen, USA
Rotiphorese Gel 40 (38% Acrylamide, 2% Bis-Acrylamide)	Roth, Karlsruhe

2.1.3 Enzymes and Enzyme Inhibitors

BamHI enzyme	New England Biolabs, USA
BamHI-HF enzyme	New England Biolabs, USA
Benzonase	Novagen, DE
DNase I	Roche, Mannheim
EDTA-free Protease Inhibitor Cocktail	Roche, Mannheim
RNase A	Ambion, Applied Biosystems, Darmstadt
RNaseT1	Ambion, Applied Biosystems, Darmstadt
RNAsin	Promega, Mannheim
RQ1 RNase-Free DNase	Promega, Mannheim
SP6 RNA polymerases	Promega, Mannheim
StyI enzyme	New England Biolabs, USA
T7 RNA polymerases	Department of Cellular Biochemistry, MPI-BPC, Goettingen
Trypsin, modified (sequencing grade)	Promega, Mannheim
XbaI enzyme	New England Biolabs, USA

2.1.4 Nucleotides

Adenosine-5'-Triphosphate (rATP, 100 mM)	Promega, Mannheim
Cytosine-5'-Triphosphate (rCTP, 100 mM)	Promega, Mannheim
Guanosine-5'-Triphosphate (rGTP, 100 mM)	Promega, Mannheim
Uridine-5'-Triphosphate (rUTP, 100 mM)	Promega, Mannheim
m^7G(5')ppp(5')G cap	Kedar, Warsaw
Uridine-$^{13}C_9$-5′-Triphosphate, sodium salt solution (100 mM)	Sigma-Aldrich, Steinheim
Uridine-$^{13}C_9$,$^{15}N_2$- 5′-Triphosphate, sodium salt solution (100 mM)	Sigma-Aldrich, Steinheim
Uridine-5'-Triphosphate, ammonium salt solution (5-D_1, ribose-3',4',5',5"-D_4) (100 mM)	Euriso-top, Saarbrücken

2.1.5 Plasmids

All plasmids were obtained by the courtesy of Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany.

pMINX
pMS2-MBP
pPM5
pT7U4
pT7U6

2.1.6 Cell line

Following *Escherichia coli* (*E.coli*) strains used during the study were provided by Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany.

Rosetta II
DH5α

HeLa S3 cell line (human cervical cancer, Computer cell culture center, BE) was kindly provided by Thomas Conrad (Facility for Cell Production, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany).

2.1.7 Cell Culture Materials

100 X L-Glutamine	PAA Laboratories, Colbe
100 X Penicillin/Streptomycin	PAA Laboratories, Colbe
Ampicillin	Sigma-Aldrich, Steinheim
Chloramphenicol	Roth, Karlsruhe
Dialyzed Fetal Bovine Serum	PAA Laboratories, Colbe
High Glucose Dulbecco's Modified Eagle's Medium (DMEM)	PAA Laboratories, Colbe
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich, Steinheim
L-Arginine (Arg0)	Sigma-Aldrich, Steinheim
LB-Agar	MP Biomedicals, USA
LB-Medium	MP Biomedicals, USA
L-Lysine (Lys0)	Sigma-Aldrich, Steinheim

2.1.8 Chromatographic Materials and Consumables

HiPrep 26/10 desalting cartridge	GE Healthcare, Freiburg
Gravity flow disposable chromatography columns	Bio-Rad, Munich
Amylose resin	New England Biolabs, USA
Heparin Sepharose HiTrap column, 5 ml	GE Healthcare, Freiburg
Reprosil-Pur basic C18-AQ, 120 Å, 5 µm	Dr. Maisch, Ammerbuch
TiO_2 titansphere 5 µm	GL Sciences, Tokyo
Superdex 75 PC 3.2/30	GE Healthcare, Freiburg

2.1.9 Solutions and Buffers

10 X PBS	1300 mM NaCl, 200 mM KPO_4 (pH 7.4)
10 X TBE	890 mM Tris base, 890 mM Boric acid, 20 mM EDTA-NaOH (pH 8)
5% Polyacrylamide gel solution	12.5 ml Rotiphorese gel 40, 8 M Urea, 1 X TBE
Amylose matrix wash buffer	20 mM HEPES-KOH (pH 7.9), 150 mM NaCl, 0.05% NP-40
Colloidal Coomassie Staining	0.12% (w/v) Coomassie Brilliant Blue G-250, 20% (v/v) Methanol, 10% (v/v) Ortho-Phosphoric acid, 10% (w/v) Ammonium sulfate
Gradient buffer	150 mM NaCl, 20 mM HEPES-KOH (pH 7.9), 1.5 mM $MgCl_2$
Heparin elution buffer	20 mM HEPES-KOH (pH 7.9), 100 mM KCl, 10% (w/v) Glycerol
MC buffer	10 mM KOAc, 10 mM HEPES-KOH (pH 7.6), 0.5 mM $Mg(OAc)_2$
MS2-MBP lysis buffer	20 mM Tris-Cl (pH 7.4), 200 mM NaCl, 1 mM DTT, 3 tablets EDTA-free protease inhibitor cocktail, 1 mM $MgCl_2$, Few corns of DNase I

Na-P buffer	5 mM Na_2HPO_4
RNA sample loading buffer	80% (w/v) Formamide, 1 mM EDTA, 0.05% (w/v) Bromophenol blue, 0.05% (w/v) Xylene cyanol
Roeder C buffer	20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA (pH 8), 420 mM NaCl, 1.5 mM $MgCl_2$, 25% (w/v) Glycerol
Roeder D buffer	20 mM HEPES-KOH (pH 7.9), 0.2 mM EDTA (pH 8), 100 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM PMSF, 10% (w/v) Glycerol
Size exclusion running buffer	20 mM Tris base, 1.5 mM $MgCl_2$, 150 mM NaCl
Tio_2 Buffer A	200 mg of 2,5-dihydroxybenzoic acid in 1 ml buffer B
Tio_2 Buffer B	80% (v/v) Acetonitrile, 5% (v/v) Trifluoroacetic acid
Tio_2 Buffer C	0.3 N Ammonium hydroxide solution pH >10.5
TNES elution buffer	20 mM Tris-Cl (pH 7.5), 150 mM NaCl, 0.5% (w/v) SDS, 0.2 mM EDTA (pH 8)

2.1.10 Softwares

Adobe Creative Suite 4	Adobe Systems, San Jose
Skyline (v.4.1.0.18169)	MacCoss Lab Software
STRING (v.10)	ELIXIR's Core Data
MaxQuant and Andromeda (v.1.5.2.8)	Max-Planck-Gesellschaft, Munich
Microsoft Office 2007	Microsoft Corporation, Redmont
OpenMS (v.1.11.0)	Eberhard-Karls Universität, Tübingen
Proteome Discoverer (v.1.10)	Thermo Fischer Scientific, Schwerte
Pymol (v.1.8)	Schrödinger, New York
Xcalibur (v.2.2)	Thermo Fischer Scientific, Schwerte

SPSS Statistics software (v.17.0)	SPSS Inc.,Chicago
Cytoscape (v.3.7.0)	Cytoscape Consortium

2.1.11 Laboratory Consumables

Black polypropylene 96-well microtiter plate	Greiner Bio-One, Frichenhausen
Bottle-top filters 0.20 µm	Sarstedt, Nümbrecht
Coffee filter paper	Melitta, Minden
Combitips	Eppendorf, Hamburg
Conical centrifuge tubes (15 ml, 50 ml)	Sarstedt, Nümbrecht
Corex glass tubes	Kendro, USA
Custom made glass dishes	MPI-BPC, Goettingen
Dialysis tube clips	Spectrum Laboratories, Rancho Domingues
Disposable cuvettes	Sarstedt, Nümbrecht
Filtropur S 0.20 µm	Sarstedt, Nümbrecht
Gel cutting device (in-house designed)	MPI-BPC, Goettingen
Kontes® Glass Dounce homogenizer	VWR International, Radnor
LC sample vials	Waters, Eschborn
Microcentrifuge tubes (0.5 ml, 1.5 ml, 2 ml, 5 ml)	Eppendorf, Hamburg
Parafilm sealing film	Bemis Company Inc., Neenah
pH paper	Roth, Karlsruhe
Pipettes	Gilson, Limburg-Offheim
Pipetus® Pipetting aid	Hirschmann Laborgeräte, Eberstadt
Spectra/Por Dialysis membrane (MWCO 6-8000)	Spectrum Laboratories, Rancho Domingues
Syringe (1 ml, 5 ml, 20 ml, 50 ml)	B. Braun Melsungen AG, Melsungen
TLC plate	Sigma-Aldrich, Steinheim

2.1.12 Laboratory Equipment and Instruments

Amersham Biosciences EPS 301 Power Supply	GE Healthcare, Freiburg
Autoclave, Varioklav Steam Sterilizer H+P	Thermo Fischer Scientific, Schwerte
BBD 6220 CO_2 incubator	Heraeus, Hanau
Bioreactor 5 L with ez-Control	Applikon, Schiedam
Cryofuge 6000i Swing Bucket Rotor	Heraeus, Hanau
EASY-nLC II System	Thermo Fischer Scientific, Dreieich
Eppendorf 5415R Refrigerated Microcentrifuge	Eppendorf, Hamburg
Eppendorf BioPhotometer	Eppendorf, Hamburg
Gel Doc 2000 System	Bio-Rad, Munich
Heating Bath	IKA®-Werke, Staufen
Heraeus BK-600 Cooled Incubator	Heraeus, Hanau
Heraeus® Megafuge® Swing Bucket Rotor 2704	Heraeus, Hanau
Heraeus® Megafuge® 1.0 R Centrifuge	Heraeus, Hanau
Heraeus® HERAsafe® KS Safety Cabinet	Heraeus, Hanau

Heraeus™ Fresco™ Microcentrifuge	Heraeus, Hanau
Heraeus™ Pico™ Microcentrifuge	Heraeus, Hanau
HPLC, 1100 Series	Agilent, USA
Infors HT Shaking Incubator	Infors HT, Einsbach
LTQ Orbitrap Velos Mass Spectrometer	Thermo Fischer Scientific, Dreieich
LTQ Orbitrap XL Mass Spectrometer	Thermo Fischer Scientific, Dreieich
Magnetic Stirrer	IKA®-Werke, Staufen
Microfluidics HC-5000 950 ml/min Lab Fluidizer	Microfluidics, Worcestershire
MPC227 Dual Purpose Conductivity/pH/T Meter	Metler-Toledo, Giesen
NanoDrop 1000 Spectrophotometer	Thermo Scientific, Wilmington
Novex® NuPAGE® SDS-PAGE Gel System	Invitrogen, USA
Perfection V700 PRO scanner	Epson, Nagano
Pharmacia SMART System/uPeak Micro-Purification System	GE Healthcare, Freiburg
Polyacrylamide Gel Electrophoresis System	MPI-BPC, Goettingen
Power Pac 200 Electrophoresis Power Supply	Bio-Rad, Munich
Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer	Thermo Fischer Scientific, Dreieich
Savant SpeedVac Concentrators	Thermo Scientific, Braunschweig
SONOREX Super Ultrasonic Bath	BANDELIN Electronic, Berlin
Sorval SA600 rotor	Kendro, USA
Sorvall Evolution RC Superspeed Centrifuge	Kendro, USA
Sorvall Hb-6 rotor	Kendro, USA
Sorvall SS-34 rotor	Kendro, USA
Sub-Cell® GT Agarose Gel Electrophoresis System	Bio-Rad, Munich
Thermomixer Comfort	Eppendorf, Hamburg
Ultrospec 3000 UV/Visible Spectrophotometer	GE Healthcare, Freiburg
UV cross-linking equipment with four 8-W germicidal lamps (G8T5)	MPI-BPC, Goettingen, Sankyo Denki, Japan
UV Transilluminator 2000	Bio-Rad, Munich
Vortex Gene 2	Thermo Fischer Scientific, Schwerte
VXR basic Vibrax®	IKA®-Werke, Staufen
Weighing balance	Sartorius, Goettingen

2.2 Methods

2.2.1 Standard Molecular Biology Methods

2.2.1.1 Preparation of Competent Cells

The chemically competent cells of *E.coli* (DH5α and Rosetta II) were prepared according to the standard method ($CaCl_2$-method) (Sambrook et al., 1989). DH5α and Rosetta II strains were used for *in vivo* plasmid amplification and protein expression respectively.

For the preparation of pre-culture, 5 ml of LB (lysogeny broth) medium was inoculated with *E.coli* strains and incubated at 37 °C overnight with constant shaking. This pre-culture was used to inoculate 250 ml of LB medium. The culture was then incubated at 37 °C with continuous shaking at 180 rpm till the OD_{600} of 0.4 was attained. The cells were left on ice for 10 min followed by harvesting by centrifugation at 3000 rpm for 15 min at 4 °C. The cell pellet was resuspended in 50 ml of cold 50 mM $CaCl_2$ and incubated on ice for 30 min. The cells were then centrifuged again at 3000 rpm for 10 min at 4 °C. The supernatant was discarded and the cells were resuspended in 3 ml of cold 50mM $CaCl_2$ containing 10% w/v sterile glycerol, aliquoted in 50 ul, frozen in liquid nitrogen and stored at -80 °C.

2.2.1.2 Transformation of DH5α Cells

Transformation of DH5α cells was carried out by the heat shock method (Sambrook et al., 1989). 350-400 ng of the plasmid DNA was incubated with 50µl of competent cells on ice for 30 min. The cells were given heat shock at 42 °C for 90 sec and immediately chilled on ice for 1 min. Then 1 ml of LB-medium was added and the cells were incubated at 37 °C for 1 h with constant shaking. The cells were centrifuged at 8000 rpm for 2 min. Approximately 850 µl of supernatant was discarded and the cell pellet was resuspended in the rest of the supernatant. The cells were then platted on the selection media containing 100 µg/ml of

ampicillin and incubated overnight at 37 °C for the isolation of transformed bacterial colonies. Single transformed bacterial colonies were selected for the inoculation of LB medium containing 100 µg/ml of ampicillin and incubated at 37°C for overnight. Plasmid DNA was isolated from the transformed strain by using Invitrogen PureLink® HiPure Filter Plasmid Maxiprep Kit according to the manufacturer's instructions.

2.2.1.3 Determination of Nucleic Acid Concentration

Concentrations of the nucleic acids were determined by measuring the absorbance of aqueous solutions at a wavelength of 260 nm against a reference on NanoDrop spectrophotometer. The equations used for the calculation of the concentrations are as follows (Sambrook et al., 1989)

1 OD_{260} = 50 µg/ml double stranded DNA

1 OD_{260} = 33 µg/ml single stranded DNA

1 OD_{260} = 40 µg/ml single stranded RNA

2.2.1.4 Restriction Digestion of Plasmid DNA

Restriction digestion of different plasmid DNAs was performed in order to generate the template DNAs for *in vitro* transcription. For the generation of template DNA, 200 µg of PM5 plasmid DNA was digested with 100 U of BamHI-HF enzyme, 200 µg of MINX plasmid DNA was digested with 200 U of XbaI enzyme, 200 µg of U6 plasmid DNA was digested with 200 U of BamHI enzyme using 140 µl of 10 X CutSmart® buffer making volume upto 1400 µl with RNase-free water, whereas 200 µg of U4 plasmid DNA was digested with 200 U of StyI enzyme using 140 µl of 10 X NEBuffer 3.1 making volume upto 1400 µl with RNase-free water. The restriction digestion reaction mix was incubated at 37 °C for overnight. Approximately 5 µl of reaction mix was separated before adding enzyme, as control. The linearized DNA was recovered by phenol chloroform isoamyl alcohol extraction (PCI). The linearization of the plasmid DNA was confirmed by agarose gel electrophoresis.

2.2.1.5 Phenol Chloroform Isoamyl Alcohol (PCI) Extraction

In order to separate the nucleic acids from proteins, phenol chloroform isoamyl alcohol extraction was performed. Phenol-chloroform-isoamyl alcohol was added in the sample in 1:1 ratio along with 1 µl of 10 µg/µl of glycogen and shaken vigorously for 15 min and then centrifuged for 5 min at 13000 rpm. The upper aqueous phase was collected in the separate eppendorf tube. Then chloroform was added in the aqueous phase in 1:1 ratio and shaken vigorously for 15 min and centrifuged at 13000 rpm for 5 min. The upper aqueous phase was collected separately and nucleic acid was recovered by ethanol precipitation.

2.2.1.6 Ethanol Precipitation

The sample was ethanol precipitated by adding 1/10 volume of 3 M NaOAc pH 5.3 and 2.5-3 volumes of 100% (v/v) ethanol for 2 h to overnight at -25 °C. The sample was then centrifuged at 4 °C for 30 min at 13000 rpm. The supernatant was discarded and the pellet was washed to remove remaining salts with 2 volumes of 80% (v/v) ethanol. Sample was then centrifuged again at 4 °C for 30min at 13000 rpm. The pellet was air dried after removing the supernatent.

2.2.1.7 Agarose Gel Electrophoresis

Agarose gel electrophoresis was performed to analyze and visualize the DNA according to the standard method (Sambrook et al., 1989). Gels were prepared by dissolving 1% w/v agarose in 1 X TBE buffer by heating. For visualization of DNA, ethidium bromide was added to the gel up to 0.5 µg/ml concentration while cooling. The DNA samples were diluted with 6 X DNA gel loading dye and were run horizontally by using 1 X TBE buffer at 120 V for 1-1.5 h along with DNA ladder. DNA was visualized on a Bio-Rad Gel Documentation System.

2.2.1.8 *In Vitro* Transcription

Different pre-mRNAs and snRNAs were synthesized by *in vitro* transcription using SP6 or T7 RNA polymerases and linearized plasmid DNAs as templates. The *in vitro* transcription reaction was prepared as follows

Components	Final Concentration	Amount (PM5 pre-mRNA/ MINX pre-mRNA)	Amount (U4 snRNA/ U6 snRNA)
5 X Transcription buffer	1 X	20 µl	20 µl
0.1 M ATP	7.5 mM	7.5 µl	7.5 µl
0.1 M UTP	7.5 mM	7.5 µl	7.5 µl
0.1 M CTP	7.5 mM	7.5 µl	7.5 µl
0.01 M GTP	1.3 mM	13 µl	13 µl
152 mM m^7GpppG cap	5 mM	3.28 µl	—
1 M $MgCl_2$	20 mM	2 µl	2 µl
1 M DTT	10 mM	1 µl	1 µl
10 mg/ml BSA	0.1 mg/ml	1 µl	1 µl
40 U/µl RNAsin	100 U	2.5 µl	2.5 µl
SP6/T7 RNA polymerase	200 U	10 µl	10 µl
DNA template		10 µg	10 µg
Make up to final volume with autoclaved deionized water	Total	100 µl	100 µl

The reaction mix was incubated at 40 °C in case of SP6 RNA polymerase or at 37 °C for T7 RNA polymerase for 3-4 h. The template DNA was digested with 10U of RQ1 RNase Free DNase at 37 °C for 30 min. In order to purify the RNA from free nucleotides and fragments of RNA, the reaction mix was mixed with equal amount of RNA sample loading buffer and loaded on the 5% or 8% denaturing polyacrylamide gel. The RNA band was visualized by UV-shadowing using flour coated thin layer chromatography (TLC) plate. The RNA band was excised from the gel. RNA was eluted from the gel band by incubating in TNES elution buffer for overnight with shaking at room temperature. The eluted RNA was then purified by PCI extraction followed by ethanol precipitation. The precipitated RNA was dissolved in RNase free water.

In order to study the interaction between Brat-NHL protein and hb RNA, the hb RNA was *in vitro* transcribed by Dr. Inga Loedige (from Dr. Gunter Meister's lab of RNA Biology, Biochemistry Center Regensburg, University of Regensburg, Germany) according to the protocol described by Loedige et al., 2014. Briefly, a transcription reaction mix was prepared containing 2 µg/ml DNA template, 0.1mg/ml T7 polymerase, 30 mM Tris pH 8, 25 mM $MgCl_2$, 0.01% Triton-X100, 1mM DTT, 10 mM each NTP, 2 U/mL pyrophosphatase (New England Biolabs) and 2 mM spermidin. The reaction mix was incubated at 37 °C for 4 h. The RNA was purified on 15% denaturing polyacrylamide gel.

2.2.1.9 Denaturing Polyacrylamide Gel Electrophoresis for RNA

Denaturing polyacrylamide gel electrophoresis was carried out in the presence of 8 M urea for the separation of RNA fragments. The concentration of polyacrylamide 5% and 8% were used according to the size of the RNA to be isolated. The gel was polymerized by adding ammonium persulfate and TEMED. The RNA samples were mixed with RNA sample loading buffer and were loaded onto 0.5 mm thick polyacrylamide gel. The gel was run vertically in 1 X TBE at 20W till the dye reaches the bottom of the gel. The RNA bands were visualized by UV-shadowing.

2.2.1.10 Silver Staining of PAGE Gels

Silver staining of the PAGE gels was done according to the modified protocol described by Merril et al., 1983. The PAGE gel was fixed in 40% (v/v) methanol, 10% (v/v) acetic acid solution for at least 30 min to overnight. The gel was washed twice with 10% (v/v) ethanol, 5% (v/v) acetic acid solution each for 15min and then rinsed briefly with deionized water. The gel was stained with 12 mM silver nitrate solution for 30 min and then briefly rinsed twice with deionized water. The gel was developed with 0.28 M Na_2CO_3 and 0.0185% (v/v) formaldehyde solution until the desired staining intensity was reached. The developing reaction was stopped by using 5% (v/v) acetic acid solution. The gel was then scanned.

2.2.2 Standard Protein Biochemical Methods

2.2.2.1 Determination of Protein Concentration

Concentrations of the protein samples were determined by Bradford colorimetric assay (Bradford, 1976). The standard curve for 0-20 µg was prepared from BSA standard stock solution (0.2 mg/ml). The protein sample was diluted with autoclaved deionized water upto 800 µl so the final concentration of the sample lies within the concentration range of the standards. Then 200 µl of Bradford solution was added so the final volume of the sample became 1 ml. The absorbance was measured at 595 nm and the protein concentration of the sample was calculated from the BSA standard curve.

Concentrations of the purified proteins were measured at 280 nm by NanoDrop spectrophotometer. The concentrations were calculated by using theoretically determined extinction coefficients generated on the basis of protein sequences (Gasteiger et al., 2005).

2.2.2.2 Denaturing Polyacrylamide Gel Electrophoresis for Proteins

Proteins were separated by using the Novex® NuPAGE® SDS-PAGE Gel System according to manufacturer's protocol under reducing conditions. The protein samples were mixed with 10 X Reducing Agent and 4 X Sample Buffer and heated at 70 °C for 10 min. The samples were loaded onto the pre-cast 4-12% Bis-Tris Gel with the thickness of 1.0 mm along with the protein ladder. The gel was run for 50 min at constant 200 V using MOPS as a running buffer with an Antioxidant in the inner chamber.

2.2.2.3 Colloidal Coomassie Staining

To visualize the separated proteins on the polyacrylamide gels, the gels were stained by using colloidal coomassie staining (Neuhoff et al., 1988). The gel was completely immersed and incubated with colloidal coomassie staining solution for overnight with gentle shaking. The gel was destained to remove the background

staining by rinsing it several times with water. The gel was scanned by Epson scanner.

2.2.3 Cell Culturing and Nuclear Extract Preparation

2.2.3.1 *HeLa* S3 Cell Culturing

HeLa S3 cells were grown in a fermenter according to the protocol described by Hartmuth et al., 2012. A cryostock of 10^8 cells was used to prepare a starting culture of 100 ml in a spinner flask. The cells were grown in High Glucose Dulbecco's Modified Eagle's Medium (DMEM) lacking Arginine, Lysine and Glutamine. The Light Arginine (Arg0) and Light Lysine (Lys0) were added into the medium upto the final concentration of 50 mg/ml. The medium was also supplemented with 10% Dialyzed Fetal Bovine Serum, 1 X Penicillin/Streptomycin and 1 X Glutamine. The cells were grown at 37 °C in 5% CO_2 and 95% relative humidity. The cell culture was expanded over 4 L in the spinner flasks and grown upto six passages before inoculating the 5 L Bioreactor. The cells were grown in the fermenter under standard conditions (2–5 x 10^6 cells/ml, with barbutation of synthetic air, keeping dissolved oxygen level to 20% by using feedback monitoring system).

2.2.3.2 *HeLa* Nuclear Extract Preparation

The cells from the fermenter were used to prepare the *HeLa* nuclear extracts. The cells were harvested freshly before the preparation of *HeLa* nuclear extract by centrifugation at 2000 rpm for 5 min in a Cryofuge 6000i, swing bucket rotor. The cells were washed twice with ice cold PBS and resuspended in 1.25-fold packed cell volume (PCV) of MC buffer containing 1/500 volume of 0.25 M DTE and EDTA-free protease inhibitor cocktail according to manufacturer's instructions. The cells were incubated on ice for 5 min and dounced 18 times to lyse in a glass douncer kept on ice. Nuclei were pelleted by centrifugation for 5min at 18000 x g at 4 °C in Sorvall SS34 rotor. The nuclei pellet was then resuspended in 1.3 fold the weight of nuclei of Roeder C buffer supplemented

with 1/500 volume of 0.25 M DTE and 1/200 volume of 0.1 M PMSF and dounced 20 times in a glass douncer on ice. The extract was then stirred for 40 min at 4°C. The nuclear debris was pelleted by centrifugation at 16000 rpm for 30 min at 4 °C in Sorvall SS34 rotor. The nuclear membrane was removed from the top of the supernatant. The supernatant was aliquoted and flash frozen in liquid nitrogen and stored at -80 °C.

2.2.3.3 *HeLa* Nuclear Extract Dialysis

The *HeLa* nuclear extract was thawed on ice. To remove traces of ethanol in which it was previously stored, the dialysis tube was washed multiple times with autoclaved deionized water. The nuclear extract was filled in the dialysis tube by clipping its one end. After filling the nuclear extract, the dialysis tube was clipped on the other end as well. The nuclear extract was dialyzed three times against 40 volumes of Roeder D buffer for 2 h each with constant stirring at 4 °C. The dialyzed nuclear extract was centrifuged at 9000 x g for 2 min at 4 °C in Sorvall Hb-6 rotor. The supernatant was aliquoted and frozen in liquid nitrogen and stored at -80 °C.

2.2.4 Expression, Isolation and Purification of Proteins and RNA-Protein Complexes

2.2.4.1 MS2-MBP Fusion Protein Overexpression and Purification

In order to affinity purify the complex assembled on MS2-tagged pre-mRNA, MS2-MBP fusion protein was overexpressed and purified from chemically competent Rosetta II cells. For the expression of MS2-MBP fusion protein, 50 µl of the Rosetta II competent cells were chemically transformed with 350 ng of the MS2-MBP fusion protein plasmid DNA and inoculated in 100 ml of LB medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and incubated for overnight at 37 °C with constant shaking at 180 rpm as pre-culture. Then 4 L of LB medium with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol was inoculated with pre-culture. The bacterial culture was then incubated at 37 °C

with constant shaking at 180 rpm till the OD_{600} reached upto 0.6-0.7. To induce the expression of the MS2-MBP fusion protein, IPTG was added to the final concentration of 0.75 mM. The culture was incubated at 37 °C with constant shaking at 180 rpm till the OD_{600} reached upto 2-2.5. Before induction approximately 1 ml of the bacterial cell culture was separated as control. The cells were harvested by centrifugation at 5000 rpm for 30 min at 4 °C. The supernatant was discarded and the cells were washed twice with PBS by centrifugation. The cells were resuspended in lysis buffer and lysed two times by microfluidizer at 80 psi. The lysate was centrifuged at 16000 rpm for 20 min at 4°C. The supernatant was incubated for 1 h at 4 °C in a rotating glass bottle with 10 ml of amylose beads equilibrated with water and lysis buffer. The incubated beads along with supernatant was loaded onto the 20 ml gravity flow column and let the column drain off. The beads were washed with 300 ml of lysis buffer, then with 200 ml of amylose matrix wash buffer and finally with 200 ml of sodium-potassium buffer. The MS2-MBP fusion protein was eluted with 50 ml of sodium-potassium buffer containing 15 mM maltose. In order to remove salts and excess maltose, the protein eluate was desalted by using HiPrep 26/10 desalting cartridge pre-equilibrated with sodium-potassium buffer. The eluate was then loaded onto a Heparin Sepharose HiTrap column equilibrated with sodium-potassium buffer. The protein was then eluted by using heparin elution buffer. The protein concentration was determined by NanoDrop spectrophotometer. The protein was aliquoted and frozen in liquid nitrogen and stored at -80 °C.

2.2.4.2 CWC2 Protein Overexpression and Purification

The CWC2 protein was expressed and purified by Monika Raabe (from Bioanalytical Mass Spectrometry Group, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany) and Dr. Jana Schmitzová (from Dr. Vladimir Pena's Lab of Macromolecular Crystallography, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany) respectively according to the protocol published by Schmitzová et al., 2012. The Rosetta II competent cells were transformed with pETM11-yCWC2 plasmid DNA for the expression of

CWC2 protein with N-terminal hexahistidine tag. The pre-culture was prepared by inoculating the transformed colonies to the 2xYT medium supplemented with 34µg/ml chloramphenicol and 50 µg/ml kanamycin and incubating at 30 °C with constant shaking overnight. The pre-culture was used to inoculate 6 L of 2xYT medium with 34 µg/ml chloramphenicol and 50 µg/ml kanamycin and incubated at 30 °C with constant shaking at 180 rpm till the OD_{600} reached upto 0.6. For the induction of CWC2 protein expression, IPTG was added to the final concentration of 0.6 mM. The culture was grown at 17 °C for 20 h with constant shaking till the OD_{600} reached upto 1.0-1.2. The cells were harvested by centrifugation at 4000rpm for 20 min at 4 °C. The cell pellet was washed with PBS and frozen in liquid nitrogen and stored at -80 °C. Later, the cell pellet was resuspended in lysis buffer (50 mM HEPES-NaOH pH 7.5, 600 mM NaCl, 4 mM β-mercaptoethanol, 15 mM imidazole, 15% (w/v) glycerol) containing EDTA-free protease inhibitor cocktail according to manufacturer's instructions by vortexing. The cells were lysed six times by a microfluidizer at 80 psi. The lysate was centrifuged at 10,000rpm for 40 min in Sorvall SS34 rotor. All purification steps were carried out at 4°C. The supernatant containing protein was incubated with Ni-NTA agarose beads (Macherey-Nagel) for 3 h. The protein was eluted from the beads by elution buffer (50 mM HEPES-NaOH pH 7.5, 600 mM NaCl, 2 mM β-mercaptoethanol, 250 mM imidazole, 10% (w/v) glycerol). Since the eluate from the Ni-NTA column has high salt concentration so it was diluted to the final concentration of 50 mM NaCl. The eluate was then applied to the HiTrap Heparin HP column (GE Healthcare). The protein eluted from the heparin beads by elution buffer (50 mM HEPES-NaOH pH 7.5, 1 M NaCl, 4 mM β-mercaptoethanol, 10 mM imidazole, 7% (w/v) glycerol) was concentrated using Centricon concentrators and was further purified by size exclusion chromatography (GE Healthcare). The protein was eluted from Superdex 75 16/600 column (GE Healthcare) by elution buffer (30 mM HEPES-NaOH pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol, 5% (w/v) glycerol). The purified protein was analyzed by SDS-PAGE. The fractions with high purity were pooled together and concentrated. The protein concentration was determined by Bradford assay.

The protein sample was then aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

2.2.4.3 Brat-NHL Protein Overexpression and Purification

The Drosophila BRAT-NHL protein provided by Dr. Inga Loedige (from Dr. Gunter Meister's lab of RNA Biology, Biochemistry Center Regensburg, University of Regensburg, Germany) was prepared according to the protocol described by Loedige et al., 2014. Briefly, the Brat-NHL protein was expressed as hexahistidine-ubiquitin fusion by using pHUE vector system (Catanzariti et al., 2004; Baker et al., 2005) in *E.coli* BL21. The culture was grown at 37 °C to an OD_{600} of 0.6. The protein expression was induced by adding 1 mM IPTG and incubated at 23 °C for overnight. The cells were lysed by incubation in His A buffer (50 mM Tris-Cl pH 8, 1 M NaCl, 5% (w/v) glycerol, 10 mM imidazol) containing 1 mg/ml lysozyme, 1 mM AEBSF, and 5 U/ml Benzonase followed by sonication. The lysate was centrifuged at 48,000 x g for 40 min at 4 °C. The supernatant was applied to HiTrap IMAC FF column charged with Ni^{2+}. The protein was eluted by His B buffer (50 mM Tris-Cl pH 8, 1 M NaCl, 200 mM imidazol). Then the hexahistidine-ubiquitin moiety was cleaved off by incubating the eluate overnight at 4 °C with the Usp2cc enzyme in the presence of 1 mM DTT. The protein was then loaded on a HiPrep Superdex 75 26/60 column equilibrated by buffer containing 20 mM Tris-Cl pH 8, 150 mM NaCl and 1 mM DTT. The fractions containing highly pure protein were pooled together. All purification steps were carried out at 4 °C. The protein concentration was determined spectrophotometrically at 280 nm. The protein sample was aliquoted and flash-frozen in liquid nitrogen and stored at -80 °C.

2.2.4.4 *In Vitro* RNA–Protein Complex Assembly from *HeLa* Nuclear Extract and Purification

The RNA-protein complex (H/E complex) from the *HeLa* nuclear extract was assembled and purified according to the protocol mentioned by Qamar et al., 2015. Briefly, 1 nmol of *in vitro* transcribed (labeled/non-labeled) MS2-tagged (PM5/MINX) pre-mRNA was incubated with 15-fold excess of MS2-MBP fusion

protein for 30 min on ice. Then the pre-mRNA bound to MS2-MBP protein was incubated with 10 ml of *HeLa* nuclear extract for 30 min on ice. In order to affinity purify the complex, the gravity flow disposable chromatographic column was packed with 1 ml of amylose beads and washed three times with 2 ml of gradient buffer without glycerol. The sample was then loaded onto the column and allowed to flow through under gravity. The column was then washed three to five times with 2 ml of gradient buffer without glycerol. The assembled complex was eluted with 2 ml of 15 mM maltose buffer by gravity flow. The protein concentration was determined by Bradford assay.

2.2.5 UV-Induced Cross-linking in RNA-Protein Complexes

2.2.5.1 UV-Cross-linking of Brat-NHL protein with hb RNA

The Brat-NHL-hb RNA complex was assembled by incubating 1 nmol of *in vitro* transcribed hb RNA with 1 nmol of recombinant Brat-NHL protein (provided by Inga Loedige from Dr. Gunter Meister's lab of RNA Biology, Biochemistry Center Regensburg, University of Regensburg, Germany) making volume upto 100 µl with the buffer (20 mM Tris-Cl pH 8, 150 mM NaCl) for 1 h on ice. For UV-cross-linking, the assembled complex was taken in a microtiter plate placed on aluminum block on ice at a distance of 1 cm from the lamps and UV-irradiated at 254 nm for 10 min. The sample was then ethanol precipitated for overnight. The pellet was dissolved in 50 µl of 4 M Urea in 50 mM Tris-Cl pH 7.9 and diluted to 1M Urea by adding 150 µl of 50 mM Tris-Cl pH 7.9. The RNA was hydrolyzed using 1 µl of RNase A (1 µg/µl) and T1 (1 U/µl) at 52 °C for 1 h followed by 1 µl of Benzonase (25 U/µl) in the presence of 1 mM $MgCl_2$ at 37 °C for 1 h with continuous shaking at 500 rpm. The protein/peptide was digested using trypsin in enzyme to protein ratio of 1:20 (w/w) at 37 °C for 16 h with shaking at 500 rpm. The sample was then desalted by C18 reversed phase chromatography and enriched by TiO_2 solid phase extraction as described by Kramer et al., 2011 (details in materials and methods heading 2.2.5.3). The control (non-UV-irradiated) sample was also processed in parallel to the UV-irradiated sample.

For mass spectrometric analysis, the samples were dried in the SpeedVac and reconstituted in 12 µl of 5% (v/v) acetonitrile, 0.1% (v/v) formic acid. The samples were analyzed by LTQ Orbitrap Velos mass spectrometer (Thermo Scientific).

2.2.5.2 UV-Cross-linking of CWC2 protein with U4 and U6 snRNAs

The UV-cross-linking analysis of CWC2 protein with U4 and U6 snRNAs was done according to the protocol published by Schmitzová et al., 2012. The CWC2-U4 snRNA and CWC2-U6 snRNA complexes were reconstituted by incubating 100 µg of CWC2 for 30 min on ice with 3 µg of U4 and U6 snRNAs separately making volume upto 100 µl with the buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT). The sample was UV-irradiated at 254 nm for 10 min in a microtiter plate kept on an aluminum block on ice at a distance of 1 cm from the light source. The sample was ethanol precipitated for overnight. The pellet was dissolved in 50 µl of 4 M Urea in 50 mM Tris-Cl pH 7.9 and adjusted to final concentration of 1 M Urea by adding 150 µl of 50 mM Tris-Cl pH 7.9. The hydrolysis of RNA was carried out by 1 µl of RNase A (1 µg/µl) and T1 (1 U/µl) at 52 °C for 1 h followed by 1 µl of Benzonase (25 U/µl) in the presence of 1 mM $MgCl_2$ at 37 °C for 1 h with shaking at 500 rpm. The protein was digested using trypsin in enzyme to protein ratio of 1:50 (w/w) at 37 °C for 16 h with continuous shaking at 500 rpm. The desalting and enrichment of the sample were carried out by C18 reversed phase chromatography and TiO_2 solid phase extraction respectively according to the protocol described by Kramer et al., 2011 (detail in materials and methods heading 2.2.5.3). The control (non-UV-irradiated) samples were also processed in parallel to the UV-irradiated samples. For mass spectrometric analysis, the samples were dried in the SpeedVac and reconstituted in 12 µl of 5% (v/v) acetonitrile, 0.1% (v/v) formic acid. The samples were analyzed by Q-Exactive mass spectrometer (Thermo Scientific).

2.2.5.3 UV-Cross-linking of RNA-Protein Complex from *HeLa* Nuclear Extract Assembled on PM5/MINX pre-mRNAs

The purified RNA-protein complex (H/E complex) from *HeLa* nuclear extract was UV-cross-linked according to the protocol described by Qamar et al., 2015. The

sample was taken in a volume of 1 ml (protein concentration 0.3 mg/ml) in pre-cooled custom-made glass dishes, with a planar surface and an inner diameter of 3.5 cm, so the depth of the sample solution was approximately 1 mm. The dishes were kept on an aluminum block on ice at a distance of 1 cm from the light source. The sample was UV-irradiated at 254 nm for 10 min. The sample was pooled in the Corex glass tube and ethanol precipitated for overnight. The pellet was dissolved in 100 µl of 1% (w/v) SDS in size exclusion (SE) running buffer by shaking and diluted to final concentration of 0.1% (w/v) SDS with size exclusion (SE) running buffer. The protein was digested with trypsin in 1:50 (w/w) enzyme to protein ratio at 37 °C for 16 h with continuous shaking at 500 rpm. The sample was then again ethanol precipitated. The pellet was re-dissolved in 5 µl of 1% (w/v) SDS in size exclusion (SE) running buffer and diluted upto 0.1% (w/v) SDS with SE running buffer. The sample was injected into the SMART system equipped with Superdex 75 PC 3.2/30 column running in SE running buffer with a flow rate of 40 µl/min at room temperature. The fractions containing RNA were pooled together and ethanol precipitated overnight. The pellet was dissolved in 4M Urea, in 50 mM Tris-Cl pH 7.9 and adjusted to final concentration of 1 M Urea by adding 150 µl of 50 mM Tris-Cl pH 7.9. Digestion of RNA was carried out by 1µl of RNase A (1 µg/µl) and T1 (1 U/µl) at 52 °C for 1 h followed by 1 µl of Benzonase (25 U/µl) in the presence of 1 mM $MgCl_2$ at 37 °C for 1 h with shaking at 500 rpm. The protein was digested by using trypsin in 1:20 (w/w) enzyme to protein ratio at 37 °C for 16 h with continuous shaking at 500 rpm. The sample was desalted and enriched by C18 reversed phase chromatography and TiO_2 solid phase extraction respectively according to the protocol described by Kramer et al., 2011. For the C18 reversed phase chromatography the columns were prepared in-house by fitting 2 mm^2 piece of coffee filter paper at the end of 10 µl pipette tip as frit material. The columns were packed with the C18 material suspended in 100% (v/v) methanol with the help of 1 ml combitip to give a bed height of 3 mm. The columns were fitted in 2 ml microcentrifuge tubes by making holes in the lid and washed by applying 60 µl of 95% (v/v) acetonotrile, 0.1% (v/v) formic acid, then 80% (v/v) acetonitrile, 0.1% (v/v) formic acid followed by 50% (v/v) acetonitrile, 0.1% (v/v) formic acid and finally by 0.1% (v/v) formic acid with

centrifugation at 5000 rpm for 5 min after each step. Meanwhile, 10 µl of 100% (v/v) acetonitrile and 2 µl of 10% (v/v) formic acid were added into the sample to make an end concentration to 5% (v/v) acetonitrile and 0.1% (v/v) formic acid respectively. The sample was vortexed and centrifuged for 2 min at 13,000 rpm at room temperature to remove precipitates. The sample was then applied onto the column in portions of 60 µl by centrifugation at 5000 rpm for 5 min after each step. The column was then washed twice with 60 µl of 0.1% (v/v) formic acid and the sample was eluted, once with 60 µl of 20% (v/v) acetonitrile, 0.1% (v/v) formic acid, twice with 50% (v/v) acetonitrile, 0.1% (v/v) formic and once with 80% (v/v) acetonitrile, 0.1% (v/v) formic acid by centrifugation at 5000 rpm for 5 min. The eluate was then dried in the SpeedVac for 45 min. In order to remove the non-cross-linked peptides TiO_2 solid phase extraction was performed. The TiO_2 columns were prepared with the TiO_2 suspension in the same manner as for C18 reversed phase chromatography. The columns were washed twice with 60 µl of buffer B by centrifugation at 3000 rpm for 5 min. Meanwhile, the sample was dissolved in 100 µl of buffer A by vortexing and ultrasonification for 1 min and loaded onto the column in the portion of 60 µl by centrifugation at 3000 rpm for 5min. The column was then washed three times with buffer A and four times with buffer B, each time with centrifugation at 3000 rpm for 5 min. The sample was eluted thrice with 40 µl of buffer C by centrifugation at 3000 rpm for 5 min. The control (non-UV-irradiated) sample was also processed in parallel to the UV-irradiated sample. For mass spectrometric analysis, the samples were dried in the SpeedVac and reconstituted in 12 µl of 5% (v/v) acetonitrile, 0.1% (v/v) formic acid. The samples were analyzed by LTQ Orbitrap Velos mass spectrometer (Thermo Scientific).

2.2.6 Quantitative Analysis of RNA-Protein Cross-links

2.2.6.1 Quantitative Analysis of CWC2-U4 snRNA and U6 snRNA Cross-links

For the quantitative analysis of CWC2-U4 snRNA and U6 snRNA cross-links, the experiment was carried out in the set of three forward and reverse biological replicates. In the forward experiment, 3 µg of U6 snRNA transcribed by using non-labeled UTP and 3 µg of U4 snRNA transcribed by using isotopically-labeled (^{13}C) UTP were pooled together in 1:1 ratio and were incubated with 200 µg of CWC2 making volume upto 200 µl with the buffer (20 mM HEPES pH 7.5, 100mM NaCl, 1 mM DTT) for 30 min on ice. For the reverse experiment 3 µg of U6 snRNA transcribed by using isotopically labeled (^{13}C) UTP and 3 µg of U4 snRNA transcribed by using non-labeled UTP were pooled together in 1:1 ratio and were incubated with 200 µg of CWC2 making volume up to 200 µl with the buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT) for 30 min on ice. The samples were UV-irradiated and processed in the same manner as described in section 2.2.5.2 of materials and methods. Finally, the samples were analyzed by Q-Exactive mass spectrometer (Thermo Scientific).

2.2.7 Mass Spectrometry Methods

2.2.7.1 In-Gel Digestion of Proteins

In-gel hydrolysis of proteins was performed according to the modified protocol of Shevchenko et al., 2007. Unless otherwise stated, all the incubation steps were carried out at 26 °C with continuous shaking at 1050 rpm for 15 min. The solutions were removed after each incubation step. Each SDS-PAGE gel lane was cut into 22 equal slices with the help of in-house designed gel cutting device (Schmidt & Urlaub, 2009). Each gel slice was cut into small pieces, washed with 150 µl of water and dehydrated with 150 µl of acetonitrile. The gel pieces were then dried by SpeedVac. The proteins were reduced by adding 100 µl of 100 mM DTT prepared in 50 mM ammonium bicarbonate pH 8 and incubating at 56 °C for

50 min at 1050 rpm. The gel pieces were then dehydrated with 150 µl of acetonitrile and the reduced cysteines were alkylated with 100 µl of 60 mM Iodoacetamide (IAA) prepared in 50 mM ammonium bicarbonate pH 8 for 20 min at 26 °C and 1050 rpm. The gel pieces were washed with 150 µl of 50 mM ammonium bicarbonate pH 8, dehydrated and dried again as described before. The dried gel pieces were rehydrated with 20 µl of trypsin digestion buffer (15 µl of 0.1 µg/µl modified trypsin making volume upto 100 µl with 25 mM ammonium bicarbonate pH 8), for 30 min on ice. The gel pieces were overlaid with 25 mM ammonium bicarbonate pH 8, if needed. Then they were incubated at 37 °C for overnight with constant shaking at 1050 rpm in thermomixer.

2.2.7.2 Extraction of Peptides

The peptides from in-gel digestion were extracted according to the protocol described by Shevchenko et al., 2007. All incubation steps were carried out at 37°C for 15 min with a constant shaking at 1050 rpm by thermomixer. The gel pieces were processed by series of extraction steps comprised of incubation with 50 µl of water and 50 µl of acetonitrile then with 50 µl of 5% (v/v) formic acid followed by twice with 50 µl of acetonitrile. The supernatants from each step were collected and pooled together in new microcentrifuge tubes. The extracted peptides were dried in SpeedVac and stored at -20 °C until subjected to LC-ESI-MS/MS analysis. For MS analysis the samples were dissolved in 20 µl of 5% (v/v) acetonitrile and 1% (v/v) formic acid by extensive vortexing and sonication for 2min each. The samples were then analyzed by LTQ Orbitrap XL mass spectrometer (Thermo Scientific).

2.2.7.3 LC-ESI-MS/MS

The mass spectrometric analysis was carried out by administering the samples to nanoflow-liquid chromatography (nano-LC) system coupled to electrospray ionisation mass spectrometer (ESI-MS). During the course of Ph.D. studies three mass spectrometers were used. The samples from in-gel digestion were analyzed on LTQ Orbitrap XL mass spectrometer (Thermo Scientific) coupled to Agilent nano-LC system (Agilent Technologies) whereas the RNA-protein cross-

linking samples were analyzed on LTQ Orbitrap Velos (Thermo Scientific) and Q-Exactive (Thermo Scientific) mass spectrometers coupled to Agilent nano-LC system (Agilent Technologies) and EASY-nLC II system (Thermo Scientific) respectively. The details regarding LC separation and MS analysis are given as follows.

(A) Nanoflow-Liquid Chromatography Separation (Nano-LC)

(i) Nano-LC Separation (Agilent nano-LC system)

The sample was applied onto trapping column (C18 AQ 120 Å material with particle size of 5 µm, 20 mm length, 0.150 mm inner diameter) at a flow rate of 10µl/min (60 min gradient) and 15 µl/min (118 min gradient) in 3% buffer B (buffer A: 0.1% (v/v) formic acid; buffer B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid) followed by elution and separation on an analytical column (C18 AQ 120 Å material with particle size of 5 µm, 150 mm length, 0.075 mm inner diameter) at a flow rate of 130 nl/min (60 min gradient) and 150 nl/min (118 min gradient) using a linear gradient of 4-37% buffer B (buffer A: 0.1% (v/v) formic acid; buffer B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid) over 37 min (60 min gradient) and 102 min (118 min gradient). The column was then washed with 90-95% buffer B and re-equilibrated with 3% buffer B. Both the columns were packed in-house by Uwe Pleßmann (Bioanalytical Mass Spectrometry Group, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany).

(ii) Nano-LC Separation (Thermo EASY-nLC II system)

The sample was injected into a trapping column (C18 AQ 120 Å material with particle size of 3 µm, 40 mm length, 0.1 mm inner diameter) in-line with the analytical column (C18 AQ 120 Å material with particle size of 3 µm, 10 cm length, 50 µm inner diameter), both packed in-house by Uwe Pleßmann (Bioanalytical Mass Spectrometry Group, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany). The sample was loaded onto trapping column at a flow rate of 15 µl/min in 3% buffer

B (buffer A: 0.1% (v/v) formic acid; buffer B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid) followed by elution and separation on an analytical column at a flow rate of 320 nl/min using a linear gradient of 4-36% buffer B (buffer A: 0.1% (v/v) formic acid; buffer B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid) over 42 min (50 min gradient) and 97 min (105 min gradient). The column was then washed with 95% buffer B and equilibrated automatically by the instrument.

(B) ESI-MS/MS Analysis

(i) LTQ Orbitrap XL Mass Spectrometer

The instrument was operated in data dependent acquisition mode with Top 8 method. The MS scans were recorded in the *m/z* range 350-1600 at a resolution setting of 30,000 FWHM at *m/z* 400 and automatic gain control (AGC) at 10^6. Fragmentation was generated by CID activation for the precursor ions having the charge state 2 and above. The MS/MS scans were recorded at normalized collision energy of 35 and a dynamic exclusion of 60 sec with a repeat count of 1.

(ii) LTQ Orbitrap Velos Mass Spectrometer

The instrument was operated in data dependent acquisition mode with Top 10 method. The MS survey scans were recorded in the *m/z* range 350-1600 at a resolution setting of 30,000 FWHM. The automatic gain control was set to 10^6. Fragmentation was generated by HCD activation for the precursor ions having the charge state 2, 3 and 4. The MS/MS scans were recorded at normalized collision energy of 35 and a dynamic exclusion of 20 sec at a resolution setting of 7500 FWHM and isolation width of 2 Th.

(iii) Q-Exactive Mass Spectrometer

The instrument was operated in data dependent acquisition mode with Top 12 method. The MS survey scans were recorded in the *m/z* range 350-1600 at a resolution setting of 70,000 FWHM. The automatic gain control was set to 10^6. Fragmentation was generated by HCD activation for the precursor ions having the charge state 2, 3 and 4. The MS/MS scans were recorded at normalized

collision energy of 28 and a dynamic exclusion of 15 sec at a resolution setting of 17,500 FWHM with a fixed first mass of *m/z* 100.

2.2.8 Data Analysis

2.2.8.1 Proteome Analysis by MaxQuant

The raw data obtained after MS analysis was further analyzed by using MaxQuant software version 1.5.2.8 incorporated with Andromeda (Cox & Mann, 2008; Cox et al., 2011). The searches were made against UniProt human protein database (23rd December 2011). Following parameters were brought under consideration for data analysis: MS/MS tolerance was set to 0.5 Da, false discovery rate (FDR) at both peptide and protein level to 1%. Minimum peptide length of 5 amino acids was used along with minimum ratio count of 2. Cysteine carbamidomethylation was used as fixed modification whereas the oxidation of methionine, N-terminal protein acetylation and phosphorylation of serine, threonine and tyrosine were used as variable modifications. The tryptic specificity with no proline restriction, allowed upto 2 missed cleavages, was set. The results obtained were used for further data analysis and interpretation.

2.2.8.2 RNA-protein Cross-linking Analysis by OpenMS

The RNA-protein cross-linking data was analyzed by using RNP^{xl} tool of TOPPAS workflow engine in an OpenMS software environment using OMSSA as search engine (Kohlbacher et al., 2007; Sturm et al., 2008) according to the details provided by Kramer et al., 2014. Briefly, the raw data files were converted into .mzML format with the help of msconvert command line tool of ProteoWizard set of Library and Tools (Chambers et al., 2012) or by using Proteome Discoverer software version 1.10 (Colaert et al., 2011). The data was then subjected to series of filter pipelines and then searched for the probable cross-linked peptides by using RNP^{xl} pipeline. The results obtained were visualized by TOPPView and manually validated. The identified cross-links were then mapped on the crystal

structure, if available from RCSB PDB (http:// www.rcsb.org) using PyMOL software version 1.8 (Schrödinger, LLC).

For the quantitative analysis, the extracted ion chromatograms (XICs) of MS spectra were generated by using Xcalibur software (Thermo Scientific) and Skyline software (MacCoss Lab).The ratios were calculated from the peak area of the extracted ions. The results obtained were used for further statistical analysis by using SPSS Statistics software (v.17.0) and data interpretation.

The other online tools used for the calculation of monoisotopic masses of peptide and RNA oligonucleotides etc. were as follows

- Peptide Mass Calculator version 3.2 (University of Leuven, Belgium)

  (http://rna.rega.kuleuven.ac.be/masspec/pepcalc)

- Mongo Oligo Mass Calculator version 2.06 (University of Leuven, Belgium)

  (http://rna.rega.kuleuven.ac.be/masspec/mongo)

- Protein Prospector version 5.14.4 (University of California, US)

  (http://prospector.ucsf.edu/prospector/mshome)

2.2.8.3 Quantitative Analysis by Skyline

For relative quantitative analysis using skyline software, the extracted ion chromatograms were generated for the spectral library and Savitzky-Golay smoothing was applied (Appendix Figure 6.33-6.47). In each window the most intense peak was selected for marking the peak boundaries (dotted black lines). IDs were from the built spectral library showing the retention times of spectra of the identified cross-links. Red, blue, brown and purple colour peaks were from doubly charged light labeled, doubly charged heavy labeled, triply charged light labeled and triply charged heavy labeled forms of cross-links respectively. In the present study, the presence of first two isotopic peaks was set mandatory for quantification. Therefore, the area for the first two isotopic peaks (M and M+1) were used for the quantitative analysis (Appendix Table 6.7 & Table 6.8).

2.2.8.4 Statistical Analysis

The compiled quantitative data from excel file was transferred to the SPSS data editor and $\log_{10}$ transformation was applied for normalization. The data normality was confirmed by using Kolmogorov-Smirnov and Shapiro-Wilk tests. The comparative studies were carried out statistically by using one-way ANOVA. For determining any significant differences, Duncan's multiple comparison test was applied at 5% level of significance.

2.2.8.5 Interactome Analysis

For protein-protein interaction network analysis, the data obtained from OpenMS and MaxQuant were analyzed by STRING database. The STRING database (Jensen et al., 2009) was queried for known protein-protein structural and functional interactions. The comma-delimited files (.csv) were imported in cytoscape (version 3.7.0). The network was analyzed by using NetworkAnalyzer plug-in (Assenov et al., 2008) and cytoHubba plug-in (Chin et al., 2014).

3 RESULTS

In the cellular context, the RNA-binding proteins (RBPs) interact with the RNA molecules to constitute ribonucleoprotein (RNP) complexes that play a crucial role in the transcriptional and post transcriptional gene regulation. The application of RNA-protein cross-linking by UV-irradiation at a wavelength of 254 nm followed by mass spectrometry has proved to be promising technique in predicting the arrangement of RNA and protein molecules in these RNP complexes. During the course of study, the RNA-protein interaction sites within various *in vitro* reconstituted RNP complexes were determined up to the peptide or even the amino acid level. The RNP complexes containing single protein were processed by the conventional cross-linking method; however, for the analysis of complex samples comprised of several proteins like the interactome from *HeLa* nuclear extract assembled on pre-mRNA, the method was developed and optimized. The RNA-protein cross-links identified from the complexes assembled on different pre-mRNAs were also quantitatively analyzed by labeling RNAs, peptides and proteins providing the information regarding the ability/affinity of various RBPs to interact with particular RNA assembled under similar conditions/treatments. The MS data obtained was evaluated and validated manually. The manual interpretation of the MS/MS spectra led to the identification of the cross-linking site up to the amino acid resolution as well as the detection of fragment ions, generated during the fragmentation of nucleic acid present in a complex, resulting in the improvement in the data analysis strategy of the software.

3.1 Identification of Cross-links from Brat-NHL-hb RNA Complex

During the early embryonic development in Drosophila, brain tumor protein (Brat) along with Pumilio (Pum) and Nanos (Nos) forms a complex for the translational repression of the hb mRNA (Wharton & Struhl, 1991; Murata & Wharton, 1995; Sonoda & Wharton, 2001). It has been thought that Pum binds to the hb RNA directly and then recruits Brat and Nos by protein-protein interactions but here it has been reported in collaboration with Inga Loedige (from Dr. Gunter Meister's lab of RNA Biology, Biochemistry Center, University of Regensburg, Germany) that Brat directly binds to the hb RNA.

Brat is a member of conserved family of TRIM-NHL proteins, which are identified by the presence of N-terminal tripartite motif (TRIM) and a C-terminal NCL-1, HT2A, LIN-41 (NHL) domain (Sardiello et al., 2008). The NHL domain of TRIM-NHL proteins has been reported as RBD (Kwon et al., 2013). It is arranged in six-bladed β propeller structure (Figure 3.1 A & B), similar to the WD40 fold (Edwards et al., 2003) and has a positively charged top surface showing the potential to bind negatively charged molecules like nucleic acids. It has been observed that for *in vivo* translational repression of the hb mRNA approximately 100 nucleotides long fragment of hb 3′ UTR containing two Nanos response elements (NREs) (termed as hb RNA hereafter) is necessary (Wharton & Struh, 1991) (Figure 3.1 C).

In order to find the RNA-binding activity of Brat, the Brat-NHL-hb RNA complex was *in vitro* assembled by incubating the *in vitro* transcribed hb RNA along with recombinantly expressed purified NHL domain of Brat (~ 32 kDa, ranging 756-1037 amino acid) provided by the collaborator (Inga Loedige). The sample was UV-cross-linked, and the cross-links were enriched according to the standard protocol as described in sections 2.2.5.1 and 2.2.5.3 of materials and methods respectively. The samples were analyzed on LTQ Orbitrap Velos mass spectrometer and the data analysis was performed by OpenMS as mentioned in

section 2.2.8.2 of materials and methods. The candidate spectra of the cross-linked peptides were then manually validated.

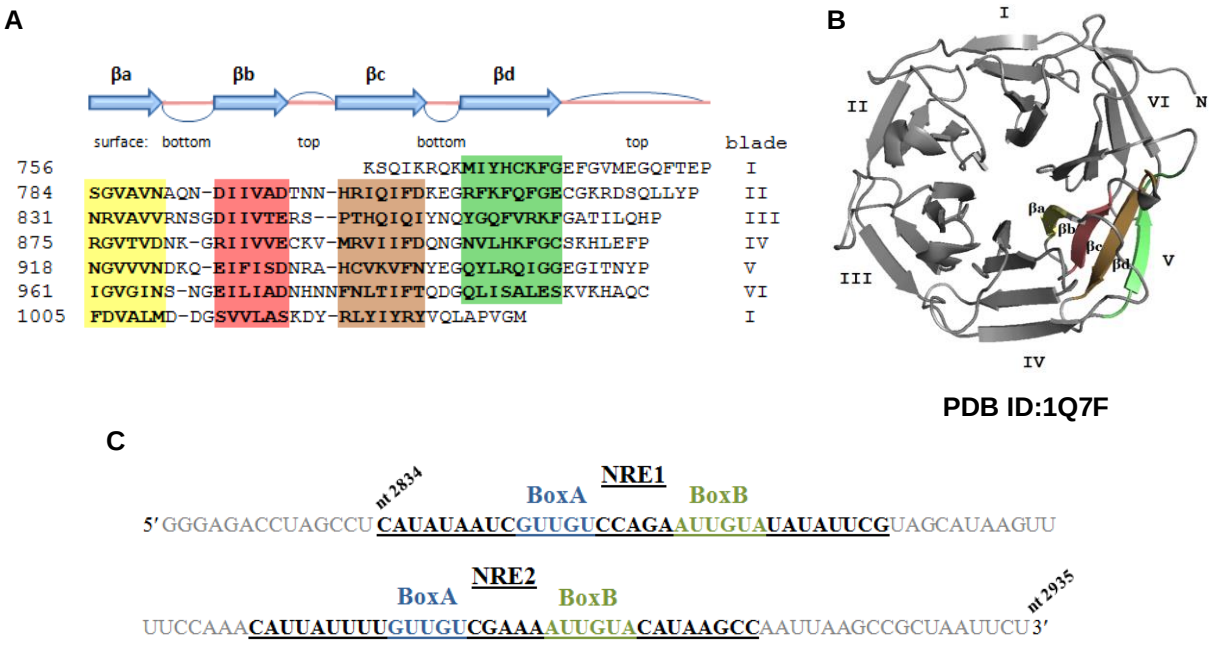


Figure 3.1: The top electropositive surface of the Drosophila Brat-NHL domain interacts with hb RNA (Figure adapted from Loedige et al., 2014 and modified). (A)The sequence alignment based on the crystal structure of Brat (PDB ID: 1Q7F). The secondary structure has been illustrated above the sequence alignment (Edwards et al., 2003). The amino acids constituting the β strands has been shown in bold and highlighted in yellow (βa), pink (βb), brown (βc) and green (βd) colours (Loedige et al., 2014). (B)The top view of Brat-NHL domain's crystal structure at the resolution of 1.95 Å (PDB ID: 1Q7F, chain A) (Edwards et al., 2003).It has six bladed β propeller structure. Each blade is comprised of βa, βb, βc and βd strands which are anti-parallel β strands that are interlinked through flexible loop regions. The strands of blade V have been coloured according to the sequence alignment shown in Fig. 3.1 A. The loops connecting βa with βb and βc with βd form the bottom surface whereas the ones connecting βb with βc and βd with βa constitute the top surface (Loedige et al., 2014). (C) Sequence of ~100 nucleotide long hb RNA fragment (nucleotide 2834-2935) containing two NREs shown in bold and underlined. Each NRE is composed of one BoxA (blue) and one BoxB (green) motif (Loedige et al., 2014).

So far, six peptides of Brat-NHL protein have been identified to be cross-linked to the nucleotides of hb RNA (Table 3.1). For all these cross-linked peptides the cross-linking site has been cut down to the amino acid resolution indicating the exact amino acids in contact with the hb RNA or are lying in close proximity to it. It has been observed that five out of six cross-linked peptides, span the top surface of the NHL domain (Figure 3.2 A). Four cross-linked amino acids (Y^{829}, C^{820}, F^{916} and C^{890}) are lying on the top surface, one (K^{809}) at the bottom surface

while the two (K^{865} and F^{866}) are located on a positively charged patch at the circumference (Figure 3.2 B).

Bringing the cross-linking data into consideration, the collaborators have performed the studies by making point mutations in Brat-NHL domain and checked the RNA-binding activity by electrophoretic mobility shift assay (EMSA). It has been found that the mutations of the top surface residues Y^{829} (cross-linked amino acid) and R^{875} (amino acid of a cross-linked peptide ^{865}KFGATILQHPR^{875}) resulted in the impaired or completely abrogated binding of Brat-NHL domain to hb RNA respectively. The mutation of K^{809} (cross-linked amino acid), lying at the bottom surface had no effect on its binding activity (Loedige et al., 2014).

In addition to EMSA, the collaborators have also investigated the effect of point mutations on Brat-mediated translational repression by using luciferase reporter assay. The mutations of C^{890}, F^{916} (cross-linked amino acids) and R^{875}, K^{891} (amino acids of the cross-linked peptides ^{865}KFGATILQHPR^{875} and ^{885}IIVVECK^{891} respectively) resulted in the impaired Brat-mediated repression whereas unlike the effect of mutation of Y^{829} (cross-linked amino acid) in EMSA, there was no effect on repression. The repression by Brat was unaffected by the mutations of top surface C^{820} (cross-linked amino acid) and the bottom surface K^{809} (cross-linked amino acid) and K^{925} (amino acid of a cross-linked peptide ^{913}HLEFPNGVVVNDK^{925}) residues (Loedige et al., 2014).

Overall, the mutations of amino acids Y^{829}, C^{890} and F^{916} which are found to be cross-linked and amino acids R^{875}, K^{891} belonging to cross-linked peptides ^{865}KFGATILQHPR^{875} and ^{885}IIVVECK^{891} respectively has led to either compromised Brat-mediated translational repression or impaired binding of Brat-NHL domain to hb RNA. On the other hand, the mutations of top surface cross-linked amino acid C^{820} and the bottom surface cross-linked amino acid K^{809} and the amino acid K^{925} of a cross-linked peptide ^{913}HLEFPNGVVVNDK^{925} showed no effect in either of the two assays. This clearly demonstrates that the positively charged top surface of Brat-NHL domain directly interacts with the hb RNA.

Table 3.1: Cross-links identified from Brat-NHL-hb RNA complex.

Surface/Blade	Peptide	Amino acid	RNA	Figure
top/II	^{824}DSQLLYPNR^{832}	Y^{829}	**U**	A.6.1
			UA	
			UAA	
			UG	
			U-H_2O	
			UU	
	^{815}FQFGECGK^{822}	C^{820}	**U**	A.6.2
			UA	
			UA-H_2O	
			UG	
			UUA	
	^{815}FQFGECGKR^{823}		U	
			UA	
			UA-H_2O	
			UG	
			UUA	
	^{813}FKFQFGECGK^{822}		UA	
	^{813}FKFQFGECGKR^{823}		UA	
			UUA	
bottom/II	^{804}IQIFDKEGR^{812}	K^{809}	**U-H_2O**	A.6.3
top/III	^{865}KFGATILQHPR^{875}	^{865}K/F^{866}	**U-H_2O**	A.6.4
top/IV	^{913}HLEFPNGVVVNDK^{925}	F^{916}	GCA	A.6.5
			U	
			UA	
			UC	
			UG	
			U-H_2O	
			UU	
			UUA	
	^{913}HLEFPNGVVVNDKQEIFISDNR^{934}		UG	
	^{885}IIVVECK^{891}	C^{890}	**U**	A.6.6
			UG	

Surface/Blade: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotides found to be cross-linked to the peptide. Nucleotides in bold are provided with corresponding MS/MS spectra (Appendix, Figure A.6.1-A.6.6); Figure: Figure numbers of MS/MS spectra of the corresponding cross-linked peptides provided in Appendix.

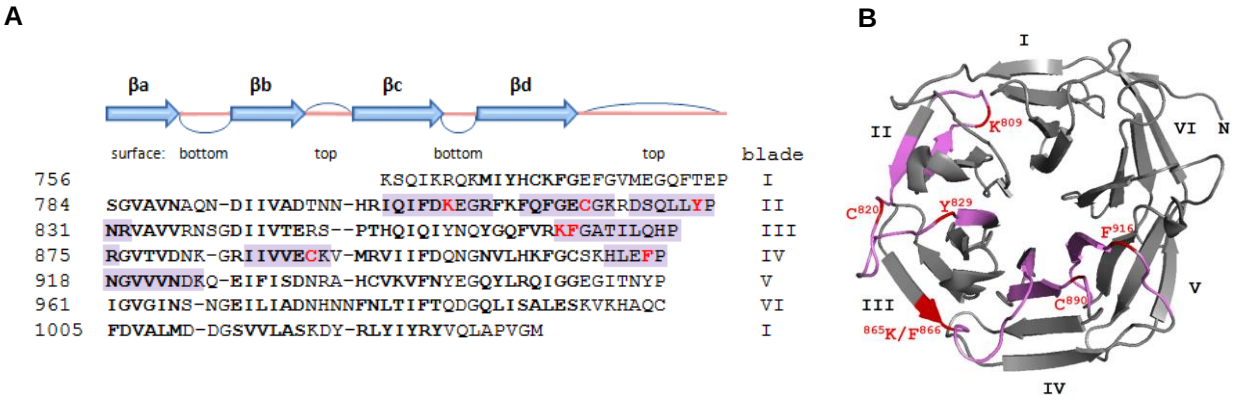


Figure 3.2: Cross-links identified from *in vitro* assembled Brat-NHL-hb RNA complex (Figure adapted from Loedige et al., 2014 and modified). (A) The identified cross-linked peptides have been highlighted in purple in a structure based sequence alignment (Edwards et al., 2003). Five out of six peptides are located on the top surface of the NHL domain. The amino acid residues found to be cross-linked to RNA adducts have been shown in bold and red (Loedige et al., 2014). (B) Crystal structure of Brat-NHL domain (PDB ID: 1Q7F, chain A) (Edwards et al., 2003) viewed from the top. The cross-linked peptides and amino acids have been displayed in purple and red colours respectively (Loedige et al., 2014).

It has also been established that BoxA motif of NRE recruits Brat. In a very recent publication by Loedige et al., 2015, the Brat-NHL domain in complex with fifteen nucleotide short RNA containing the consensus UUGUUG motif at the 5′ end followed by an oligo stretch of nine Us (Figure 3.3 A) was crystalized at resolution of 2.3 Å (PDB ID: 4ZLR) (Figure 3.3 B). In a crystal structure, the RNA was found to be interacting to three pockets, each located at the interface of two neighboring blades (between blades II and III, IV and V, VI and I) across positively charged top surface of the Brat-NHL domain. All identified RNA-protein cross-links were mapped on the crystal structure. The cross-linked amino acids Y^{829} and C^{820} were located in the first binding pocket along with U1 and U2 whereas C^{890} and F^{916} were lying in the second binding pocket along with G3 of RNA motif (Figure 3.3 B). It has been observed that amino acids Y^{829} and F^{916} were situated in closer proximity to U1 and G3-U4 respectively (Figure 3.3 C & D) than C^{820} and C^{890} in their respective binding pockets. The amino acids K^{891} and R^{875} from the cross-linked peptides, used for the mutation analysis were also mapped on the crystal structure. Both these amino acids were found to be located closer to G3 and U4 of RNA respectively (Figure 3.3 D & E). Similarly, a second Brat-NHL domain (Chain B) was bound to 3′ oligo-U tail of RNA showing the same arrangements of

cross-linked amino acids with oligo-U stretch as with consensus RNA motif at 5′ end (Figure 3.3 B).

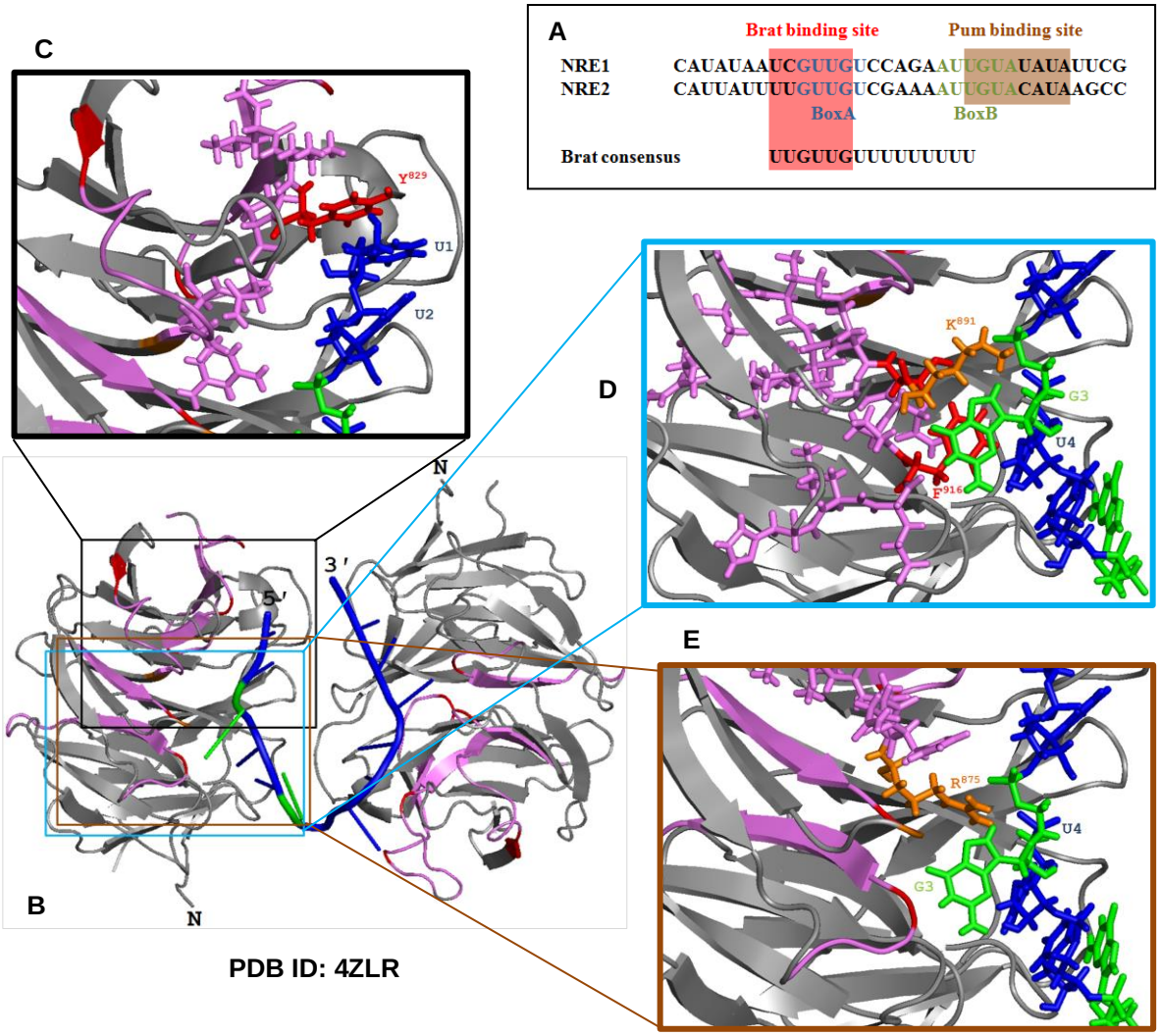


Figure 3.3: Mapping of cross-linked peptides on the crystal structure of Brat-NHL domain with small stretch of RNA. (A) Sequence of NREs of hb RNA containing BoxA and BoxB sites indicated in blue and green letters respectively. The Brat and Pum binding sites have been accentuated in light red and brown colours respectively. The Brat consensus sequence UUGUUG has been highlighted in light red followed by an oligo stretch of nine Us also crystallized with Brat-NHL domain as shown in Fig. 3.3 B (Figure adapted from Loedige et al., 2015and modified). (B) Crystal structure of both Brat-NHL monomers bound to the consensus motif RNA shown in Fig. 3.3 A (PDB ID: 4ZLR) (Loedige et al., 2015). The cross-linked peptides and amino acids have been shown in purple and red colours respectively. The amino acids (K^{891} and R^{875}) used for mutation analysis from the cross-linked peptides other than the cross-linked amino acids have been indicated in orange colour. The RNA has been shown in blue (Us) and green (Gs) colours. (C) The zoomed in crystal structure showing the cross-linked peptide in sticks (purple) with cross-linked amino acid Y^{829} (red) lying in close proximity to nucleotide U1 (blue). (D) The zoomed in structure presenting the cross-linked peptide in sticks (purple) with F^{916} (red) lying closer to G3-U4 (green-blue) and K^{891} (orange) near to G3 (green). (E) The zoomed in view of the cross-linked peptide in sticks (purple) with R^{875} (orange) lying closer to U4 (blue).

3.2 Identification of Cross-links from CWC2-U4 and U6 snRNAs Complexes

Splicing is an essential step in eukaryotic pre-mRNA processing. In yeast, this process is catalysed by the spliceosomes and nineteen complex (NTC) consisting of small nuclear ribonucleoproteins and Prp19 along with number of associated splicing factors respectively. Among these splicing factors, CWC2 has been reported to be capable of binding RNA (Schmitzová et al., 2012). It has been shown that CWC2 contacts the U6 snRNA during splicing in yeast extracts. Furthermore, it can also interact with U1, U4 and U5 snRNAs, *in vitro* (McGrail et al., 2009). Although few identified interaction sites have already been published by Schmitzová et al., 2012 by UV-cross-linking, but here additional interaction sites of CWC2 have been reported in addition to the previously reported ones which will later help in quantitative analysis of the CWC2-U4 and U6 snRNA complexes.

The crystal structure of CWC2 at a resolution of 2.4 Å shows that it is comprised of CCCH-type zinc finger (ZnF), a RNA recognition motif (RRM) domain and a Torus domain. The N-terminal part of the RRM domain is connected to a C-terminal appendage of Torus domain by a positively charged connector element showing the tendency to interact with RNA (Figure 3.4). Here the U4 and U6 snRNAs have been used for the interaction studies of CWC2 with snRNAs and will provide a better comparison to the previously published data by Schmitzová et al., 2012.

In order to have detailed insight into CWC2 binding activity with U4 and U6 snRNAs, the CWC2 protein (~39 kDa) was incubated with *in vitro* transcribed U4 (162 nt) and U6 (121 nt) snRNAs separately. The complex was UV-cross-linked, followed by enrichment of cross-links according to the protocol, described in sections 2.2.5.2 and 2.2.5.3 of materials and methods respectively. The samples were analyzed on a Q-Exactive mass spectrometer. The data obtained was then analyzed by OpenMS as mentioned in section 2.2.8.2 of materials and methods. The candidate spectra of the cross-linked peptides were then manually validated.

During the current studies, eighteen peptides of CWC2 protein have been found to be cross-linked with U4/U6 snRNAs. Except for one cross-linked peptide, the cross-linking site has been curtailed to the amino acid level, showing the probable cross-linked amino acid interacting with U6/U4 snRNAs (Table 3.2). It has been observed that three cross-linked amino acids (K^{10}, W^{37} and C^{111}) are lying in the Torus domain, three (K^{116}, Y^{120} and S^{129}) in the connector element and eight (Y^{138}, K^{152}, F^{162}, R^{172}, Y^{188}, L^{222}, K^{224} and C^{181}) in the RRM domain. One cross-linked peptide ^{87}CEYLHHIPDEEDIGK^{101} occupies the ZnF domain (Figure 3.5 A & B). On comparison it has been noticed that seven of the identified cross-linked peptides are the same as reported by Schmitzová et al., 2012 and eleven are newly identified peptides that are found to be cross-linked with U4/U6 snRNAs.

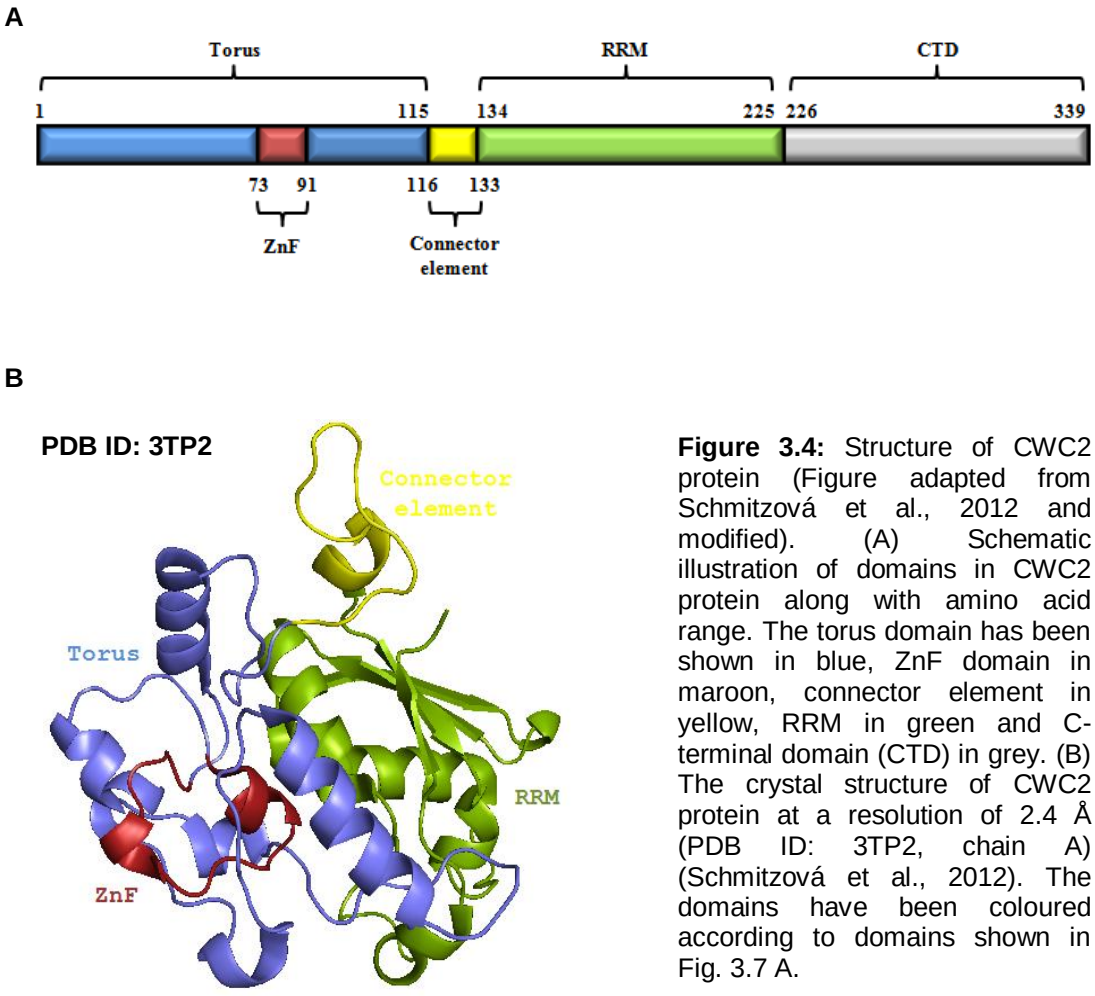


Figure 3.4: Structure of CWC2 protein (Figure adapted from Schmitzová et al., 2012 and modified). (A) Schematic illustration of domains in CWC2 protein along with amino acid range. The torus domain has been shown in blue, ZnF domain in maroon, connector element in yellow, RRM in green and C-terminal domain (CTD) in grey. (B) The crystal structure of CWC2 protein at a resolution of 2.4 Å (PDB ID: 3TP2, chain A) (Schmitzová et al., 2012). The domains have been coloured according to domains shown in Fig. 3.7 A.

Table 3.2: Cross-links identified from CWC2-U4 snRNA and CWC2-U6 snRNA complexes.

Domain	Peptide	Amino acid	RNA	Figure
Torus	^{8}SAKVQVK^{14}	K^{10}	**U-H$_2$O**	A.6.7
	^{37}WSQGFAGNTR^{46}	W^{37}	**U**	A.6.8
	^{106}TEVLDCFGR^{114}	C^{111}	U	A.6.9
			U+152	
			UA	
ZnF	^{87}CEYLHHIPDEEDIGK^{101}	-	**U+152**	A.6.10
			UA	
			UG	
connector element	^{115}EKFADYR^{121}	K^{116}	**U-H$_2$O**	A.6.11
	^{117}FADYR^{121}	Y^{120}	U	A.6.12
			UA	
			UG	
	^{122}EDM(Oxidation)GGIGSFR^{131}	S^{129}	**U**	A.6.13
			UA	
			UG	
			UU	
RRM/RNP2	^{134}NKTLYVGGIDGALNSK^{149}	Y^{138}	U-H$_2$O	A.6.14
	^{136}TLYVGGIDGALNSK^{149}		**U**	
			UA	
			UAA	
			UAG	
			UCG	
			UG	
			UU	
			UUA	
			UUG	

Continued.....

Domain	Peptide	Amino acid	RNA	Figure
RRM	^{150}HLKPAQIESR^{159}	K^{152}	**U-H$_2$O**	A.6.15
			UU	
	^{160}IRFVFSR^{166} ^{162}FVFSR^{166}	F^{162}	U	A.6.16
			U	
			U-H$_2$O	
	^{167}LGDIDRIR^{174}	R^{172}	**UA**	A.6.17
			UG	
	^{186}FKYQANAEFAK^{196} ^{188}YQANAEFAK^{196}	Y^{188}	U	A.6.18
			U	
	^{217}EGTGLLVK^{224}	L^{222}	**U-H$_2$O**	A.6.19
	^{217}EGTGLLVKWANEDPDPAAQK^{236}	K^{224}	**U-H$_2$O**	A.6.20
RRM/RNP1	^{180}NCGFVK^{185}	C^{181}	U	A.6.21
			U+152	
			UA	
			UA+152	
			UAA	
			UAA+152	
			UG	
CTD	^{225}WANEDPDPAAQK^{236} ^{225}WANEDPDPAAQKR^{237}	W^{225}	**U**	A.6.22
			UA	
			UG	
			U	
			UA	
			UG	
	^{276}TFPEASVDNVK^{286}	F^{277}	**U**	A.6.23
	^{315}ENISSKPSVGK^{325}	K^{320}	UA-H$_2$O	A.6.24
			U-H$_2$O	

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence. The peptides in yellow have also been reported by Schmitzová et al., 2012; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids in yellow have also been reported by Schmitzová et al., 2012; RNA: Nucleotides found to be cross-linked to the peptide. The mass increment of 152 Da ($C_4H_8S_2O_2$) is due to the involvement of DTT in cysteine-uracil cross-links (Zaman et al., 2015). Nucleotides in bold are provided with corresponding MS/MS spectra (Appendix, Figure A.6.7-A.6.24); Figure: Figure numbers of MS/MS spectra of the corresponding cross-linked peptides provided in Appendix.

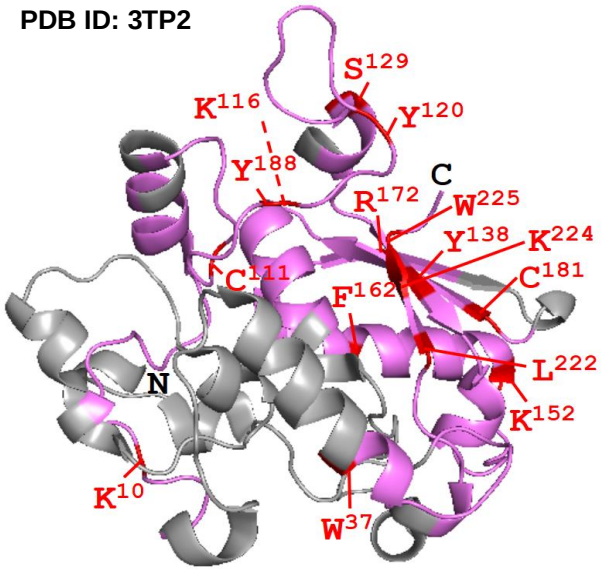
The RRM, ZnF and Torus domains have been shown to have tendencies to interact with RNA previously (Schmitzová et al., 2012). The RRM domain has two well conserved RNP1 and RNP2 motifs. Among seven identified cross-linked peptides, two peptides ^{180}NCGFVK^{185} and ^{136}TLYVGGIDGALNSK^{149} have been found to be lying in these RNP1 and RNP2 motifs respectively (Figure 3.5 C). It has been reported by Schmitzová et al., 2012 that the double mutants Y^{138}/Y^{120} and Y^{138}/C^{181} depicted impaired binding of CWC2 protein with U6 snRNA whereas Y^{138}/K^{152} showed the insignificance of K^{152} for RNA binding by EMSA. The C-terminal domain of CWC2 have been reported to be interacting with the WD40 domain of Prp19 protein (Ohi & Gould, 2002; Vander Kooi et al., 2010) during splicing and is showing no detectable RNA-binding activity by EMSA (Schmitzová et al., 2012). During the interaction analysis of CWC2 with U4 and U6 snRNAs by UV-cross-linking, three peptides of the C-terminal domain ^{225}WANEDPDPAAQK^{236}, ^{276}TFPEASVDNVK^{286} and ^{315}ENISSKPSVGK^{325} with amino acid residues W^{225}, F^{277} and K^{320} have been found to be cross-linked to U4/U6 snRNAs.

A

CWC2 Protein Sequence

```
MTSWRDKSAKVQVKESELPSSIPAQTGLTFNIWYNKWSQGFAGNTRFVSPFALQPQLHSG 60
KTRGDNDGQLFFCLFFAKGMCCLGPKCEYLHHIPDEEDIGKLALRTEVLDCFGREKFADY 120
REDMGGIGSFRKKNKTLYVGGIDGALNSKHLKPAQIESRIRFVFSRLGDIDRIRYVESKN 180
CGFVKFKYQANAEFAKEAMSNQTLLLPSDKEWDDRREGTGLLVKWANEDPDPAAQKRLQE 240
ELKLESLNMMVHLINNNTNSAGTEVNNKNNERLDRTFPEASVDNVKKRLLPLDNGMESDD 300
FIEKLKKVKKNISRENISSKPSVGKLGGPLLDYLSSDED                        339
```

B



C

Consensus RNP1: (R/K) – G – (F/Y/Q) – (G/A) – (F/Y) – (V/I/L) – X – (F/Y)
CWC2 RNP1 : K^{179} N C G F V K F

Consensus RNP2: (I/V/L) – (F/Y) – (I/V/L) – X – Q – L
CWC2 RNP2 : L^{137} Y V G G I

Figure 3.5: Identified cross-linked peptides of CWC2-U4 snRNA and CWC2-U6 snRNA complexes. (A) The identified cross-linked peptides have been highlighted in purple in a sequence of CWC2 protein (Schmitzová et al., 2012). The cross-linked amino acid residues have been indicated in red. (B) The crystal structure of CWC2 protein at a resolution of 2.4 Å (PDB ID: 3TP2, chain A) (Schmitzová et al., 2012). The identified cross-linked peptides and amino acid residues have been illustrated in purple and red colours respectively. (C) Comparison of consensus RNP1 and RNP2 sequences (Maris et al., 2005) (Figure adapted from Schmitzová et al., 2012). The amino acids shown in red colour are important for RNA binding. X depicts any amino acid.

3.3 Quantitative Analysis of CWC2-U4 snRNA and U6 snRNA Cross-links

The CWC2 protein has been reported to bind with RNAs non-specifically *in vitro* (McGrail et al., 2009). The previous studies have shown that CWC2 interacts with RNAs through RRM, torus, connector element and ZnF domains (Schmitzová et al., 2012). In order to determine whether the CWC2 as a whole protein and its domains individually possess a preference for any RNA to interact or bind indiscriminately, the comparative studies have been carried out by relative quantification approach. For relative comparison of different conditions, relative quantification through label based approach proves to be extremely useful (Nikolov et al., 2012).

For relative quantitative analysis of CWC2-U4 snRNA and U6 snRNA cross-links, the U6 and U4 snRNAs, transcribed by using non-labeled (light (^{12}C)) and isotopically labeled (heavy (^{13}C)) UTP respectively were pooled in 1:1 ratio and were incubated with CWC2 protein. In another set of identical experiment, the labels were swapped for RNA transcription. The former set of experiment was referred as “forward” and the later one as “reverse” respectively. The complexes were UV irradiated and the cross-links were enriched according to the protocol described in sections 2.2.6.1 and 2.2.5.3 of materials and methods respectively. The samples were run on Q-Exactive mass spectrometer (Figure 3.6). The data obtained was then analysed by OpenMS. The candidate spectra of the cross-linked peptides were then manually validated (Figure 3.7 A & B).

The relative quantification was first carried out by generating extracted ion chromatograms (XICs) of the identified cross-linked peptides precursors *m/z* at their respective retention time in Xcalibur software (Thermo Scientific) as mentioned in section 2.2.8.2 of materials and methods. In case the cross-linked peptides have been identified in one condition and if the intensity of the cross-linked peptide is not sufficient to trigger the MS/MS fragmentation in another, the XIC was generated by using expected precursor *m/z* at its particular retention time. The ratios were calculated from the peak area of the extracted ion

chromatograms. The results obtained were used for further statistical analysis and data interpretation.

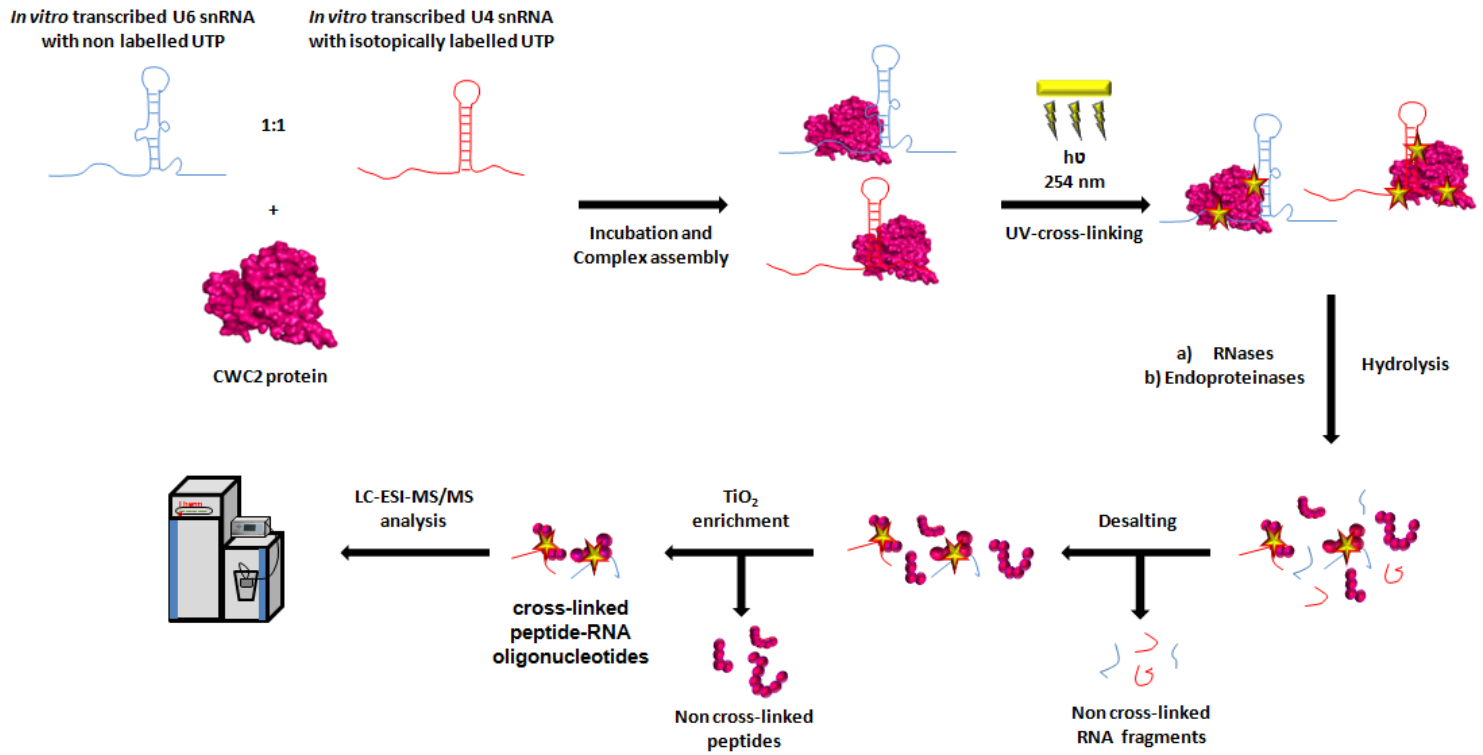


Figure 3.6: Workflow for relative quantification of U4/U6 snRNA-CWC2 cross-links. The differentially labeled U4 and U6 snRNAs are pooled in 1:1 ratio and incubated with CWC2 protein. The sample is UV-irradiated and hydrolyzed by RNases and endoproteinases. The non-cross-linked RNA fragments are removed by desalting and the cross-links are further enriched by TiO_2 chromatography. The isolated cross-links are analyzed by LC-ESI-MS/MS. The shown workflow is of forward experiment and for the reverse experiment only labels are swapped for transcription of snRNAs.

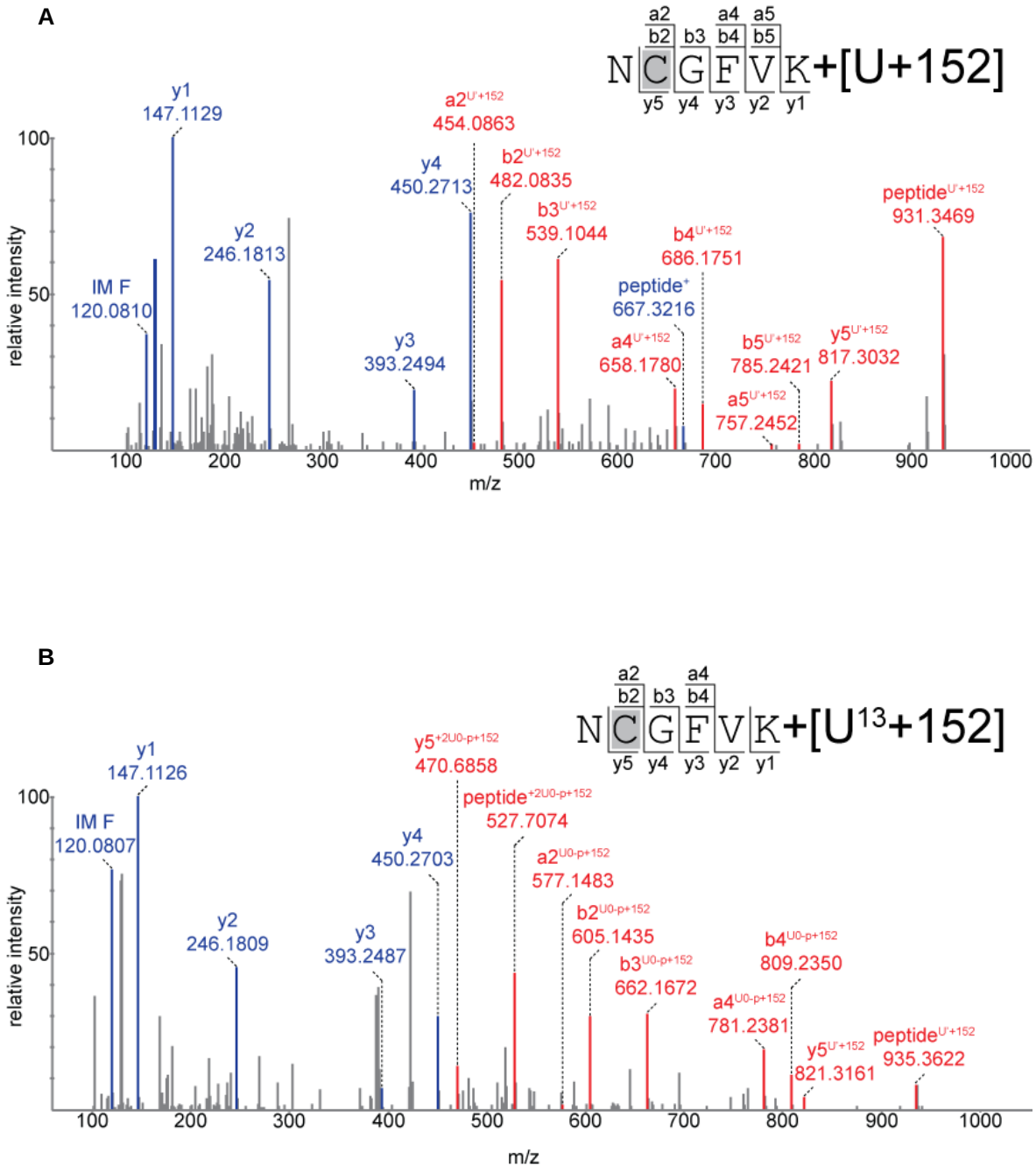


Figure 3.7: MS/MS spectrum of peptide NCGFVK cross-linked to U+152. (A) MS/MS spectrum of peptide cross-linked to [U+152] having light uracil. The mass shift of b-series from b2 ion and of y-series from y5 ion shows that Cysteine is a cross-linked amino acid. (B) MS/MS spectrum of peptide cross-linked to [U+152] having heavy uracil. The mass shift of b-series from b2 ion and of y-series from y5 ion depicts that Cysteine is a cross-linked amino acid.

Table 3.3: U4/U6 snRNA-CWC2 cross-links identified in relative quantification experiments.

Domain	Peptide	Amino acid	RNA	Figure
Torus	^{8}SAKVQVK^{14}	K^{10}	**U-H$_2$O**	A.6.7
			UA-H$_2$O	
	^{37}WSQGFAGNTR^{46}	W^{37}	**U**	A.6.25
	^{47}FVSPFALQPQLHSGK^{61}	F^{47}	**U-H$_2$O**	A.6.26
	^{106}TEVLDCFGR^{114}	C^{111}	U	A.6.27
			U+152	
			UA+152	
ZnF	^{87}CEYLHHIPDEEDIGK^{101}	-	U	A.6.28
			U+152	
connector element	^{115}EKFADYR^{121}	K^{116}	**U-H$_2$O**	A.6.11
	^{122}EDM(Oxidation)GGIGSFR^{131}	S^{129}	**U**	A.6.13
RRM/RNP2	^{134}NKTLYVGGIDGALNSK^{149} ^{136}TLYVGGIDGALNSK^{149}	Y^{138}	U-H$_2$O	A.6.29
			UU-H$_2$O-HPO$_3$	
			U	
			UA	
			UAA	
			UG	
			UU	
RRM	^{150}HLKPAQIESR^{159}	K^{152}	**U-H$_2$O**	A.6.15
			UA-H$_2$O	
	^{186}FKYQANAEFAK^{196} ^{188}YQANAEFAK^{196}	Y^{188}	U-H$_2$O	A.6.30
			U	
			UA	
	^{217}EGTGLLVK^{224}	L^{222}	**U-H$_2$O**	A.6.19
	^{217}EGTGLLVKWANEDPDPAAQK^{236}	K^{224}	**U-H$_2$O**	A.6.20

Continued.....

Domain	Peptide	Amino acid	RNA	Figure
RRM/RNP1	^{180}NCGFVK^{185}	C^{181}	U	A.6.31
			U+152	
			UA	
			UA+152	
			UAA	
			UAA+152	
			UG	
			UG+152	
CTD	^{225}WANEDPDPAAQK^{236}	W^{225}	**U**	A.6.22
			U-H_2O	
	^{310}KNISR^{314}	K^{310}	**UA-H_2O**	A.6.32
	^{315}ENISSKPSVGK^{325}	K^{320}	**U-H_2O**	A.6.24
			UA-H_2O	

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence. Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. RNA: Non-labeled/labeled uracil along with other nucleotides found to be cross-linked to the peptide. The cross-linked peptides along with the isotopically labeled (heavy (^{13}C)) uracil, identified at the MS2 level are indicated by highlighting the nucleotides in yellow. The mass increment of 152 Da ($C_4H_8S_2O_2$) is due to the involvement of DTT in cysteine-uracil cross-links (Zaman et al., 2015). Nucleotides in bold are provided with corresponding MS/MS spectra (Appendix, Figure A.6.7-A.6.32); Figure: Figure numbers of MS/MS spectra of the corresponding cross-linked peptides provided in Appendix.

Total sixteen peptides have found to be cross-linked to U4/U6 snRNAs in forward and reverse replicates (Table 3.3). The *m/z* of the identified cross-links with light and heavy labeled uracil have been theoretically calculated for the charge states of +2 and +3. The area under the curve (AUC) have been estimated for all the cross-linked peptides by generating the XICs. The area of each cross-linked peptide along with its miss-cleavage state and all of the charge states (+2 and +3) and RNA moiety combinations have been added and used for further calculations. For statistical analysis log transformation has been applied to normalize the peak areas of the cross-links and for normality confirmation

Kolmogorov-Smirnov and Shapiro-Wilk tests have been applied (mentioned in section 2.2.8.4 of materials and methods). (Figure 3.8 A & B, Appendix Table 6.6).

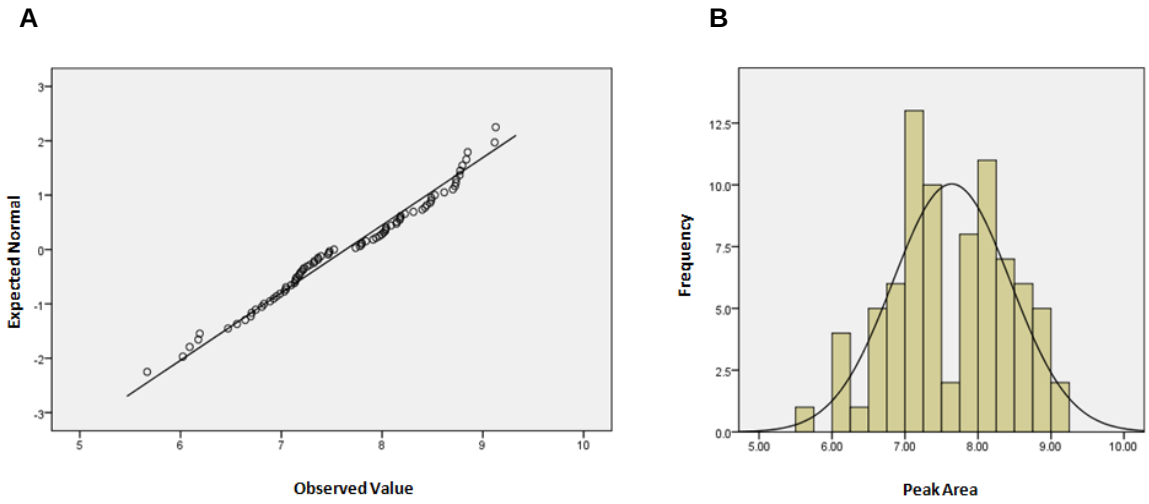


Figure 3.8: Normalization graphs. (A) The Q-Q plot of observed logarithmic (log_{10}) peak area vs. expected normal value. (B) The histogram of observed logarithmic (log_{10}) peak area values plotted against their frequencies confirm normality.

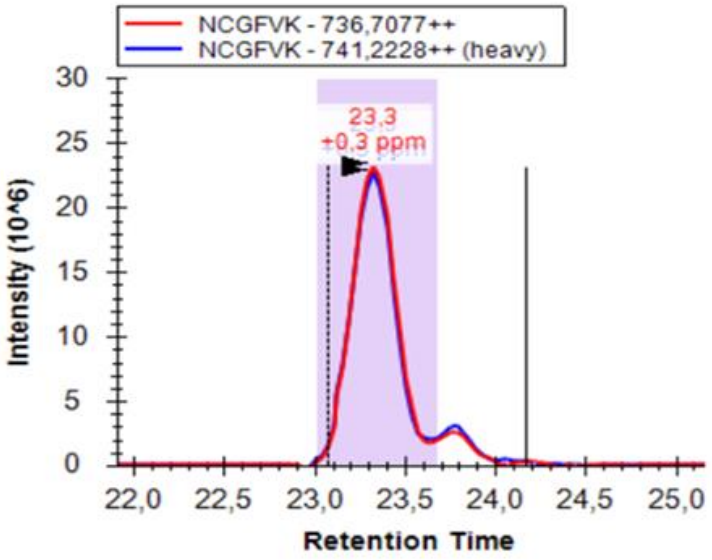
Keeping the selection stringent for relative quantification, three cross-linked peptides, ^{106}TEVLDCFGR^{114} of torus domain, ^{136}TLYVGGIDGALNSK^{149} of RNP2 motif of RRM domain and ^{180}NCGFVK^{185} of RNP1 motif of RRM domain have been selected for further studies. These cross-linked peptides have been identified in most of the forward and reverse replicates and used for calculating heavy to light (H/L) and light to heavy (L/H) ratios respectively. Few cross-links peak area values have been found to be missing in their light-heavy pair so their ratios cannot be calculated and are omitted from comparative studies (Appendix Table 6.6).

The relative quantitation was also carried out by generating the extracted ion chromatograms using Skyline software (as described in section 2.2.8.3 of materials and methods). The spectral library of above mentioned three cross-linked peptides along with different oligonucleotide combinations (Appendix Table 6.7) was built. The identified oligonucleotide combinations were added as

modifications. The Savitzky-Golay smoothing was applied to the extracted ion chromatograms (XICs). The most intense peak among the light and heavy pair was selected for marking the peak boundaries manually after carefully inspecting the isotopic pattern (Figure 3.9 A & B, Appendix Figures 6.33-6.47). The ratios were then calculated from the peak areas of the extracted ion chromatograms which were further used for the statistical analysis and data interpretation (Appendix Table 6.7).

It has been reported that the CWC2 protein binds more efficiently to U4 snRNA than U6 snRNA *in vitro* (McGrail et al., 2009). The current comparative study of cross-linking tendency of CWC2 protein to U4 and U6 snRNA by taking individual ratio (ratio of each peptide cross-linked to per RNA moiety) into consideration has shown that the CWC2 is more inclined towards U4 snRNA than U6 snRNA for cross-linking (Appendix Table 6.6 & Table 6.8). However, the combined effect of the three cross-linked peptides (^{106}TEVLDCFGR^{114}, ^{136}TLYVGGIDGALNSK^{149} and ^{180}NCGFVK^{185}) along with all the combinations of RNA moiety shows that both U4 and U6 snRNAs cross-link with almost similar efficiency with CWC2 protein in all the individual forward and reverse replicates. The statistical analysis by one-way ANOVA (Duncan's test) at p=0.05 (mention in section 2.2.8.4 of materials and methods) proves that there is no significant difference in the cross-linking efficiency of CWC2 protein to both U4 and U6 snRNAs (Figure 3.10).

A



B

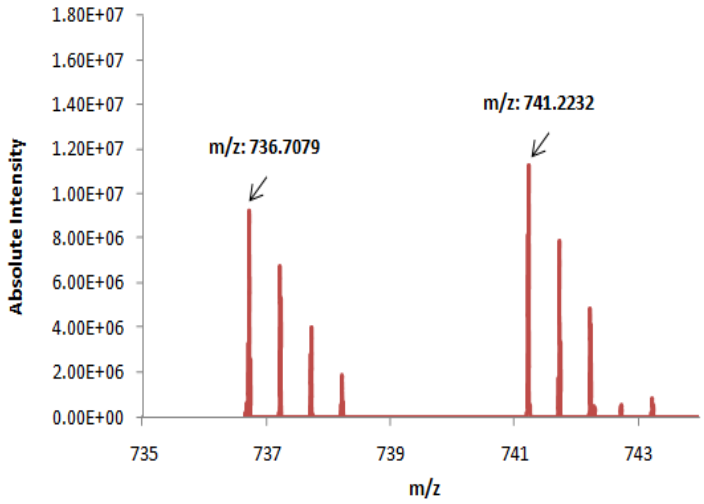


Figure 3.9: Example of Extracted ion chromatogram and isotopic distribution of a cross-link showing the mass difference due to labeling. (A) The extracted ion chromatograms (XICs) of both light-heavy pair of a peptide NCGFVK cross-linked to UA+152 with the charge state of +2. The XIC in red and blue are from doubly charged light labeled and doubly charged heavy labeled forms of the cross-link respectively. (B) The isotopic distribution of NCGFVK + [UA+152] cross-link, with non-labeled (light (^{12}C)) UTP at *m/z* 736.7079 and with isotopically labeled (heavy (^{13}C)) UTP at *m/z* 741.2232 with the charge state of +2. The difference between the two monoisotopic peaks corresponds to the number of ^{13}C labeled atoms.

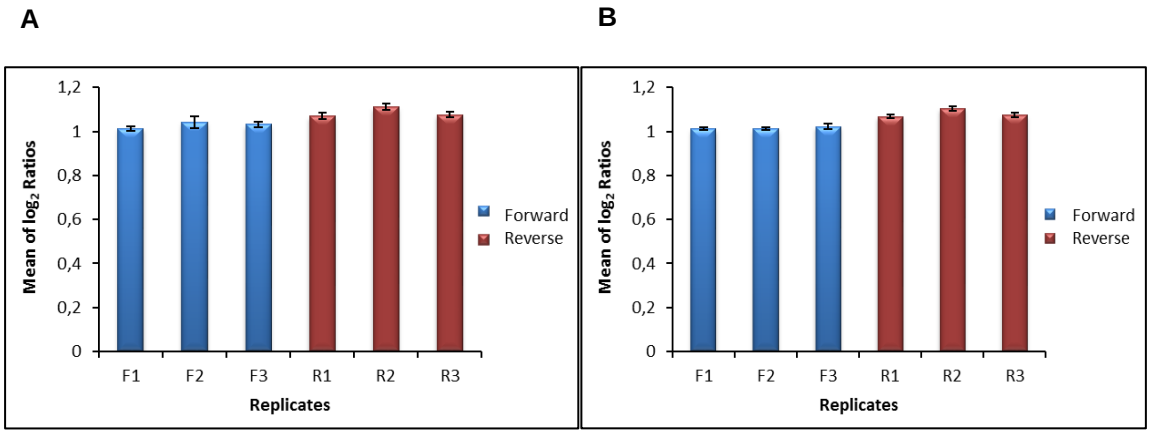


Figure 3.10: Bar diagrams of Mean of log_2 ratios of CWC2-U4/U6 snRNA cross-links. (A) The means are calculated from the logarithmic (log_2) ratios of peak area from Xcalibur of all the three cross-linked peptides of each of the forward and reverse replicates. (B) The means are calculated from the logarithmic (log_2) ratios of peak area from Skyline of all the three cross-linked peptides of each of the forward and reverse replicates. The statistical analysis shows that there is no significant difference in the cross-linking efficiency of CWC2 protein to U4 and U6 snRNAs among the replicates of forward experiment as well as of the reverse experiment.

In order to further investigate if any domain/motif of CWC2 protein preferentially cross-links to U4 or U6 snRNA, the relative quantification was carried out by using the ratios of the cross-linked peptides of Torus domain and RNP1 and RNP2 motifs in both forward and reverse experiments. The analysis has revealed that all the domains/motifs of CWC2 protein cross-link to U4 and U6 snRNAs almost indiscriminately (Figure 3.11).

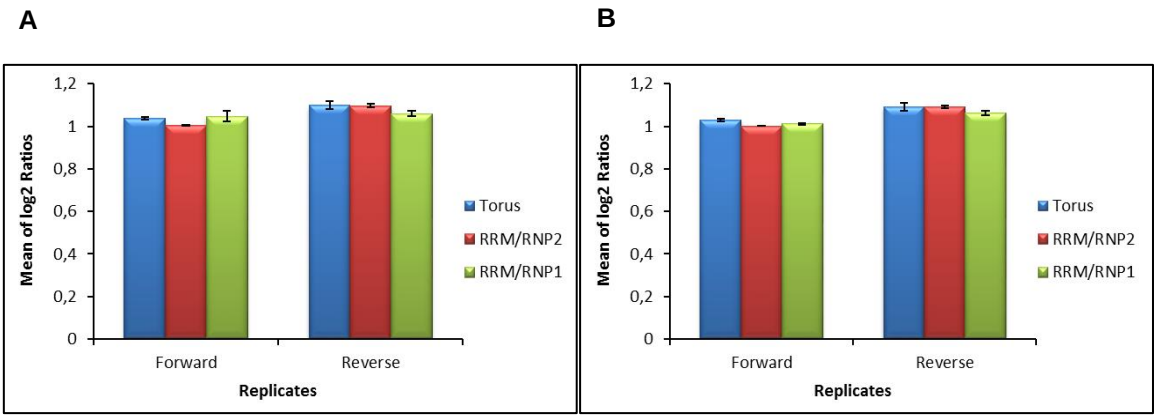


Figure 3.11: Bar diagram of Mean of log_2 ratios of individual cross-linked domain/motif to U4/U6 snRNA. (A) The means are calculated from the logarithmic (log_2) ratios of peak area from Xcalibur of all the three forward and reverse replicates of each domain/motif cross-link. (B) The means are calculated from the logarithmic (log_2) ratios of peak area from Skyline of all the three forward and reverse replicates of each domain/motif cross-link.

3.4 Identification of Cross-links from MS2-MBP Protein

The ribonucleoprotein complexes play a significant role in all fundamental cellular processes. There are several methods to purify these complexes. For the current studies, the H/E complex from *HeLa* nuclear extract was purified by MS2-MBP affinity purification method (Zhou & Reed, 2003). The method is based on the affinity of the bacteriophage MS2 coat protein for its target MS2 RNA stem-loops. For the isolation of *in vitro* assembled RNP complex, the bait RNA (PM5 pre-mRNA/MINX pre-mRNA) was tagged with three MS2 RNA stem-loops. The MS2-MBP fusion protein was bound to these MS2 hairpin loops. The RNA-MS2-MBP complex was then used to assemble the H/E complex by incubating with *HeLa* nuclear extract. The assembled complex was isolated by affinity purification using amylose beads and maltose in elution buffer as described in section 2.2.4.4 of materials and methods. The evident binding of MS2-MBP fusion protein to the MS2 RNA stem-loops of target RNA (Figure 3.12) explicitly require the detailed account of probable interaction sites of this protein with RNA.

In order to find the interactions of MS2-MBP fusion protein, the data obtained by the MS analysis of the UV-cross-linked RNP complex from *HeLa* nuclear extract on LTQ Orbitrap Velos mass spectrometer (as described in section 2.2.5.3 of materials and methods) was analyzed by OpenMS as mentioned in section 2.2.8.2 of materials and methods using the database containing MS2-MBP sequence. The candidate spectra of the cross-linked peptides identified were then manually validated.

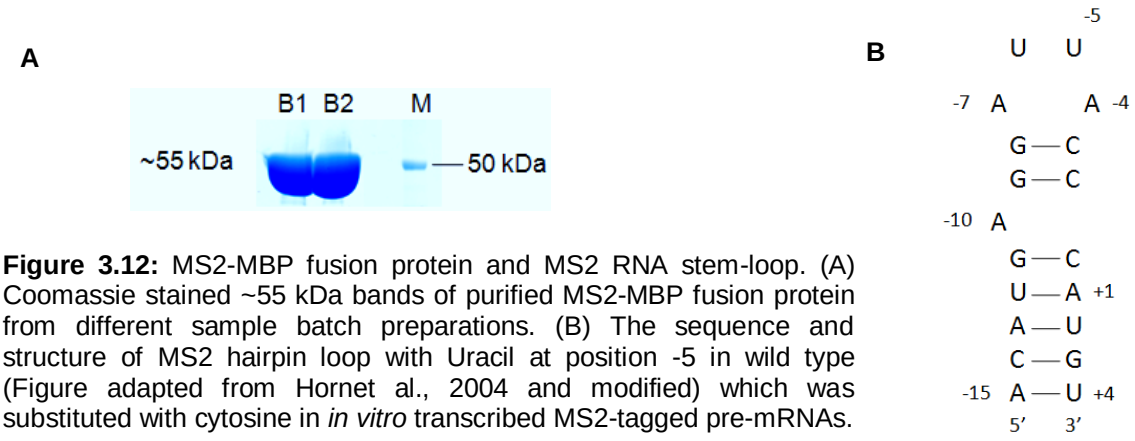


Figure 3.12: MS2-MBP fusion protein and MS2 RNA stem-loop. (A) Coomassie stained ~55 kDa bands of purified MS2-MBP fusion protein from different sample batch preparations. (B) The sequence and structure of MS2 hairpin loop with Uracil at position -5 in wild type (Figure adapted from Hornet al., 2004 and modified) which was substituted with cytosine in *in vitro* transcribed MS2-tagged pre-mRNAs.

Twelve peptides of maltose binding protein and two of MS2 coat protein have been found to be cross-linked to the nucleotides (Figure 3.13 A, Table 3.4). Overall, out of fourteen peptides, the probable cross-linked amino acid residues have been screened for seven peptides only. For seven of the cross-linked peptides of MBP protein, the exact cross-linked amino acid could not be identified however the cross-linking regions have been narrowed down upto few amino acids as shown in Table 3.4.

Table 3.4: Cross-links from MS2-MBP protein.

Protein (UniProt ID)	Domain/ Segment	Peptide	Amino acid	RNA	Figure
MBP (P0AEX9)	N	^{27}KFEKDTGIK^{35} ^{28}FEKDTGIK^{35}	K^{30}	**U-H_2O** U-H_2O UA-H_2O	A.6.48
	N	^{31}DTGIKVTVEHPDK^{43}	K^{35}	**U-H_2O**	A.6.49
	N	^{36}VTVEHPDKLEEK^{47}	K^{43}	U-H_2O **UA-H_2O**	A.6.50
	N	^{68}FGGYAQSGLLAEITPDKAFQDK^{89}	K^{84}	**U-H_2O**	A.6.51
	C	^{121}DLLPNPPKTWEEIPALDK^{138}	^{124}P-T^{129}	**U-H_2O**	A.6.52
	C	^{129}TWEEIPALDKELK^{141}	^{138}K/E^{139}	**U-H_2O**	A.6.53
	N	^{275}ELAKEFLENYLLTDEGLEAVNK^{296}	^{275}E-L^{285}	**U-H_2O**	A.6.54
	N	^{279}EFLENYLLTDEGLEAVNKDKPLGAVALK^{306}	^{285}L-K^{296}	**U-H_2O**	A.6.55
	N	^{297}DKPLGAVALK^{306}	^{297}D/K^{298}	**U-H_2O** UA-H_2O	A.6.56
	3	^{307}SYEEELAKDPR^{317}	^{314}K/D^{315}	**U-H_2O**	A.6.57
	C	^{356}QTVDEALKDAQTNSSSVPGR^{375}	K^{363}	**U-H_2O**	A.6.58
	C	^{364}DAQTNSSSVPGRGSIEGR^{381}	^{370}S-R^{375}	UA **UCA**	A.6.59
MS2 (P03612)		^{438}KYTIKVEVPK^{447}	K^{442}	**U-H_2O**	A.6.60
		^{443}VEVPKGAWR^{451}	K^{447}	**U-H_2O**	A.6.61

Protein: Protein name along with its UniProt ID; Domain/Segment: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids that are found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotides found to be cross-linked to the peptide. Nucleotides in bold are provided with corresponding MS/MS spectra (Appendix, Figure A.6.48-A.6.61); Figure: Figure numbers of MS/MS spectra of the corresponding cross-linked peptides provided in Appendix.

Maltose binding protein has previously been known to play role in several biological processes like carbohydrate transport and cellular response to DNA damage stimulus etc. but its direct interaction with nucleic acid has not been reported until now. In the current study, twelve peptides of MBP have been found to be cross-linked to uracil. In most of these peptides, lysine has been found to be a cross-linked amino acid residue. These cross-linked peptides have been mapped on the 1.8 Å crystal structure of monomeric maltose binding protein (Figure 3.13 B). The protein has two distinct globular N (residues 1-109 and 264-309) and C (114-258 and 316-370) domains, joined by three segments 1 (residues 110-113), 2 (residues 259-263) and 3 (residues 310-315). There is a deep groove between the two domains containing the oligosaccharide-binding site. Each domain is comprised of central β-pleated sheet, flanked on both sides by α-helices (Quiocho et al., 1997). Most of the cross-linked peptides are found to be lying in the region of α-helices. Out of twelve cross-linked peptides, seven have been occupying N domain, four occupying C domain and one occupying the third segment joining the two domains.

A

MS2-MBP Fusion protein Sequence

MKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGD 56
GPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVE 112
ALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGY 168
AFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETA 224
MTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEF 280
LENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQ 336
MSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSVPGRGSIEGRASNFTQFVLVD 392
NGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG 448
AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY 497

B

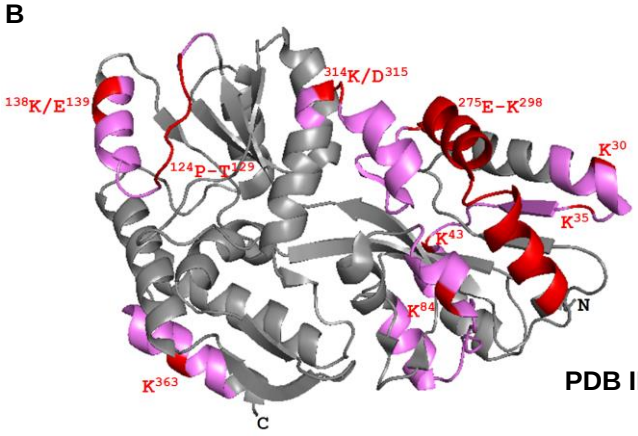


Figure 3.13: Identified cross-linked peptides of MS2-MBP fusion protein. (A) The identified cross-linked peptides have been highlighted in purple in a sequence of MS2-MBP fusion protein (Sharff et al., 1992; Grahn et al., 1999). The cross-linked amino acid residues and regions have been indicated in red. The MBP sequence has been shown in bold and blue colour whereas the MS2 coat protein sequence has been presented in bold and brown colour. (B) Crystal structure of MBP viewed from the top (PDB ID: 1OMP) (Sharff et al., 1992). The identified cross-linked peptides and amino acid residues/regions have been illustrated in purple and red colours respectively.

Much has been reported about the interaction of RNA bacteriophage coat protein MS2 with specific stem-loop structure of viral RNA. In the current studies, two peptides of MS2 protein have been found to be cross-linked to uracil. It has also been observed that lysine is a cross-linked amino acid in the peptides which is in concordance to the observation of Budowsky et al., 1976 in which lysine residues in coat protein were found to be cross-linked to the genomic nucleic acid upon UV-irradiation. The cross-linked peptides have been mapped on 2.86 Å crystal structure of MS2 coat protein with a stretch of RNA (Figure 3.14). The structure is comprised of three chains each containing 129 amino acids. The structure has been shown in three coat protein dimers. Each monomer is composed of five stranded β-sheet and two α-helical segments. The cross-linked lysine residues K^{442} and K^{447} of peptides ^{438}KYTIKVEVPK^{447} and ^{443}VEVPKGAWR^{451} respectively are found to be lying in β-sheet in closer proximity to cytosine at position −5 of nucleotides in RNA in the crystal structure. This has also been reported by Grahn et al., 1999 that K^{442} is among the probable amino acids that formed cross-links with nucleotides and the substitution of wild type uracil by cytosine at position −5 resulted in stronger binding of RNA hairpin to MS2 coat protein so it can be anticipated that in the present study, K^{442} is the probable amino acid and uracil (in case of contamination)/cytosine (in case of bait RNA) at position −5 is the probable nucleotide taking part in cross-linking. The study of the crystal structure of MS2 coat protein revealed the partial mapping of the cross-linked peptide ^{443}VEVPKGAWR^{451}, reason being the difference arising in the MS2 coat protein sequence of MS2-MBP fusion protein from the original sequence of the available MS2 coat protein alone with the deletion of ^{67}VATQTVGGVELPVA^{80} region.

In short the studies showed that the MS2 coat protein interacts with the specific stem-loop structure of the RNA. In addition to this the current study also revealed that the MBP part of the MS2-MBP fusion protein may undergo cross-linking with RNA. Mostly the lysine residue of the MS2-MBP protein get cross-linked to uracil with the loss of water molecule.

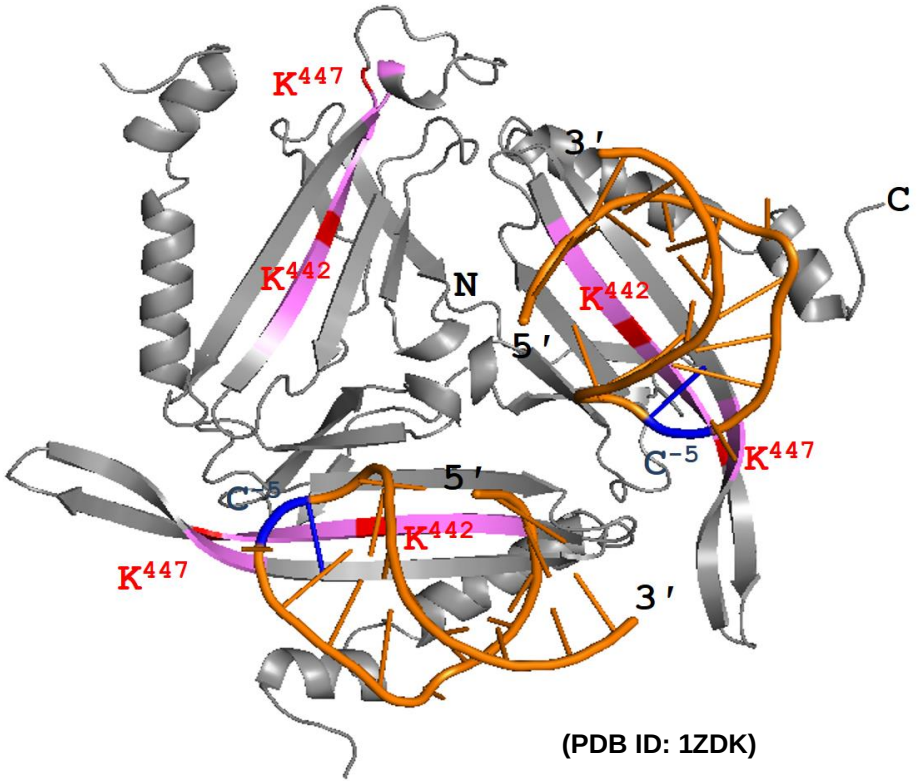


Figure 3.14: Crystal structure of MS2 coat protein along with the MS2 stem-loop RNAs (PDB ID: 1ZDK) (Grahn et al., 1999). Cross-linked peptides have been highlighted in purple whereas the cross-linked amino acids have been indicated in red colour. The RNA has been illustrated in orange colour with the probable nucleotide (Cytosine in this case) shown in blue at position -5 of RNA that can take part in cross-linking.

3.5 Identification of Uracil Fragments and Adducts

Cross-linking in combination with mass spectrometry (MS) has proven to be a powerful approach for the structural investigation of RNA-protein cross-links. It not only provides the information about the part of the biomolecule undergone cross-linking but also helps in the identification of specific site of cross-linking. The UV induced cross-linking utilizes the natural sensitivity of nucleobases and amino acid residues to generate a zero-length cross-link directly between the adjacent molecules without any intervening linker (Meisenheimer & Koch, 1997). The subsequent mass spectrometric analysis of the peptide-oligonucleotide heteroconjugates is used to determine the exact cross-linking site.

By taking the fact into consideration that the cross-links are additive in nature, the peptide-oligonucleotide cross-links are identified by database search (Kramer et al., 2014). The MS2 spectra of the cross-links are usually dominated by the peptide signals but one can also observe the characteristic marker ions produced by the CID fragmentation of the cross-linked nucleotides. In addition to these, the nucleotide fragments can be observed as adducts cross-linked to an amino acid residue resulting in the shifting of ion series.

Several studies have been conducted on the nucleobases and nucleosides along their isotopically labeled analogues and modified species by CID fragmentation. The fragmentation occurs by the opening of the base ring and the loss of low molecular weight neutral components. The common losses detected during fragmentation are of ammonia (NH_3), carbon monoxide (CO), cyanic acid NHCO, hydrogen cyanide (HCN) and water (H_2O) (Nelson & McCloskey, 1994; Liu et al., 2008). However, different nucleosides and their modified species show significant differences in their fragmentation.

During the cross-linking studies, conducted by using labeled/non-labeled PM5 pre-mRNA/MINX pre-mRNA complexed with *HeLa* nuclear extract proteins (as described in section 2.2.4.4 & 2.2.5.3 of materials and methods), the MS2 spectra obtained, has the sequence information of the cross-linked peptides along with the signals of fragment ions of the cross-linked nucleotides. Uracil

nucleotide is found to be the most frequently cross-linked nucleotide with the peptide moieties. Several uracil nucleotide fragments that have been reported before (Table 3.5 & 3.6) were also been observed in the MS2 spectra (Newton et al., 1986; Nelson & McCloskey, 1994; Liu et al., 2008). However, there have been few unidentified signals in the MS2 spectra. Since the peptide and oligonucleotides have different fractional mass depending upon their different molecular composition therefore on the basis of fractional mass these unidentified signals are assumed to be generated by the fragmentation of oligonucleotides (Pourshahian & Limbach, 2008) i.e. uracil nucleotide in the current case.

In order to estimate the elemental composition of the generated fragments, the ion signals from unlabeled and different isotopically labeled uracil nucleotide i.e. ^{13}C, $^{13}C^{15}N$ labeled and site specific deuterated uracil nucleotides (Figure 3.15) have been compared for the same peptide-oligonucleotide cross-link.

Table 3.5: List of Uracil nucleotide (fragment) adducts with their assigned abbreviations and calculated masses.

*The RNA (fragment) adducts and marker ions highlighted in yellow have been identified during the current studies.

RNA (fragment) adducts	Abbreviation in annotated spectra	Calculated mass(Da.)
C_3O	#	51.9949
Uracil-H_2O	U'^0	94.0167
Uracil	U'	112.0273
Uridin-H_2O	$U^{0\text{-}p}$	226.0590
Uridin	U^p	244.0695
Uridinmonophosphate-H_2O	U^0	306.0253
Uridinmonophosphate	U	324.0359
Uridin-H_2O-H_2O-CH_2O	$U^{0\text{-}p\text{-}0\text{-}CH_2O}$	178.0378
Uridin-H_2O-C_3O	$U^{0\text{-}p\text{-}\#}$	174.0640
Uridin-H_2O-H_2O	$U^{0\text{-}p\text{-}0}$	208.0484

Table 3.6: List of Uracil nucleotide marker ions with their symbols and calculated *m/z*.

RNA marker ions	Symbol	Calculated *m/z*
Uracil	U'	113.0351
Uridine-H_2O	$U^{0\text{-}p}$	227.0667
Uridinmonophosphate-H_2O	U^0	307.0331
Uridinmonophosphate	U	325.0437
Uridin-H_2O-H_2O-CH_2O	$U^{0\text{-}p\text{-}0\text{-}CH_2O}$	179.0456
Uridin-H_2O-C_3O	$U^{0\text{-}p\text{-}\#}$	175.0718

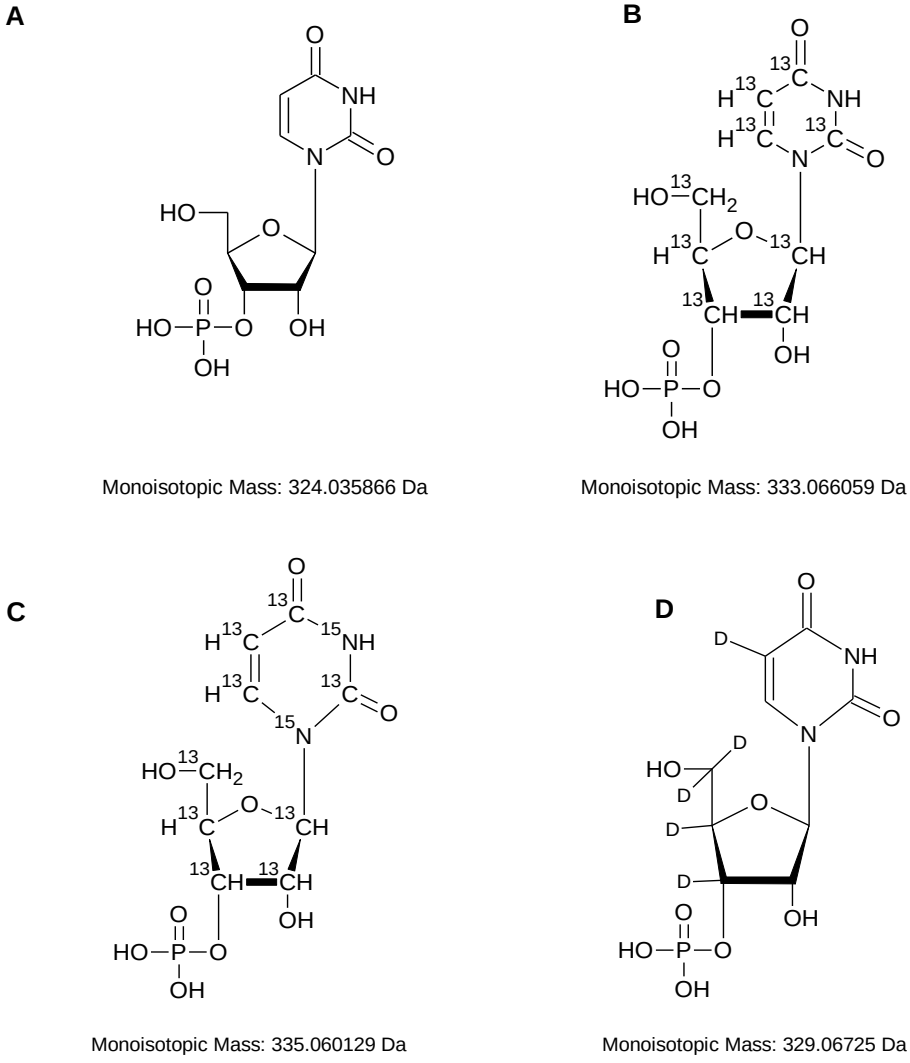


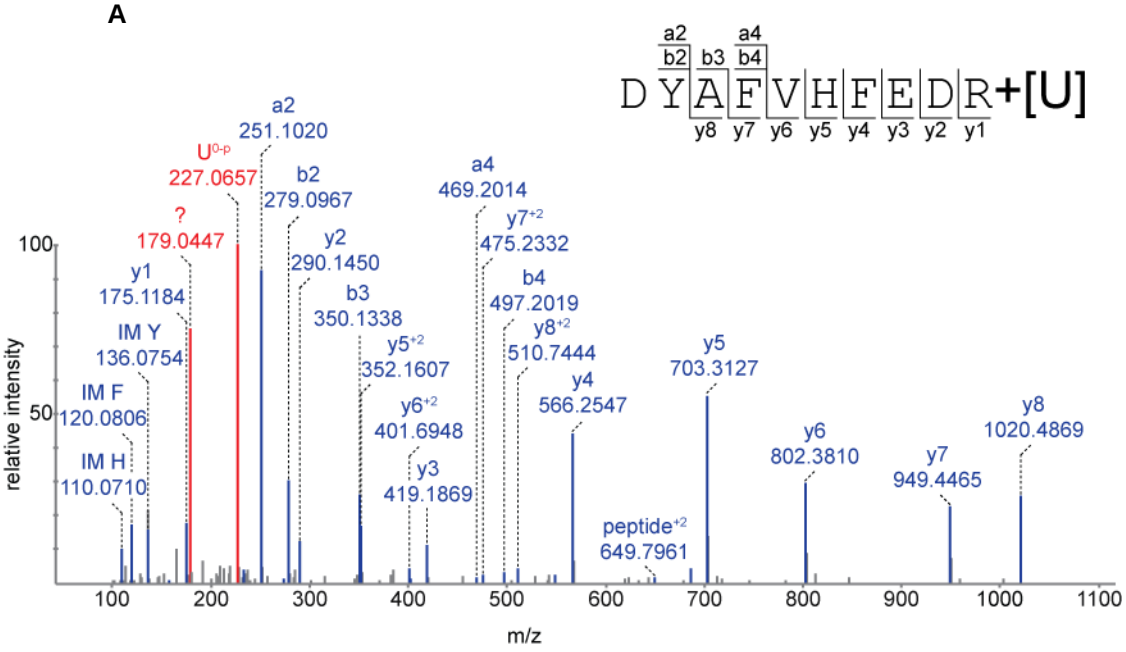
Figure 3.15: Uracil nucleotides. (A) Unlabeled uracil nucleotide. (B) $^{13}C_9$ labeled uracil nucleotide. (C) $^{13}C_9{}^{15}N_2$ labeled uracil nucleotide. (D) 5-D_1, ribose-3‘, 4‘, 5‘, 5‘-D_4 labeled uracil nucleotide.

The peptide ^{375}DYAFVHFEDR^{384} derived from heterogeneous nuclear ribonucleoprotein R protein was found to be cross-linked to uracil nucleotide. In MS2 spectrum, the ion signal at *m/z* 179.0447 (Figure 3.16 A) with relatively high intensity is assumed to be generated by the fragmentation of the cross-linked uracil nucleotide. This ion signals have been compared with its corresponding ion signals in other spectra of the same cross-linked peptide with the isotopically labeled uracil nucleotides ($^{13}C_9$, $^{13}C_9{}^{15}N_2$ and 5-D_1, ribose-3‘, 4‘, 5‘, 5‘-D_4 labeled uracil nucleotide). The difference in mass helps in predicting the elemental

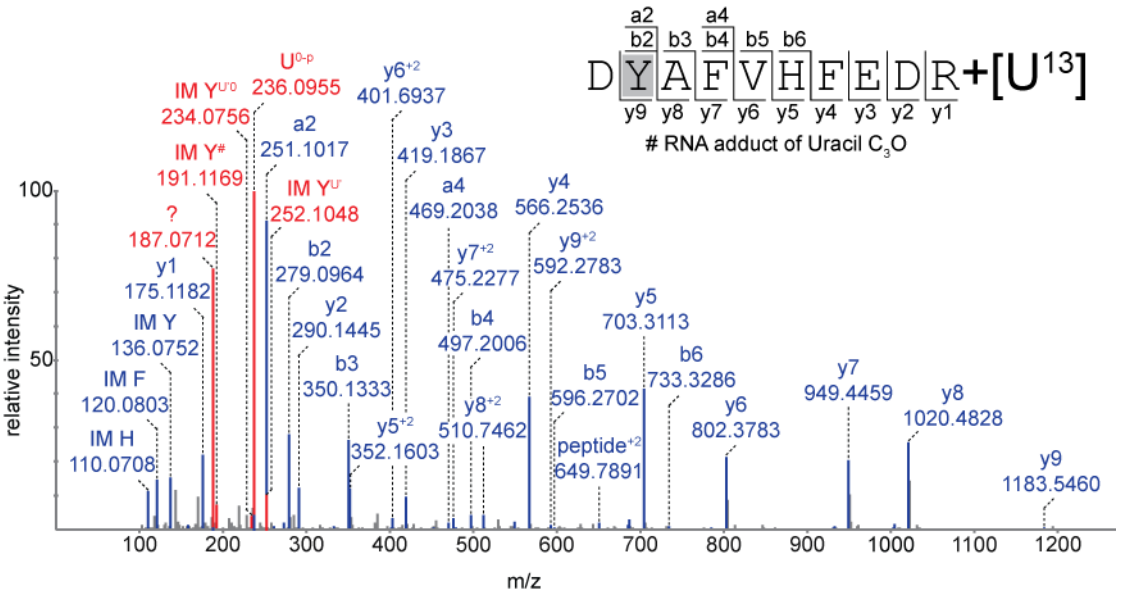
composition of the uracil nucleotide fragment. Since in ^{13}C-labeled uracil nucleotide all carbon atoms are labeled so the difference of the ion signal at *m/z* 187.0712 in spectrum B (Figure 3.16 B) with the corresponding ion signal at *m/z* 179.0447 in spectrum A (Figure 3.16 A) shows the presence of eight carbon atoms in the fragment. Similarly, in case of $^{13}C^{15}N$-labeled uracil nucleotide, the difference of ion signal at *m/z* 189.0654 in spectrum C (Figure 3.16 C) with the corresponding ion signal at *m/z* 187.0712 in spectrum B (Figure 3.16 B) depicts the presence of two nitrogen atoms in the fragment. When the site specific deuterated uracil nucleotide was used, the difference of ion signal at *m/z* 182.0636 in spectrum D (Figure 3.16 D) with its corresponding ion signal at *m/z* 179.0447 in spectrum A (Figure 3.16 A) indicates the loss of two deuterium atoms and the presence remaining 3 deuterium atoms in the uracil nucleotide fragment of interest.

Upon CID fragmentation of nucleotides, the loss of phosphate group and neutral loss of water are commonly observed. So by keeping all these points into consideration the elemental composition of the uracil nucleotide fragment of interest can be assumed i.e. $C_8H_6N_2O_3$ generated via combined loss of two water (H_2O) molecules, one phosphate group (HPO_3) and formaldehyde (CH_2O) molecule. This result is in concordance with the previous studies conducted on the pseudouridine fragmentation in which the loss of two water molecules and CH_2O from the sugar moiety give rise to a product ion at *m/z* 179 (Dudley et al., 2000). The structure of uracil nucleotide fragment of interest can be predicted as a furanyl derivative fitting to the exact monoisotopic mass of 178.0378 Da (Figure 3.17).

A


a2 b2 b3 a4 b4
D Y A F V H F E D R+[U]
y8 y7 y6 y5 y4 y3 y2 y1
a2 251.1020
U0-p 227.0657
b2 279.0967
a4 469.2014
y7+2 475.2332
? 179.0447
y2 290.1450
y1 175.1184
b4 497.2019
b3 350.1338
IM Y 136.0754
y8+2 510.7444
y5+2 352.1607
y5 703.3127
IM F 120.0806
y4 566.2547
y6+2 401.6948
y8 1020.4869
IM H 110.0710
y6 802.3810
y3 419.1869
y7 949.4465
peptide+2 649.7961
relative intensity
m/z

B


a2 b2 b3 a4 b4 b5 b6
D Y A F V H F E D R+[U13]
y9 y8 y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
U0-p 236.0955
y6+2 401.6937
IM Y U'0 234.0756
a2 251.1017
y3 419.1867
IM Y# 191.1169
IM Y U- 252.1048
a4 469.2038
y4 566.2536
? 187.0712
y9+2 592.2783
b2 279.0964
y7+2 475.2277
y1 175.1182
y5 703.3113
y2 290.1445
b4 497.2006
IM Y 136.0752
b6 733.3286
b3 350.1333
b5 596.2702
y7 949.4459
IM F 120.0803
y8+2 510.7462
y6 802.3783
y8 1020.4828
IM H 110.0708
y5+2 352.1603
peptide+2 649.7891
y9 1183.5460
relative intensity
m/z

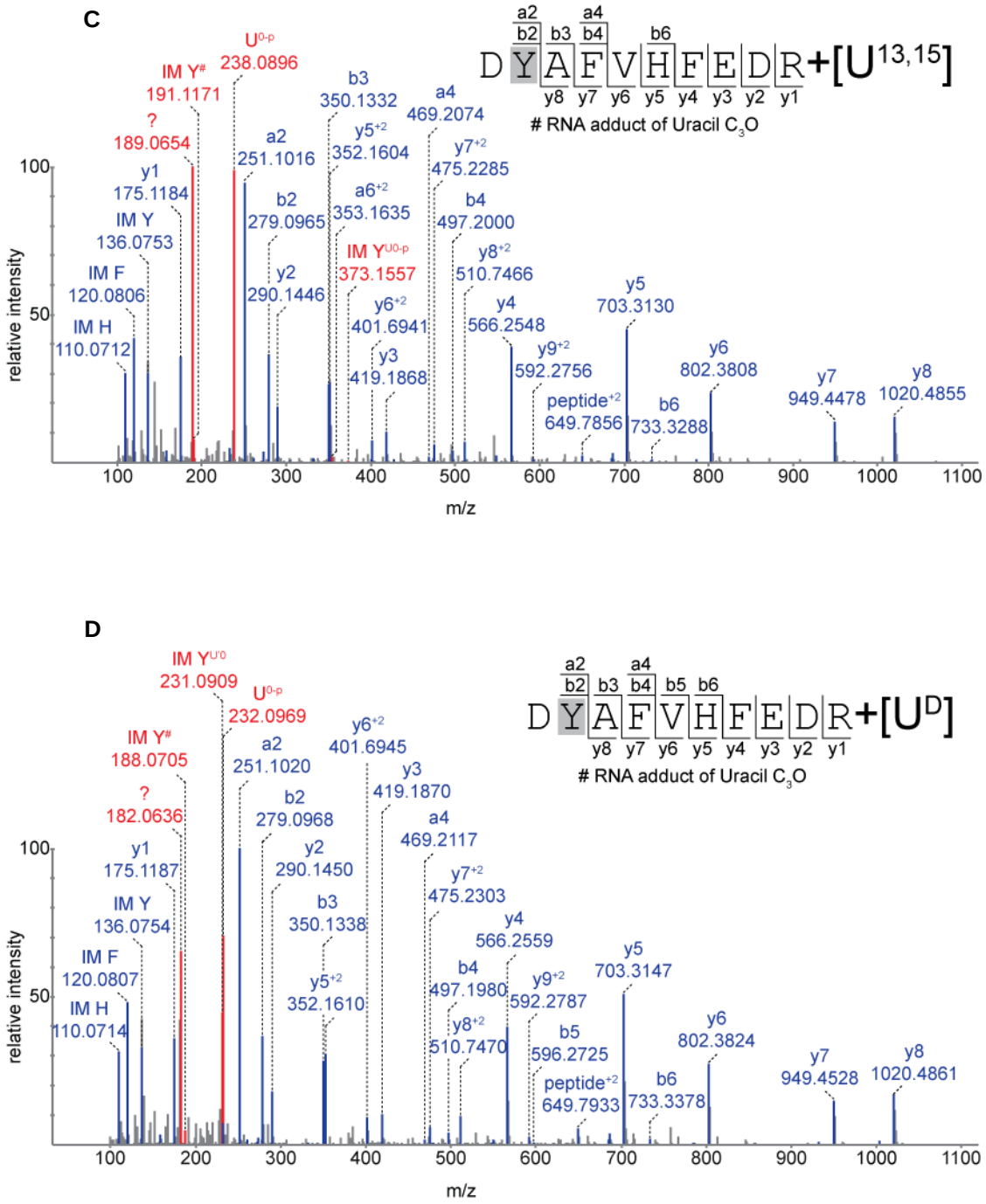
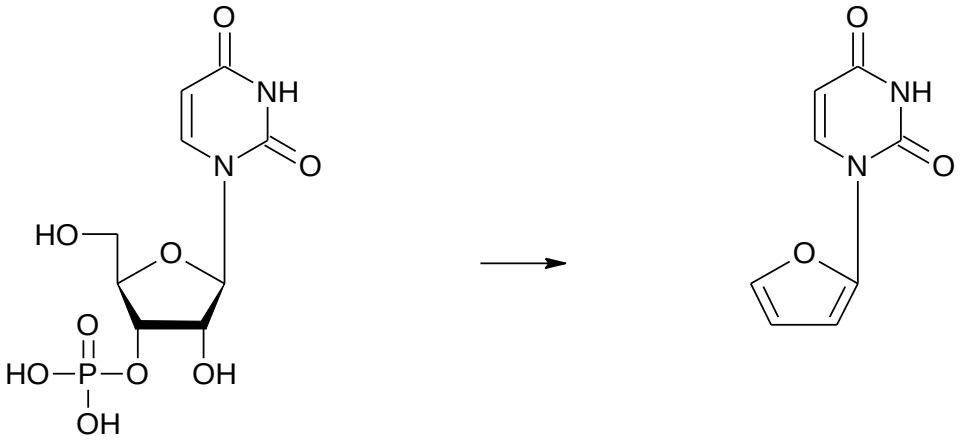


Figure 3.16: The MS2 spectra of the peptide ^{375}DYAFVHFEDR^{384} cross-linked to U. (A) The MS2 spectrum of before mentioned cross-link with unlabeled uracil nucleotide. The spectrum has ion signals at *m/z* 179.0447 of uracil nucleotide fragment of interest and *m/z* 227.0657 (U-H_3PO_4) showing that the unlabeled U is cross-linked with the peptide. (B) The MS2 spectrum of before mentioned cross-link with $^{13}C_9$ labeled uracil nucleotide. The spectrum has ion signals at *m/z* 187.0712 of uracil nucleotide fragment of interest, *m/z* 236.0955 (U-H_3PO_4) and of tyrosine immonium ion with the U adducts (#: C_3O, U^0: U-H_2O, U': Uracil marker ion) showing that the ^{13}C-labeled U is cross-linked to Y. (C) The MS2 spectrum of before mentioned cross-link with $^{13}C_9{}^{15}N_2$ labeled uracil nucleotide. The spectrum has ion signals at *m/z* 189.0654 of uracil nucleotide fragment of interest, *m/z* 238.0896 (U-H_3PO_4) and of tyrosine immonium ion with the U adducts

(#: C_3O, $U^{0\text{-}p}$: U-H_3PO_4) showing that the $^{13}C^{15}N$-labeled U is cross-linked to Y. (D) The MS2 spectrum of before mentioned cross-link with 5-D_1, ribose-3‘, 4‘, 5‘, 5‘-D_4labeled uracil nucleotide. The spectrum has ion signals at *m/z* 182.0636 of uracil nucleotide fragment of interest, *m/z* 232.0969 (U-H_3PO_4) and of tyrosine immonium ion with the U adducts (#: C_3O, U^0: U-H_2O) showing that the D-labeled U is cross-linked to Y.

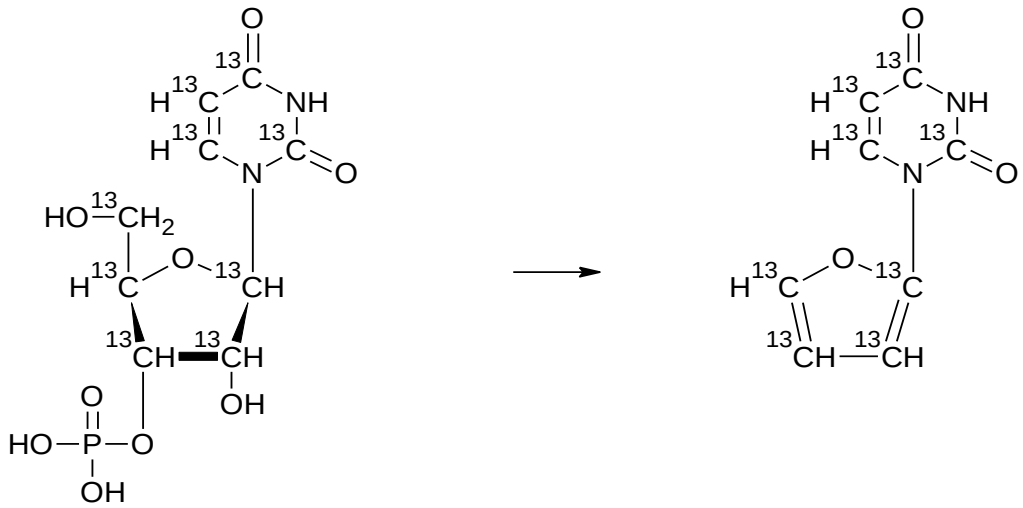
A



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 324.035866 Da

Molecular Formula: $C_8H_6N_2O_3$
Monoisotopic Mass: 178.037842 Da

B



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 333.066059 Da

Molecular Formula: $C_8H_6N_2O_3$
Monoisotopic Mass: 186.064681 Da

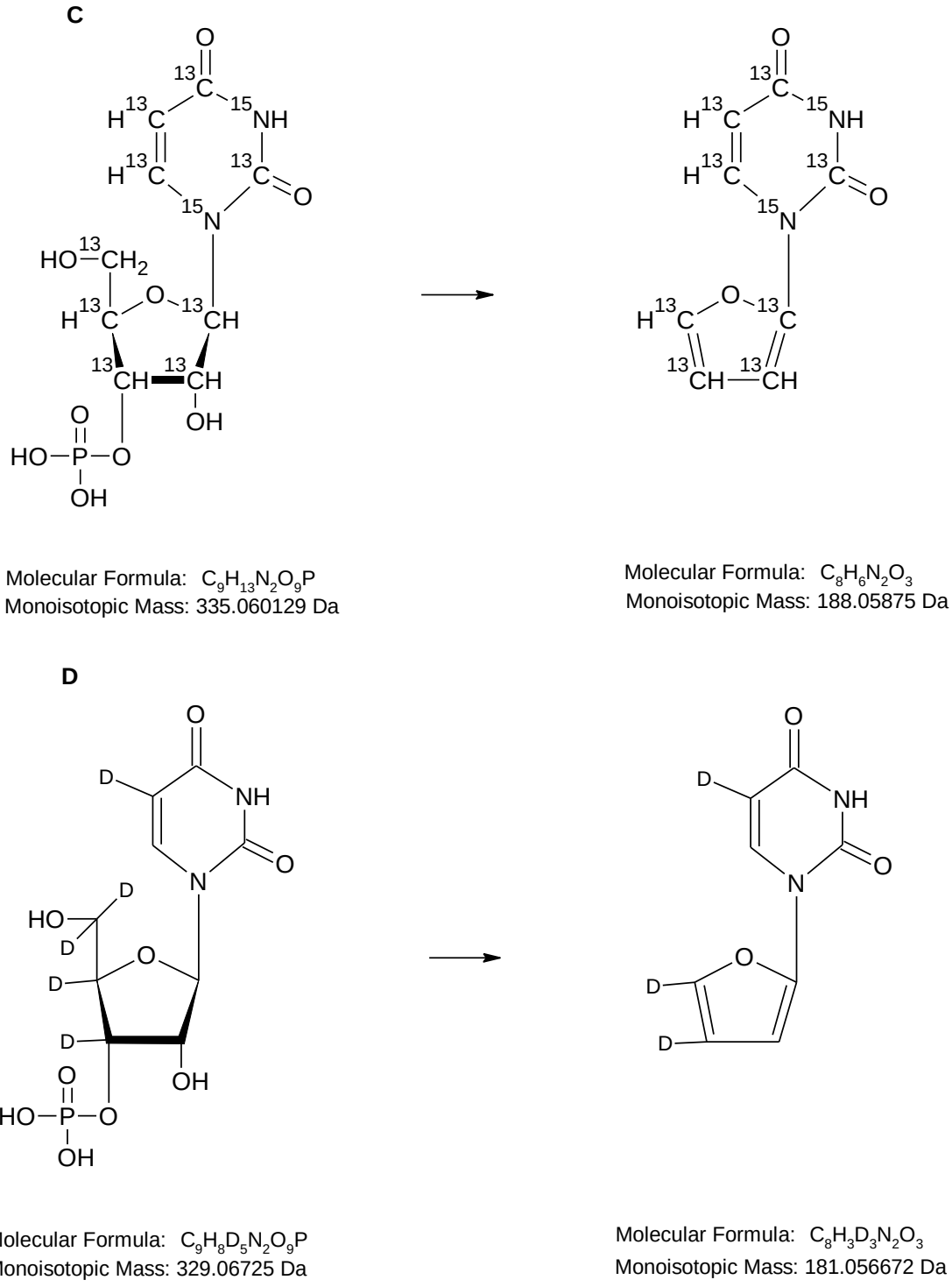
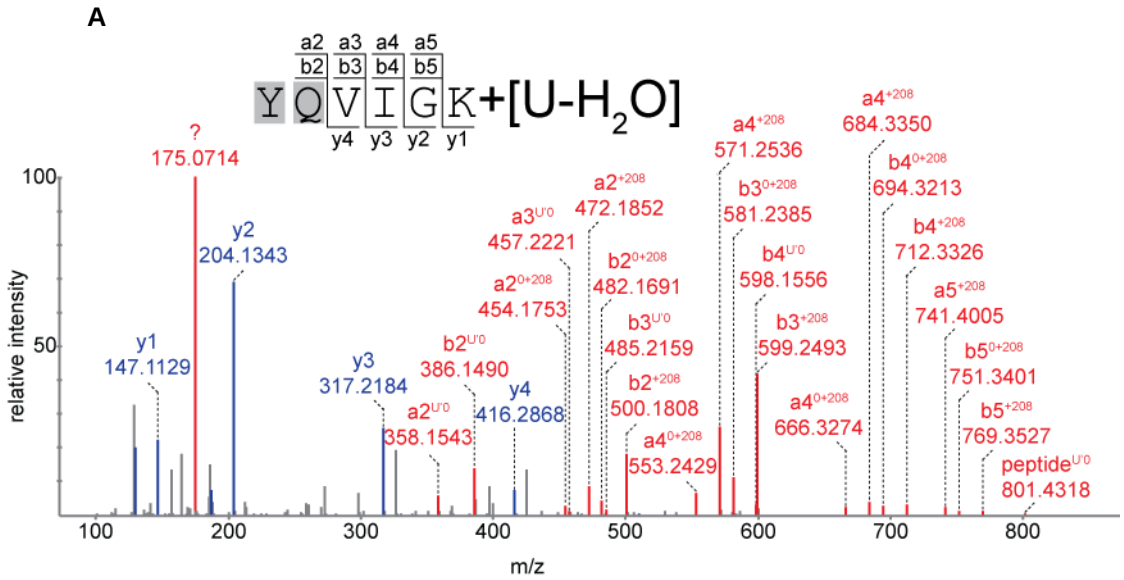
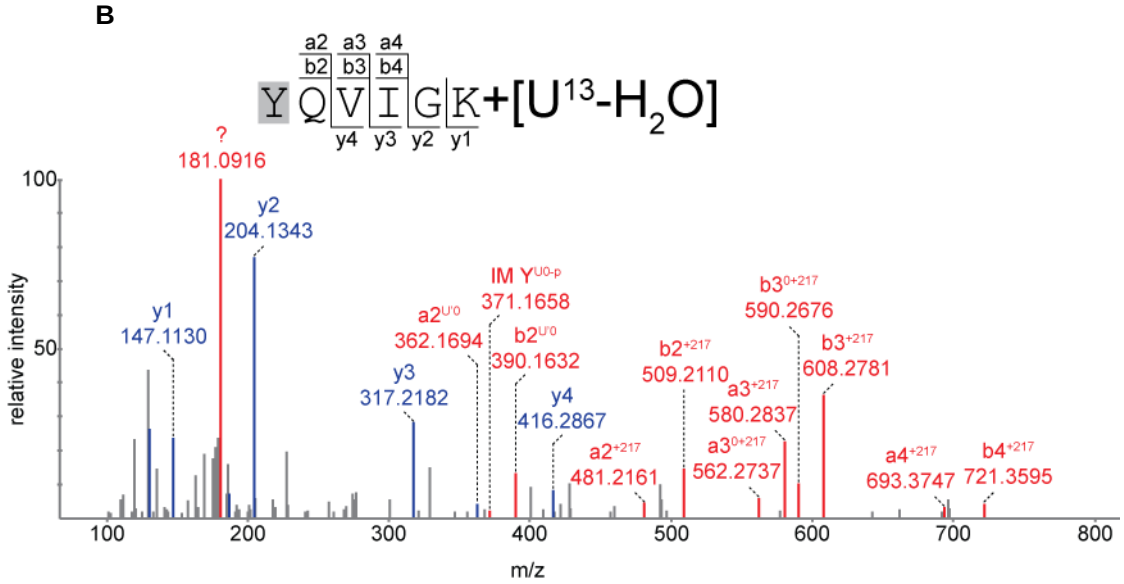


Figure 3.17: The predicted structural formulae of the Uracil nucleotide fragment (*m/z* 179.0447). (A) The predicted structural formula of the unlabeled uracil nucleotide fragment generated by the loss of two H_2O, one HPO_3 and one CH_2O molecules. (B) The predicted structural formula of the ^{13}C-labeled uracil nucleotide fragment generated by the loss of two H_2O, one HPO_3 and one CH_2O molecules. (C) The predicted structural formula of the $^{13}C^{15}N$-labeled uracil nucleotide fragment generated by the loss of two H_2O, one HPO_3 and one CH_2O molecules. (D) The predicted structural formula of the D-labeled uracil nucleotide fragment generated by the loss of two H_2O, one HPO_3 and one CH_2O molecules.

Similarly, the peptide ^{215}YQVIGK^{220} of mitochondrial endonuclease G is found to be cross-linked to uracil nucleotide with the loss of H_2O molecule. Its MS2 spectrum shows the ion signal with relatively high intensity at *m/z* 175.0714 (Figure 4 A), which is expected to be produced as a result of fragmentation of cross-linked uracil nucleotide. In order to deduce the elemental composition of this uracil nucleotide fragment of interest, the ion signal at *m/z* 175.0714 (Figure 3.18 A) is compared with its corresponding ion signals in other spectra of the same cross-linked peptide with the isotopically labeled uracil nucleotides. The MS2 spectrum B (Figure 3.18 B) of cross-linked peptide ^{15}YQVIGK^{220} with ^{13}C-labeled uracil nucleotide with the loss of H_2O shows the ion signal of the uracil nucleotide fragment of interest at *m/z* 181.0916 which is different from the corresponding ion signal in spectrum A by six carbon atoms. Likewise, ion signal at *m/z* 183.0851 in the MS2 spectrum C (Figure 3.18 C) of cross-linked peptide ^{15}YQVIGK^{220} with $^{13}C^{15}N$-labeled uracil nucleotide with the loss of H_2O in comparison with the ion signal at *m/z* 181.0916 in spectrum B (Figure 3.18 B) indicates the presence of two nitrogen atoms. Finally, the elemental composition of the uracil nucleotide fragment of interest, deduced by the above mentioned observations is $C_6H_{10}N_2O_4$, generated by the loss of one HPO_3 and C_3O molecules. On the basis of this molecular formula, the probable structure of uracil nucleotide fragment of interest can be predicted, fitting to the exact monoisotopic mass of 174.0640 Da (Figure 3.19).


A
a2 a3 a4 a5
b2 b3 b4 b5
YQVIGK+[U-H2O]
y4 y3 y2 y1
relative intensity
m/z
? 175.0714
y2 204.1343
y1 147.1129
y3 317.2184
b2U'0 386.1490
a2U'0 358.1543
y4 416.2868
a3U'0 457.2221
a2+208 472.1852
a2O+208 454.1753
b2O+208 482.1691
b3U'0 485.2159
b2+208 500.1808
a4O+208 553.2429
a4+208 571.2536
b3O+208 581.2385
b4U'0 598.1556
b3+208 599.2493
a4O+208 666.3274
a4+208 684.3350
b4O+208 694.3213
b4+208 712.3326
a5+208 741.4005
b5O+208 751.3401
b5+208 769.3527
peptideU'0 801.4318


B
a2 a3 a4
b2 b3 b4
YQVIGK+[U13-H2O]
y4 y3 y2 y1
relative intensity
m/z
? 181.0916
y2 204.1343
y1 147.1130
IM YU0-p 371.1658
a2U'0 362.1694
b2U'0 390.1632
y3 317.2182
y4 416.2867
a2+217 481.2161
b2+217 509.2110
a3+217 580.2837
a3O+217 562.2737
b3O+217 590.2676
b3+217 608.2781
a4+217 693.3747
b4+217 721.3595

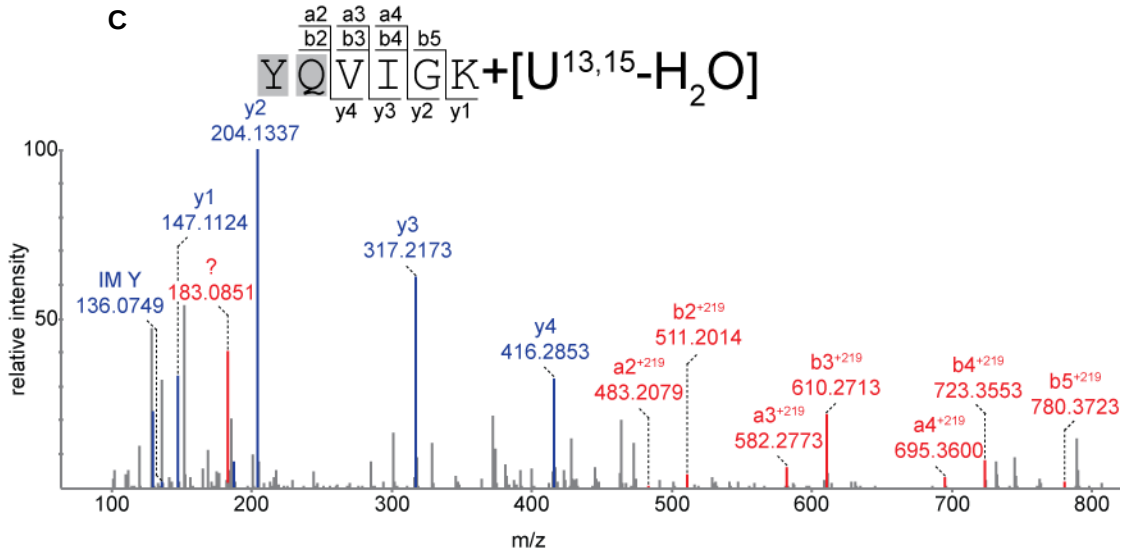
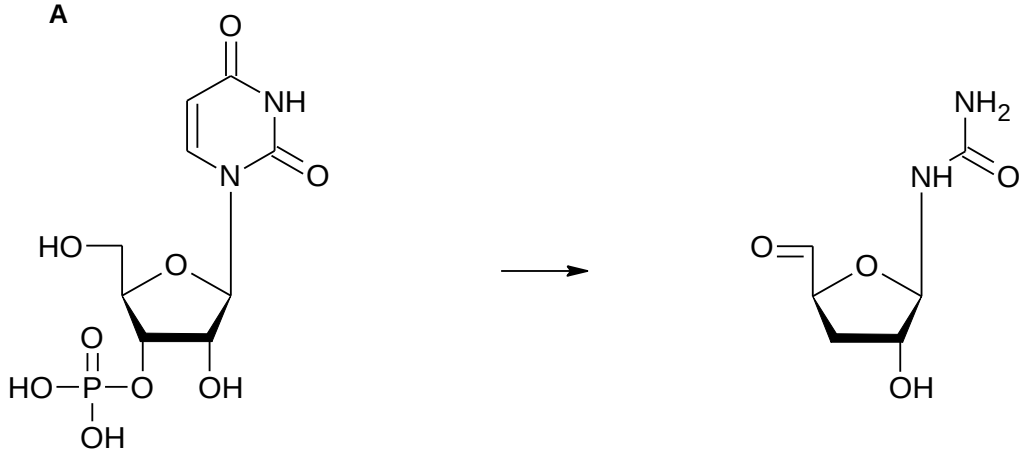


Figure 3.18: The MS2 spectra of the peptide ^{215}YQVIGK^{220} cross-linked to U-H_2O. (A) The MS2 spectrum of the mentioned cross-link with unlabeled uracil nucleotide with the loss of water. The spectrum has ion signal at *m/z* 175.0714 of uracil nucleotide fragment of interest and a shift in b-series from b2 ion by U adduct (U'^{0}: Uracil marker ion-H_2O and predicted uracil nucleotide adduct of 208 Da) showing that the unlabeled U is cross-linked with the Y/Q of a peptide. (B) The MS2 spectrum of before mentioned cross-link with $^{13}C_9$ labeled uracil nucleotide. The spectrum has ion signals at *m/z* 181.0916 of uracil nucleotide fragment of interest and of tyrosine immonium ion with the U adduct (U$^{0\text{-}p}$: U-H_3PO_4) and a shift of b-series from b2 ion by U adduct (U'^{0}: Uracil marker ion-H_2O and predicted uracil nucleotide adduct of 217 Da) showing that the ^{13}C-labeled U is cross-linked to Y. (C) The MS2 spectrum of before mentioned cross-link with $^{13}C_9{}^{15}N_2$ labeled uracil nucleotide. The spectrum has ion signals at *m/z* 183.0851 of uracil nucleotide fragment of interest and a shift of b-series from b2 ion by U adduct (predicted uracil nucleotide adduct of 219 Da) showing that the $^{13}C^{15}N$-labeled U is cross-linked to Y/Q of a peptide

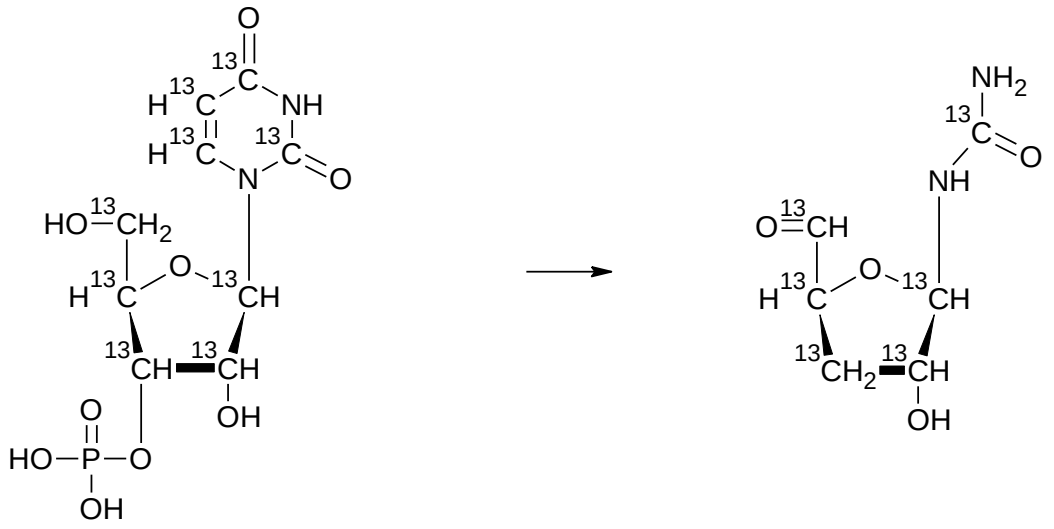
A



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 324.035866 Da

Molecular Formula: $C_6H_{10}N_2O_4$
Monoisotopic Mass: 174.064057 Da

B



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 333.066059 Da

Molecular Formula: $C_6H_{10}N_2O_4$
Monoisotopic Mass: 180.084186 Da

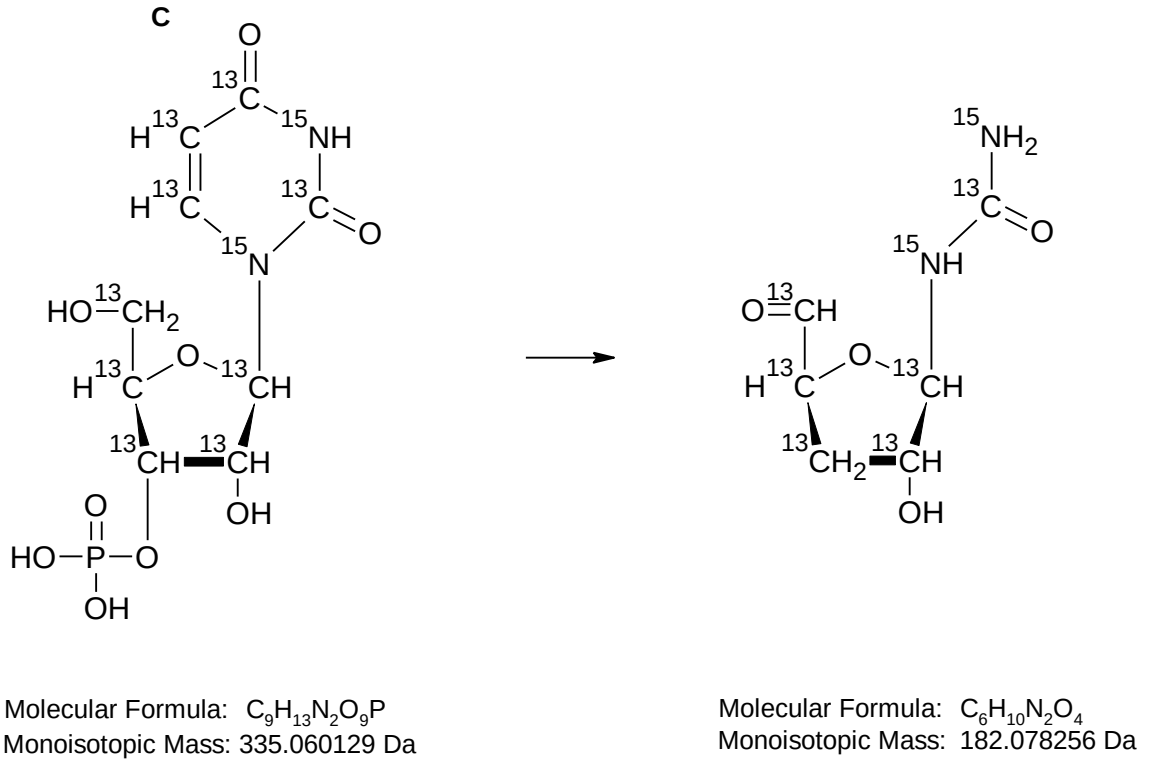
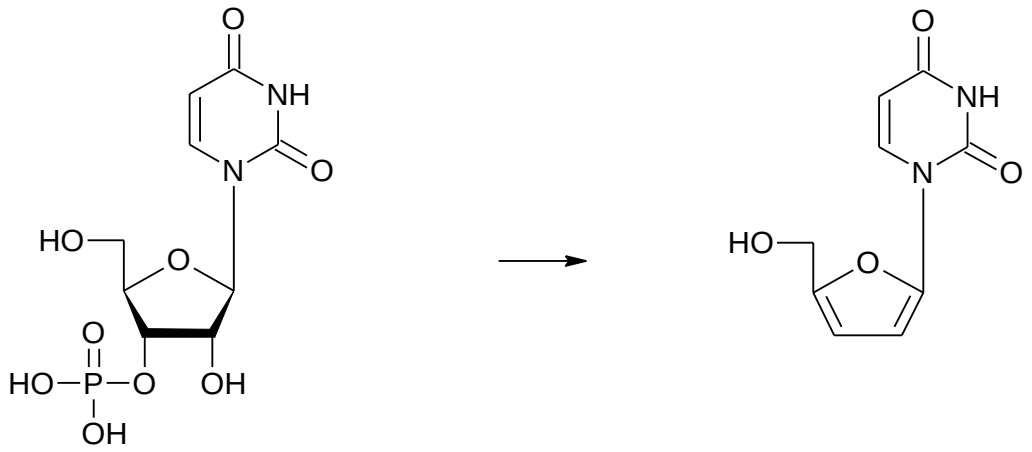


Figure 3.19: The predicted structural formulae of the Uracil nucleotide fragment (*m/z* 175.0714). (A) The predicted structural formula of the unlabeled uracil nucleotide fragment generated by the loss of one H_2O, one HPO_3 and one C_3O molecules. (B) The predicted structural formula of the ^{13}C-labeled uracil nucleotide fragment generated by the loss of one H_2O, one HPO_3 and one C_3O molecules. (C) The predicted structural formula of the $^{13}C^{15}N$-labeled uracil nucleotide fragment generated by the loss of one H_2O, one HPO_3 and one C_3O molecules.

The MS/MS spectrum of the cross-linked peptide ^{215}YQVIGK^{220} with U-H_2O shows a shift in b ion series with the mass of 208 Da, 217 Da and 219 Da when the peptide is cross-linked with unlabeled, ^{13}C-labeled and $^{13}C^{15}N$-labeled uracil nucleotides with the loss of water respectively. The difference in the mass shift of b ion series with the corresponding ion series in other spectra of the same cross-linked peptide with the isotopically labeled uracil nucleotides (Figure 3.18) helps in predicting the elemental composition of the expected uracil nucleotide adduct. By comparing the mass shifts, it has been concluded that the elemental composition of the predicted uracil nucleotide fragment observed as an adduct in b ion series is $C_9H_8N_2O_4$. This fragment is assumed to be generated by the loss of two H_2O and one HPO_3 molecules from the intact uracil nucleotide. Based on the molecular formula, the probable structure of the uracil nucleotide fragment

observed as an adduct can be predicted which fits to its monoisotopic mass of 208.0484 Da (Figure 3.20). Similar fragment of 208 Da was also observed by the fragmentation of pseudouridine in the studies conducted by Rice & Dudek, 1969.

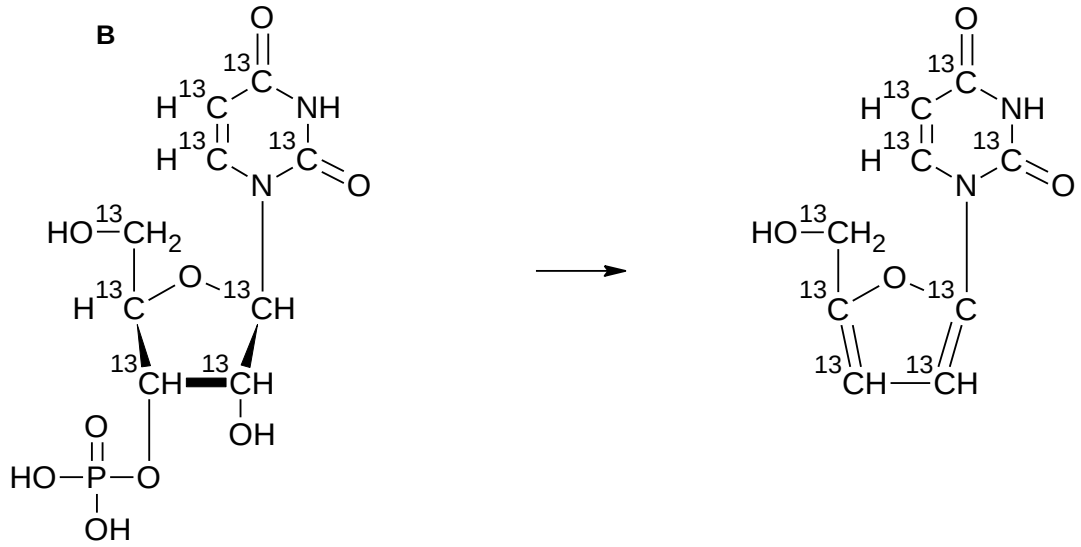
A



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 324.035866 Da

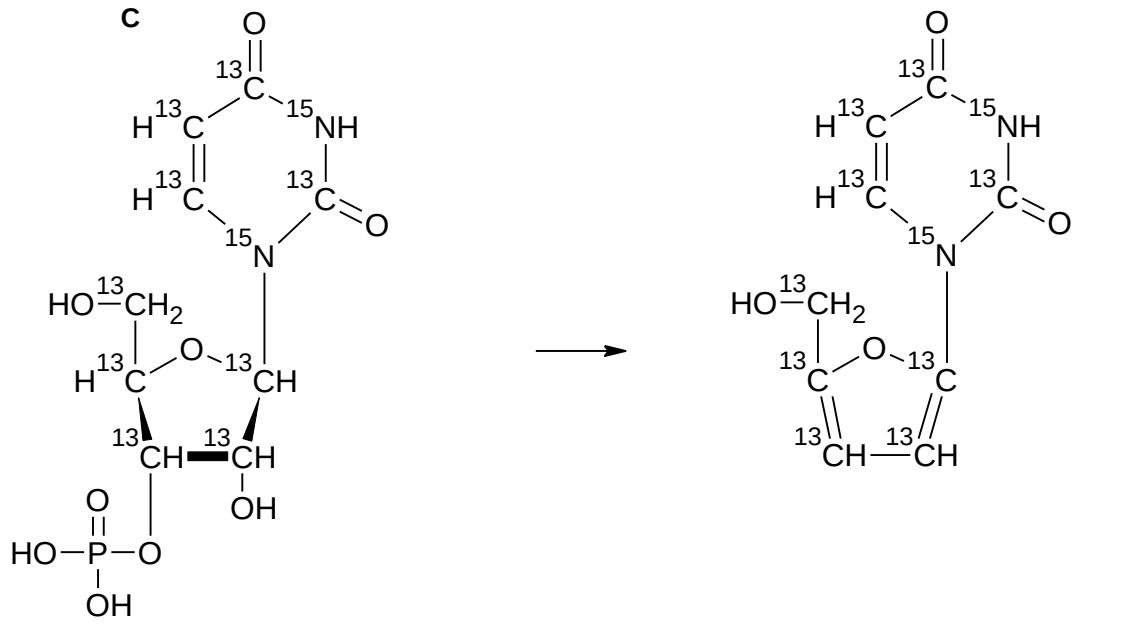
Molecular Formula: $C_9H_8N_2O_4$
Monoisotopic Mass: 208.048407 Da

B



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 333.066059 Da

Molecular Formula: $C_9H_8N_2O_4$
Monoisotopic Mass: 217.0786 Da



Molecular Formula: $C_9H_{13}N_2O_9P$
Monoisotopic Mass: 335.060129 Da

Molecular Formula: $C_9H_8N_2O_4$
Monoisotopic Mass: 219.07267 Da

Figure 3.20: The predicted structural formulae of the Uracil nucleotide fragment observed as a U adduct in Figure 3.18 resulting in the shift of b ion series by the mass of 208 Da. (A) The predicted structural formula of the unlabeled uracil nucleotide fragment generated by the loss of two H_2O and one HPO_3 molecules. (B) The predicted structural formula of the ^{13}C-labeled uracil nucleotide fragment generated by the loss of two H_2O and one HPO_3 molecules. (C) The predicted structural formula of the $^{13}C^{15}N$-labeled uracil nucleotide fragment generated by the loss of two H_2O and one HPO_3 molecules.

3.6 Identification of RNP Complexes Isolated from *HeLa* Nuclear Extract and their Cross-linking Analysis

The life cycle of mRNA is predominantly regulated by the interaction of RNA binding protein with the RNAs (Keene, 2007; Glisovic et al., 2008). The complexity of this regulation has increased with that of the organisms. In the mammalian cells, more than 1300 RNA binding proteins interact with miscellaneous RNA with varying specificity and affinity through their diverse range of RNA binding domains (RBDs) and motifs (Castello et al., 2012; Munschauer, 2015). These include polypyrimidine tract-binding protein 1 (PTBP1), Y-box-binding protein 1 (YBX1) and serine-rich (SR) proteins etc. Till now more than 50 RBDs have been reported (Anantharaman et al., 2002; Glisovic et al., 2008; Munschauer, 2015; Castello et al., 2016). The RBPs interact with RNAs by using one or multiple RBDs. The most widely studied RBDs are RNA recognition motif (RRM), K homology (KH) domain, zink finger (ZnF) domain and cold shock domain (CSD) etc. For the better understanding of these interactions, high-throughput *in vitro* (Van Nostrand et al., 2016) and *in vivo* (Ray et al., 2013) techniques have been established. These techniques can be either protein-centric (Ingolia et al., 2009; Darnell, 2010) or RNA-centric (Hartmuth et al., 2002; Castello et al., 2012) followed by mass spectrometry or RNA sequencing respectively. However, all these methods involve few technical challenges for example the isolation of non-specific RNA-protein interactions, the low abundance of target biomolecules and low yield of RNP complexes etc.

To overcome the limitations occurring during interaction analysis of the large RNP complexes, the optimization was carried out in the existing protocols described by Deckert et al., 2006 for isolation of RNP complex and by Urlaub et al., 2002 and Luo et al., 2008 for purification and enrichment of cross-links. For the current studies, RNA-centric approach has been adopted. The (PM5/MINX) pre-mRNA tagged with three MS2-binding RNA stem-loops was transcribed by using non-labeled/labeled UTP (Figure 3.21). To assemble the RNP complex the tagged pre-mRNA was pre-incubated with MS2-MBP fusion protein and later with the *HeLa* nuclear extract on ice to prevent any protein degradation and RNA

processing. After several washes to remove the non-specific interacting proteins, the RNP complex was affinity purified by using amylose beads and maltose in elution buffer without glycerol (as described in detail in materials and method section 2.2.4.4) (Jurica et al., 2002; Deckert et al., 2006; Bessonov et al., 2008).

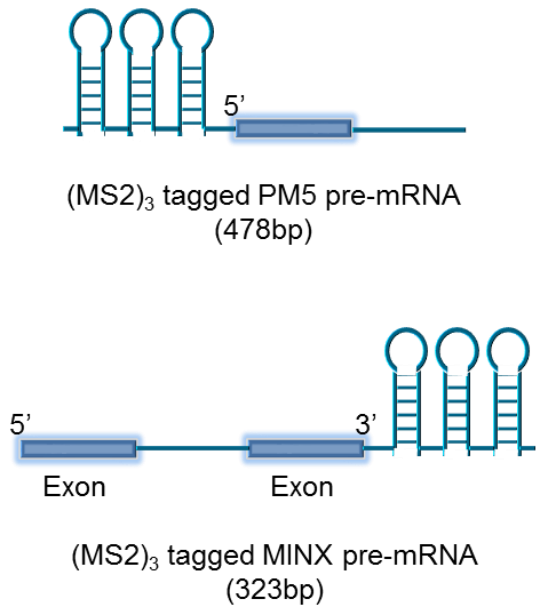


Figure 3.21: The structure of MS2-tagged (PM5 & MINX) pre-mRNA. The PM5 pre-mRNA comprised of 478 bp and has single exon whereas MINX pre-mRNA has 323bp and two exons (Deckert et al., 2006; Bessonov et al., 2008).

The purified RNP complex was UV-irradiated at 254 nm to generate a zero length cross-linking between the nucleotide of RNA molecule and the amino acid of a protein lying in close proximity to each other. In parallel, the non-UV-irradiated sample was also processed as control. In order to perform the interactome analysis, the proteins of the RNP complex were separated by using pre-cast 4-12% Bis-Tris Gel. The gel was subsequently stained with coomassie blue followed by cutting out 22 slices from each lane (as mentioned in section 2.2.7.1 of materials and methods) (Figure 3.22). The proteins were hydrolyzed within the gel by using trypsin and the peptides were extracted to perform the mass spectrometric analysis for the identification of the proteins (as described in materials and methods section 2.2.7). The results obtained after data searches

were used to construct an interactome by using Cytoscape (version 3.7.0) and NetworkAnalyser plug-in (described in section 2.2.8.5 of material and methods) (Appendix Figure 6.62 & 6.63).

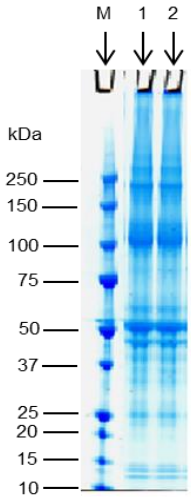


Figure 3.22: Coomassie stained 4-12% Bis-Tris Gel. The proteins of the UV-cross-linked RNP complex were separated on the gel and the coomassie stained. Each lane was then cut into 23 slices. M: Protein marker; Lane1: UV-cross-linked RNP complex assembled on PM5 pre-mRNA; Lane 2: UV-cross-linked RNP complex assembled on MINX pre-mRNA.

Protein-Protein interaction network was also constructed for the identified cross-linked proteins (described in section 2.2.8.5 of materials and methods) in order to have a clear overview of the isolated RNP complex and to find out the structural as well as functional relationships between the proteins. Network construction was based on known interactions deduced on experimental inference, literature survey and databases. Every protein was presented as an individual node. Edges represented the interactions between the proteins. The interactome analysis of the cross-linked proteins showed that the cross-linked RNP complex was mainly H/E complex that under suitable conditions can lead to spliceosomal A, B, and C complexes (Wahl et al., 2008).

Network analysis for RNP complexes assembled on two pre-mRNAs MINX and PM5 were carried out separately. Interactome for RNP complex assembled by using MINX pre-mRNA revealed 52 nodes and 227 edges with good connectedness (3 weakly connected components and 23 strongly connected components) (Figure 3.24). Average degree (number of edges connected to

nodes) was 8.73. The network diameter was 9, whereas average path length was found to be 2.74. Among other parameters the graph density was calculated to be 0.086. Connected component attribute revealed that there were 23 strongly connected components. The Average clustering coefficient was 0.288.

With 44 nodes and 163 edges, the attribute values for interaction revealed by RNP proteins assembled on PM5 pre-mRNA network analysis also showed good connectedness (Figure 3.23). Average degree was calculated to be 7.244 with network diameter of 7. Graph density was calculated to be 0.082. Connected components overview showed 26 strongly connected components. Average clustering coefficient was found to be 0.197.

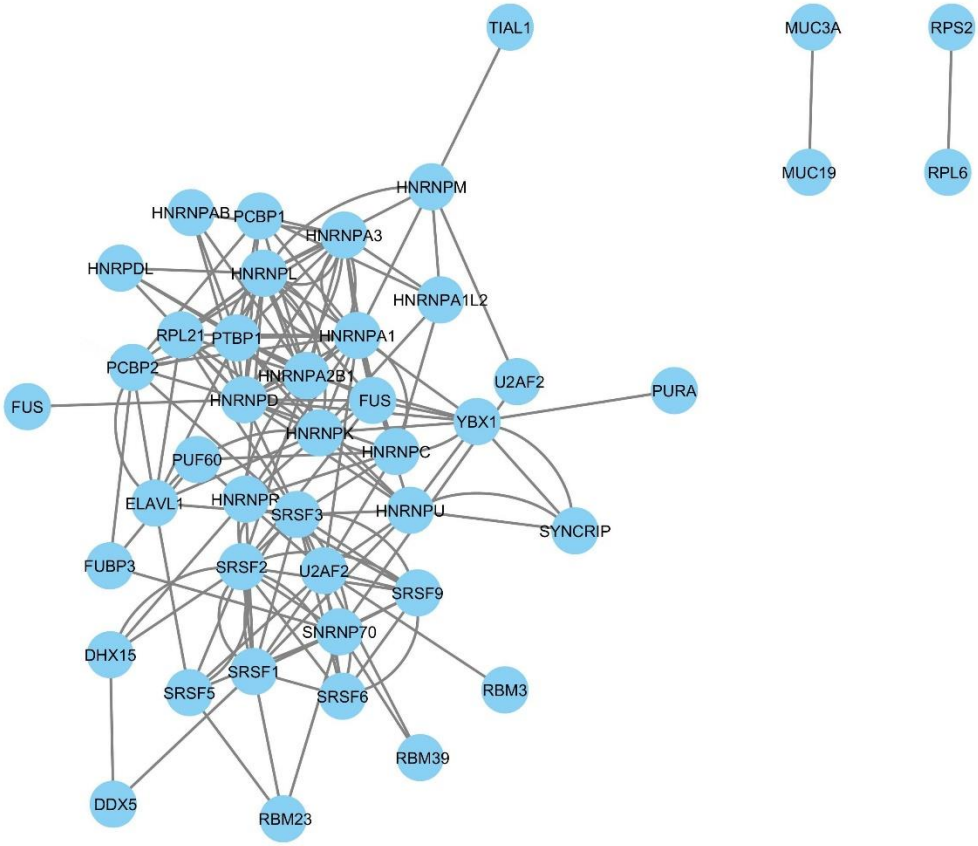


Figure 3.23: Interactome of RNP complex (after cross-linking, purification and enrichment) from *HeLa* nuclear extract assembled on PM5 pre-mRNA. Interactome network was constructed based on the information from the experiments. The structural and functional relationships are shown as edges that are connecting individual proteins (nodes). Nodes were labeled according to the gene names as reported by STRING database (Jensen et al., 2009).

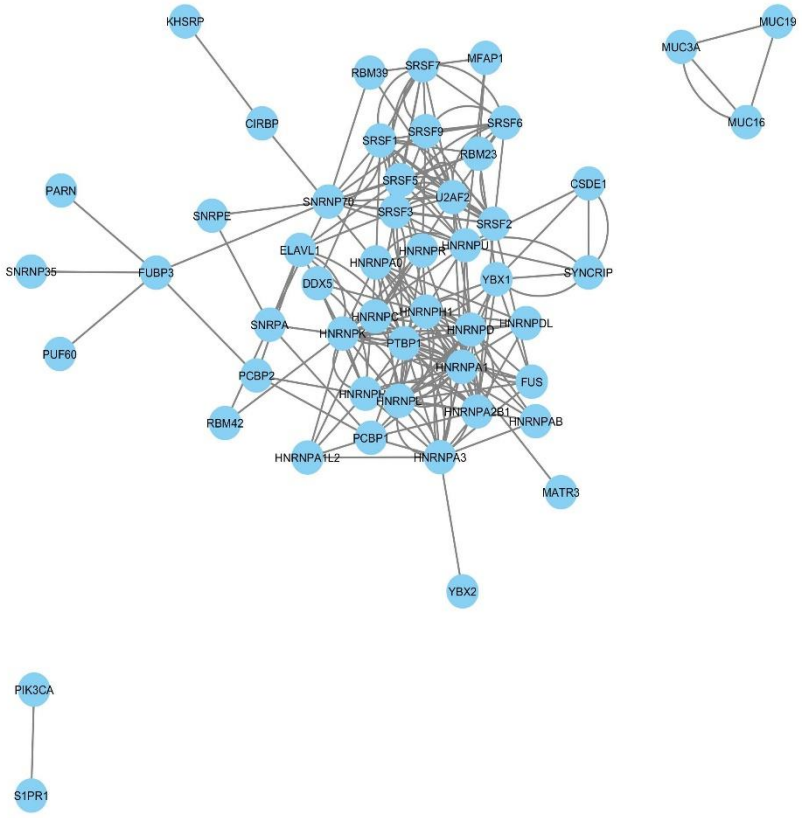


Figure 3.24: Interactome of RNP complex (after cross-linking, purification and enrichment) from *HeLa* nuclear extract assembled on MINX pre-mRNA. Interactome network was constructed based on the information from the experiments. The structural and functional relationships are shown as edges that are connecting individual proteins (nodes). Nodes were labeled according to the gene names as reported by STRING database (Jensen et al., 2009).

After ethanol precipitation of the UV-cross-linked RNP complex, the pellets were dissolved in 1% (w/v) SDS in size exclusion buffer (as mentioned in section 2.2.5.3 of materials and methods). Such high percentage of SDS, inhibit the activity of the enzymes which were used later in the protocol. So different percentages of SDS by w/v (1%, 0.1% and 0.05%) were tested for the efficient activity of trypsin. The *HeLa* nuclear extract proteins were ethanol precipitated and the pellets were diluted in the above mentioned SDS concentrations in SE running buffer and later digested with trypsin. The peptides were identified by mass spectrometry. The increasing number of identified peptides was found to be in the order of 0.05%>0.1%>1% (w/v) SDS. Keeping the dilution volume under consideration, the pulled-down H/E complex pellet dissolved in 1% (w/v) SDS in size exclusion buffer was diluted to 0.1% (w/v) SDS with SE running buffer. The

proteins were digested with trypsin and the samples were again ethanol precipitated. The pellets were re-dissolved in 1% (w/v) SDS and diluted up to 0.1% (w/v) SDS with size exclusion (SE) running buffer. The intact non-cross-linked pre-mRNA and the pre-mRNA with cross-linked peptides were separated from the non-cross-linked peptides by size exclusion (SE) chromatography (as mentioned in section 2.2.5.3 of materials and methods). The elution profile of the size exclusion chromatography of both control a UV-irradiated samples did not display any noticeable difference due to the low cross-linking yield and relatively low resolution of the column. The fractions that contain RNA were pooled together and then administered to ethanol precipitation. The fractions were also analyzed by running them on the Bis-Tris gel followed by silver staining (as mentioned in section 2.2.1.10 of materials and methods). The silver stained gel of control sample showed sharp silver stained bands of PM5 pre-mRNA (lanes containing fractions 4-6) and peptides (lanes containing fractions 10-16) where as that of UV-irradiated sample displayed a smear of PM5 pre-mRNA along with the cross-linked peptides (lane containing fraction 4). A smear was also observed in the lane containing fraction 8 presumably due to the cross-linked peptides with smaller RNA species (Figure 3.25).

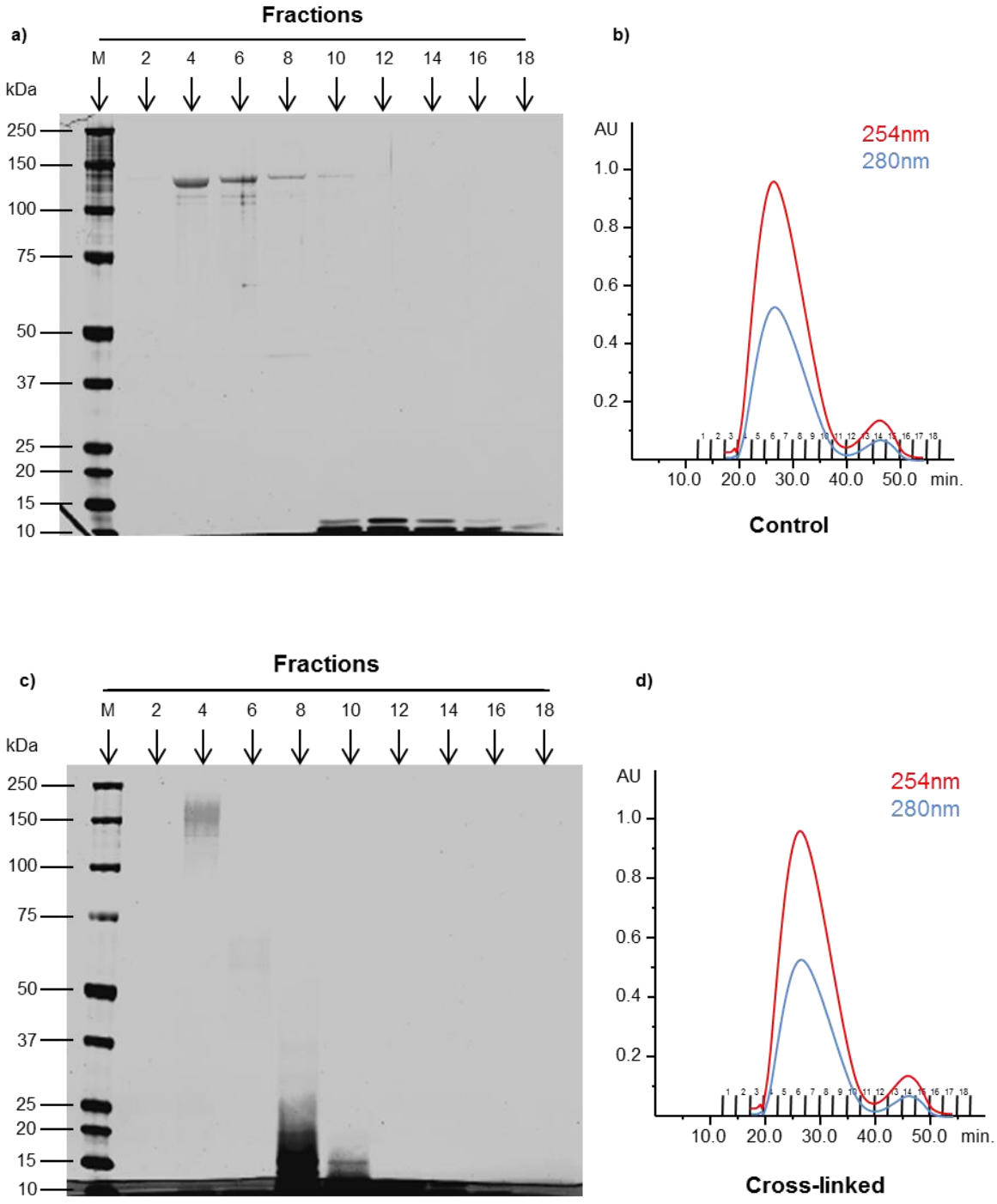


Figure 3.25: Silver stained gels with their respective size exclusion chromatograms (Figure adopted from Qamar et al., 2015). (a) Silver stained gel of control sample fractions; (c) Silver stained gel of cross-linked sample fractions; (b) Size exclusion chromatogram of control sample fractions; (d) Size exclusion chromatogram of cross-linked sample fractions. In comparison of the 4th fraction of control and cross-linking samples, a smear due to cross-linking of RNA is clearly visible in silver stained gel of cross-linking sample.

The fractions containing RNA were pooled together and ethanol precipitated. For the unambiguous identification of the cross-linked peptide by mass spectrometry, the corresponding cross-linked nucleotide moiety was generated as small as possible. The pellets were dissolved in urea and the RNA was digested by using combination of nucleases. First by RNase A and T1 and then by benzonase so the RNA oligonucleotide should not be more than four nucleotides. The samples were then subjected to C18 reversed-phase chromatography to desalt the sample and to remove the non-cross-linked RNA from the cross-linked peptide-oligonucleotide heteroconjugates. Making use of the phosphate backbone of the peptide-oligonucleotide heteroconjugate, the cross-links were enriched from residual non-cross-linked peptides by TiO_2 solid phase extraction as mentioned in section 2.2.5.3 of materials and methods. The cross-links were analyzed by mass spectrometry (Figure 1.4). The data obtained was then analyzed by using OpenMS. The cross-linked peptide hits were identified by comparing with the control and after removing the true peptide hits as mentioned in section 2.2.8.2 of materials and methods. The candidate spectra of the cross-linked peptides were then manually validated.

During the current studies, more than 3000 candidate cross-linked peptide hits generated by PM5 pre-mRNA and MINX pre-mRNA each, have been manually validated in order to filter out the false positive results from the true cross-linked peptide hits. More than 290 cross-links belonging to 123 peptides along with respective oligonucleotide moieties have been identified. In 87 of the cross-linked peptides the cross-linking site has been cut down to amino acid resolution. These cross-linked peptides belonged to 54 different proteins. The positions of these cross-linked peptides within their respective proteins revealed that most of them belong to the RBDs like RRMs and KH domains etc. (Table 3.7, Figure 3.26). This adds more authenticity to the results obtained from the present study. In addition to this the cross-linked peptides who have not been assigned to any RBDs can also be considered as a strong candidate for the discovery of novel RNA-binding motifs.

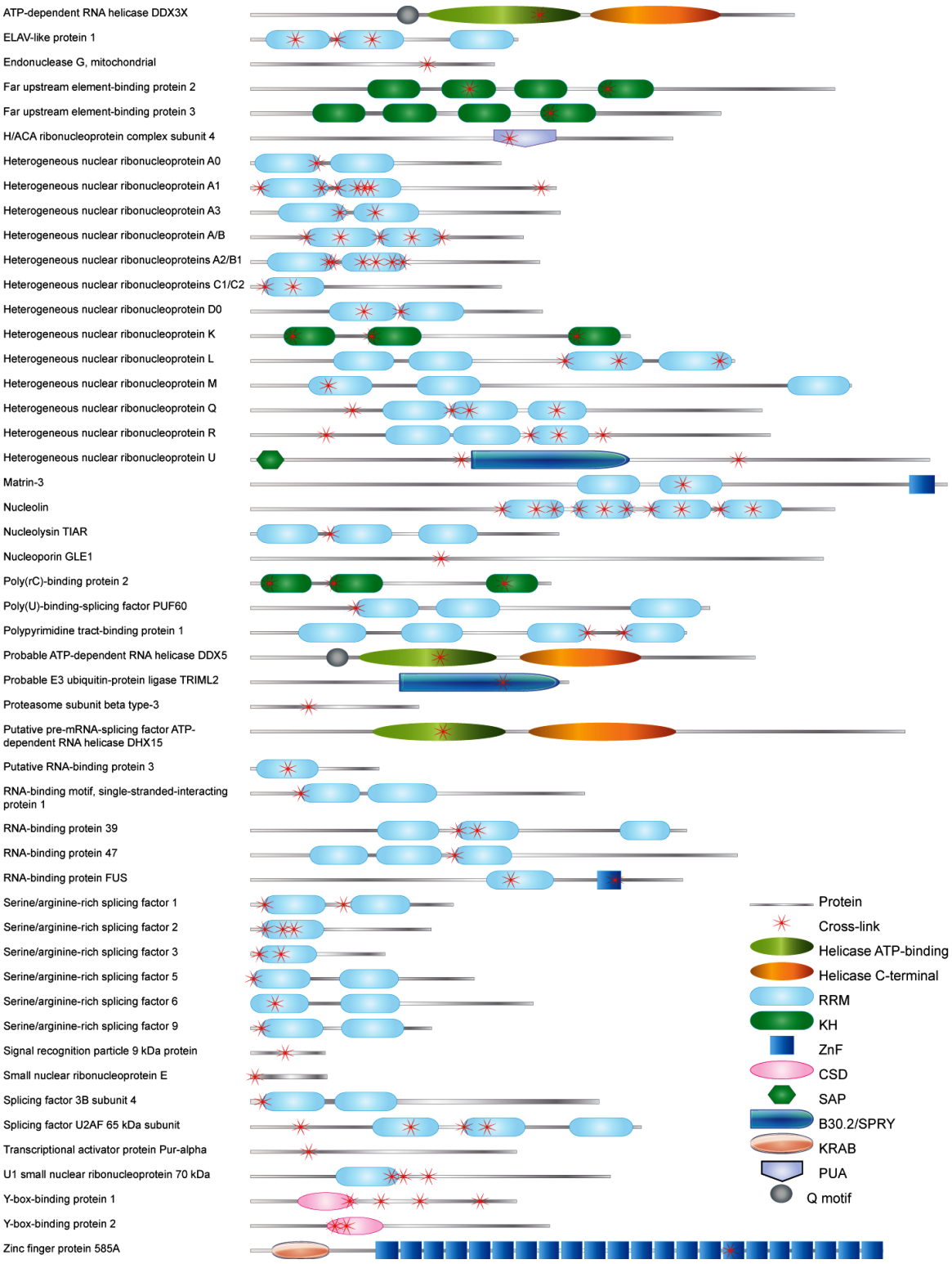


Figure 3.26: The graphical representation of distribution of the cross-linking sites within RNA-binding proteins along with their domains. The structures have been derived from UniProt/PROSITE (human) database. The ribosomal proteins are not included within the figure. The annotation of the symbols representing the domains/sites/protein are given on the lower right side of the figure.

Table 3.7: Cross-links from RNP complex from *HeLa* nuclear extract.

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Nucleoporin GLE1/Terminal uridylyltransferase 4/Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1	Q53GS7/ Q5TAX3/ Q8N1G2	-	^{482}EKFVK^{486}/^{565}EKFVK^{569}/ ^{684}EKFVK^{688}	F^{484}/F^{567}/F^{686}	U GU
ELAV-like protein 1	Q15717	RRM1	^{56}VAGHSLGYGFVNYVTAK^{72}	-	U CU UU CUU UUU
		RRM2	^{105}DANLYISGLPR^{115}	Y^{109}	U AU UU
		RRM2	^{148}GVAFIR^{153}	F^{151}	UU UUU
Endonuclease G, mitochondrial	Q14249	-	^{215}YQVIGK^{220}	Y^{215}-Q^{216}	U-H_2O AU-H_2O
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform/Zinc finger protein 585A/585B/420	Q15172/ Q6P3V2/ Q52M93/ Q8TAQ5	ZnF-16	^{222}RAFIR^{226}/^{585}RAFIR^{589}/^{585}RAFIR^{589}/^{342}RAFIR^{346}	F^{224}/F^{587}/ F^{587}/F^{344}	UU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Far upstream element-binding protein 2	Q92945	KH2	^{267}M(Ox)ILIQDGSQNTNVDKPLR^{284}	G^{273}-S^{274}	U
					UU
		KH4	^{436}CGLVIGR^{442}	C^{436}	U-H_3PO_4
					U
					UU
					AU
Far upstream element-binding protein 3	Q96I24	KH4	^{366}CGLVIGK^{372}	C^{366}	U
H/ACA ribonucleoprotein complex subunit 4	O60832	PUA	^{315}IMLPGVLR^{322}	M^{316}	AU-H_2O
					AU
					AAU
					GU
			^{315}IM(Ox)LPGVLR^{322}	-	AU
Heterogeneous nuclear ribonucleoprotein A0	Q13151	RRM1	^{82}AVSREDSARPGAHAK^{96}	-	UGA
Heterogeneous nuclear ribonucleoprotein A/B	Q99729	RRM2	^{156}IFVGGLNPEATEEK^{169}	F^{157}	U
					AU
					GU
					UU
		-	^{233}EVYQQQQYGSGGR^{245}	Y^{240}	U
					AU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoprotein A/B	Q99729	RRM1	^{111}GFGFILFK^{118}	-	U
					AU
					UU
		RRM2	^{196}GFVFITFK^{203}	F^{197} or F^{199} or F^{202}	UU
		RRM1	^{71}M(Ox)FVGGLSWDTSK^{82}	W^{78}	U
					AU
Heterogeneous nuclear ribonucleoprotein A1	P09651	-	^{353}NQGGYGGSSSSSSYGSGR^{370}	Y^{366}	U
					UA
Heterogeneous nuclear ribonucleoprotein A1/A1-like 2	P09651/ Q32P51	RRM1	^{15}KLFIGGLSFETTDESLR^{31}/^{15}KLFIGGLSFETTDESLR^{31}	-	U
					U-H_2O
					AU
			^{16}LFIGGLSFETTDESLR^{31}/^{16}LFIGGLSFETTDESLR^{31}	F^{23}/F^{23}	U-H_2O
					U
					AU-H_2O
					CU
					AGU
					UU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoprotein A1/A1-like 2	P09651/Q32P51	RRM2	^{147}GFAFVTFDDHDSVDK^{161}/^{147}GFAFVTFDDHDSVDK^{161}	-	U
					AU
					AUU
		RRM1	^{89}AVSREDSQRPGAHLTVK^{105}/^{89}AVSREDSQRPGAHLTVK^{105}	-	U
		RRM2	^{131}IEVIEIM(Ox)TDR^{140}/^{131}IEVIEIM(Ox)TDR^{140}	-	U
					AU
					CU
			^{131}IEVIEIM(Ox)TDRGSGK^{144}/^{131}IEVIEIM(Ox)TDRGSGK^{144}	R^{140}/R^{140}	U
					GU
			^{131}IEVIEIMTDR^{140}/^{131}IEVIEIMTDR^{140}	M^{137}/M^{137}	AU
					UU
Heterogeneous nuclear ribonucleoprotein A1/A1-like 2/A3	P09651/ Q32P51/ P51991	RRM2	^{107}IFVGGIK^{113}/^{107}IFVGGIK^{113}/^{128}IFVGGIK^{134}	F^{108}/F^{108}/ F^{129}	U
			^{106}KIFVGGIK^{113}/^{106}KIFVGGIK^{113}/^{127}KIFVGGIK^{134}	-	U-H_2O
					UA-H_2O
Heterogeneous nuclear ribonucleoprotein A3	P51991	RRM2	^{152}IETIEVMEDR^{161}	M^{158}	U
					UU
					UG
			^{152}IETIEVM(Ox)EDR^{161}	-	U
		RRM 1	^{110}AVSREDSVKPGAHLTVK^{126}	-	U
					UGA

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	RRM1	^{96}AVAREESGKPGAHVTVK^{112}	R^{99}	U
					CU
					ACGU
			^{100}EESGKPGAHVTVK^{112}	G^{103}-K^{104}	U-H_2O
		RRM2	^{154}GFGFVTFDDHDPVDK^{168}	-	U
					AU
		RRM2	^{174}YHTINGHNAEVR^{185}	Y^{174}	U
					AU
					UU
		RRM2	^{187}ALSRQEM(Ox)QEVQSSR^{200}	-	AU
					AUU
		RRM2	^{138}IDTIEIITDR^{147}	-	U-H_2O
			^{138}IDTIEIITDRQSGK^{151}	R^{147}	U
					CU
Heterogeneous nuclear ribonucleoproteins C1/C2	P07910	RRM	^{51}GFAFVQYVNER^{61}	F^{52} or F^{54}	U
					UU
					UUU
Heterogeneous nuclear ribonucleoproteins C1/C2/C-like 1	P07910/ O60812	RRM	^{18}VFIGNLNTLVVK^{29}/^{18}VFIGNLNTLVVK^{29}	F^{19}/F^{19}	UU
					UUU
					U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoprotein D0/D-like	Q14103/ O14979	RRM1	^{139}GFGFVLFK^{146}/^{190}GFGFVLFK^{197}	F^{142}/F^{193}	U
					AU
					CU
					UU
					AUU
Heterogeneous nuclear ribonucleoprotein D0	Q14103	RRM2	^{184}IFVGGLSPDTPEEK^{197}	-	U
Heterogeneous nuclear ribonucleoprotein K	P61978	KH2	^{149}LLIHQSLAGGIIGVK^{163}	G^{157}	U
					CU
					UU
		KH3	^{397}DLAGSIIGK^{405}	G^{400}	U
		KH1	^{53}NAGAVIGK^{60}	G^{55}	AU
					GU
Heterogeneous nuclear ribonucleoprotein L	P14866	RRM3	^{359}YGPQYGHPPPPPPPPEYGPHADSPVLM(Ox)VYGLDQSK^{393}	Y^{386}	U
					AU
		RRM4	^{569}NPNGPYPYTLK^{579}	Y^{574}	U
Heterogeneous nuclear ribonucleoprotein L/L-like	P14866/ Q8WVV9	RRM3	^{449}LNVCVSK^{455}/^{402}LNVCVSK^{408}	-	U
Heterogeneous nuclear ribonucleoprotein M	P52272	RRM1	^{95}VGEVTYVELLM(Ox)DAEGK^{110}	Y^{100}	U
Heterogeneous nuclear ribonucleoprotein Q	O60506	RRM2	^{266}VTEGLTDVILYHQPDDK^{282}	Y^{276}	U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoprotein Q	O60506	RRM3	^{370}LKDYAFIHFDER^{381}	-	U
					AU
					UU
					UUU
			^{372}DYAFIHFDER^{381}		U
					AU
					GU
					UU
Heterogeneous nuclear ribonucleoprotein R	O43390	RRM3	^{373}LKDYAFVHFEDR^{384}	-	U
					AU
					UU
			^{375}DYAFVHFEDR^{384}		U
					AU
					UU
		-	^{428}STAYEDYYYHPPPR^{441}	-	U
Heterogeneous nuclear ribonucleoprotein Q/R	O60506/ O43390	-	^{124}IKALLER^{130}	K^{125}	U-H_2O
		RRM2	^{245}LFVGSIPK^{252}/^{248}LFVGSIPK^{255}	S^{249}/S^{252}	U-H_3PO_4
					U-H_2O
					U
					AU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Heterogeneous nuclear ribonucleoprotein Q/R	O60506/ O43390	RRM2	^{245}LFVGSIPK^{252}/^{248}LFVGSIPK^{255}	S^{249}/S^{252}	CU
					GU
					UU
					AAU
					ACU
					CUU
		RRM3	^{339}VLFVR^{343}/^{342}VLFVR^{346}	F^{341}/F^{344}	U
					AU
					CU
					UU
		-	^{92}SAFLCGVM(Ox)K^{100}/^{95}SAFLCGVM(Ox)K^{103}	C^{96}/C^{99}	U
					UU
Heterogeneous nuclear ribonucleoprotein U	Q00839	-	^{593}M(Ox)CLFAGFQR^{601}	C^{594}	U
		-	^{256}GYFEYIEENK^{265}	Y^{257}	U
Nucleolin/Origin recognition complex subunit 1/DNA-binding protein RFX8/Titin	P19338/ Q13415/ Q6ZV50/ Q8WZ42	RRM4/BAH/-/Ig-like38	^{573}TLFVK^{577}/^{153}TLFVK^{157}/^{283}TLFVK^{287}/^{5781}TLFVK^{5785}	F^{575}/F^{155}/F^{285}/F^{5783}	UGG

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Nucleolin	P19338	RRM1	^{298}VEGTEPTTAFNLFVGNLNFNK^{318} ^{296}QKVEGTEPTTAFNLFVGNLNFNK^{318}	-	U AU GU GUU U
		RRM1	^{348}KFGYVDFESAEDLEK^{362} ^{349}FGYVDFESAEDLEK^{362}	Y^{351}	U GU-H_2O U GU-H_2O
		RRM2	^{399}NLPYKVTQDELK^{410}	Y^{402}	U AU GU
		RRM2	^{430}GIAYIEFK^{437}	Y^{433}	U AU UU GU-H_2O
		RRM3	^{524}GYAFIEFASFEDAK^{537}	F^{527}	U AU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Nucleolin	P19338	RRM4	^{611}GFGFVDFNSEEDAK^{624}	F^{612} or F^{614} or F^{617}	U
					AU
					GU
					UU
		RRM3	^{487}TLVLSNLSYSATEETLQEVFEK^{508}	Y^{495}	U
					AU
		RRM1	^{371}VFGNEIKLEKPK^{382}	K^{377}	U-H_2O
		RRM2	^{458}SISLYYTGEK^{467}	Y^{462}	U
Nucleolysin TIAR	Q01085	RRM2	^{92}DTSNHFHVFVGDLSPEITTEDIK^{114}	-	U
					UU
					UUU
Poly(rC)-binding protein 1/2/3	Q15365/ Q15366/ P57721	KH3	^{302}INEIRQM(Ox)SGAQIK^{314}/^{310}INEIRQM(Ox)SGAQIK^{322}/ ^{316}INEIRQM(Ox)SGAQIK^{328}	-	U-H_2O
Poly(rC)-binding protein 2/3	Q15366/ P57721	KH2	^{102}LVVPASQCGSLIGK^{115}/^{134}LVVPASQCGSLIGK^{147}	C^{109}/C^{141}	U
					AU
					GU
					UU
					U-H_3PO_4

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Poly(rC)-binding protein 1/2/3/4	Q15365/ Q15366/ P57721/P 57723	KH1	^{24}EVGSIIGK^{31}/^{24}EVGSIIGK^{31}/^{56}EVGSIIGK^{63}/^{28}EVGSIIGK^{35}	E^{24}-G^{26}/ E^{24}-G^{26}/ E^{56}-G^{58}/ E^{28}-G^{30}	U
Poly(U)-binding-splicing factor PUF60	Q9UHX1	RRM1	^{131}VYVGSIYYELGEDTIR^{146}	Y^{138}	U-H_3PO_4
					U
					AU
					GU
					CU
					UU
					AUU
					UUU
Polypyrimidine tract-binding protein 1	P26599	RRM3	^{411}HQNVQLPR^{418}	H^{411}	U
		RRM4	^{445}NFQNIFPPSATLHLSNIPPSVSEEDLK^{471}	-	U-H_2O
Probable ATP-dependent RNA helicase DDX5	P17844	Helicase ATP-binding	^{228}LIDFLECGK^{236}	C^{234}	U
Probable E3 ubiquitin-protein ligase TRIML2	Q8N7C3	B30.2/ SPRY	^{310}RLFLEK^{315}	-	UU
Proteasome subunit beta type-3	P49720	-	^{71}LNLYELK^{77}	Y^{74}	U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	O43143	Helicase ATP-binding	^{235}YM(Ox)TDGM(Ox)LLREAM(Ox)NDPLLER^{253}	-	ACU U
Putative RNA-binding protein 3	P98179	RRM	^{48}GFGFITFTNPEHASVAM(Ox)R^{65}	-	U
RNA-binding protein 39	Q14498	RRM2	^{276}IESIQLMM(Ox)DSETGR^{289} ^{276}IESIQLM(Ox)M(Ox)DSETGR^{289}	M^{282} -	U U
RNA-binding protein 39/Probable RNA-binding protein 23	Q14498/ Q86U06	RRM2	^{252}LYVGSLHFNITEDM(Ox)LR^{267}/ ^{265}LYVGSLHFNITEDM(Ox)LR^{280}	-	U UU
RNA-binding protein 47/ APOBEC1 complementation factor/Mitotic spindle assembly checkpoint protein MAD2B	A0AV96/ Q9NQ94/ Q9UI95	RRM3/ RRM3/HORMA	^{247}ILYVR^{251}/^{232}ILYVR^{236}/^{30}ILYVR^{34}	Y^{249}/Y^{234}/ Y^{32}	AU
RNA-binding protein FUS	P35637	RRM	^{317}TGQPMINLYTDR^{328} ^{317}TGQPM(Ox)INLYTDR^{328}	-	AU U
		RANBP 2-type zinc finger	^{449}APKPDGPGGGPGGSHM(Ox)GGNYGDDR^{472}	-	U
Serine/arginine-rich splicing factor 1	Q07955	RRM1	^{18}IYVGNLPPDIR^{28}	Y^{19}	U AU
		-	^{112}YGPPSR^{117}	Y^{112}	U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Serine/arginine-rich splicing factor 2/8	Q01130/ Q9BRL6	RRM	^{56}GFAFVR^{61}	F^{59}	U
					UU
		RRM	^{18}VDNLTYR^{24}	Y^{23}	U
		RRM	^{40}VGDVYIPR^{47}	Y^{44}	U
Serine/arginine-rich splicing factor 3	P84103	RRM	^{38}SVWVAR^{43}	W^{40}	UU
			^{38}SVWVARNPPGFAFVEFEDPR^{57}	-	U
					AU
					CU
					ACU
		RRM	^{12}VYVGNLGNNGNK^{23}	Y^{13}	U
Serine/arginine-rich splicing factor 5	Q13243	RRM1	^{6}VFIGRLNPAAR^{16}	G^{9}-R^{10}	U
					CU
					ACU
					CGU
Serine/arginine-rich splicing factor 6	Q13247	RRM1	^{35}NGYGFVEFEDSR^{46}	-	U
Serine/arginine-rich splicing factor 9	Q13242	RRM1	^{16}IYVGNLPTDVR^{26}	-	AU
Signal recognition particle 9 kDa protein	P49458	-	^{42}VTDDLVCLVYK^{52}	C^{48}	U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Splicing factor U2AF 65 kDa subunit	P26368	RRM1	^{196}NFAFLEFR^{203}	F^{199}	U
					CU
		-	^{68}GAKEEHGGLIR^{78}	K^{70}	AU-H_2O
		RRM2	^{261}LFIGGLPNYLNDDQVK^{276}	L^{261}-F^{262}	U
					AU
					CU
					UU
					AUU
					CUU
					UUU
		RRM2	^{287}AFNLVKDSATGLSK^{300}	D^{293}-S^{294}	UU
Transcriptional activator protein Pur-alpha/beta/Purine-rich element-binding protein gamma	Q00577/ Q96QR8/ Q9UJV8	-	^{73}FYLDVK^{78}/^{46}FYLDVK^{51}/^{72}FYLDVK^{77}	F^{74}/F^{47}/ F^{73}	AU
U1 small nuclear ribonucleoprotein 70 kDa	P08621	RRM	^{173}RVLVDVER^{180}	L^{175}	U-H_3PO_4
					U
					AU-H_3PO_4
					AU-H_2O
					AU
					ACU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
U1 small nuclear ribonucleoprotein 70 kDa	P08621	RRM	^{174}VLVDVER^{180}	L^{175}	U
					AU
					AU-H_2O
					AU-H_3PO_4
		-	^{192}LGGGLGGTR^{200}	L^{196}	ACU
		-	^{219}YDERPGPSPLPHR^{231}	Y^{219}	U
Y-box-binding protein 1	P67809	-	^{205}RPQYSNPPVQGEVM(Ox)EGADNQGAGEQGRPVR^{234}	Y^{208}-N^{210}	U
					UG
		CSD	^{119}GAEAANVTGPGGVPVQGSKYAADR^{142}	S^{136}-Y^{138}	U-H_2O
		-	^{157}NYQQNYQNSESGEK^{170}	Y^{158}	UA
			^{157}NYQQNYQNSESGEKNEGSESAPEGQAQQR^{185}	-	U
ATP-dependent RNA helicase DDX3X/ DDX3Y/Probable ATP-dependent RNA helicase DDX5/DDX17	O00571/ O15523/ P17844/ Q92841	Helicase ATP-binding	^{352}M(Ox)LDMGFEPQIR^{362}/^{350}M(Ox)LDMGFEPQIR^{360}/ ^{253}M(Ox)LDMGFEPQIR^{263}/^{330}M(Ox)LDMGFEPQIR^{340}	M^{355}/M^{353}/ M^{356}/M^{333}	U
Splicing factor 3B subunit 4	Q15427	RRM1	^{10}NQDATVYVGGLDEK^{23}	Y^{16}	U
Small nuclear ribonucleoprotein E	P62304	-	^{10}VQKVM(Ox)VQPINLIFR^{23}	K^{12}	U-H_2O
RNA-binding motif, single-stranded-interacting protein 1/2/3	P29558/ Q15434/ Q6XE24/	RRM1	^{62}TNLYIR^{67}/^{56}TNLYIR^{61}/^{61}TNLYIR^{66}	Y^{65}/Y^{59}/Y^{64}	U
Matrin-3	P43243	RRM2	^{525}NYILM(Ox)R^{530}	Y^{526}	U

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Y-box-binding protein 1/2/3	P67809/ Q9Y2T7/ P16989	CSD	^{70}NGYGFINR^{77}/^{105}NGYGFINR^{112}/^{102}NGYGFINR^{109}	F^{74}/F^{109}/ F^{106}	U
					UU
			^{65}WFNVRNGYGFINR^{77}/^{100}WFNVRNGYGFINR^{112}/^{97}WFNVRNGYGFINR^{109}	-	UU
		CSD	^{78}NDTKEDVFVHQTAIK^{92}/^{113}NDTKEDVFVHQTAIK^{127}/	-	U
			^{110}NDTKEDVFVHQTAIK^{124}		AU
					CU
					UU
					ACU
					UUU
			^{110}NDTKEDVFVHQTAIKK^{125}		CU
			^{78}N(Carbamyl)DTKEDVFVHQTAIK^{92}/		U
			^{113}N(Carbamyl)DTKEDVFVHQTAIK^{127}/		AU
			^{110}N(Carbamyl)DTKEDVFVHQTAIK^{124}		U
			^{82}EDVFVHQTAIK^{92}/^{117}EDVFVHQTAIK^{127}/^{114}EDVFVHQTAIK^{124}		AU
					CU
					GU
					GU-H_2O
					ACU
					UU
					UUU

Continued.....

Protein	UniProt	Motif/ Domain	Peptide	Amino acid	RNA
Y-box-binding protein 1/2/3	P67809/ Q9Y2T7/ P16989	CSD	^{82}EDVFVHQTAIKK^{93}/^{117}EDVFVHQTAIKK^{128}/^{114}EDVFVHQTAIKK^{125}	-	U AU CU GU UU
Y-box-binding protein 1/Protein TASOR/Proteoglycan 3	P67809/ Q9UK61/ Q9Y2Y8	-/-/C-type lectin	^{284}NFNYR^{288}/^{142}NFNYR^{146}/^{143}NFNYR^{147}	Y^{287}/ Y^{145}/ Y^{146}	AU
40S ribosomal protein S2	P15880	-	^{247}TYSYLTPDLWK^{257}	Y^{248} or Y^{250}	U UU
60S ribosomal protein L5	P46777	-	^{148}HIM(Ox)GQNVADYMR^{209} ^{148}HIM(Ox)GQNVADYM(Ox)R^{209}	Y^{207}	AU U AU
60S ribosomal protein L6	Q02878	-	^{238}EKYEITEQR^{246}	E^{238}-K^{239}	U-H_2O
60S ribosomal protein L34	P49207	-	^{94}AFLIEEQK^{101}	I^{97}	U GU

Protein: Protein name; UniProt ID; UniProt ID of the identified protein; Motif/Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence, the ones which were identified by using PM5 pre-mRNA their first position number was highlighted in grey and the ones which were identified by using MINX pre-mRNA their sequence was highlighted in yellow; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotides found to be cross-linked to the peptide.

4 DISCUSSION

Soon after the emergence of pre-mRNA from its transcription site, it undergoes splicing, editing and polyadenylation, mediated by RBPs (RNA binding proteins) and trans-acting RNAs, present as RNPs. RNA in concert with the protein partners performs functions of considerable subtlety. Binding of proteins with RNA in fact determines the fate of the RNA from its synthesis to decay (Castello et al., 2012; Beckmann et al., 2016).

Studies on RNA binders have corroborated the presence of protein domains with dual functionality, RNA binding activity and enzymatic properties (Walden et al., 2006), foldings of unexplored function (Jia et al., 2008) and unbound protein sectors with ill-defined tertiary structure (Phan et al., 2011). Advanced scientific approaches have led to the exploration of entire repertoire of RBPs (Cléry & Allain, 2013). There are multiple copies of RNA binding domains possessing varied structural configurations to broaden the functional repertoire of RNA binding proteins, this in turn helps to accommodate enormous diversified substrates (Lunde et al., 2007).

The work presented in the current study focused on two main objectives:

1. (a) Qualitative analysis of protein-RNA cross-links involving Brat-NHL-hb RNA complex and CWC2-U6/U4 snRNAs complex using conventional cross-linking method.
   (b) Qualitative analysis of protein-RNA cross-links derived from highly complex RNP assemblies isolated from *HeLa* nuclear extract by using the modified and optimized protocol.
2. Quantitative studies of CWC2-U6/U4 snRNAs cross-links.

In this chapter the implication of qualitative approach for the analysis of Brat-NHL-hb RNA cross-links and CWC2-U6/U4 snRNAs cross-links has been discussed in conjunction with the strategy adopted for the quantitative studies of CWC2-U6/U4 snRNAs cross-links by using isotope labeled RNAs.

Structural investigation through optimized method used in this study helped to elucidate interaction sites within large RNP complexes using UV induced cross-linking approach.

This investigation may help in determining the significance of mass spectrometric methods for identification and characterization of protein-RNA interactions in RNP complexes as well as in the prospective quantitative analysis of protein-RNA cross-links

During the course of this research several RNP complexes studied have been discussed as follows

4.1 Cross-links from Brat-NHL-hb RNA Complex

Various TRIM-NHL proteins like Drosophila Brat protein, *C. elegans* NHL-2 and LIN-41 proteins and mammalian TRIM71 etc. have been found in different RNP complexes (Duchaine et al., 2006; Hammell et al., 2009; Rybak et al., 2009; Wulczyn et al., 2010; Chang et al., 2012; Loedige et al., 2013). The interactions within these RNPs are either dependent on RNA (Hammell et al., 2009; Chang et al., 2012; Li et al., 2012) or the relevant NHL domain of the protein (Neumüller et al., 2008; Schwamborn et al., 2009; Chang et al., 2012; Loedige et al., 2013). Defining domain of the TRIM-NHL is a set of the NHL repeats. The NHL domain of TRIM-NHL proteins has been reported as RBD (Kwon et al., 2013).

Previously it has been believed that the Pum protein directly interacts with the hb RNA and then along with Nos recruits the Brat via protein-protein interactions. During our studies it has been shown that the Brat protein directly binds to the hb RNA. The NHL domain of brain tumor protein (Brat), member of conserved family of TRIM-NHL proteins, has been identified as RBD. In order to find the RNA-binding activity of Brat, the BRAT-NHL-hb RNA complex was *in vitro* assembled by incubating the *in vitro* transcribed hb RNA along with recombinantly expressed purified NHL domain of Brat.

Using conventional UV-cross-linking method followed by mass spectrometry (MS), positively charged top surface of the NHL domain has been identified to contact RNA (Table 3.1 & Figure 3.2). The mutations of the cross-linked and selected amino acids residues located within the cross-linked peptides on this surface abrogated Brat-NHL binding to the hb mRNA *in vitro* and impaired translational repression by Brat *in vivo* (Loedige et al., 2014).

Cross-linking site when cut down to the amino acid resolution, the exact amino acids in contact with the hb RNA or lying in close proximity to it are indicated. Five out of six identified cross-linked peptides are located on the top surface of the NHL domain. Within these peptides, seven amino acid residues were identified to be cross-linked, indicating their direct contact with or very close proximity to RNA. Among these, four residues were located on the top surface,

whereas two others lied in a positively charged stretch at the circumference (Table 3.1 & Figure 3.2).

Recently, it has been established that the Brat protein is recruited by the BoxA motif of NRE of hb RNA (Figure 3.3) (Loedige et al., 2015). The NHL domain of Brat is a six-bladed β propeller structure, similar to WD40 fold (Slack & Ruvkun, 1998; Stirnimann et al., 2010; Edwards et al., 2003) and has a positively charged top surface showing the potential to bind negatively charged molecules like nucleic acids (Loedige et al., 2014). Numerous aromatic and positively charged residues that have the tendency to interact with the RNA, project from this top surface. This surface has been reported to be affected by the mutations which lead to the Brat mutant phenotype or abrogate Brat function (Arama et al., 2000; Sonoda & Wharton, 2001; Harris et al., 2011).

4.2 Cross-links from CWC2-U4 and U6 snRNAs Complexes

4.2.1 Identification of Cross-links from CWC2-U4 and U6 snRNAs Complexes

Removal of non-coding introns from pre-mRNA and ligation of flanking coding regions called exons, are the essential events of splicing process for eukaryotic gene expression (Hayduk et al., 2012). Accurate and original protein coding capacity of the gene is conditional with the fact that how precisely and accurately the introns are removed to retain the originality of the gene (McGrail et al., 2009).

Process of splicing is catalysed by spliceosomes that are comprised of numerous protein splicing factors as well as five small nuclear RNAs namely U1, U2, U4, U5 and U6. A set of orderly interactions engenders the assembly of snRNPs onto pre-mRNA. The NineTeen Complex (NTC) which serves an important assistance for the spliceosomal snRNPs, has been considered as an essential component of active spliceosomes from yeast to humans (Villa & Guthrie, 2005). The release of U1 and U4 and the incorporation of NTC is indicative of the change of inactive state of spliceosome to active state (Hogg et al., 2010). Among the splicing factors, CWC2 is exclusively significant NTC associated protein that carries RRM (RNA recognition motif) and CCCH-type Zinc finger (ZnF) motifs for binding RNA (Ohi & Gould, 2002) in addition to Torus domain with motifs that has been reported to be capable of binding RNA.

Data generated by using UV-cross-linking approach along with mass spectrometry for the CWC2-U4/U6 snRNAs complex, revealed total eighteen peptides of CWC2 protein cross-linked to U4/U6 snRNAs (Table 3.2 & Figure 3.5). Seven of the identified cross-linked peptides have already been reported by Schmitzová et al., 2012 and eleven are the novel peptides that are found to be cross-linked with U4/U6 snRNAs during the current studies. Previously the C-

terminal domain of CWC2 protein has not been known for any RNA-binding activity but in the present studies, three peptides of the C-terminal domain i.e. ^{225}WANEDPDPAAQK^{236}, ^{276}TFPEASVDNVK^{286} and ^{315}ENISSKPSVGK^{325} with amino acid residues W^{225}, F^{277} and K^{320} have been tracked down to be cross-linked with U4/U6 snRNAs. With the exception of one cross-linked peptide ^{87}CEYLHHIPDEEDIGK^{101}, the cross-linking site has been reduced to the amino acid level in rest of the identified cross-linked peptides (Table 3.2).

In the future, the mutation analysis of the identified cross-linked amino acid residues, if performed, can help in adding to the information regarding the association of CWC2 with any of the U4/U6 snRNAs of the spliceosome.

4.2.2 Quantification of Cross-links from CWC2-U4 and U6 snRNAs Complexes

A relative quantification approach by using isotopically labeled RNAs was adopted during the present study to figure out whether the CWC2 as a single protein, and its domains in their individual entity, possess interaction affinity for any RNA or an impartial binding takes place. Relative quantification approach in this regard is quite amenable (Nikolov, et al., 2012). *In vitro* studies on CWC2 have revealed nonspecific binding of RNAs with CWC2 (McGrail et al., 2009). RRM, Torus, ZnF domains and connector elements are the key factors involved during CWC2 interaction with RNAs (Schmitzová et al., 2012).

Two sets of reciprocative experiments (Figure 3.6) elucidated sixteen peptides cross-linked to U4/U6 snRNAs in forward and reverse replicates (Table 3.3). Xcalibur software (mentioned in section 2.2.8.2 of materials and methods) has been used for relative quantitation of tracked down peptides. Three cross-linked peptides i.e. ^{106}TEVLDCFGR^{114} of torus domain, ^{136}TLYVGGIDGALNSK^{149} of RNP2 motif of RRM domain and ^{180}NCGFVK^{185} of RNP1 motif of RRM domain, identified in all forward and reverse replicates were selected for further quantitative analysis. Under the defined parameters, $\log_{10}$ of peak areas

calculated from XICs of individual cross-linked peptides of CWC2, along with all combinations of RNA moiety and charge states show that CWC2 is more inclined towards U4 snRNA for cross-linking as compared to U6 snRNA (Appendix Table 6.6). This was inconcordance to previous studies by McGrail et al., 2009. However, the ratios calculated for previously mentioned three cross-linked peptides pointed to the similar cross-linking tendency of CWC2 toward both U4 and U6 snRNAs in both sets of experiments (Figure 3.10).

In an attempt to adopt a semi-automated approach, Skyline software tool was used for quantitative analysis. The log_2 ratio of peak areas of the three individual cross-linked peptides, ^{106}TEVLDCFGR^{114} of torus domain, ^{136}TLYVGGIDGALNSK^{149} of RNP2 motif of RRM domain and ^{180}NCGFVK^{185} of RNP1 motif of RRM domain along with separate RNA moiety combinations but combined charge state of each RNA moiety, showed that CWC2 cross-links more efficiently with U4 snRNA as compared to U6 snRNA (Appendix Table 6.7 & 6.8).

In an effort to explore the preference of any domain of CWC2 protein for U4 or U6 snRNAs, the relative quantification studies performed on log_2 ratios of cross-linked peptides of RNP1 and RNP2 motifs and Torus domains for both sets of forward and reverse experiments, by using Xcalibur and Skyline software revealed that there was no evidence of any cross-link preference by any of the domains/motifs of CWC2 protein for U4 and U6 snRNAs (Figure 3.11).

4.3 Cross-links from *HeLa* Nuclear Extract

4.3.1 Identification of Cross-links from MS2-MBP Protein

Isolation of functional ribonucleoproteins *in vitro* is obtained through the combination of approaches like gel filtration with affinity chromatography employing bacteriophage MS2 coat protein. Spliceosome obtained so is used to determine the protein components of the spliceosome by mass spectrometry (Zhou & Reed, 2003). During the course of current studies, the RNP complex assembled by using *HeLa* nuclear extract was purified by MS2-MBP affinity purification method (Zhou & Reed, 2003). Data obtained by the MS analysis of the UV-cross-linked RNP complexes on LTQ Orbitrap Velos mass spectrometer was analysed by OpenMS. The affinity of the bacteriophage MS2 coat protein for its target MS2 RNA stem loops formed the basis of this method. The binding of MS2-MBP fusion protein to the MS2 RNA stem loops of target RNA has opened possibility to study the probable interaction sites of the aforementioned protein with RNA.

Direct interaction of MBP with nucleic acids has not been reported till now. Two peptides of MS2 coat protein and twelve peptides of MBP were found to be involved in cross-linking with oligonucleotides (Table 3.4 & Figure 3.13). Cross-linked amino acid residues for only seven peptides were screened out of fourteen cross-linked peptides. For the first time, the present studies have demonstrated that twelve peptides of MBP cross-link to uracil, whereby the lysine has been designated as the cross-linked amino acid residue.

Crystal structure of monomeric MBP (Figure 3.13) shows two distinct globular N and C domains, joined by three segments. Each domain is composed of central β-pleated sheet, flanked on both sides by α-helices (Quiocho et al., 1997). Most of the cross-linked peptides are found to be lying in the region of α-helices. Out of twelve cross-linked peptides, seven have been occupying N domain, four

occupying C domain and one occupying the third segment joining the two domains.

In the current studies, two peptides of MS2 protein were found to be cross-linked to uracil. The results are concomitant to those reported by Budowsky et al., 1976, according to which lysine residues in coat protein were found to cross-link to the genomic nucleic acid upon UV-irradiation.

Crystal structure of MS2 coat protein is composed of three coat protein dimers with a stretch of RNA (Figure 3.14). Each monomer is composed of five stranded β-sheet and two α-helical segments. The cross-linked lysine residues K^{442} and K^{447} of peptides ^{438}KYTIKVEVPK^{447} and ^{443}VEVPKGAWR^{451} respectively are found to be lying in β-sheet in closer proximity to cytosine at position −5 of nucleotides in RNA in the crystal structure. Grahn et al., 1999 reported that K^{442} is among the probable amino acids that formed cross-links with nucleotides and the substitution of wild type uracil by cytosine at position −5 resulted in stronger binding of RNA hairpin to MS2 coat protein so from the present study, it can be anticipated that K^{442} is the probable amino acid and uracil (in case of contamination)/cytosine (in case of bait RNA) at position −5 is the probable nucleotide taking part in cross-linking.

4.3.2 Identification of Uracil Fragments and Adducts

To determine the specific and definite cross-linking sites, mass spectrometric analysis of peptide oligonucleotide heteroconjugates is carried out. Cross-linking in combination to mass spectrometry provides an ideal paradigm to elucidate uncharacterized and unidentified RNA-protein cross-links (Kühn-Holsken et al., 2005). UV-induced cross-linking assists in generating direct zero-length cross links without intervening linker.

Keeping in view that the cross-links are additive in nature, the peptide-oligonucleotide cross-links were identified by using OpenMS. In addition to the peptide signals, the signals of characteristic marker ions produced by

fragmentation of the cross-linked nucleotides can also be observed in the MS2 spectra. Furthermore, the nucleotide fragments can be observed as adducts cross-linked to an amino acid residue resulting in the shifting of ion series within the spectra. The identification of these fragments and adducts may add to the authenticity of cross-linking data generated by the mass spectrometric analysis of the cross-links.

During the cross-linking analysis of RNP complex assembled by incubating labeled/non-labeled (PM5/MINX) pre-mRNA with *HeLa* nuclear extract proteins followed by the mass spectrometric and data analysis, revealed that the uracil nucleotide is the most frequently cross-linked nucleotide with the peptide moieties. In the previous studies numerous uracil nucleotide fragments have been reported (Table 3.5 & 3.6). However, in the current studies, there have been few signals in the MS2 spectra, which are assumed to be generated by the fragmentation of uracil. For this, the target signals generated by using isotopically labeled uracil (^{13}C, $^{13}C^{15}N$ and 5-D_1, ribose-3‘, 4‘, 5‘, 5‘-D4 labeled uracil) were compared to estimate the elemental composition.

The peptide ^{375}DYAFVHFEDR^{384} derived from heterogeneous nuclear ribonucleoprotein R protein was found to be cross-linked to uracil nucleotide. Upon CID fragmentation of nucleotides, the loss of phosphate group and neutral loss of water are usually observed. By keeping all these points in view the elemental composition of the uracil nucleotide fragment of interest can be predicted i.e. $C_8H_6N_2O_3$ generated via combined loss of two water (H_2O) molecules, one phosphate group (HPO_3) and formaldehyde (CH_2O) molecule (Figure 3.17). This result is in accordance with the previous studies conducted on the pseudouridine fragmentation in which the loss of two water molecules and CH_2O from the sugar moiety give rise to a product ion at *m/z* 179.

The peptide ^{215}YQVIGK^{220} of mitochondrial endonuclease G is found to be cross-linked to uracil nucleotide with the loss of H_2O molecule. Its MS2 spectrum shows relatively high intensity ion signal at *m/z* 175.0714, which is expected to be generated by the fragmentation of cross-linked uracil nucleotide. The

comparative analysis among the MS2 spectra of the same cross-linked peptide with labeled and non-labeled uracil shows that the elemental composition of the target ion signal is $C_6H_{10}N_2O_4$ (Figure 3.18 & 3.19).

The MS2 spectrum of the cross-linked peptide ^{215}YQVIGK^{220} with U-H_2O shows a shift in b ion series with the mass of 208 Da. When the MS2 spectra of the same cross-linked peptide along with differentially labeled and non-labeled uracil have been compared, the elemental composition of the target adduct was deduced to be $C_9H_8N_2O_4$ which was probably generated due to the loss of two H_2O and one HPO_3 molecules from the intact uracil nucleotide (Figure 3.18 & 3.20). In 1969, Rice & Dudek have also reported a similar fragment of 208 Da generated by the fragmentation of pseudouridine however in the current studies it has been observed by the fragmentation of uracil.

4.3.3 Identification of RNPs Isolated from *HeLa* Nuclear Extract and their Cross-linking Analysis

Biological and mechanistic functions of RNA-proteins complexes are crucial to many biological processes (Kramer et al., 2014). UV-induced cross-linking approach coupled with mass spectrometry analysis have paved way for the identification of proteins in direct contact with RNA (Castello et al., 2012; Klass et al., 2013). Progress and advancements in resolving different elements of RNA-protein complexes will contribute to the identification of different RBPs as well as will help in predicting the binding time of a set of given RBPs to an individual RNA (Klass et al., 2013). Such approaches will ultimately provide more precise information of the function of RNA-protein complexes, including how does the assemblage of the complexes take place and how they modulate cellular function.

Previous studies have established the specificity of combining UV-induced cross-linking and mass spectrometry in structural investigations of moderately complex RNA-protein complexes (Urlaub et al., 1997; Kühn-Hölsken et al., 2010; Kramer

et al., 2011; Mozaffari-Jovin et al., 2012) and also through mutagenesis studies of identified cross-linked amino acids in single RNPs (Ghalei et al., 2010; Müller et al., 2011).

In the present studies, the emphasis has been laid in establishing a method for the interaction analysis of highly complex RNP complexes, as in the current case the RNP complex isolated from the *HeLa* nuclear extract. For this the already existing protocols used for the analysis of small and less complex samples have been modified and optimized. For the isolation of the RNP complex, RNA-centric methodology has been adopted by using pre-mRNAs tagged with MS2-binding RNA stem loops (Deckert et al., 2006). The isolated complex has been cross-linked by UV-irradiation at 254 nm for 10 min. In order to improve the identification by mass spectrometry, the cross-linked peptides have been enriched from the non-cross-linked RNA and peptides by using size exclusion chromatography, C18 reversed-phase chromatography and TiO_2 solid phase extraction (Urlaub et al., 1995; Urlaub et al., 2002; Luo et al., 2008). Later on the enriched peptide-RNA oligonucleotide heteroconjugates have been analyzed by mass spectrometry.

Our results emphasize on the use of the established method to identify direct interaction sites within proteins-RNA complexes, and this has considerably broadened the scope of preceding studies in which mass spectrometry had been able to identify the entire proteins rather than the specific cross-linking sites (Baltz et al., 2012; Castello et al., 2012; Klass et al., 2013; Mitchell et al., 2013). The comparative study involving the cross-linking sites and their relative 3D structures signifies the structural relevance of cross-linking approach as well as providing an opportunity for predicting the novel RNA-binding sites.

There are considerable differences between previously reported methods and our adopted approach. The interactions that are described by computationally predicted RNA-binding sites, may not be recognized in a particular RNA-protein complex.

The current experimental approach helps in improving computational speculation of RNA-binding motifs in proteins that contain multiple RNA interaction sites (Schmitzová et al., 2012) as well as the proteins that constitute the composite RNA interaction site via inter protein interactions (Urlaub et al., 2001). The cross-linked proteins isolated from *HeLa* nuclear extract during this study are found to belong to mainly H/E complex (Wahl et al., 2008). The number of cross-links identified i.e. more than 290 belonging to 123 peptides along with respective RNA oligonucleotide combinations belonging to more than 54 proteins mostly located in the RBDs, has provided the new perspectives in studying the RNP complexes (Table 3.7 & Figure 3.26).

4.4 Conclusion and Future Perspectives

While much of the work has been done during the past decade to establish the methodologies for the identification and investigation of the interactions in RNA-protein complexes, there exist considerable challenges that are still needed to be addressed to identify cross-linked peptides, cross-linking sites and cross-linked RNA oligonucleotide moieties with RNA binding proteins.

Generating the accurate predictions about protein complexes that interact with RNA is an important step in understanding the biological functions and mechanisms. One of the main objectives of current study was to combine the results of different approaches for identification of the cross-linked regions within RNP complexes and explore the caveats and considerations required regarding adoption of method to identify cross-linked regions up to peptide or amino acid level. This was required to build a clear picture of the RNA-protein complex functionality, considering the assemblage of these complexes and their role in the modulation of cellular functions.

Vast heterogeneity exists in the RNA protein interactions within RNP complexes. Novel approaches to identify and validate intermolecular cross-links in RNPs may not only help in predicting the exact protein entity (peptide/amino acid) in already described RBDs/RRMs/RBMs involved in interacting with RNA but will also help in predicting the regions or domains that interact with RNA and have not been reported before. Relative quantitative analysis of RNA-protein cross-links has been carried out during the course of the study that helped to quantitatively analyse the interactions of the protein as well as of its domains with the cross-linked RNA. Manual interpretation of the MS/MS spectra has contributed promising improvement in data analysis strategy.

The studies conducted during the research work have contributed in the identification and characterization of protein-RNA interactions within the aforementioned complexes and also provided the quantitative insight into the protein-RNA interactions. The methods adapted will assist in the qualitative as well as quantitative interaction analysis of both *in vivo* and *in vitro* assembled

large RNP complexes. Fully automated system for quantitative data analysis of RNA-protein cross-links is required to handle the large cross-linking data sets. The results will also contribute in improving the data analysis approach for protein-RNA cross-links and will serve as an outlook to future directions of this project.

5 REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., ... & George, R. A. (2000). The genome sequence of *Drosophila melanogaster. Science*, *287*(5461), 2185-2195.

Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, *537*(7620), 347-355.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J. C., Mann, M., & Lamond, A. I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *The EMBO Journal*, *19*(23), 6569-6581.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, *92*(3), 291-294.

Anantharaman, V., Koonin, E. V., & Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic acids Research*, *30*(7), 1427-1464.

Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., & Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, *41*(10), 3912-3928.

Arama, E., Dickman, D., Kimchie, Z., Shearn, A., & Lev, Z. (2000). Mutations in the β-propeller domain of the Drosophila brain tumor (brat) protein induce neoplasm in the larval brain. *Oncogene*, *19*(33), 3706-3716.

Assenov, Y., Ramírez, F., Schelhorn, S. E., Lengauer, T., & Albrecht, M. (2007). Computing topological parameters of biological networks. *Bioinformatics*, *24*(2), 282-284.

Aston, F.W (1933). Mass Spectra and Isotopes. Longmans, Green & Company, London, UK.

Baker, R. T., Catanzariti, A. M., Karunasekara, Y., Soboleva, T. A., Sharwood, R., Whitney, S., & Board, P. G. (2005). Using deubiquitylating enzymes as research tools. *Methods in Enzymology*, *398*, 540-554.

Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., ... & Wyler, E. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell*, *46*(5), 674-690.

Bauer, M., Ahrné, E., Baron, A. P., Glatter, T., Fava, L. L., Santamaria, A., ... & Schmidt, A. (2014). Evaluation of data-dependent and-independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. *Journal of Proteome Research*, *13*(12), 5973-5988.

Beckmann, B. M., Castello, A., & Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflügers Archiv-European Journal of Physiology*, *468*(6), 1029-1040.

Bertsch, A., Gröpl, C., Reinert, K., & Kohlbacher, O. (2011). OpenMS and TOPP: open source software for LC-MS data analysis. In *Data Mining in Proteomics, Methods in Molecular Biology*, *696*, 353-367.

Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., & Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, *452*(7189), 846-850.

Betschinger, J., Mechtler, K., & Knoblich, J. A. (2006). Asymmetric segregation of the tumor suppressor brat regulates self-renewal in Drosophila neural stem cells. *Cell*, *124*(6), 1241-1253.

Beynon, R. J., & Pratt, J. M. (2005). Metabolic labeling of proteins for proteomics. *Molecular & Cellular Proteomics*, *4*(7), 857-872.

Biemann, K. (1990). Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods in Enzymology*, *193*, 886-887.

Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., & Heck, A. J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, *4*(4), 484-494.

Borland, K., & Limbach, P. A. (2017). Applications and advantages of stable isotope phosphate labeling of RNA in mass spectrometry. *Topics in Current Chemistry*, *375*(2), 33.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2), 248-254.

Bruce, C., Stone, K., Gulcicek, E., & Williams, K. (2013). Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. *Current Protocols in Bioinformatics*, *41*(1), 13-21.

Budowsky, E. I., Simukova, N. A., Turchinsky, M. F., Boni, I. V., & Skoblov, Y. M. (1976). Induced formation of covalent bonds between nucleoprotein components. V. UV or bisulfite induced polynucleotide-protein crosslinkage in bacteriophage MS2. *Nucleic acids Research*, *3*(1), 261-276.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., ... & Krijgsveld, J. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, *149*(6), 1393-1406.

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A. M., Foehr, S., … & Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell, 63*(4), 696-710.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., ... & Hentze, M. W. (2013). System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols*, *8*(3), 491-500.

Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G., & Baker, R. T. (2004). An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science*, *13*(5), 1331-1339.

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... & Hoff, K. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, *30*(10), 918-920.

Chan, S. P., Kao, D. I., Tsai, W. Y., & Cheng, S. C. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, *302*(5643), 279-282.

Chang, H. M., Martinez, N. J., Thornton, J. E., Hagan, J. P., Nguyen, K. D., & Gregory, R. I. (2012). Trim71 cooperates with microRNAs to repress Cdkn1a expression and promote embryonic stem cell proliferation. *Nature Communications*, *3*, 923.

Chen, C. H., Kao, D. I., Chan, S. P., Kao, T. C., Lin, J. Y., & Cheng, S. C. (2006). Functional links between the Prp19-associated complex, U4/U6 biogenesis, and spliceosome recycling. *RNA*, *12*(5), 765-774.

Chen, X., Wei, S., Ji, Y., Guo, X., & Yang, F. (2015). Quantitative proteomics using SILAC: principles, applications, and developments. *Proteomics*, *15*(18), 3175-3192.

Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., ... & Cramer, P. (2010). Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal*, *29*(4), 717-726.

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(4), S11.

Cléry, A., & Allain, F. H. T. (2013). From structure to function of RNA binding domains. Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK63528/.

Colaert, N., Barsnes, H., Vaudel, M., Helsens, K., Timmerman, E., Sickmann, A., ... & Martens, L. (2011). Thermo-msf-parser: an open source Java library to parse and visualize Thermo Proteome Discoverer msf files. *Journal of Proteome Research*, *10*(8), 3840-3843.

Conrads, T. P., Alving, K., Veenstra, T. D., Belov, M. E., Anderson, G. A., Anderson, D. J., ... & Thrall, B. D. (2001). Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. *Analytical Chemistry*, *73*(9), 2132-2139.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*, *10*(4), 1794-1805.

Cuomo, A., & Bonaldi, T. (2010). Systems biology “on-the-fly”: SILAC-based quantitative proteomics and RNAi approach in *Drosophila melanogaster.* In *Systems Biology in Drug Discovery and Development*, *Methods in Molecular Biology*, *662*, 59-78.

Darnell, R. B. (2010). HITS-CLIP: panoramic views of protein–RNA regulation in living cells. *Wiley Interdisciplinary Reviews: RNA*, *1*(2), 266-286.

Dawson, P. H. (2013). Quadrupole mass spectrometry and its applications. Elsevier Scientific Publishing Company, Amsterdam, Netherland.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., ... & Lührmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Molecular and Cellular Biology*, *26*(14), 5528-5543.

Doerr, A. (2015). DIA Mass Spectrometry*. Nature Methods, 12*(1), 35-35.

De Hoffman, E., & Stroobant, V. (2007). Mass Spectroscopy: Principles and Applications. John Wiley & Sons: West Sussex, UK.

Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., & Alice, M. B. (1968). Molecular beams of macroions. *The Journal of Chemical Physics*, *49*(5), 2240-2249.

Dubois, F., Knochenmuss, R., Zenobi, R., Brunelle, A., Deprun, C., & Le Beyec, Y. (1999). A comparison between ion-to-photon and microchannel plate detectors. *Rapid Communications in Mass Spectrometry*, *13*(9), 786-791.

Duchaine, T. F., Wohlschlegel, J. A., Kennedy, S., Bei, Y., Conte, D., Pang, K., ... & Yates, J. R. (2006). Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell*, *124*(2), 343-354.

Dudley, E., El-Sharkawi, S., Games, D. E., & Newton, R. P. (2000). Analysis of urinary nucleosides. I. Optimisation of high performance liquid chromatography/electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry*, *14*(14), 1200-1207.

Duval, M., Marenna, A., Chevalier, C., & Marzi, S. (2017). Site-Directed Chemical Probing to map transient RNA/protein interactions. *Methods*, *117*, 48-58.

Edwards, T. A., Wilkinson, B. D., Wharton, R. P., & Aggarwal, A. K. (2003). Model of the brain tumor–Pumilio translation repressor complex. *Genes & Development*, *17*(20), 2508-2513.

Eliuk, S., & Makarov, A. (2015). Evolution of orbitrap mass spectrometry instrumentation. *Annual Review of Analytical Chemistry*, *8*, 61-80.

Faini, M., Stengel, F., & Aebersold, R. (2016). The evolving contribution of mass spectrometry to integrative structural Biology. *Journal of the American Society for Mass Spectrometry*, *27*(6), 966-974.

Faoro, C., & Ataide, S. F. (2014). Ribonomic approaches to study the RNA-binding proteome. *FEBS letters*, *588*(20), 3649-3664.

Fredens, J., Engholm-Keller, K., Giessing, A., Pultz, D., Larsen, M. R., Højrup, P., ... & Færgeman, N. J. (2011). Quantitative proteomics by amino acid labeling in *C. elegans*. *Nature Methods*, *8*(10), 845-847.

Frese, C. (2013). Development and Application of Novel Electron Transfer Dissociation-based Technologies for Proteomics. (Doctoral dissertation, Utrecht University, Utrecht, Netherland), ProefschriftMaken, Utrecht, Netherland.

Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., & Domon, B. (2012). Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, *11*(12), 1709-1723.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*, (pp. 571-607). Humana Press, NJ, USA.

Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., ... & Bryant, S. H. (2004). Open mass spectrometry search algorithm. *Journal of Proteome Research*, *3*(5), 958-964.

Ghalei, H., Hsiao, H. H., Urlaub, H., Wahl, M. C., & Watkins, N. J. (2010). A novel Nop5–sRNA interaction that is required for efficient archaeal box C/D sRNP formation. *RNA*, *16*(12), 2341-2348.

Gillet, L. C., Leitner, A., & Aebersold, R. (2016). Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. *Annual Review of Analytical Chemistry*, *9*(1), 449-472.

Girolamo, F. D., Lante, I., Muraca, M., & Putignani, L. (2013). The role of mass spectrometry in the "omics" era. *Current Organic Chemistry*, *17*(23), 2891-2905.

Glish, G. L., & Burinsky, D. J. (2008). Hybrid Mass Spectrometers for tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *19*(2), 161-172.

Glish, G. L., & Vachet, R. W. (2003). The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery*, *2*(2), 140-150.

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*, *582*(14), 1977-1986.

Gordiyenko, Y., & Robinson, C. V. (2008). The emerging role of MS in structure elucidation of protein–nucleic acid complexes. *Biochemical Society Transactions, 36(*Pt 4), 723-731.

Grahn, E., Stonehouse, N. J., Murray, J. B., Van Den Worm, S. J. O. E. R. D., Valegaard, K. A. R. I. N., Fridborg, K., ... & Liljas, L. (1999). Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. *RNA*, *5*(1), 131-138.

Grassucci, R. A., Taylor, D. J., & Frank, J. (2007). Preparation of macromolecular complexes for cryo-electron microscopy. *Nature Protocols*, *2*(12), 3239-3246.

Griffiths, W. J., Jonsson, A. P., Suya, L. I. U., & Yuqin, W. A. N. G. (2001). Electrospray and tandem mass spectrometry in Biochemistry. *Biochemical Journal*, *355*(3), 545-561.

Guilhaus, M. (1995). Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry*, *30*(11), 1519-1532.

Hager, J. W. (2002). A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry*, *16*(6), 512-526.

Hale, J. E. (2013). Advantageous uses of mass spectrometry for the quantification of proteins. *International Journal of Proteomics*, *2013*.

Hall, K. B. (2002). RNA–protein interactions. *Current Opinion in Structural Biology*, *12*(3), 283-288.

Hammell, C. M., Lubin, I., Boag, P. R., Blackwell, T. K., & Ambros, V. (2009). NHL-2 Modulates microRNA activity in *Caenorhabditis elegans*. *Cell*, *136*(5), 926-938.

Harris, R. E., Pargett, M., Sutcliffe, C., Umulis, D., & Ashe, H. L. (2011). Brat promotes stem cell differentiation via control of a bistable switch that restricts BMP signaling. *Developmental Cell*, *20*(1), 72-83.

Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M., & Lührmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proceedings of the National Academy of Sciences*, *99*(26), 16719-16724.

Hartmuth, K., van Santen, M. A., Rösel, T., Kastner, B., & Lührmann, R. (2012). The preparation of HeLa cell nuclear extracts. *Alternative pre-mRNA Splicing: Theory and Protocols*, 311-319.

Hebert, A. S., Merrill, A. E., Bailey, D. J., Still, A. J., Westphall, M. S., Strieter, E. R., ... & Coon, J. J. (2013). Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods*, *10*(4), 332-334.

Hégarat, N., François, J. C., & Praseuth, D. (2008). Modern tools for identification of nucleic acid-binding proteins. *Biochimie*, *90*(9), 1265-1272.

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols*, *2*(8), 1849-1861.

Heyduk, T., Ma, Y., Tang, H., & Ebright, R. H. (1996). Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods in Enzymology*, *274*, 492-503.

Hillenkamp, F., & Karas, M. (1990). Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods in Enzymology*, *193*, 280-295.

Hogg, R., McGrail, J. C., & O'Keefe, R. T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochemical Society Transactions*, *38*(4), 1110-1115.

Horn, W. T., Convery, M. A., Stonehouse, N. J., Adams, C. J., Liljas, L., Phillips, S. E., & Stockley, P. G. (2004). The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: further challenges in the modeling of ligand–RNA interactions. *RNA*, *10*(11), 1776-1782.

Hsu, J. L., & Chen, S. H. (2016). Stable isotope dimethyl labeling for quantitative proteomics and beyond. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2079), 20150364.

Hsu, J. L., Huang, S. Y., Chow, N. H., & Chen, S. H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry*, *75*(24), 6843-6852.

Hu, A., Noble, W. S., & Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000 Research*, *5*(419).

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, *40*(4), 430-443.

Huber, C. G., & Oberacher, H. (2001). Analysis of nucleic acids by on-line liquid chromatography–Mass spectrometry. *Mass Spectrometry Reviews*, *20*(5), 310-343.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, *324*(5924), 218-223.

Iribarne, J. V., & Thomson, B. A. (1976). On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, *64*(6), 2287-2294.

Irish, V., Lehmann, R., & Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature*, *338*(6217), 646-648.

Jennebach, S., Herzog, F., Aebersold, R., & Cramer, P. (2012). Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic acids Research*, *40*(12), 5591-5601.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., ... & Bork, P. (2008). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids Research*, *37*(suppl_1), D412-D416.

Jia, J., Arif, A., Ray, P. S., & Fox, P. L. (2008). WHEP domains direct noncanonical function of glutamyl-Prolyl tRNA synthetase in translational control of gene expression. *Molecular Cell*, *29*(6), 679-690.

Jiang, H., & English, A. M. (2002). Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of Proteome Research*, *1*(4), 345-350.

Jurica, M. S., Licklider, L. J., Gygi, S. P., Grigorieff, N., & Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, *8*(4), 426-439.

Kapp, E. A., Schütz, F., Reid, G. E., Eddes, J. S., Moritz, R. L., O'Hair, R. A., ... & Simpson, R. J. (2003). Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Analytical Chemistry*, *75*(22), 6251-6264.

Karas, M., Bahr, U., & Dülcks, T. (2000). Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius' Journal of Analytical Chemistry*, *366*(6-7), 669-676.

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, *60*(20), 2299-2301.

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics*, *8*(7), 533-543.

Kelstrup, C. D., Young, C., Lavallee, R., Nielsen, M. L., & Olsen, J. V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *Journal of Proteome Research*, *11*(6), 3487-3497.

Kirpekar, F., & Krogh, T. N. (2001). RNA fragmentation studied in a matrix-assisted laser desorption/ionisation tandem quadrupole/orthogonal time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, *15*(1), 8-14.

Kito, K., & Ito, T. (2008). Mass spectrometry-based approaches toward absolute quantitative proteomics. *Current Genomics*, *9*(4), 263-274.

Klass, D. M., Scheibe, M., Butter, F., Hogan, G. J., Mann, M., & Brown, P. O. (2013). Quantitative proteomic analysis reveals concurrent RNA–protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Research*, *23*(6), 1028-1038.

Kohlbacher, O., Reinert, K., Gröpl, C., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., & Sturm, M. (2007). TOPP—the OpenMS proteomics pipeline. *Bioinformatics*, *23*(2), e191-e197.

Kovanich, D., Cappadona, S., Raijmakers, R., Mohammed, S., Scholten, A., & Heck, A. J. (2012). Applications of stable isotope dimethyl labeling in quantitative proteomics. *Analytical and Bioanalytical Chemistry*, *404*(4), 991-1009.

Kovtoun, S. V. (1997). Mass-correlated Delayed Extraction in Linear Time-of-Flight Mass Spectrometers. *Rapid Communications in Mass Spectrometry*, *11*(7), 810-815.

Kovtoun, S. V., & Cotter, R. J. (2000). Mass-correlated pulsed extraction: Theoretical analysis and implementation with a linear matrix-assisted laser desorption/ionization time of flight mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *11*(10), 841-853.

Kramer, K., Hummel, P., Hsiao, H. H., Luo, X., Wahl, M., & Urlaub, H. (2011). Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil-and 5-bromo-uracil-substituted RNA. *International Journal of Mass Spectrometry*, *304*(2-3), 184-194.

Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., ... & Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods*, *11*(10), 1064-1070.

Krüger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., ... & Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell*, *134*(2), 353-364.

Kühn-Hölsken, E., Lenz, C., Dickmanns, A., Hsiao, H. H., Richter, F. M., Kastner, B., ... & Urlaub, H. (2010). Mapping the binding site of snurportin 1 on native U1 snRNP by cross-linking and mass spectrometry. *Nucleic acids Research*, *38*(16), 5581-5593.

Kühn-Hölsken, E., Lenz, C., Sander, B., Lührmann, R., & Urlaub, H. (2005). Complete MALDI-ToF MS analysis of cross-linked peptide–RNA oligonucleotides derived from non labeled UV-irradiated ribonucleoprotein particles. *RNA*, *11*(12), 1915-1930.

Kvaratskhelia, M., & Le Grice, S. F. (2008). Structural analysis of protein-RNA interactions with mass spectrometry. In *RNA-Protein Interaction Protocols*, *Methods in Molecular Biology*, *488*, 213-219.

Kwon, M. H., Callaway, H., Zhong, J., & Yedvobnick, B. (2013). A targeted genetic modifier screen links the SWI2/SNF2 protein domino to growth and autophagy genes in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics*, *3*(5), 815-825.

Lahm, H. W., & Langen, H. (2000). Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis: An International Journal*, *21*(11), 2105-2114.

Lane, C. S. (2005). Mass spectrometry-based proteomics in the life sciences. *Cellular and Molecular Life Sciences CMLS*, *62*(7-8), 848-869.

Larance, M., Bailly, A. P., Pourkarimi, E., Hay, R. T., Buchanan, G., Coulthurst, S., ... & Lamond, A. I. (2011). Stable-isotope labeling with amino acids in nematodes. *Nature Methods*, *8*(10), 849-851.

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., & Jørgensen, T. J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics*, *4*(7), 873-886.

Lau, H.-T., Suh, H. W., Golkowski, M., & Ong, S.-E. (2014). Comparing SILAC-and stable isotope dimethyl-labeling approaches for quantitative proteomics. *Journal of Proteome Research*, *13*(9), 4164-4174.

Law, K. P., & Lim, Y. P. (2013). Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Review of Proteomics*, *10*(6), 551-566.

Lee, C. Y., Wilkinson, B. D., Siegrist, S. E., Wharton, R. P., & Doe, C. Q. (2006). Brat is a Miranda cargo protein that promotes neuronal differentiation and inhibits neuroblast self-renewal. *Developmental Cell*, *10*(4), 441-449.

Li, H. (2008). Unveiling substrate RNA binding to H/ACA RNPs: one side fits all. *Current Opinion in Structural Biology*, *18*(1), 78-85.

Li, Y., Maines, J. Z., Tastan, Ö. Y., McKearin, D. M., & Buszczak, M. (2012). Mei-P26 regulates the maintenance of ovarian germline stem cells by promoting BMP signaling. *Development*, *139*, 1547-1556.

Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H. E., Wolters, D., & Marcus, K. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biological Chemistry*, *398*(5-6), 687-699.

Liu, R., Ye, Y., Qiang, L., Liao, X., & Zhao, Y. (2008). The fragmentation pathway of the nucleosides under the electrospray ionization multi-stage mass spectrometry. *Life Scences Journal*, *5*, 37-40.

Loedige, I., Jakob, L., Treiber, T., Ray, D., Stotz, M., Treiber, N., ... & Engelmann, J. C. (2015). The crystal structure of the NHL domain in complex with RNA reveals the molecular basis of Drosophila brain-tumor-mediated gene regulation. *Cell Reports*, *13*(6), 1206-1220.

Loedige, I., Stotz, M., Qamar, S., Kramer, K., Hennig, J., Schubert, T., ... & Meister, G. (2014). The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes & Development*, *28*(7), 749-764.

Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews*, *16*(1), 1-23.

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology*, *8*(6), 479-490.

Luo, X., Hsiao, H. H., Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., ... & Wahl, M. C. (2008). Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex. *Molecular Cell*, *32*(6), 791-802.

Madeira, P. J. A., & Florêncio, M. H. (2012). Applications of Tandem Mass Spectrometry: From Structural Analysis to Fundamental Studies. In *Tandem Mass Spectrometry-Applications and Principles*. InTechOpen.

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., & Horning, S. (2006a). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, *78*(7), 2113-2120.

Makarov, A., Denisov, E., Lange, O., & Horning, S. (2006b). Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *17*(7), 977-982.

Mamyrin, B. A. (1994). Laser assisted reflectron time-of-flight mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes*, *131*, 1-19.

Mamyrin, B. A. (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, *206*(3), 251-266.

March, R. E. (2000). Quadrupole ion trap mass spectrometry: a view at the turn of the century. *International Journal of Mass Spectrometry*, *200*(1-3), 285-312.

March, R. E., & Todd, J. F. (2005). Quadrupole ion trap mass spectrometry, *165*, John Wiley & Sons: West Sussex, UK.

Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS Journal*, *272*(9), 2118-2131.

Marshall, A. G., & Chen, T. (2015). 40 years of Fourier transform ion cyclotron resonance mass spectrometry. *International Journal of Mass Spectrometry*, *377*, 410-420.

Marshall, A. G., Hendrickson, C. L., & Jackson, G. S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Review, 17*(1), 1-35.

McGrail, J. C., Krause, A., & O'Keefe, R. T. (2009). The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic acids Research*, *37*(13), 4205-4217.

McHugh, C. A., Russell, P., & Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biology*, *15*(1), 203.

McLuckey, S. A., Van Berker, G. J., & Glish, G. L. (1992). Tandem mass spectrometry of small, multiply charged oligonucleotides. *Journal of the American Society for Mass Spectrometry*, *3*(1), 60-70.

Meisenheimer, K. M., & Koch, T. H. (1997). Photocross-linking of nucleic acids to associated proteins. *Critical Reviews in Biochemistry and Molecular Biology*, *32*(2), 101-140.

Meng, Z., & Limbach, P. A. (2005). Quantitation of ribonucleic acids using 18O labeling and mass spectrometry. *Analytical Chemistry*, *77*(6), 1891-1895.

Merril, C. R., Goldman, D., & Van Keuren, M. L. (1983). Silver staining methods for polyacrylamide gel electrophoresis. *Methods in Enzymology, 96,* 230-239.

Merrill, A. E., Hebert, A. S., MacGilvray, M. E., Rose, C. M., Bailey, D. J., Bradley, J. C., ... & Coon, J. J. (2014). NeuCode labels for relative protein quantification. *Molecular & Cellular Proteomics*, *13*(9), 2503-2512.

Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., ... & Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, *10*(9), M111.011015.

Mitchell, S. F., & Parker, R. (2014). Principles and properties of eukaryotic mRNPs. *Molecular Cell*, *54*(4), 547-558.

Mitchell, S. F., Jain, S., She, M., & Parker, R. (2013). Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*, *20*(1), 127-133.

Mozaffari-Jovin, S., Santos, K. F., Hsiao, H. H., Will, C. L., Urlaub, H., Wahl, M. C., & Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes & Development*, *26*(21), 2422-2434.

Müller, M., Heym, R. G., Mayer, A., Kramer, K., Schmid, M., Cramer, P., ... & Niessing, D. (2011). A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. *PLoS Biology*, *9*(4), e1000611.

Munschauer, M. (2015). High-Resolution Profiling of Protein-RNA Interactions. Springer Theses.

Murata, Y., & Wharton, R. P. (1995). Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. *Cell*, *80*(5), 747-756.

Nakamura, R. L., Landt, S. G., Mai, E., Nejim, J., Chen, L., & Frankel, A. D. (2012). A cell-based method for screening RNA-protein interactions: identification of constitutive transport element-interacting proteins. *PLoS One*, *7*(10), e48194.

Nelson, C. C., & McCloskey, J. A. (1994). Collision-induced dissociation of uracil and its derivatives. *Journal of the American Society for Mass Spectrometry*, *5*(5), 339-349.

Nesvizhskii, A. I., Vitek, O., & Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*, *4*(10), 787-797.

Neuhoff, V., Arold, N., Taube, D., & Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, *9*(6), 255-262.

Neumüller, R. A., Betschinger, J., Fischer, A., Bushati, N., Poernbacher, I., Mechtler, K., ... & Knoblich, J. A. (2008). Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage. *Nature*, *454*(7201), 241-245.

Newton, R. P., Brenton, A. G., Smith, C. J., & Dudley, E. (2004). Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, *65*(11), 1449-1485.

Newton, R. P., Kingston, E. E., Hakeem, N. A., Salih, S. G., Beynon, J. H., & Moyse, C. D. (1986). Extraction, purification, identification and metabolism of 3′, 5′-cyclic UMP, 3′, 5′-cyclic IMP and 3′, 5′-cyclic dTMP from rat tissues. *Biochemical Journal*, *236*(2), 431-439.

Nguyen, S., & Fenn, J. B. (2007). Gas-phase ions of solute species from charged droplets of solutions. *Proceedings of the National Academy of Sciences*, *104*(4), 1111-1117.

Niessen, W. M., & Falck, D. (2015). Introduction to mass spectrometry, a tutorial. In *Analyzing Biomolecular Interactions by Mass Spectrometry*, 1-54.

Nikolov, M., Schmidt, C., & Urlaub, H. (2012). Quantitative mass spectrometry-based proteomics: an overview. In *Quantitative Methods in Proteomics*, (pp. 85-100). Humana Press, NJ, USA.

Oda, Y., Huang, K., Cross, F. R., Cowburn, D., & Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences*, *96*(12), 6591-6596.

Ohi, M. D., & Gould, K. L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA*, *8*(6), 798-815.

Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., ... & Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics*, *4*(12), 2010-2021.

Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., ... & Wouters, E. R. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Molecular & Cellular Proteomics*, *8*(12), 2759-2769.

Ong, S. E. (2012). The expanding field of SILAC. *Analytical and Bioanalytical Chemistry*, *404*(4), 967-976.

Ong, S. E., & Mann, M. (2005). Mass spectrometry–based proteomics turns quantitative. *Nature Chemical Biology*, *1*(5), 252-262.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, *1*(5), 376-386.

Paulines, M. J., & Limbach, P. A. (2017). Stable isotope labeling for improved comparative analysis of RNA digests by mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *28*(3), 551-561.

Phan, J., Yamout, N., Schmidberger, J., Bottomley, S. P., & Buckle, A. M. (2011). Refolding your protein with a little help from REFOLD. In *Protein Folding, Misfolding, and Disease*, *Methods in Molecular Biology*, *752*, 45-57.

Platner, C. B. (2013). Mass Spectrometry-based Proteomics Experiments and Peptide Fragmentation Studies of Lysine and its Homologues. Under graduate Honors Theses. Paper 628. Available from: https://scholarworks.wm.edu/honorstheses/628

Popova, A. M., & Williamson, J. R. (2014). Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. *Journal of the American Chemical Society*, *136*(5), 2058-2069.

Pourshahian, S., & Limbach, P. A. (2008). Application of fractional mass for the identification of peptide–oligonucleotide cross-links by mass spectrometry. *Journal of Mass Spectrometry*, *43*(8), 1081-1088.

Qamar, S., Kramer, K., & Urlaub, H. (2015). Studying RNA–Protein Interactions of Pre-mRNA Complexes by Mass Spectrometry. In *Methods in Enzymology*, *558*, 417-463.

Quiocho, F. A., Spurlino, J. C., & Rodseth, L. E. (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure*, *5*(8), 997-1015.

Rasche, N., Dybkov, O., Schmitzova, J., Akyildiz, B., Fabrizio, P., & Lührmann, R. (2012). Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *The EMBO Journal*, *31*(6), 1591-1604.

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., ... & Na, H. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, *499*(7457), 172-177.

Rice, J. M., & Dudek, G. O. (1969). Mass spectra of uridine and pseudouridine: fragmentation patterns characteristic of a carbon-carbon nucleosidic bond. *Biochemical and Biophysical Research Communications*, *35*(3), 383-388.

Richter, F. M., Hsiao, H. H., Plessmann, U., & Urlaub, H. (2009). Enrichment of protein–RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns. *Biopolymers: Original Research on Biomolecules*, *91*(4), 297-309.

Robinson, C. V., Sali, A., & Baumeister, W. (2007). The molecular sociology of the cell. *Nature*, *450*(7172), 973-982.

Roepstorff, P. (1984). Proposal for a nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, *11*(11), 601.

Rybak, A., Fuchs, H., Hadian, K., Smirnova, L., Wulczyn, E. A., Michel, G., ... & Wulczyn, F. G. (2009). The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nature Cell Biology*, *11*(12), 1411-1420.

Sali, A., Glaeser, R., Earnest, T., & Baumeister, W. (2003). From words to literature in structural proteomics. *Nature*, *422*(6928), 216-225.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, USA.

Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A., & Meroni, G. (2008). Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC Evolutionary Biology*, *8*(1), 225.

Scherer, S., Altwegg, K., Balsiger, H., Fischer, J., Jäckel, A., Korth, A., ... & Wurz, P. (2006). A novel principle for an ion mirror design in time-of-flight mass spectrometry. *International Journal of Mass Spectrometry*, *251*(1), 73-81.

Schmidt, A., Forne, I., & Imhof, A. (2014). Bioinformatic analysis of proteomics data. *BMC Systems Biology*, *8*(2), S3.

Schmidt, C., & Urlaub, H. (2009). iTRAQ-labeling of in-gel digested proteins for relative quantification. In *Proteomics*, *Methods in Molecular Biology, 564,* 207-226.

Schmidt, C., Kramer, K., & Urlaub, H. (2012). Investigation of protein–RNA interactions by mass spectrometry—Techniques and applications. *Journal of Proteomics*, *75*(12), 3478-3494.

Schmitzova, J., Rasche, N., Dybkov, O., Kramer, K., Fabrizio, P., Urlaub, H., ... & Pena, V. (2012). Crystal structure of Cwc2 reveals a novel architecture of a multipartite RNA-binding protein. *The EMBO Journal*, *31*(9), 2222-2234.

Schueler, M., Munschauer, M., Gregersen, L. H., Finzel, A., Loewer, A., Chen, W., ... & Dieterich, C. (2014). Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biology*, *15*(1), R15.

Schulze, W. X., & Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. *Annual Review of Plant Biology*, *61*, 491-516.

Schwamborn, J. C., Berezikov, E., & Knoblich, J. A. (2009). The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors. *Cell*, *136*(5), 913-925.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., ... & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337-342.

Scigelova, M., & Makarov, A. (2006). Orbitrap mass analyzer–overview and applications in proteomics. *Proteomics*, *6*(S2), 16-21.

Seidler, J., Zinn, N., Boehm, M. E., & Lehmann, W. D. (2010). *De novo* sequencing of peptides by MS/MS. *Proteomics*, *10*(4), 634-649.

SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., & Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *Proceedings of the National Academy of Sciences*, *93*(16), 8496-8501.

Sharff, A. J., Rodseth, L. E., Spurlino, J. C., & Quiocho, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry*, *31*(44), 10657-10663.

Shetlar, M. D., Carbone, J., Steady, E., & Hom, K. (1984b). Photochemical addition of amino acids and peptides to polyuridylic acid. *Photochemistry and Photobiology*, *39*(2), 141-144.

Shetlar, M. D., Hom, K., Carbone, J., Moy, D., Steady, E., & Watanabe, M. (1984a). Photochemical addition of amino acids and peptides to homopolyribonucleotides of the major DNA bases. *Photochemistry and Photobiology*, *39*(2), 135-140.

Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, *1*(6), 2856-2860.

Sjöberg, A., Önnerfjord, P., Mörgelin, M., Heinegård, D., & Blom, A. M. (2005). The Extracellular Matrix and Inflammation fibromodulin activates the classical pathway of complement by directly binding C1q. *Journal of Biological Chemistry*, *280*(37), 32301-32308.

Slack, F. J., & Ruvkun, G. (1998). A novel repeat domain that is often associated with RING finger and B-box motifs. *Trends in Biochemical Sciences*, *23*(12), 474-475.

Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, *136*(4), 731-745.

Song, Y., Wu, G., Song, Q., Cooks, R. G., Ouyang, Z., & Plass, W. R. (2006). Novel linear ion trap mass analyzer composed of four planar electrodes. *Journal of the American Society for Mass Spectrometry*, *17*(4), 631-639.

Sonoda, J., & Wharton, R. P. (2001). Drosophila Brain Tumor is a translational repressor. *Genes & Development*, *15*(6), 762-773.

Soufi, B., Kumar, C., Gnad, F., Mann, M., Mijakovic, I., & Macek, B. (2010). Stable isotope labeling by amino acids in cell culture (SILAC) applied to quantitative proteomics of Bacillus subtilis. *Journal of Proteome Research*, *9*(7), 3638-3646.

Steen, H., & Jensen, O. N. (2002). Analysis of protein–nucleic acid interactions by photochemical cross-linking and mass spectrometry. *Mass Spectrometry Reviews*, *21*(3), 163-182.

Stirnimann, C. U., Petsalaki, E., Russell, R. B., & Müller, C. W. (2010). WD40 proteins propel cellular networks. *Trends in Biochemical Sciences*, *35*(10), 565-574.

Struhl, G. (1989). Differing strategies for organizing anterior and posterior body pattern in Drosophila embryos. *Nature*, *338*(6218), 741-744.

Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E., ... & Kohlbacher, O. (2008). OpenMS–an open-source software framework for mass spectrometry. *BMC Bioinformatics*, *9*(1), 163.

Sury, M. D., Chen, J. X. X., & Selbach, M. (2010). The SILAC fly allows for accurate protein quantification in vivo. *Molecular & Cellular Proteomics*, *9*(10), 2173-2183.

Svergun, D. I., Barberato, C., Koch, M. H., Fetler, L., & Vachette, P. (1997). Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins: Structure, Function, and Bioinformatics*, *27*(1), 110-117.

Tabert, A. M., Goodwin, M. P., & Cooks, R. G. (2006). Co-occurrence of boundary and resonance ejection in a multiplexed rectilinear ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, *17*(1), 56-59.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., & Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *2*(8), 151-153.

Tenenbaum, S. A., Carson, C. C., Lager, P. J., & Keene, J. D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proceedings of the National Academy of Sciences*, *97*(26), 14085-14090.

Thomson, B. A., & Iribarne, J. V. (1979). Field induced ion evaporation from liquid surfaces at atmospheric pressure. *The Journal of Chemical Physics*, *71*(11), 4451-4463.

Thomson, J. J. (1921). Rays of Positive Electricity and their Application to Chemical Analyses. Longmans, Green and Company, London, UK.

Tretyakova, N. Y., Groehler IV, A., & Ji, S. (2015). DNA–protein cross-links: formation, structural identities, and biological outcomes. *Accounts of Chemical Research*, *48*(6), 1631-1644.

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., & Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science*, *302*(5648), 1212-1215.

Urban, P. L. (2016). Quantitative mass spectrometry: an overview. *Philosophical Transactions Royal Society, 374*, 20150382.

Urlaub, H., Hartmuth, K., & Lührmann, R. (2002). A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*, *26*(2), 170-181.

Urlaub, H., Kruft, V., Bischof, O., Müller, E. C., & Wittmann-Liebold, B. (1995). Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *The EMBO Journal*, *14*(18), 4578-4588.

Urlaub, H., Kühn-Hölsken, E., & Lührmann, R. (2008). Analyzing RNA-protein crosslinking sites in unlabeled ribonucleoprotein complexes by mass spectrometry. In *RNA-Protein Interaction Protocols*, *Methods in Molecular Biology*, *488*, 221-245.

Urlaub, H., Raker, V. A., Kostka, S., & Lührmann, R. (2001). Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *The EMBO Journal*, *20*(1-2), 187-196.

Urlaub, H., Thiede, B., Müller, E. C., Brimacombe, R., & Wittmann-Liebold, B. (1997). Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *Journal of Biological Chemistry*, *272*(23), 14547-14555.

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., ... & Stanton, R. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*, *13*(6), 508-514.

Vander Kooi, C. W., Ren, L., Xu, P., Ohi, M. D., Gould, K. L., & Chazin, W. J. (2010). The Prp19 WD40 domain contains a conserved protein interaction region essential for its function. *Structure*, *18*(5), 584-593.

Villa, T., & Guthrie, C. (2005). The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes & Development*, *19*(16), 1894-1904.

Waghmare, S. P., & Dickman, M. J. (2011). Characterization and quantification of RNA post-transcriptional modifications using stable isotope labeling of RNA in conjunction with mass spectrometry analysis. *Analytical Chemistry*, *83*(12), 4894-4901.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. C*ell*, *136*(4), 701-718.

Walden, W. E., Selezneva, A. I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J. C., Theil, E. C., & Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science*, *314*(5807), 1903-1908.

Walzthoeni, T., Leitner, A., Stengel, F., & Aebersold, R. (2013). Mass spectrometry supported determination of protein complex structure. *Current Opinion in Structural Biology*, *23*(2), 252-260.

Wang, C., & Lehmann, R. (1991). Nanos is the localized posterior determinant in Drosophila. *Cell*, *66*(4), 637-647.

Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., & Yates, J. R. (2002). Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Analytical Chemistry*, *74*(7), 1650-1657.

Watson, J. T., & Sparkman, O. D. (2007). Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation. John Wiley & Sons: West Sussex, UK.

Wharton, R. P., & Struhl, G. (1991). RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. *Cell*, *67*(5), 955-967.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., ... & Williams, K. L. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology*, *14*(1), 61-65.

Wilm, M. (2009). Quantitative proteomics in biological research. *Proteomics*, *9*(20), 4590-4605.

Wower, I., Wower, J., Meinke, M., & Brimacombe, R. (1981). The use of 2-iminothiolane as an RNA-protein cross-linking agent in *Escherichia coli* ribosomes, and the localisation on 23S RNA of sites cross-linked to proteins L4, L6, L21, L23, L27 and L29. *Nucleic Acids Research*, *9*(17), 4285-4302.

Wu, Y., Wang, F., Liu, Z., Qin, H., Song, C., Huang, J., ... & Zou, H. (2014). Five-plex isotope dimethyl labeling for quantitative proteomics. *Chemical Communications*, *50*(14), 1708-1710.

Wulczyn, F. G., Cuevas, E., Franzoni, E., & Rybak, A. (2010). miRNAs Need a Trim. In *Regulation of microRNAs*, *Advances in Experimental Medicine and Biology*, *700*, 85-105.

Yamashita, M., & Fenn, J. B. (1984). Negative ion production with the electrospray ion source. *The Journal of Physical Chemistry*, *88*(20), 4671-4675.

Yan, W., & Chen, S. S. (2005). Mass spectrometry-based quantitative proteomic profiling. *Briefings in Functional Genomics*, *4*(1), 27-38.

Zaman, U., Richter, F. M., Hofele, R., Kramer, K., Sachsenberg, T., Kohlbacher, O., ... & Urlaub, H. (2015). Dithiothreitol (DTT) acts as a specific, UV-inducible cross-linker in elucidation of protein-RNA interactions. *Molecular & Cellular Proteomics*, *14*(12), 3196-3210.

Zhang, G., Fenyö, D., & Neubert, T. A. (2009). Evaluation of the variation in sample preparation for comparative proteomics using stable isotope labeling by amino acids in cell culture. *Journal of Proteome Research*, *8*(3), 1285-1292.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates III, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, *113*(4), 2343-2394.

Zhou, Z., & Reed, R. (2003). Purification of Functional RNA-Protein Complexes using MS 2-MBP. *Current Protocols in Molecular Biology*, *63*(1), 27.23. 21-27.23. 27.

Zubarev, R. A., & Makarov, A. (2013). Orbit rap mass spectrometry. *Analytical Chemistry*, *85*(11), 5288-5296.

6 APPENDIX

ANNOTATION OF MS/MS SPECTRA

The MS/MS spectra of cross-linked heteroconjugates were annotated according to the established nomenclature for peptide fragments described previously (Roepstorff & Fohlman, 1984; Biemann, 1990). In addition, neutral loss of water from both peptide and RNA fragments was annotated with a superscript "0", the immonium ion with "IM" and the mass increment of 151.9966 Da due to DTT ($C_4H_8S_2O_2$) (Zaman et al., 2015) with "152". Annotations of RNA adducts and RNA marker ions are given in the tables 6.1 and 6.2 respectively together with calculated masses and *m/z* values.

Table 6.1: List of RNA (fragment) adducts with their assigned abbreviations and calculated masses.

RNA (fragment) adducts	Abbreviation in annotated spectra	Calculated mass (Da.)
C_3O	#	51.9949
Uracil-H_2O	U'^0	94.0167
Uracil	U'	112.0273
Uridin-H_2O	U^{0-p}	226.0590
Uridin	U^p	244.0695
Uridinmonophosphate-H_2O	U^0	306.0253
Uridinmonophosphate	U	324.0359

Table 6.2: List of RNA marker ions with their symbols and calculated *m/z*.

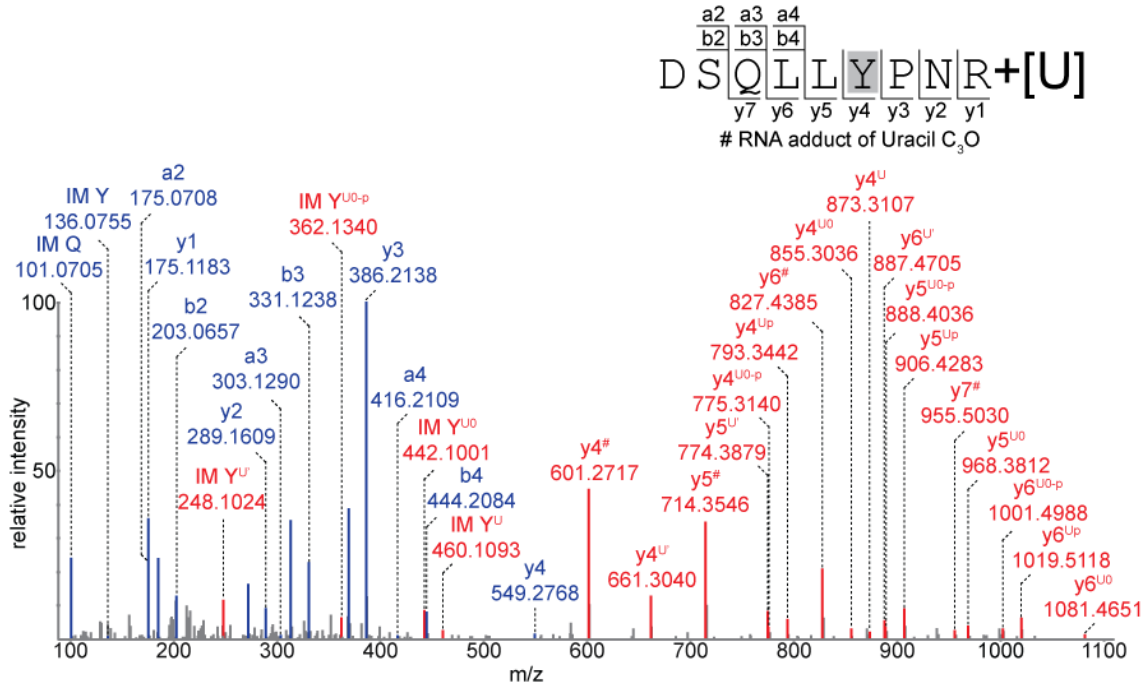
RNA marker ions	Symbol	Calculated *m/z*
Adenin	A'	136.0623
Adenosinmonophosphate-H_2O	A^0	330.0603
Adenosinmonophosphate	A	348.0709
Cytosin	C'	112.0511
Cytidinmonophosphate-H_2O	C^0	306.0491
Cytidinmonophosphate	C	324.0597
Guanin	G'	152.0572
Guanosinmonophosphate-H_2O	G^0	346.0553
Guanosinmonophosphate	G	364.0658
Uracil	U'	113.0351
Uridine-H_2O	U^{0-p}	227.0667
Uridinmonophosphate-H_2O	U^0	307.0331
Uridinmonophosphate	U	325.0437

Table 6.3: RNA-protein cross-links identified from Brat-NHL-hb RNA complex followed by corresponding MS/MS spectra.

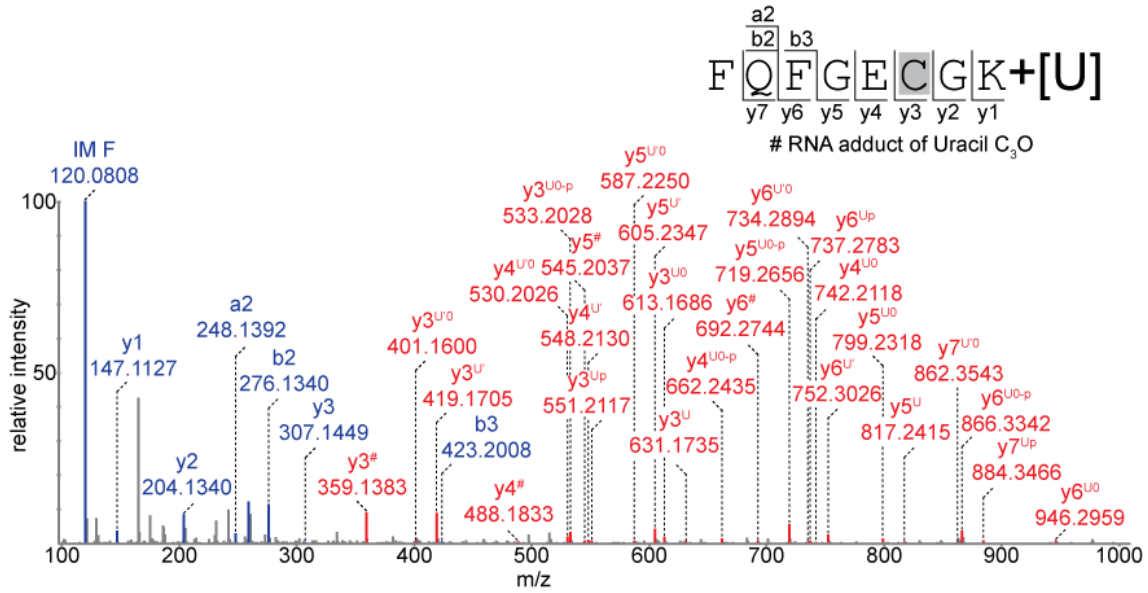
Surface/Blade	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
top/II	^{824}DSQLLYPNR^{832}	Y^{829}	U	1104.5563	324.0359	1428.5922	2	715.3039	715.3053	1.96
top/II	^{815}FQFGECGK^{822}	C^{820}	U	914.3956	324.0359	1238.4315	2	620.2236	620.2236	0
bottom/II	^{804}IQIFDKEGR^{812}	K^{809}	U-H_2O	1104.5927	306.0253	1410.618	2	706.3168	706.3162	-0.85
top/III	^{865}KFGATILQHPR^{875}	^{865}K/F^{866}	U-H_2O	1266.7196	306.0253	1572.7449	3	525.2561	525.2550	-2.10
top/IV	^{913}HLEFPNGVVVNDK^{925}	F^{916}	U	1466.7517	324.0359	1790.7874	2	896.4015	896.3995	-2.23
top/IV	^{885}IIVVECK^{891}	C^{890}	U	802.4622	324.0359	1126.4981	2	564.2569	564.2564	-0.89

Surface/Blade: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotide found to be cross-linked to the peptide; m(calc) Peptide: Theoretical mass of the cross-linked peptide; m(calc) RNA: Theoretical mass of the cross-linked RNA; m(calc) Cross-link: Calculated mass of the RNA-protein cross-link; z: Charge state of the cross-link; *m/z*(calc): Calculated *m/z* of the cross-link by using formula m+z(mH)/z; *m/z*(exp): Observed *m/z* of the cross-link; Δm (ppm): Calculated mass error in ppm by using formula $[m/z(\text{exp}) - m/z(\text{calc})/m/z(\text{calc})] \times 10^6$.

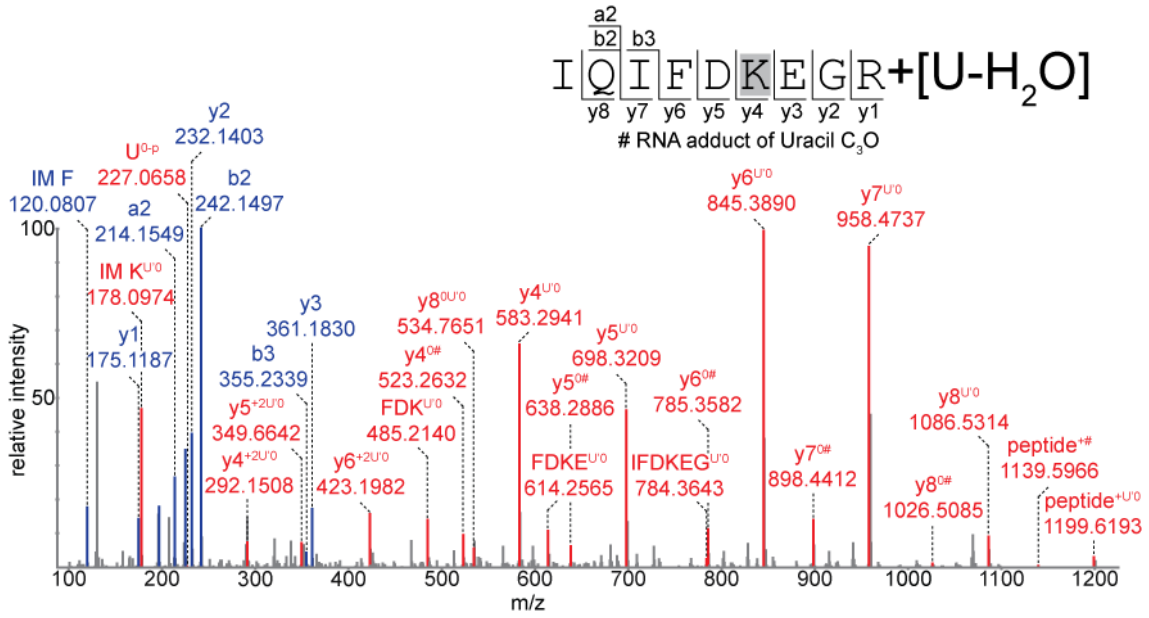
A.6.1


a2 a3 a4
b2 b3 b4
D S Q L L Y P N R+[U]
y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
IM Y 136.0755
IM Q 101.0705
a2 175.0708
y1 175.1183
b2 203.0657
a3 303.1290
y2 289.1609
IM Y' 248.1024
IM Y U0-p 362.1340
b3 331.1238
y3 386.2138
a4 416.2109
IM Y U0 442.1001
b4 444.2084
IM Y U 460.1093
y4 549.2768
y4# 601.2717
y4U' 661.3040
y4U 873.3107
y4U0 855.3036
y6U' 887.4705
y6# 827.4385
y5U0-p 888.4036
y4Up 793.3442
y5Up 906.4283
y4U0-p 775.3140
y7# 955.5030
y5U' 774.3879
y5U0 968.3812
y5# 714.3546
y6U0-p 1001.4988
y6Up 1019.5118
y6U0 1081.4651
relative intensity
m/z

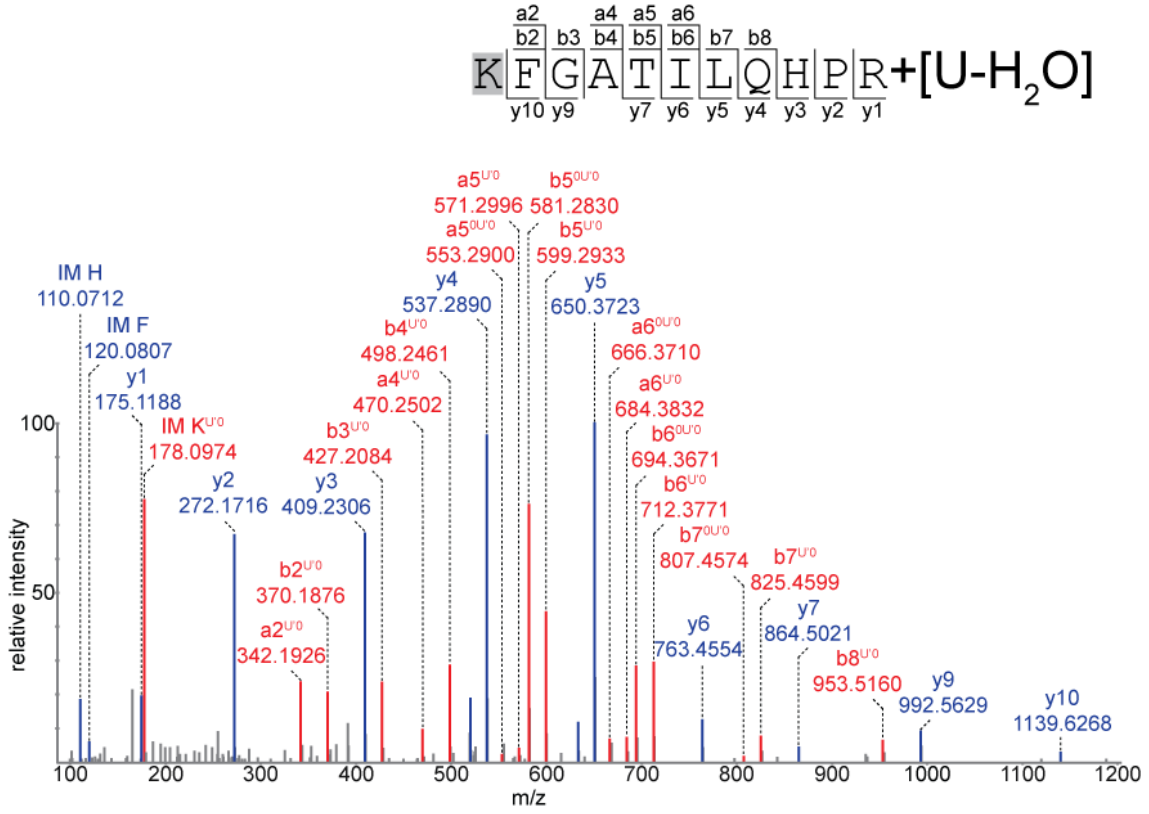
A.6.2


a2
b2 b3
F Q F G E C G K+[U]
y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
IM F 120.0808
y1 147.1127
a2 248.1392
b2 276.1340
y3 307.1449
y2 204.1340
y3# 359.1383
y3U'0 401.1600
y3U' 419.1705
b3 423.2008
y4# 488.1833
y3U0-p 533.2028
y4U'0 545.2037
530.2026
y4U' 548.2130
y3Up 551.2117
y5U'0 587.2250
y5U' 605.2347
y5# 613.1686
y3U0 631.1735
y3U 662.2435
y6U'0 734.2894
y6Up 737.2783
y5U0-p 719.2656
y4U0 742.2118
y6# 692.2744
y5U0 799.2318
y6U' 752.3026
y5U 817.2415
y7U'0 862.3543
y6U0-p 866.3342
y7Up 884.3466
y6U0 946.2959
relative intensity
m/z

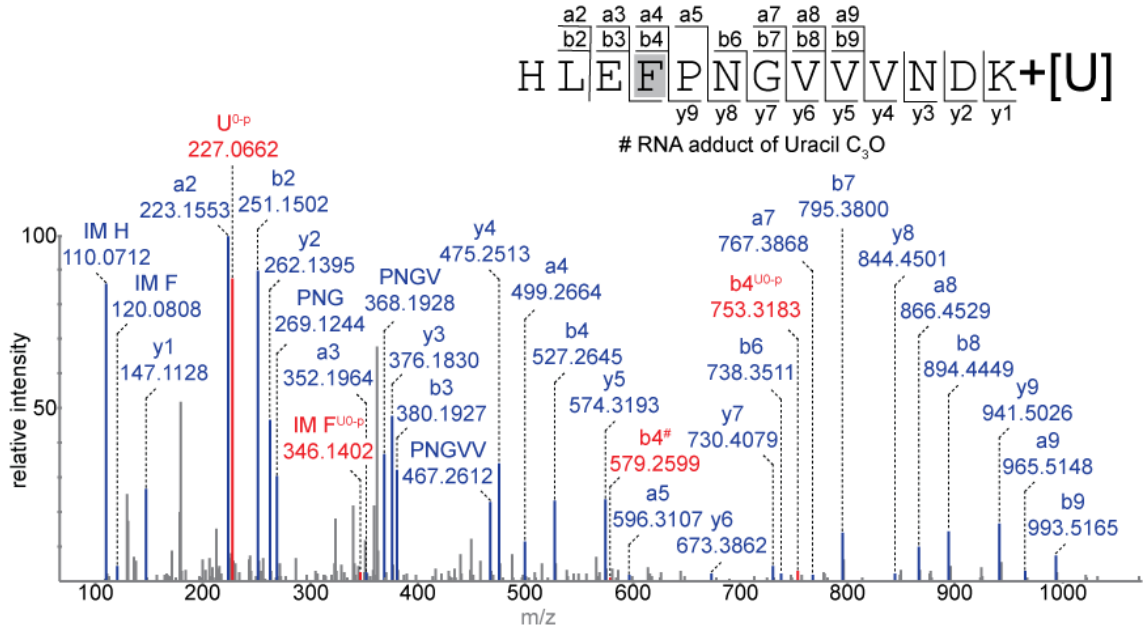
A.6.3


a2
b2 b3
I Q I F D K E G R+[U-H2O]
y8 y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
relative intensity
m/z
IM F 120.0807
U0-p 227.0658
y2 232.1403
a2 214.1549
b2 242.1497
IM KU'0 178.0974
y1 175.1187
b3 355.2339
y3 361.1830
y5+2U'0 349.6642
y4+2U'0 292.1508
y6+2U'0 423.1982
y8 0U'0 534.7651
y4 0# 523.2632
FDKU'0 485.2140
y4U'0 583.2941
y5 0# 638.2886
FDKEU'0 614.2565
y5U'0 698.3209
IFDKEGU'0 784.3643
y6 0# 785.3582
y6U'0 845.3890
y7 0# 898.4412
y7U'0 958.4737
y8 0# 1026.5085
y8U'0 1086.5314
peptide+# 1139.5966
peptide+U'0 1199.6193

A.6.4


a2 a4 a5 a6
b2 b3 b4 b5 b6 b7 b8
K F G A T I L Q H P R+[U-H2O]
y10 y9 y7 y6 y5 y4 y3 y2 y1
relative intensity
m/z
IM H 110.0712
IM F 120.0807
y1 175.1188
IM KU'0 178.0974
y2 272.1716
b3U'0 427.2084
y3 409.2306
b2U'0 370.1876
a2U'0 342.1926
a4U'0 470.2502
b4U'0 498.2461
y4 537.2890
a5 0U'0 553.2900
a5U'0 571.2996
b5 0U'0 581.2830
b5U'0 599.2933
y5 650.3723
a6 0U'0 666.3710
a6U'0 684.3832
b6 0U'0 694.3671
b6U'0 712.3771
b7 0U'0 807.4574
y6 763.4554
b7U'0 825.4599
y7 864.5021
b8U'0 953.5160
y9 992.5629
y10 1139.6268

A.6.5


a2 a3 a4 a5 a7 a8 a9
b2 b3 b4 b6 b7 b8 b9
H L E F P N G V V V N D K+[U]
y9 y8 y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
U0-p 227.0662
a2 223.1553
b2 251.1502
IM H 110.0712
IM F 120.0808
y1 147.1128
y2 262.1395
PNG 269.1244
PNGV 368.1928
y4 475.2513
a4 499.2664
a3 352.1964
y3 376.1830
b3 380.1927
IM F U0-p 346.1402
PNGVV 467.2612
b4 527.2645
y5 574.3193
b4# 579.2599
a5 596.3107
y6 673.3862
y7 730.4079
a7 767.3868
b7 795.3800
b4U0-p 753.3183
b6 738.3511
y8 844.4501
a8 866.4529
b8 894.4449
y9 941.5026
a9 965.5148
b9 993.5165
relative intensity
m/z

A.6.6

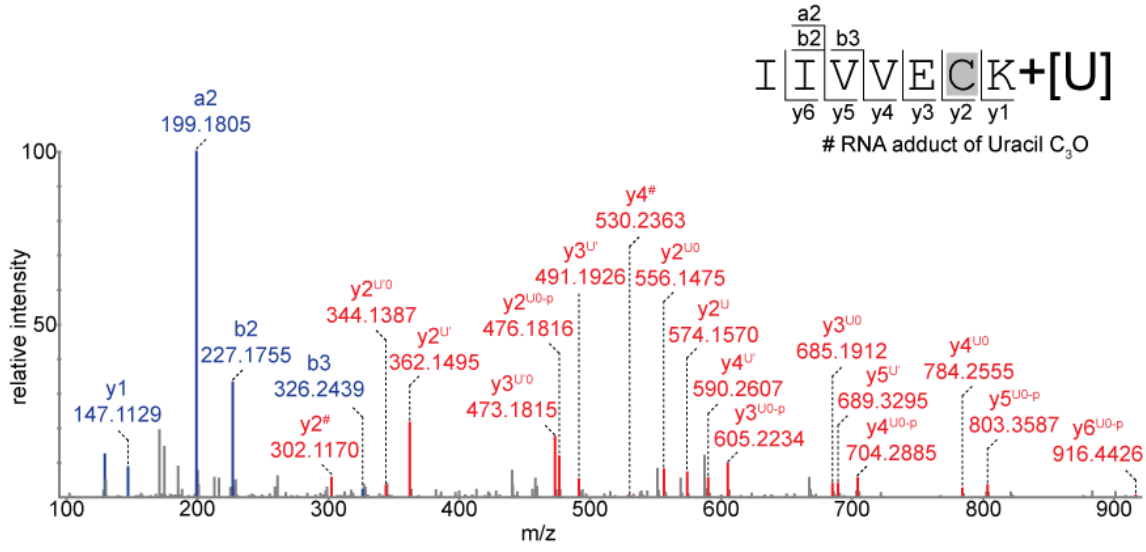
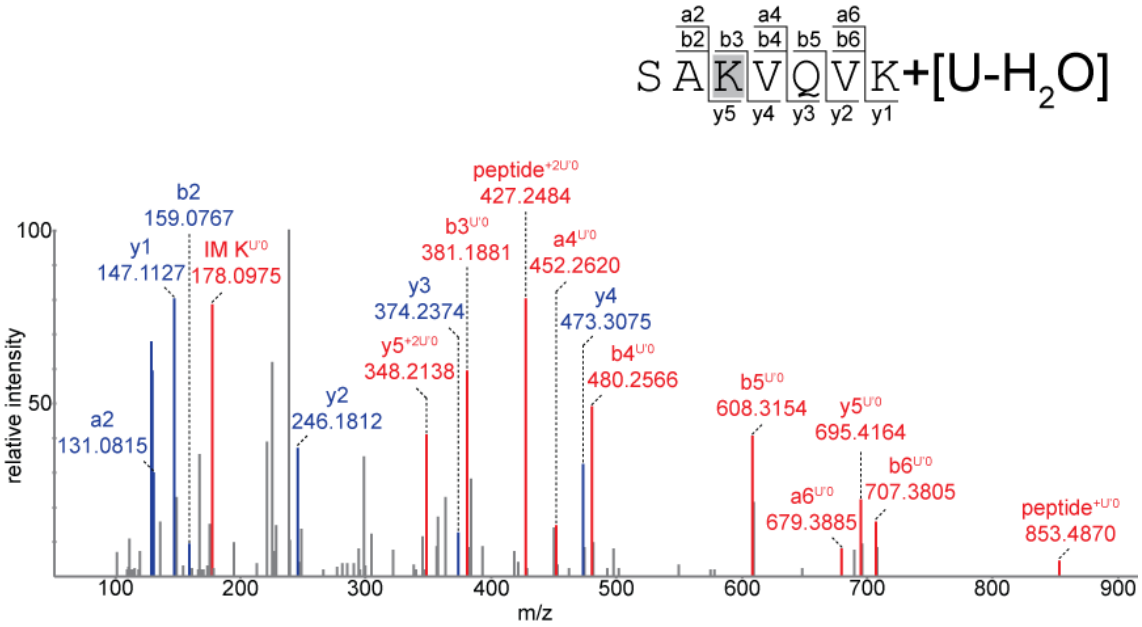

a2
b2 b3
I I V V E C K+[U]
y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
a2 199.1805
y1 147.1129
b2 227.1755
b3 326.2439
y2# 302.1170
y2U'0 344.1387
y2U' 362.1495
y3U'0 473.1815
y2U0-p 476.1816
y3U' 491.1926
y4# 530.2363
y2U0 556.1475
y2U 574.1570
y4U' 590.2607
y3U0-p 605.2234
y3U0 685.1912
y5U' 689.3295
y4U0-p 704.2885
y4U0 784.2555
y5U0-p 803.3587
y6U0-p 916.4426
relative intensity
m/z

Table 6.4: RNA-protein cross-links identified from CWC2-U4 snRNA and CWC2-U6 snRNA complexes followed by corresponding MS/MS spectra.

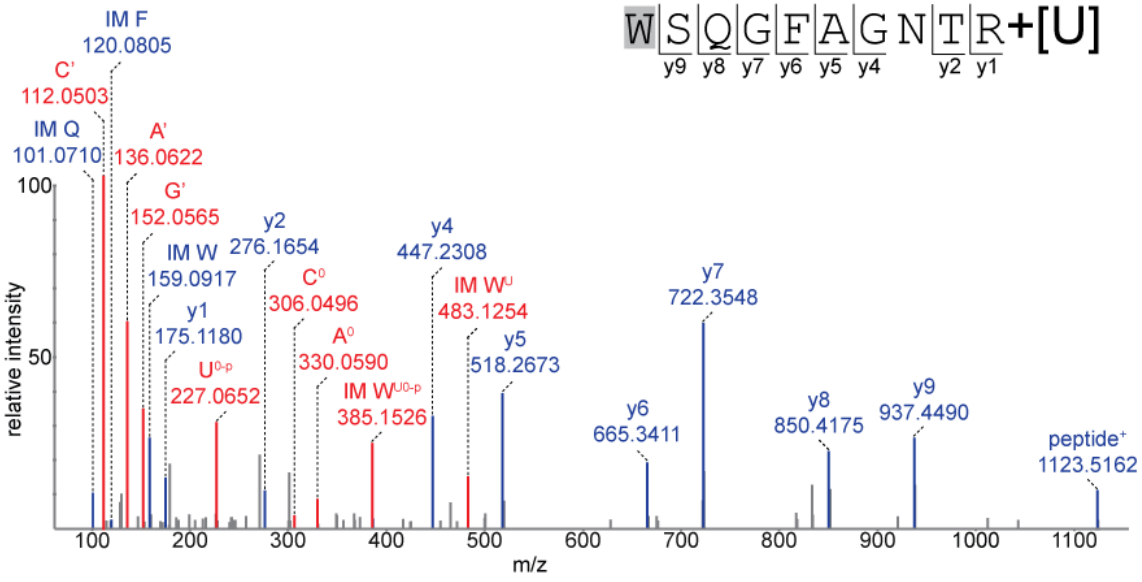
Domain	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Torus	^{8}SAKVQVK^{14}	K^{10}	U-H_2O	758.4650	306.0253	1064.4903	2	533.2530	533.2529	-0.19
Torus	^{37}WSQGFAGNTR^{46}	W^{37}	U	1122.5206	324.0359	1446.5565	2	724.2861	724.2900	5.38
Torus	^{106}TEVLDCFGR^{114}	C^{111}	U+152	1038.4804	476.0325	1514.5129	2	758.2643	758.2612	-4.09
ZnF	^{87}CEYLHHIPDEEDIGK^{101}	-	U+152	1796.8039	476.0325	2272.8364	3	758.6199	758.6200	0.13
connector element	^{115}EKFADYR^{121}	K^{116}	U-H_2O	927.4450	306.0253	1233.4703	2	617.7430	617.7452	3.56
connector element	^{117}FADYR^{121}	Y^{120}	UA	670.3074	653.088	1323.3954	2	662.7055	662.7051	-0.61
connector element	^{122}EDM(Oxidation)GGIGSFR^{131}	S^{129}	U	1083.4654	324.0359	1407.5013	2	704.7585	704.7581	-0.57
RRM/RNP2	^{136}TLYVGGIDGALNSK^{149}	Y^{138}	U	1406.7405	324.0359	1730.7764	2	866.3960	866.3917	-4.96
RRM	^{150}HLKPAQIESR^{159}	K^{152}	U-H_2O	1177.6567	306.0253	1483.6820	3	495.5685	495.5700	3.03
RRM	^{162}FVFSR^{166}	F^{162}	U-H_2O	654.3489	306.0253	960.3742	2	481.1949	481.1945	-0.83
RRM	^{167}LGDIDRIR^{174}	R^{172}	UA	956.5402	653.088	1609.6282	3	537.5505	537.5502	-0.56
RRM	^{188}YQANAEFAK^{196}	Y^{188}	U	1040.4926	324.0359	1364.5285	2	683.2721	683.2708	-1.91
RRM	^{217}EGTGLLVK^{224}	L^{222}	U-H_2O	815.4752	306.0253	1121.5005	2	561.7581	561.7579	-0.36
RRM	^{217}EGTGLLVKWANEDPDPAAQK^{236}	K^{224}	U-H_2O	2138.0643	306.0253	2444.0896	3	815.7043	815.7024	-2.33
RRM/RNP1	^{180}NCGFVK^{185}	C^{181}	U+152	666.3159	476.0325	1142.3484	2	572.1820	572.1807	-2.27
CTD	^{225}WANEDPDPAAQK^{236}	W^{225}	U	1340.5996	324.0359	1664.6355	2	833.3256	833.3300	5.28
CTD	^{276}TFPEASVDNVK^{286}	F^{277}	U	1205.5927	324.0359	1529.6286	2	765.8221	765.8209	-1.56
CTD	^{315}ENISSKPSVGK^{325}	K^{320}	U-H_2O	1144.6087	306.0253	1450.6340	2	726.3248	726.3245	-0.42

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence; RNA: Nucleotide found to be cross-linked to the peptide; m(calc) Peptide: Theoretical mass of the cross-linked peptide; m(calc) RNA: Theoretical mass of the cross-linked RNA; m(calc) Cross-link: Calculated mass of the RNA-protein cross-link; z: Charge state of the cross-link; *m/z*(calc): Calculated *m/z* of the cross-link by using formula m+z(mH)/z; *m/z*(exp): Observed *m/z* of the cross-link; Δm (ppm): Calculated mass error in ppm by using formula $[m/z(\text{exp}) - m/z(\text{calc})/m/z(\text{calc})] \times 10^6$.

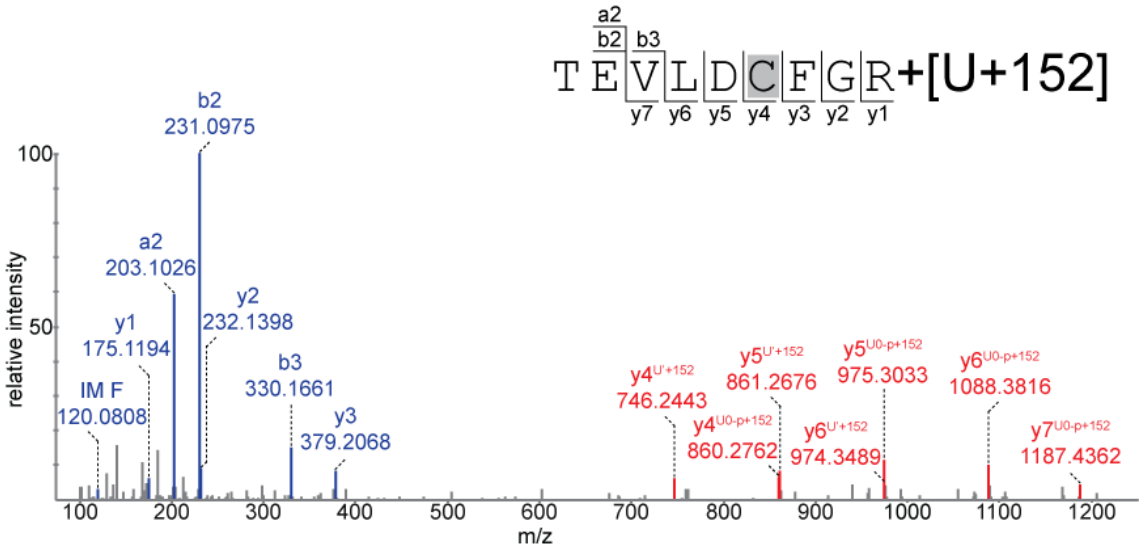
A.6.7


a2 b2 b3 a4 b4 b5 a6 b6
S A K V Q V K+[U-H2O]
y5 y4 y3 y2 y1
peptide+2U'0 427.2484
b2 159.0767
y1 147.1127
IM KU'0 178.0975
b3U'0 381.1881
a4U'0 452.2620
y3 374.2374
y4 473.3075
y5+2U'0 348.2138
b4U'0 480.2566
y2 246.1812
b5U'0 608.3154
y5U'0 695.4164
a2 131.0815
b6U'0 707.3805
a6U'0 679.3885
peptide+U'0 853.4870
relative intensity
m/z

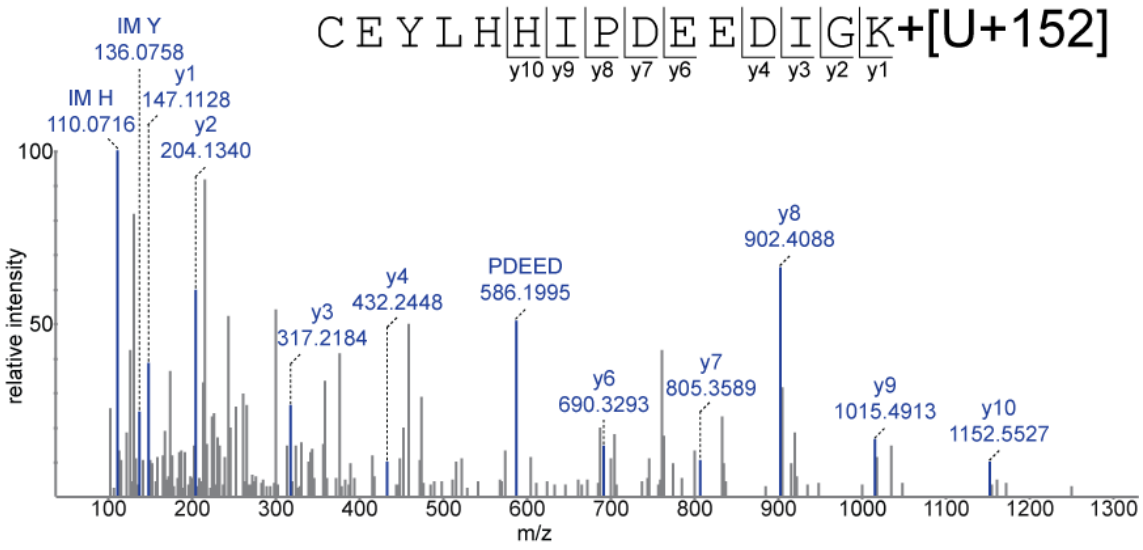
A.6.8


W S Q G F A G N T R+[U]
y9 y8 y7 y6 y5 y4 y2 y1
IM F 120.0805
C' 112.0503
IM Q 101.0710
A' 136.0622
G' 152.0565
IM W 159.0917
y2 276.1654
y4 447.2308
C0 306.0496
IM WU 483.1254
y1 175.1180
A0 330.0590
y5 518.2673
U0-p 227.0652
IM W U0-p 385.1526
y7 722.3548
y6 665.3411
y8 850.4175
y9 937.4490
peptide+ 1123.5162
relative intensity
m/z

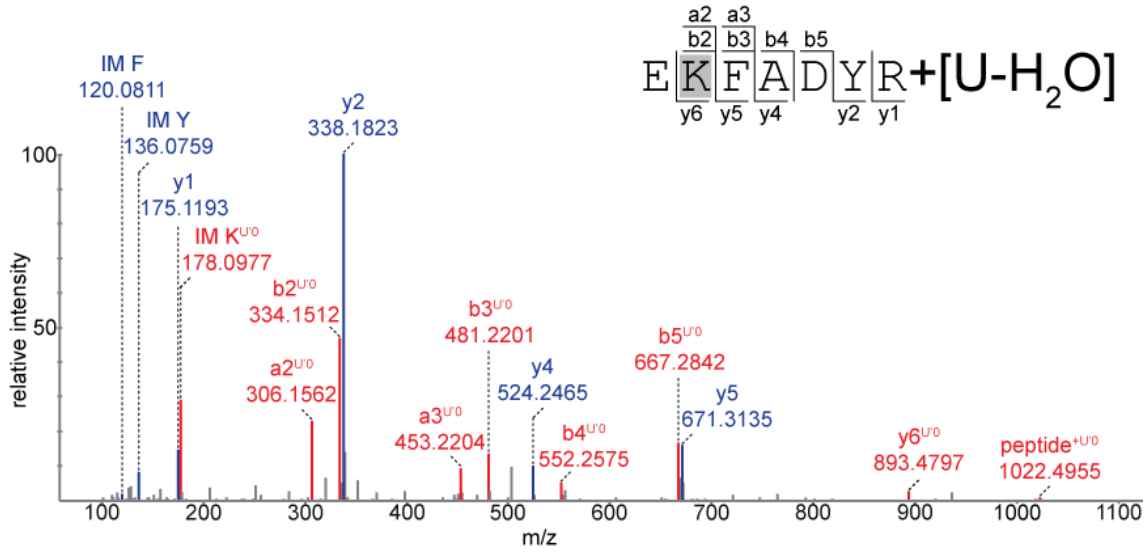
A.6.9


a2
b2 b3
T E V L D C F G R+[U+152]
y7 y6 y5 y4 y3 y2 y1
b2
231.0975
a2
203.1026
y1
175.1194
IM F
120.0808
y2
232.1398
b3
330.1661
y3
379.2068
y4U'+152
746.2443
y5U'+152
861.2676
y4U0-p+152
860.2762
y5U0-p+152
975.3033
y6U'+152
974.3489
y6U0-p+152
1088.3816
y7U0-p+152
1187.4362
relative intensity
m/z

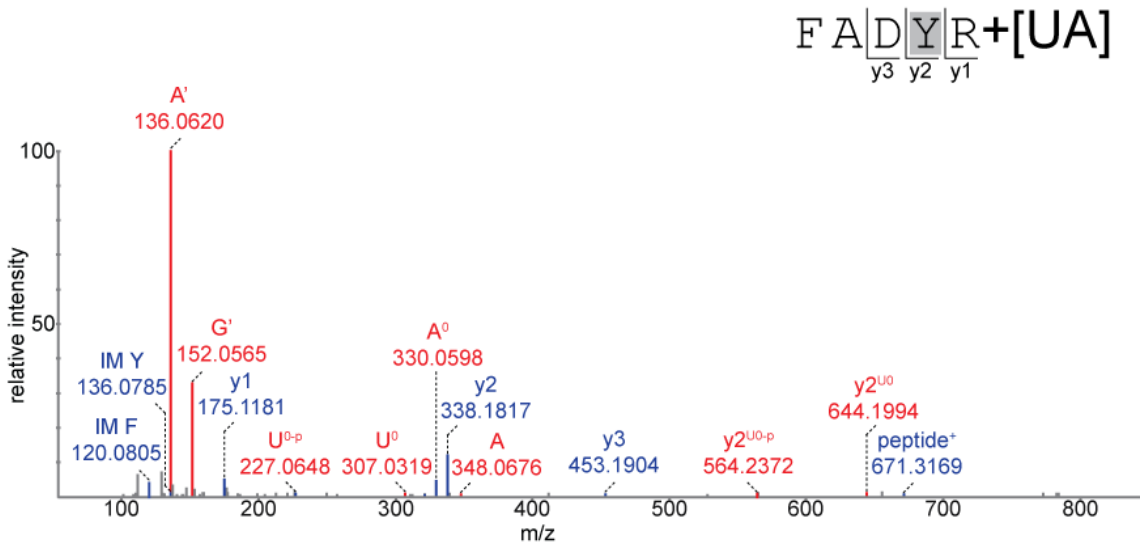
A.6.10


C E Y L H H I P D E E D I G K+[U+152]
y10 y9 y8 y7 y6 y4 y3 y2 y1
IM Y
136.0758
IM H
110.0716
y1
147.1128
y2
204.1340
y3
317.2184
y4
432.2448
PDEED
586.1995
y6
690.3293
y7
805.3589
y8
902.4088
y9
1015.4913
y10
1152.5527
relative intensity
m/z

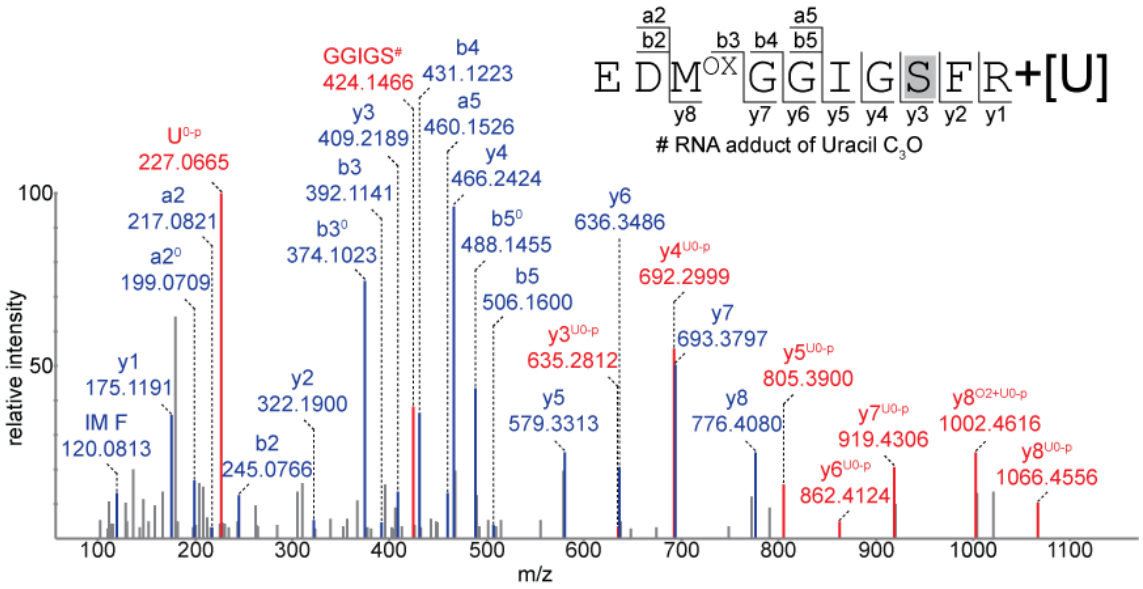
A.6.11


a2 a3
b2 b3 b4 b5
EKFADYR+[U-H2O]
y6 y5 y4 y2 y1
IM F
120.0811
IM Y
136.0759
y1
175.1193
IM K^U'0
178.0977
y2
338.1823
b2^U'0
334.1512
a2^U'0
306.1562
b3^U'0
481.2201
a3^U'0
453.2204
y4
524.2465
b4^U'0
552.2575
b5^U'0
667.2842
y5
671.3135
y6^U'0
893.4797
peptide+^U'0
1022.4955
relative intensity
m/z

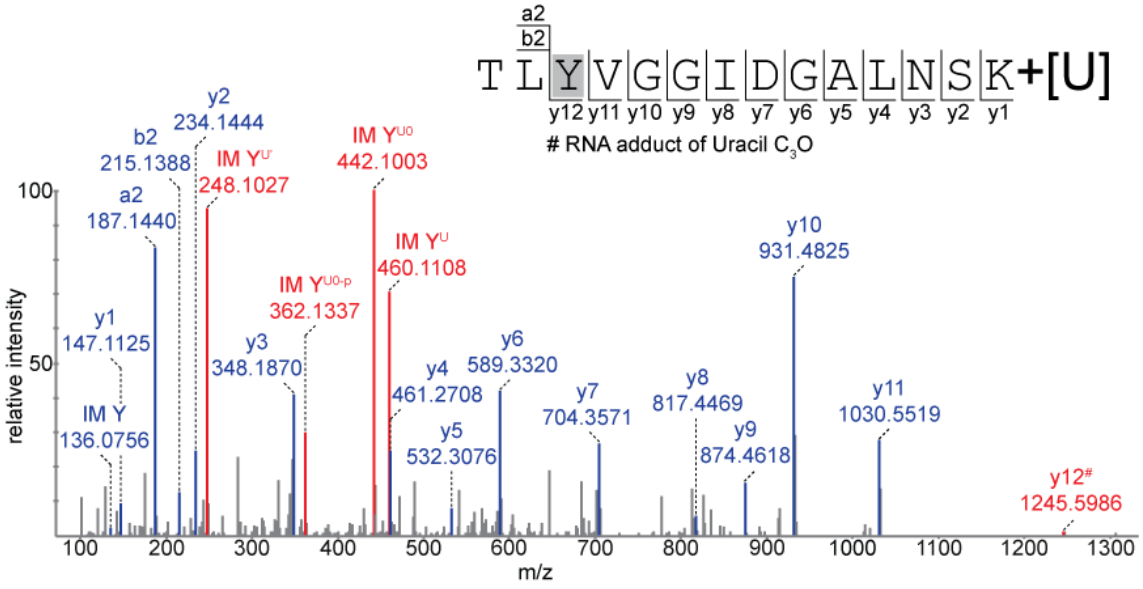
A.6.12


FADYR+[UA]
y3 y2 y1
A'
136.0620
G'
152.0565
IM Y
136.0785
y1
175.1181
IM F
120.0805
U^0-p
227.0648
U^0
307.0319
A^0
330.0598
y2
338.1817
A
348.0676
y3
453.1904
y2^U0-p
564.2372
y2^U0
644.1994
peptide+
671.3169
relative intensity
m/z

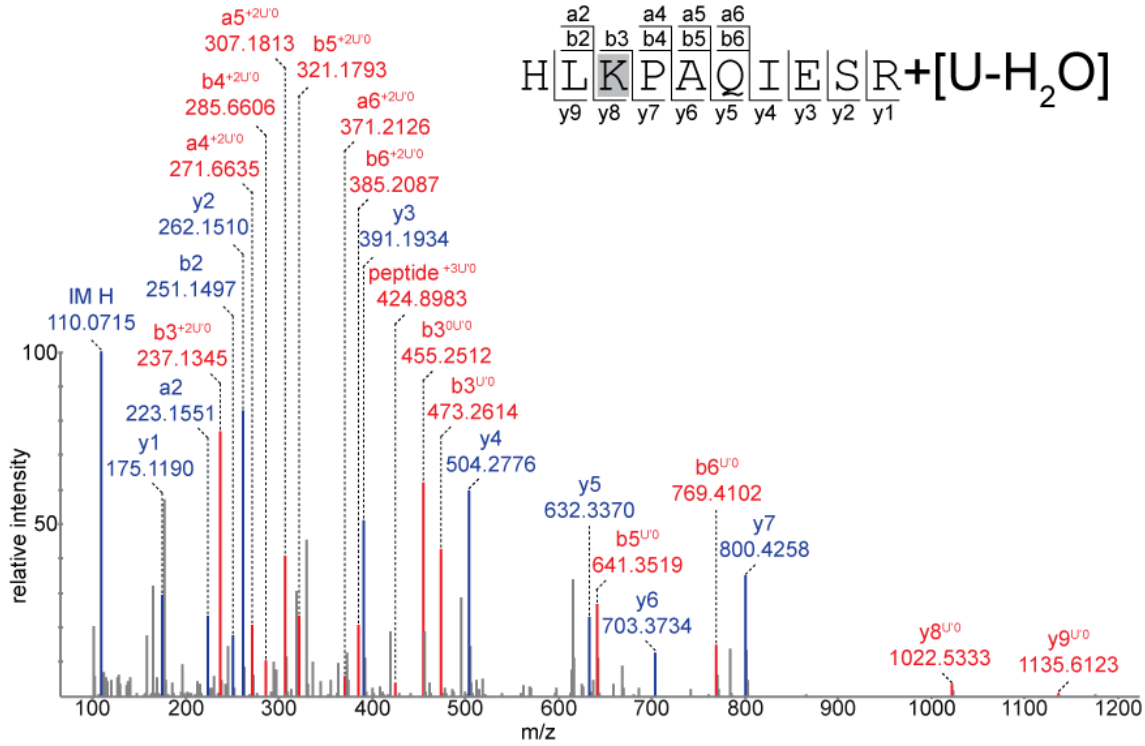
A.6.13


E D M^OX G G I G S F R+[U]
RNA adduct of Uracil C3O
relative intensity
m/z
U0-p 227.0665
a2 217.0821
a2^0 199.0709
y1 175.1191
IM F 120.0813
b2 245.0766
y2 322.1900
GGIGS# 424.1466
b4 431.1223
y3 409.2189
a5 460.1526
b3 392.1141
y4 466.2424
b3^0 374.1023
b5^0 488.1455
y6 636.3486
b5 506.1600
y4U0-p 692.2999
y3U0-p 635.2812
y7 693.3797
y5 579.3313
y5U0-p 805.3900
y8 776.4080
y7U0-p 919.4306
y8O2+U0-p 1002.4616
y6U0-p 862.4124
y8U0-p 1066.4556

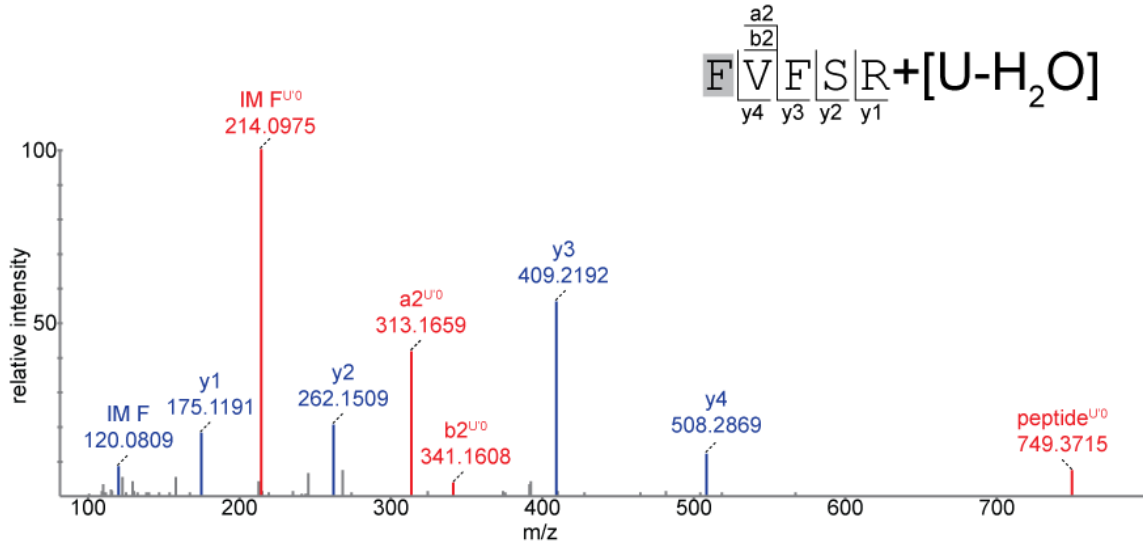
A.6.14


T L Y V G G I D G A L N S K+[U]
RNA adduct of Uracil C3O
relative intensity
m/z
y2 234.1444
b2 215.1388
IM YU0 442.1003
IM YU 248.1027
a2 187.1440
IM YU 460.1108
IM YU0-p 362.1337
y10 931.4825
y1 147.1125
y3 348.1870
y6 589.3320
y4 461.2708
y8 817.4469
y7 704.3571
y11 1030.5519
IM Y 136.0756
y5 532.3076
y9 874.4618
y12# 1245.5986

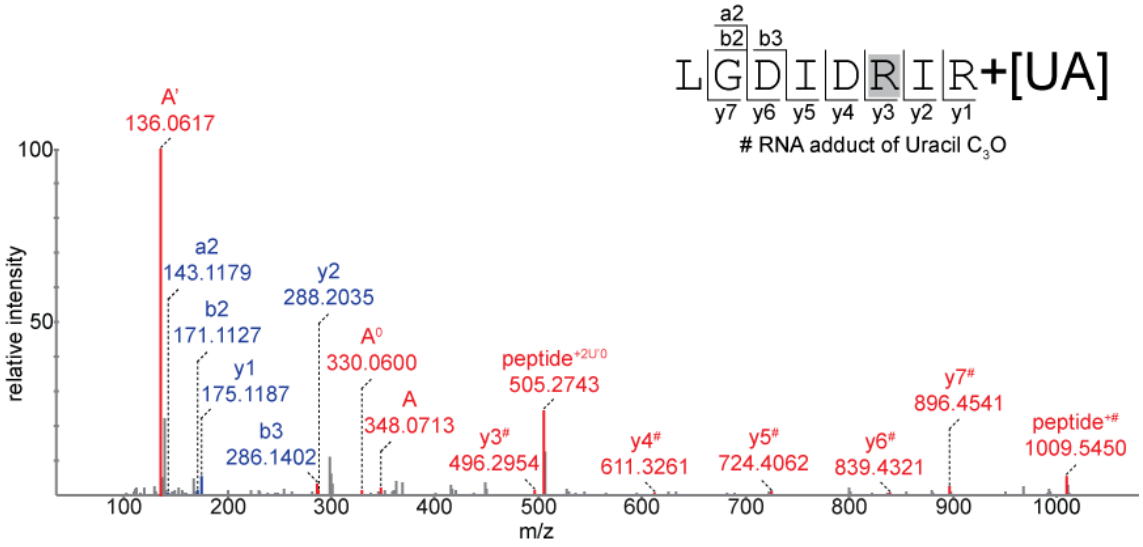
A.6.15


a2 b2 b3 a4 b4 a5 b5 a6 b6
H L K P A Q I E S R+[U-H2O]
y9 y8 y7 y6 y5 y4 y3 y2 y1
a5+2U'0 307.1813
b5+2U'0 321.1793
b4+2U'0 285.6606
a6+2U'0 371.2126
a4+2U'0 271.6635
b6+2U'0 385.2087
y2 262.1510
y3 391.1934
b2 251.1497
peptide +3U'0 424.8983
IM H 110.0715
b3+2U'0 237.1345
b30U'0 455.2512
b3U'0 473.2614
a2 223.1551
y4 504.2776
y1 175.1190
y5 632.3370
b6U'0 769.4102
b5U'0 641.3519
y7 800.4258
y6 703.3734
y8U'0 1022.5333
y9U'0 1135.6123
relative intensity
m/z

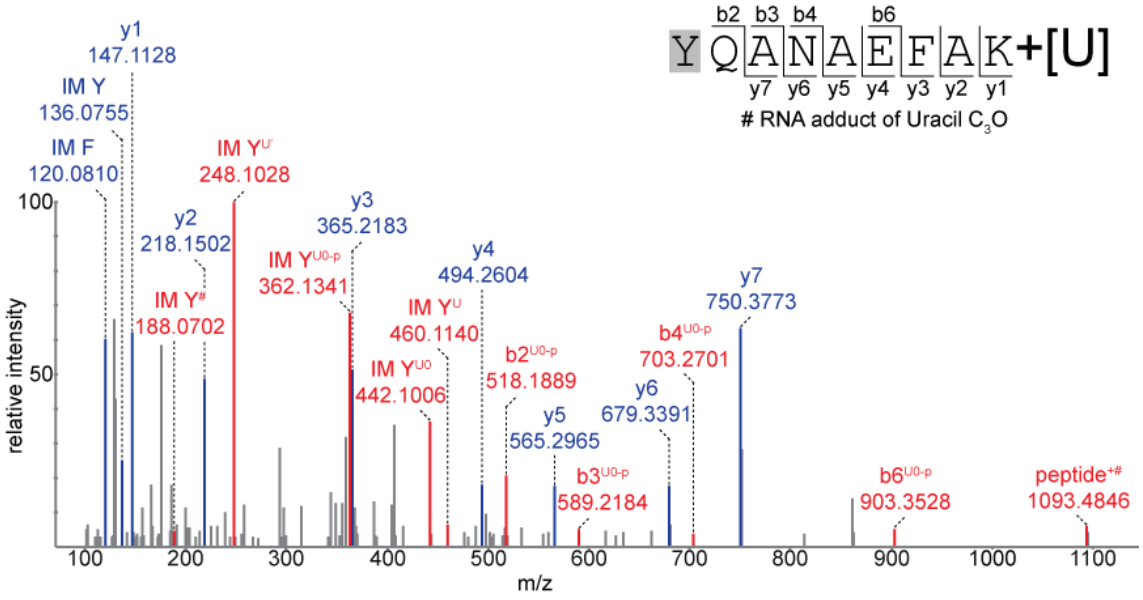
A.6.16


a2 b2
F V F S R+[U-H2O]
y4 y3 y2 y1
IM FU'0 214.0975
y3 409.2192
a2U'0 313.1659
y1 175.1191
y2 262.1509
IM F 120.0809
b2U'0 341.1608
y4 508.2869
peptideU'0 749.3715
relative intensity
m/z

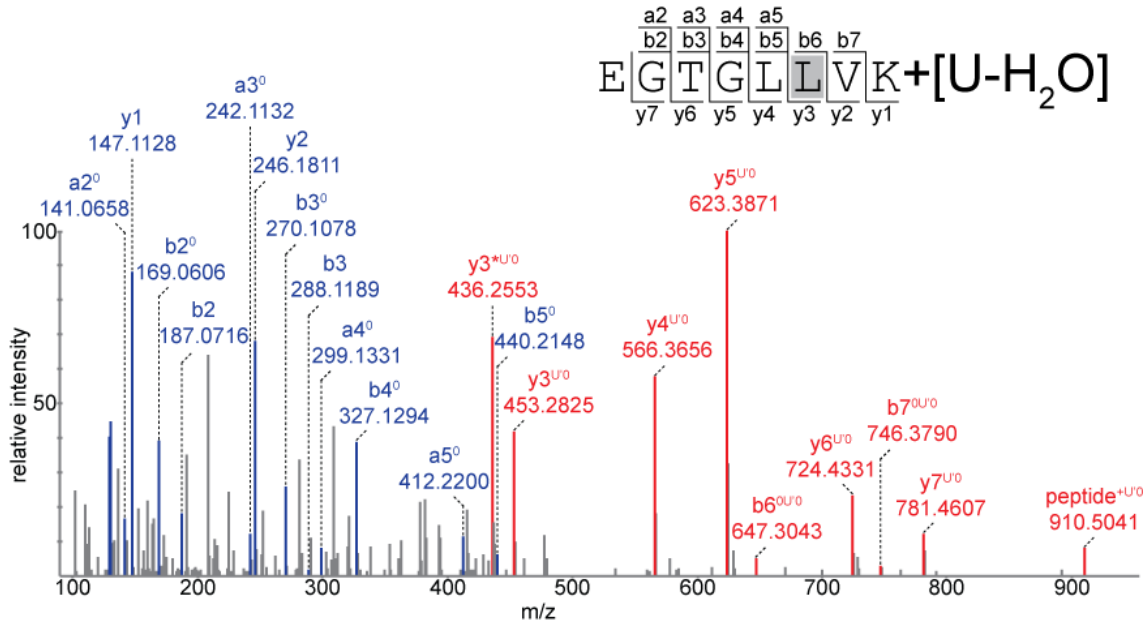
A.6.17


a2
b2 b3
L G D I D R I R+[UA]
y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
A'
136.0617
a2
143.1179
b2
171.1127
y1
175.1187
b3
286.1402
y2
288.2035
A0
330.0600
A
348.0713
y3#
496.2954
peptide+2U'0
505.2743
y4#
611.3261
y5#
724.4062
y6#
839.4321
y7#
896.4541
peptide+#
1009.5450
relative intensity
m/z

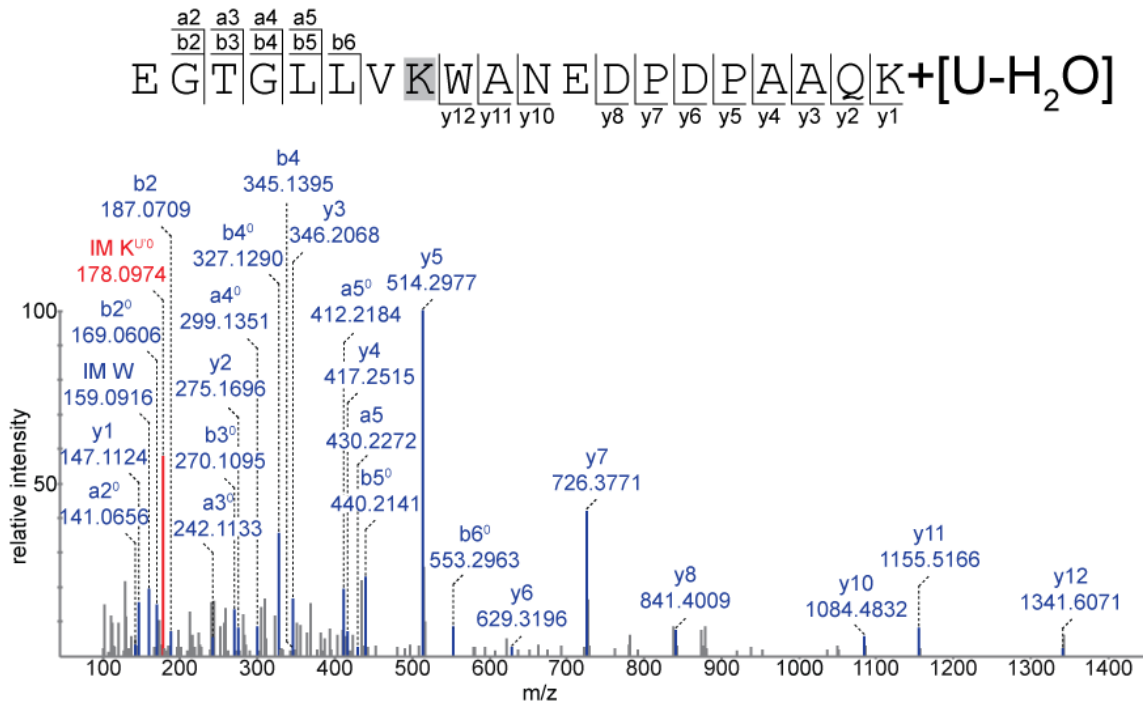
A.6.18


b2 b3 b4 b6
Y Q A N A E F A K+[U]
y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
y1
147.1128
IM Y
136.0755
IM F
120.0810
IM YU'
248.1028
y2
218.1502
IM YU0-p
362.1341
y3
365.2183
y4
494.2604
IM Y#
188.0702
IM YU
460.1140
IM YU0
442.1006
b2U0-p
518.1889
y5
565.2965
y6
679.3391
b4U0-p
703.2701
y7
750.3773
b3U0-p
589.2184
b6U0-p
903.3528
peptide+#
1093.4846
relative intensity
m/z

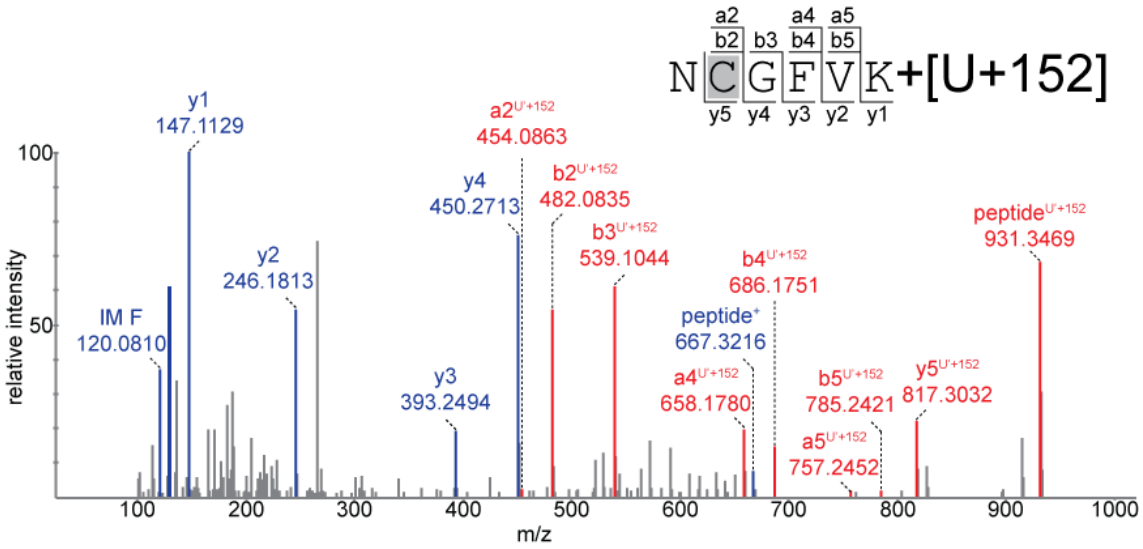
A.6.19


a2 a3 a4 a5
b2 b3 b4 b5 b6 b7
E G T G L L V K+[U-H2O]
y7 y6 y5 y4 y3 y2 y1
y1 147.1128
a3⁰ 242.1132
y2 246.1811
a2⁰ 141.0658
b3⁰ 270.1078
b2⁰ 169.0606
b3 288.1189
b2 187.0716
a4⁰ 299.1331
b4⁰ 327.1294
a5⁰ 412.2200
y3*U'0 436.2553
b5⁰ 440.2148
y3U'0 453.2825
y5U'0 623.3871
y4U'0 566.3656
b6⁰U'0 647.3043
y6U'0 724.4331
b7⁰U'0 746.3790
y7U'0 781.4607
peptide*U'0 910.5041
relative intensity
m/z

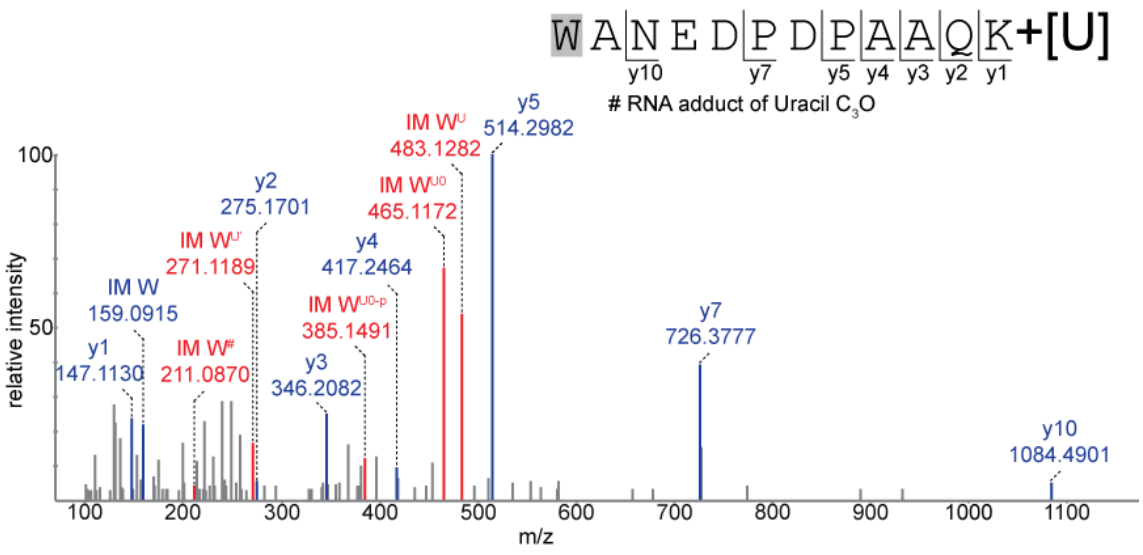
A.6.20


a2 a3 a4 a5
b2 b3 b4 b5 b6
E G T G L L V K W A N E D P D P A A Q K+[U-H2O]
y12 y11 y10 y8 y7 y6 y5 y4 y3 y2 y1
b4 345.1395
b2 187.0709
y3 346.2068
b4⁰ 327.1290
IM KU'0 178.0974
y5 514.2977
a5⁰ 412.2184
a4⁰ 299.1351
b2⁰ 169.0606
y4 417.2515
IM W 159.0916
y2 275.1696
a5 430.2272
y1 147.1124
b3⁰ 270.1095
b5⁰ 440.2141
a2⁰ 141.0656
a3⁰ 242.1133
y7 726.3771
b6⁰ 553.2963
y11 1155.5166
y8 841.4009
y10 1084.4832
y6 629.3196
y12 1341.6071
relative intensity
m/z

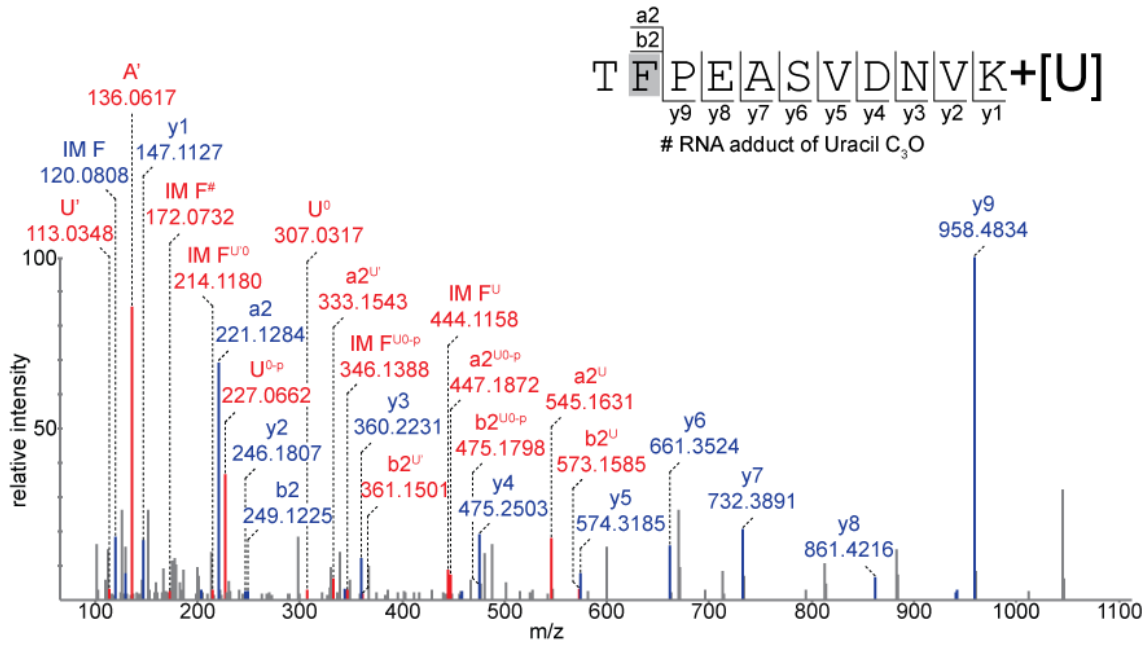
A.6.21


a2 a4 a5
b2 b3 b4 b5
NCGFVK+[U+152]
y5 y4 y3 y2 y1
y1 147.1129
a2U'+152 454.0863
y4 450.2713
b2U'+152 482.0835
b3U'+152 539.1044
peptideU'+152 931.3469
y2 246.1813
b4U'+152 686.1751
IM F 120.0810
peptide+ 667.3216
y3 393.2494
a4U'+152 658.1780
b5U'+152 785.2421
y5U'+152 817.3032
a5U'+152 757.2452
relative intensity
m/z

A.6.22


WANEDPDPAAQK+[U]
y10 y7 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
IM WU 483.1282
y5 514.2982
y2 275.1701
IM WU0 465.1172
IM WU' 271.1189
y4 417.2464
IM W 159.0915
IM WU0-p 385.1491
y7 726.3777
y1 147.1130
IM W# 211.0870
y3 346.2082
y10 1084.4901
relative intensity
m/z

A.6.23


a2
b2
T F P E A S V D N V K+[U]
y9 y8 y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
A'
136.0617
y1
147.1127
IM F
120.0808
IM F#
172.0732
U'
113.0348
IM FU'0
214.1180
U0
307.0317
a2U'
333.1543
a2
221.1284
IM FU
444.1158
IM FU0-p
346.1388
a2U0-p
447.1872
U0-p
227.0662
a2U
545.1631
y3
360.2231
b2U0-p
475.1798
y6
661.3524
y2
246.1807
b2U
573.1585
y7
732.3891
b2U'
361.1501
y4
475.2503
b2
249.1225
y5
574.3185
y8
861.4216
y9
958.4834
relative intensity
m/z

A.6.24

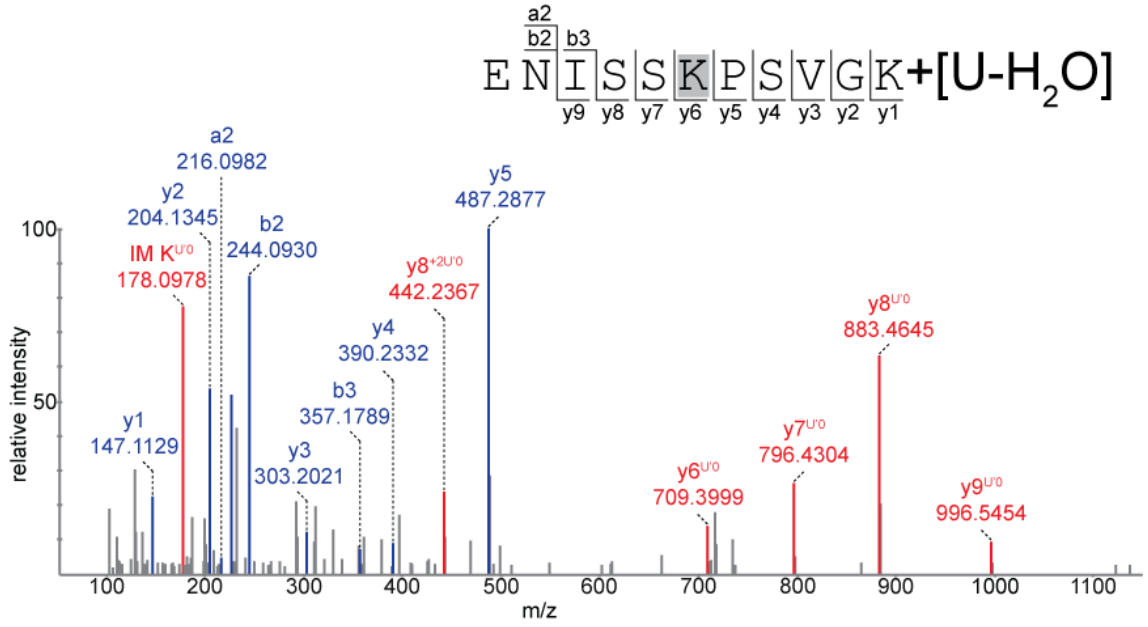
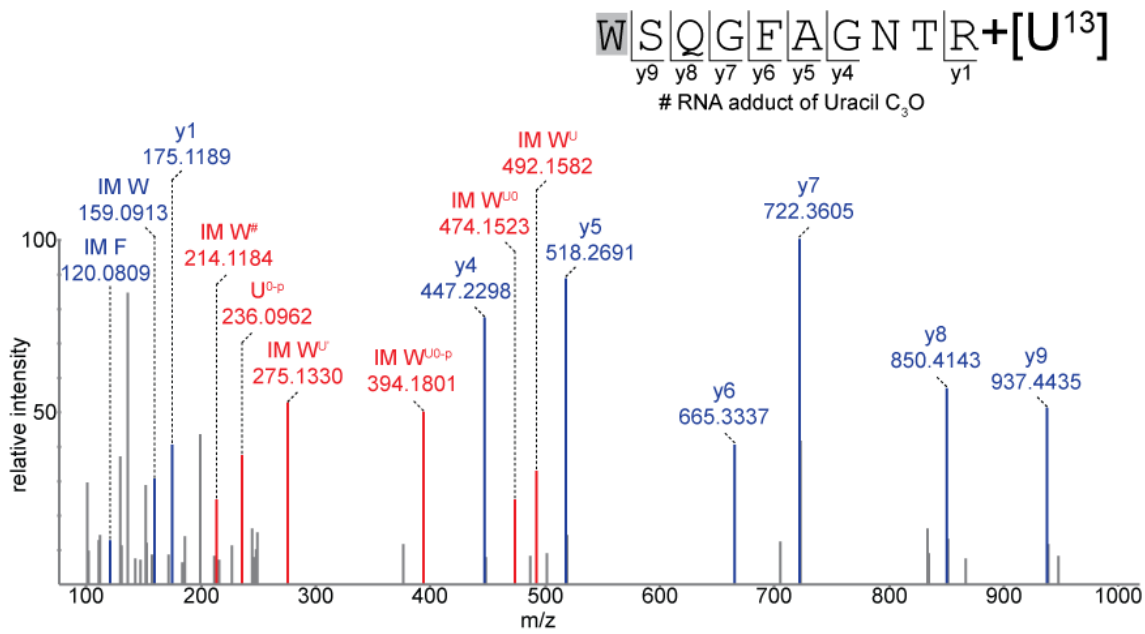

a2
b2 b3
E N I S S K P S V G K+[U-H2O]
y9 y8 y7 y6 y5 y4 y3 y2 y1
a2
216.0982
y2
204.1345
y5
487.2877
b2
244.0930
IM KU'0
178.0978
y8+2U'0
442.2367
y8U'0
883.4645
y4
390.2332
b3
357.1789
y7U'0
796.4304
y1
147.1129
y3
303.2021
y6U'0
709.3999
y9U'0
996.5454
relative intensity
m/z

Table 6.5: U4/U6 snRNA-CWC2 cross-links identified in quantitative mass spectrometric analysis followed by corresponding MS/MS spectra.

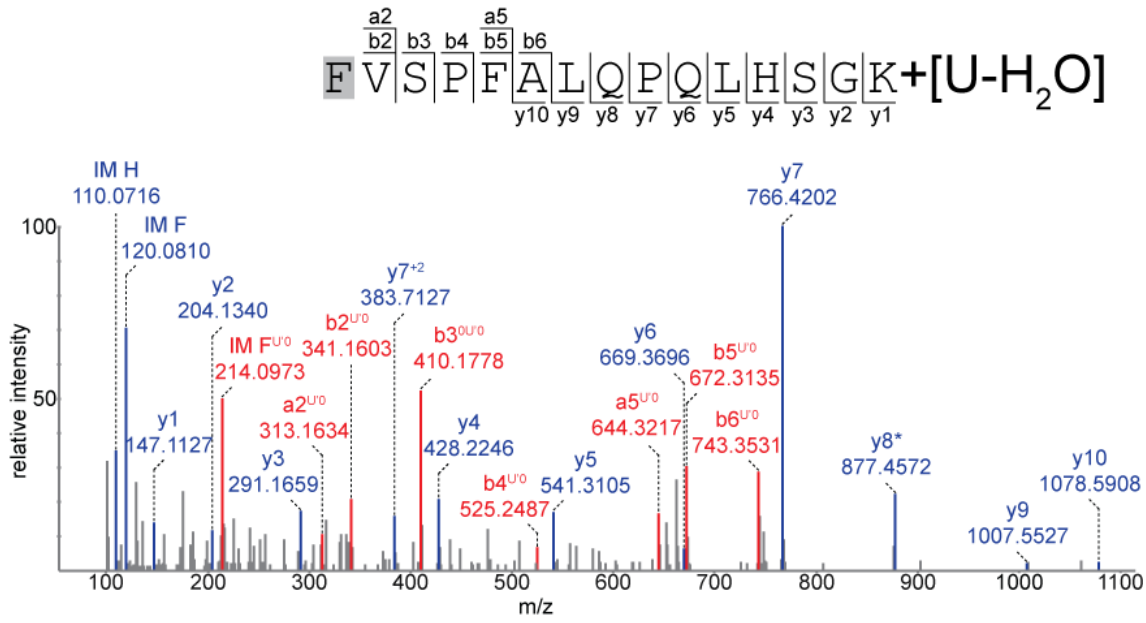
Domain	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Torus	^{37}WSQGFAGNTR^{46}	W^{37}	U	1122.5206	333.0661	1455.5867	2	728.8012	728.8007	-0.69
Torus	^{47}FVSPFALQPQLHSGK^{61}	F^{47}	U-H_2O	1654.8830	306.0253	1960.9083	3	654.6439	654.6422	-2.6
Torus	^{106}TEVLDCFGR^{114}	C^{111}	U+152	1038.4804	485.0627	1523.5431	2	762.7793	762.7789	-0.53
ZnF	^{87}CEYLHHIPDEEDIGK^{101}	-	U+152	1796.8039	485.0627	2281.8666	3	761.6300	761.6309	1.18
RRM/RNP2	^{136}TLYVGGIDGALNSK^{149}	Y^{138}	U	1406.7405	333.0661	1739.8066	2	870.9111	870.9103	-0.92
RRM	^{188}YQANAEFAK^{196}	Y^{188}	U	1040.4926	333.0661	1373.5587	2	687.7872	687.7857	-2.18
RRM/RNP1	^{180}NCGFVK^{185}	C^{181}	U+152	666.3159	485.0627	1151.3786	2	576.6971	576.6959	-2.08
CTD	^{310}KNISR^{314}	K^{310}	UA-H_2O	616.3656	635.078	1251.4436	2	626.7296	626.7290	-0.96

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence; RNA: Nucleotide found to be cross-linked to the peptide; m(calc) Peptide: Theoretical mass of the cross-linked peptide; m(calc) RNA: Theoretical mass of the cross-linked RNA; m(calc) Cross-link: Calculated mass of the RNA-protein cross-link; z: Charge state of the cross-link; *m/z*(calc): Calculated *m/z* of the cross-link by using formula m+z(mH)/z; *m/z*(exp): Observed *m/z* of the cross-link; Δm (ppm): Calculated mass error in ppm by using formula $[m/z(\text{exp}) - m/z(\text{calc})/m/z(\text{calc})] \times 10^6$.

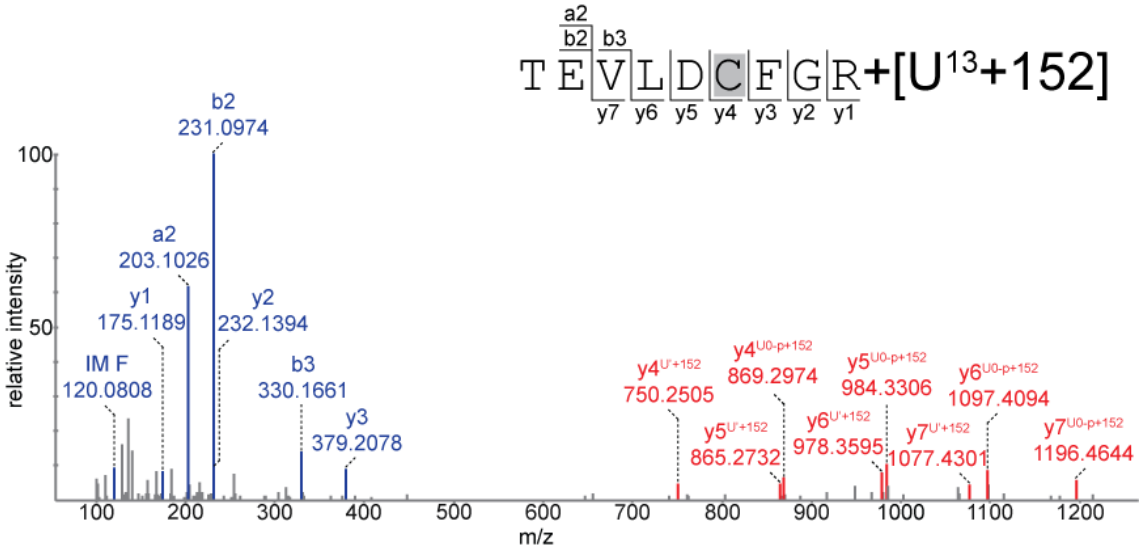
A.6.25


WSQGFAGNTR+[U13]
y9 y8 y7 y6 y5 y4 y1
RNA adduct of Uracil C3O
y1 175.1189
IM W 159.0913
IM F 120.0809
IM W# 214.1184
U0-p 236.0962
IM WU 275.1330
IM WU0-p 394.1801
y4 447.2298
IM WU0 474.1523
IM WU 492.1582
y5 518.2691
y6 665.3337
y7 722.3605
y8 850.4143
y9 937.4435
relative intensity
m/z

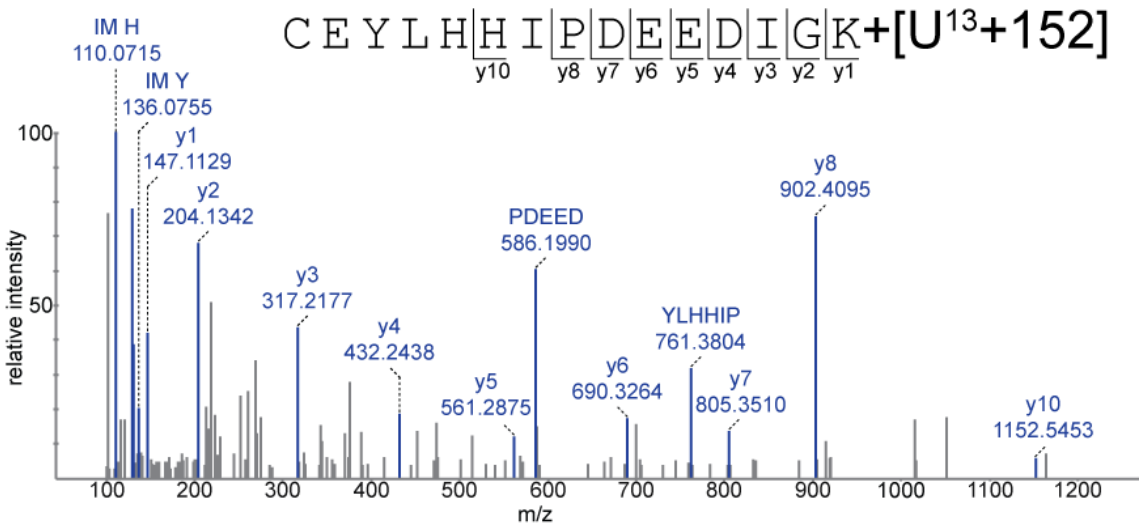
A.6.26


FVSPFALQPQLHSGK+[U-H2O]
a2 b2 b3 b4 a5 b5 b6
y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
IM H 110.0716
IM F 120.0810
y2 204.1340
IM FU'0 214.0973
y1 147.1127
a2U'0 313.1634
y3 291.1659
b2U'0 341.1603
y7+2 383.7127
b3U'0 410.1778
y4 428.2246
b4U'0 525.2487
y5 541.3105
y7 766.4202
y6 669.3696
a5U'0 644.3217
b5U'0 672.3135
b6U'0 743.3531
y8* 877.4572
y9 1007.5527
y10 1078.5908
relative intensity
m/z

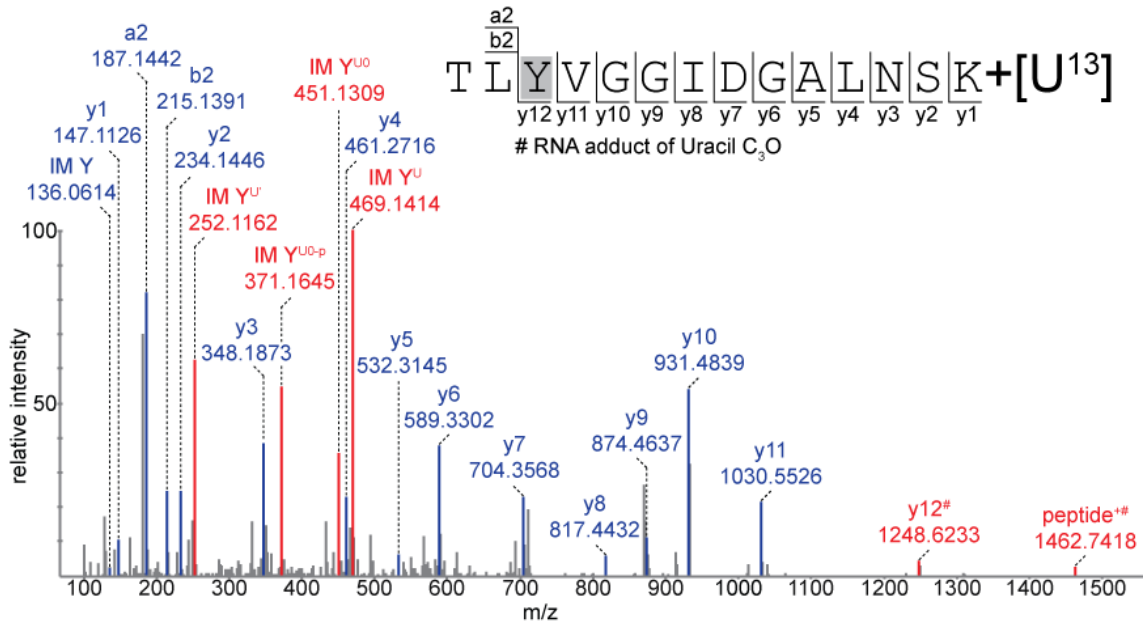
A.6.27


a2
b2 b3
T E V L D C F G R+[U13+152]
y7 y6 y5 y4 y3 y2 y1
b2
231.0974
a2
203.1026
y1
175.1189
y2
232.1394
IM F
120.0808
b3
330.1661
y3
379.2078
y4U'+152
750.2505
y4U0-p+152
869.2974
y5U'+152
865.2732
y5U0-p+152
984.3306
y6U'+152
978.3595
y6U0-p+152
1097.4094
y7U'+152
1077.4301
y7U0-p+152
1196.4644
relative intensity
100
50
100 200 300 400 500 600 700 800 900 1000 1100 1200
m/z

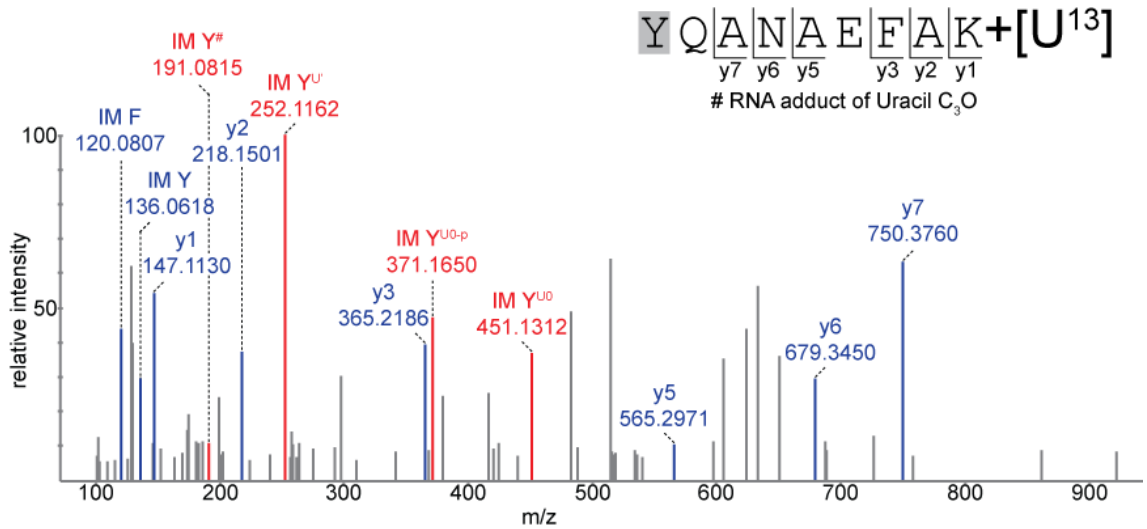
A.6.28


C E Y L H H I P D E E D I G K+[U13+152]
y10 y8 y7 y6 y5 y4 y3 y2 y1
IM H
110.0715
IM Y
136.0755
y1
147.1129
y2
204.1342
y3
317.2177
y4
432.2438
y5
561.2875
PDEED
586.1990
y6
690.3264
YLHHIP
761.3804
y7
805.3510
y8
902.4095
y10
1152.5453
relative intensity
100
50
100 200 300 400 500 600 700 800 900 1000 1100 1200
m/z

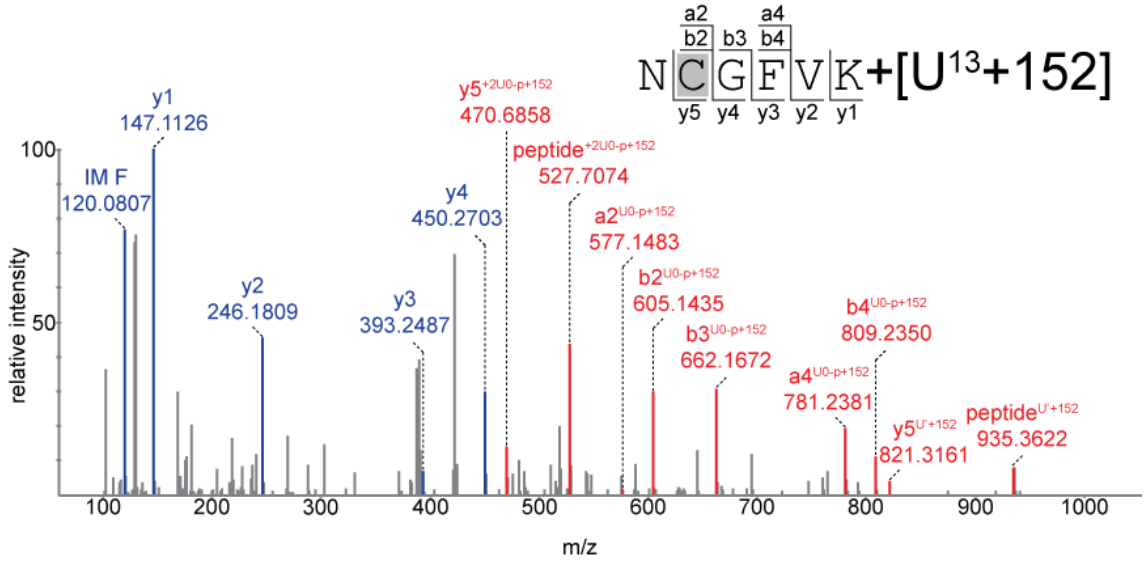
A.6.29


a2
b2
T L Y V G G I D G A L N S K+[U13]
y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
a2 187.1442
b2 215.1391
y1 147.1126
y2 234.1446
IM Y 136.0614
IM YU' 252.1162
IM YU0 451.1309
y4 461.2716
IM YU 469.1414
IM YU0-p 371.1645
y3 348.1873
y5 532.3145
y6 589.3302
y7 704.3568
y8 817.4432
y9 874.4637
y10 931.4839
y11 1030.5526
y12# 1248.6233
peptide+# 1462.7418
relative intensity
m/z

A.6.30


Y Q A N A E F A K+[U13]
y7 y6 y5 y3 y2 y1
RNA adduct of Uracil C3O
IM Y# 191.0815
IM YU' 252.1162
IM F 120.0807
y2 218.1501
IM Y 136.0618
y1 147.1130
IM YU0-p 371.1650
y3 365.2186
IM YU0 451.1312
y5 565.2971
y6 679.3450
y7 750.3760
relative intensity
m/z

A.6.31


a2 b2 b3 a4 b4
NCGFVK+[U13+152]
y5 y4 y3 y2 y1
y1 147.1126
y5+2U0-p+152 470.6858
peptide+2U0-p+152 527.7074
IM F 120.0807
y4 450.2703
a2U0-p+152 577.1483
b2U0-p+152 605.1435
y2 246.1809
y3 393.2487
b3U0-p+152 662.1672
b4U0-p+152 809.2350
a4U0-p+152 781.2381
y5U'+152 821.3161
peptideU'+152 935.3622
relative intensity
m/z

A.6.32

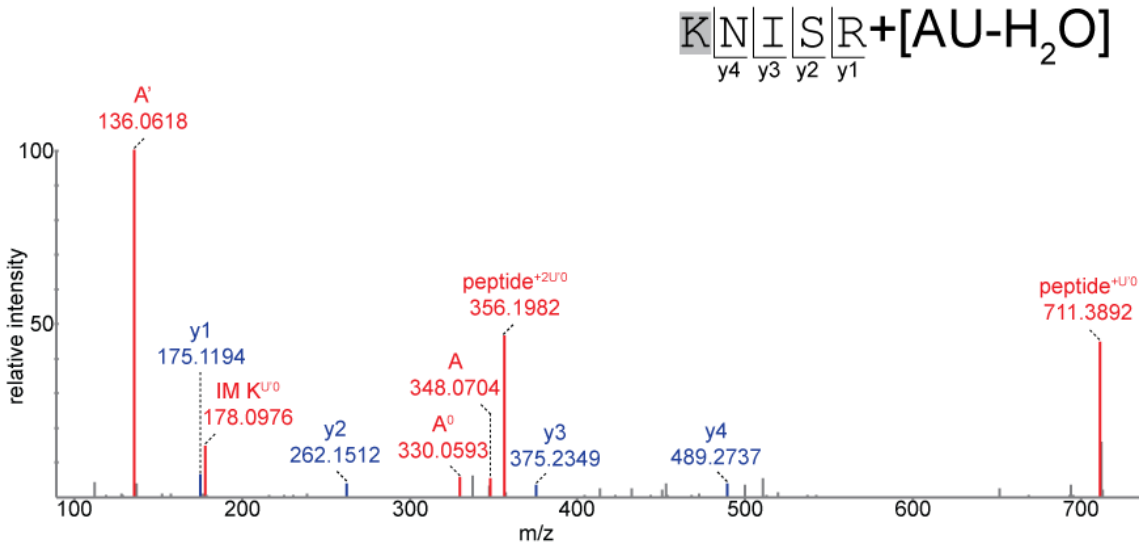

KNISR+[AU-H2O]
y4 y3 y2 y1
A' 136.0618
peptide+2U'0 356.1982
peptide+U'0 711.3892
y1 175.1194
IM KU'0 178.0976
A 348.0704
A0 330.0593
y2 262.1512
y3 375.2349
y4 489.2737
relative intensity
m/z

Table 6.6: The logarithmic (log_{10}) peak areas of the U4/U6 snRNA-CWC2 cross-links in quantitative mass spectrometric analysis, calculated by using Xcalibur.

Domain	Peptide	Replica_1				Replica_2				Replica_3			
		F-L	F-H	R-L	R-H	F-L	F-H	R-L	R-H	F-L	F-H	R-L	R-H
Torus	^{8}SAKVQVK^{14}	7.2	-	7.3	-	6.6	-	7	-	7	-	7.2	-
Torus	^{37}WSQGFAGNTR^{46}	-	-	-	-	-	6.6	-	-	-	-	-	-
Torus	^{47}FVSPFALQPQLHSGK^{61}	-	-	-	-	5.7	-	-	-	-	-	-	-
Torus	^{106}TEVLDCFGR^{114}	7.9	8.2	8.4	7.8	7.9	8.2	8.3	7.3	7.1	7.5	8	7.4
ZnF	^{87}CEYLHHIPDEEDIGK^{101}	-	7.5	-	-	-	6.8	6.7	-	-	-	-	-
connector element	^{122}EDM(Oxidation)GGIGSFR^{131}	-	-	5.7	-	-	-	6.9	-	-	-	6.8	-
connector element	^{115}EKFADYR^{121}	-	-	7.2	-	7.3	-	-	-	-	-	-	-
RRM/RNP2	^{136}TLYVGGIDGALNSK^{149}	8.5	8.5	8.7	8	8.7	8.8	9.1	8.2	8.4	8.5	8.8	8.1
RRM	^{150}HLKPAQIESR^{159}	-	-	-	-	8	-	8.2	-	-	-	-	-
RRM	^{188}YQANAEFAK^{196}	-	-	7.2	-	7.1	6.1	8	-	6	7.3	7.8	-
RRM	^{217}EGTGLLVK^{224}	-	-	-	-	-	-	7.2	-	6.7	-	7.4	-
RRM	^{217}EGTGLLVKWANEDPDPAAQK^{236}	6.9	-	7.2	-	7	-	7	-	-	-	-	-
RRM/RNP1	^{180}NCGFVK^{185}	8.6	8.7	9.1	8.7	8	8.8	8.8	8.1	8.2	8.5	8.8	8.4
CTD	^{225}WANEDPDPAAQK^{236}	6.2	-	7	-	6.2	-	7.1	-	-	-	7.5	-
CTD	^{310}KNISR^{314}	6.7	-	7.1	-	6.5	-	-	-	-	-	-	-
CTD	^{315}ENISSKPSVGK^{325}	7.5	-	7.8	-	7.8	-	8.1	-	7.4	-	7.7	-

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Replica: Three sets of replicates with forward and reverse labeled samples in each set; F-L: The log_{10} of peak areas of the U4/U6 snRNA-CWC2 cross-links in forward replicate with light label; F-H: The log_{10} of peak areas of the U4/U6 snRNA-CWC2 cross-links in forward replicate with heavy label; R-L: The log_{10} of peak areas of the U4/U6 snRNA-CWC2 cross-links in reverse replicate with light label; R-H: The log_{10} of peak areas of the U4/U6 snRNA-CWC2 cross-links in reverse replicate with heavy label.

*The area of each cross-linked peptide along with its miss-cleavage state and all of the charge states (+2 and +3) and RNA moiety combinations have been added and then their log_{10} have been calculated.

Table 6.7: The logarithmic (log_2) ratio of peak areas (Appendix A.6.33-A.6.47) of the U4/U6 snRNA-CWC2 cross-links in quantitative mass spectrometric analysis, calculated by using Skyline.

Domain	Peptide	RNA	z	Replica_1		Replica_2		Replica_3		Median	SD
				F	R	F	R	F	R		
Torus	^{106}TEVLDCFGR^{114}	U	2	0	0	1.48	0	0	0	0	0.60
		U+152	2	0.61	1.76	0.28	2.45	0.55	2.06	1.18	0.92
RRM/RNP2	^{136}TLYVGGIDGALNSK^{149}	U	2	-0.07	1.72	0.20	2.44	0.01	1.90	0.96	1.11
		UA	2	0.03	3.34	-0.10	3.53	-1.56	0	0.02	1.88
			3	-0.10	3.37	-0.07	3.83	0.43	0	-	-
		UAA	2	0	0	0.86	0	0	0	0	1.12
			3	0.46	0	0.28	3.91	0	0	-	-
		UG	2	1.26	3.01	0.32	2.85	0.32	2.35	1.80	1.27
			3	1.68	2.83	0.30	3.37	0.57	3.14	-	-
	^{134}NKTLYVGGIDGALNSK^{149}	U-H_2O	2	0	0	0	0	0	0	0	0
			3	0	0	0	0	0	0	-	-
RRM/RNP1	^{180}NCGFVK^{185}	U	2	-0.37	0	0	0	0.90	2.42	0	1.03
		U+152	2	-0.13	0.97	0.08	1.60	0.55	1.22	0.76	0.67
		UA	2	1.07	0.34	0.46	3.13	1.14	1.89	1.10	1.04
		UA+152	2	-0.11	2.29	0	2.42	0.30	1.52	0.91	1.15
		UAA	2	0	0	0	0	0	0	0	0
		UAA+152	2	2.24	0	1.02	3	0	0	0	1.07
			3	0	0	1.62	0	0	0	-	-
		UG	2	0	2.84	0	0	0.28	2.45	0.14	1.34
		UG+152	2	0	0	0.19	2.62	0.85	1.82	0.52	1.09

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; RNA: Nucleotide found to be cross-linked to the peptide; Replica: Three sets of replicates with forward and reverse labeled samples in each set; F: The log_2 ratio of peak areas of the U4/U6 snRNA-CWC2 cross-links in forward replicate; R: The log_2 ratio of peak areas of the U4/U6 snRNA-CWC2 cross-links in reverse replicate; Median: Median of the log_2 ratio of peak areas of the peptide in forward and reverse replicates with all charge states (+2 and +3) with each cross-linked oligonucleotide combination; SD: Standard deviation of the median calculated.

Table 6.8: The logarithmic (log_2) ratio of peak areas (Appendix A.6.33-A.6.47) of the U4/U6 snRNA-CWC2 cross-links in quantitative mass spectrometric analysis, calculated by using Skyline. (With Imputation)

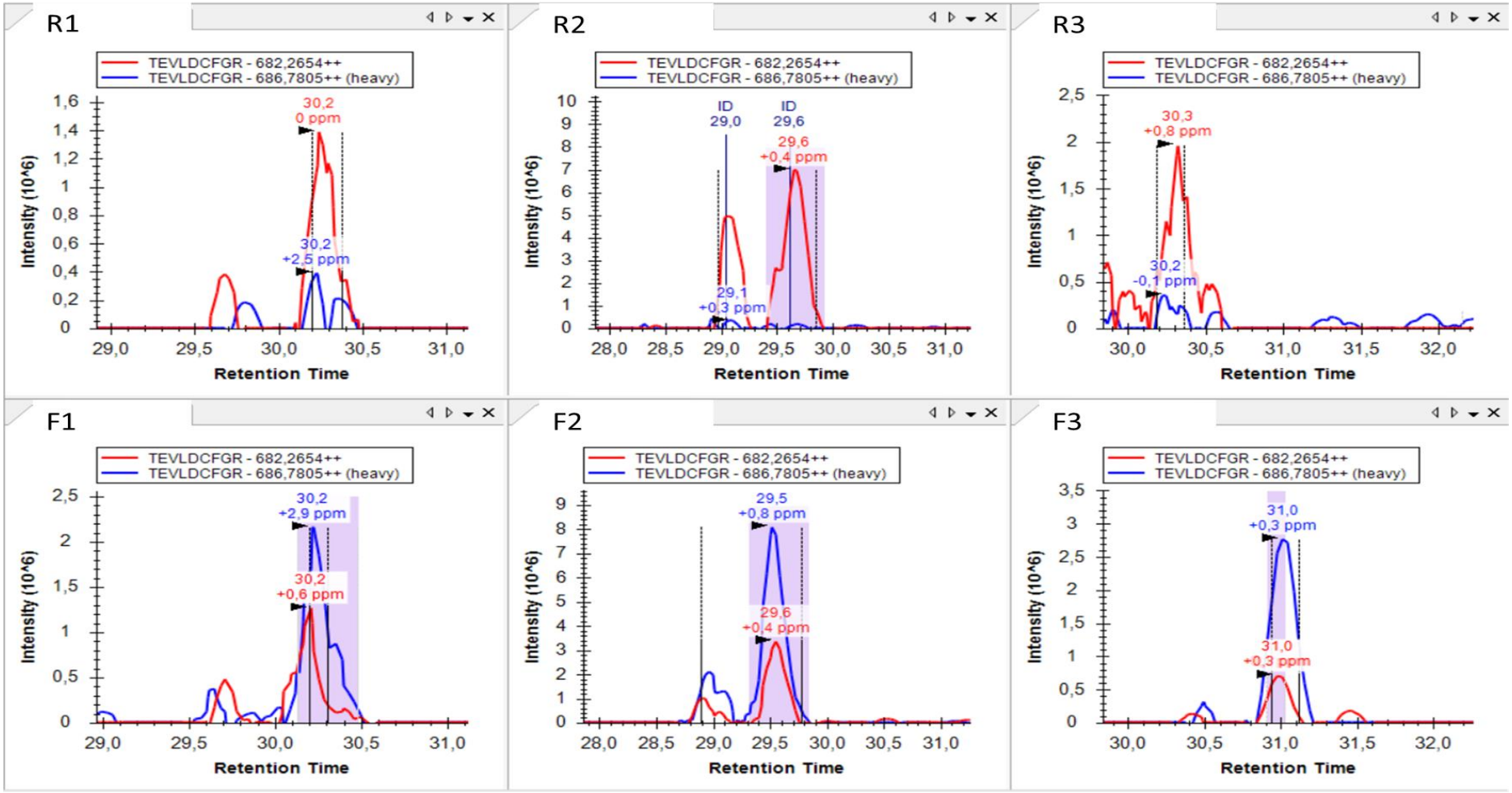
Domain	Peptide	RNA	z	Replica_1		Replica_2		Replica_3	
				F	R	F	R	F	R
Torus	^{106}TEVLDCFGR^{114}	U	2	20.17	21.12	1.48	26.87	24.12	22.58
		U+152	2	0.61	1.76	0.28	2.45	0.55	2.06
RRM/RNP2	^{136}TLYVGGIDGALNSK^{149}	U	2	-0.07	1.72	0.20	2.44	0.01	1.90
		UA	2	0.03	3.34	-0.10	3.53	-1.56	27
			3	-0.10	3.37	-0.07	3.83	0.43	27.27
		UAA	2	0	0	0.86	25.96	0	0
			3	0.46	25.41	0.28	3.91	0	23.73
		UG	2	1.26	3.01	0.32	2.85	0.32	2.35
			3	0.68	2.83	0.30	3.37	0.57	3.14
	^{134}NKTLYVGGIDGALNSK^{149}	U-H_2O	2	-21.71	23.58	-24.83	25.31	0	19.34
			3	-22.58	24.45	-25.11	25.28	0	0
RRM/RNP1	^{180}NCGFVK^{185}	U	2	-0.37	22.37	21.31	24	0.90	2.42
		U+152	2	-0.13	0.97	0.08	1.60	0.55	1.22
		UA	2	1.07	0.34	0.46	3.13	1.14	1.89
		UA+152	2	-0.11	2.29	0	2.42	0.30	1.52
		UAA	2	0	17.40	23.87	23.97	0	0
		UAA+152	2	2.24	27.55	1.02	3	0	0
			3	24.73	26.62	1.62	24.48	0	0
		UG	2	0	2.82	0	26.82	0.28	2.45
		UG+152	2	26.51	26.50	0.19	2.62	0.85	1.82

Domain: Location of the peptide within the protein structure; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; RNA: Nucleotide found to be cross-linked to the peptide; Replica: Three sets of replicates with forward and reverse labeled samples in each set; F: The log_2 ratio of peak areas of the U4/U6 snRNA-CWC2 cross-links in forward replicate; R: The log_2 ratio of peak areas of the U4/U6 snRNA-CWC2 cross-links in reverse replicate.

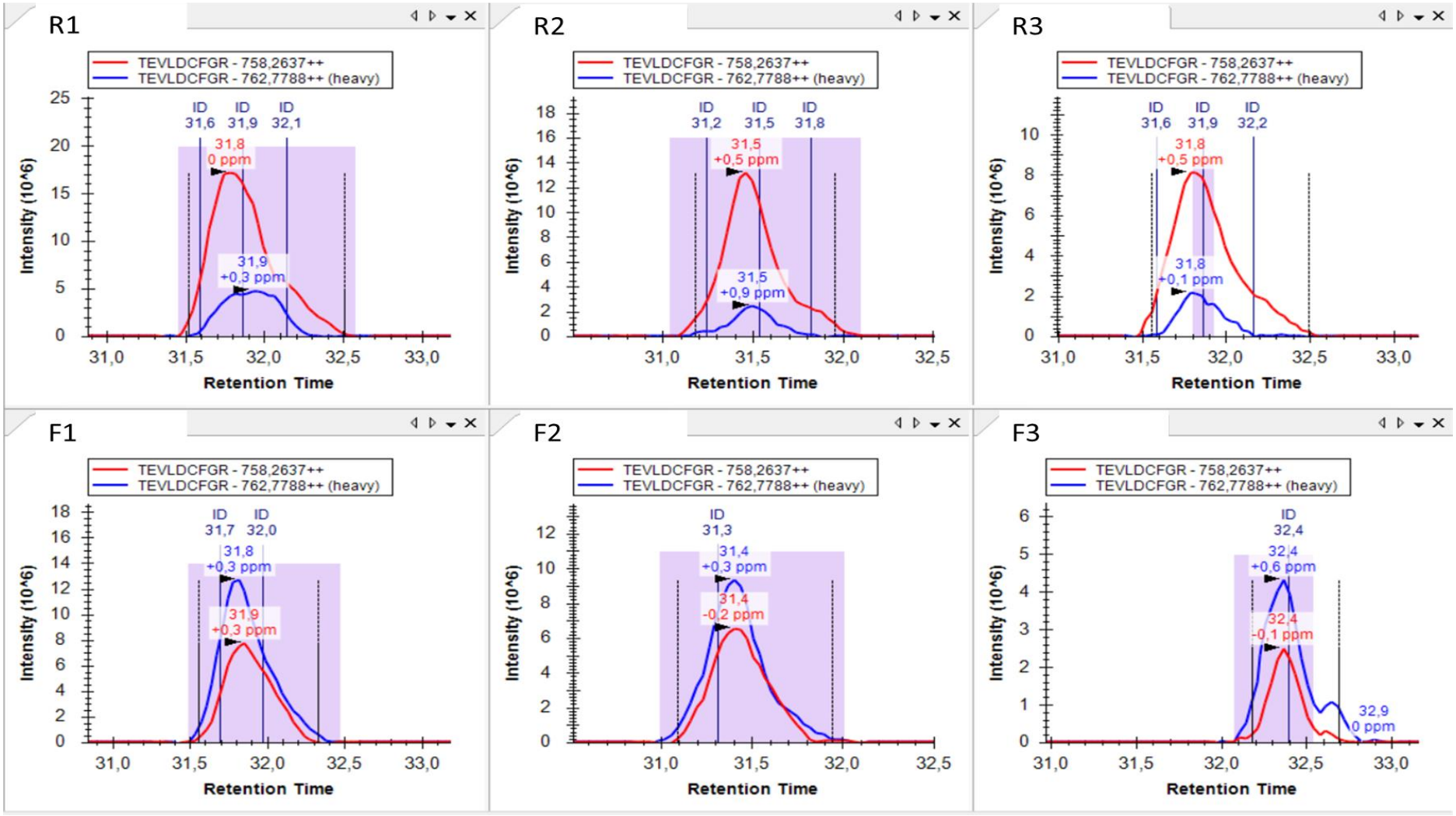
* The missing values of peak area were imputed with “1” and then ratios were calculated.

The extracted ion chromatograms (XICs) of the U4/U6 snRNA-CWC2 cross-links in quantitative mass spectrometric analysis by using Skyline.

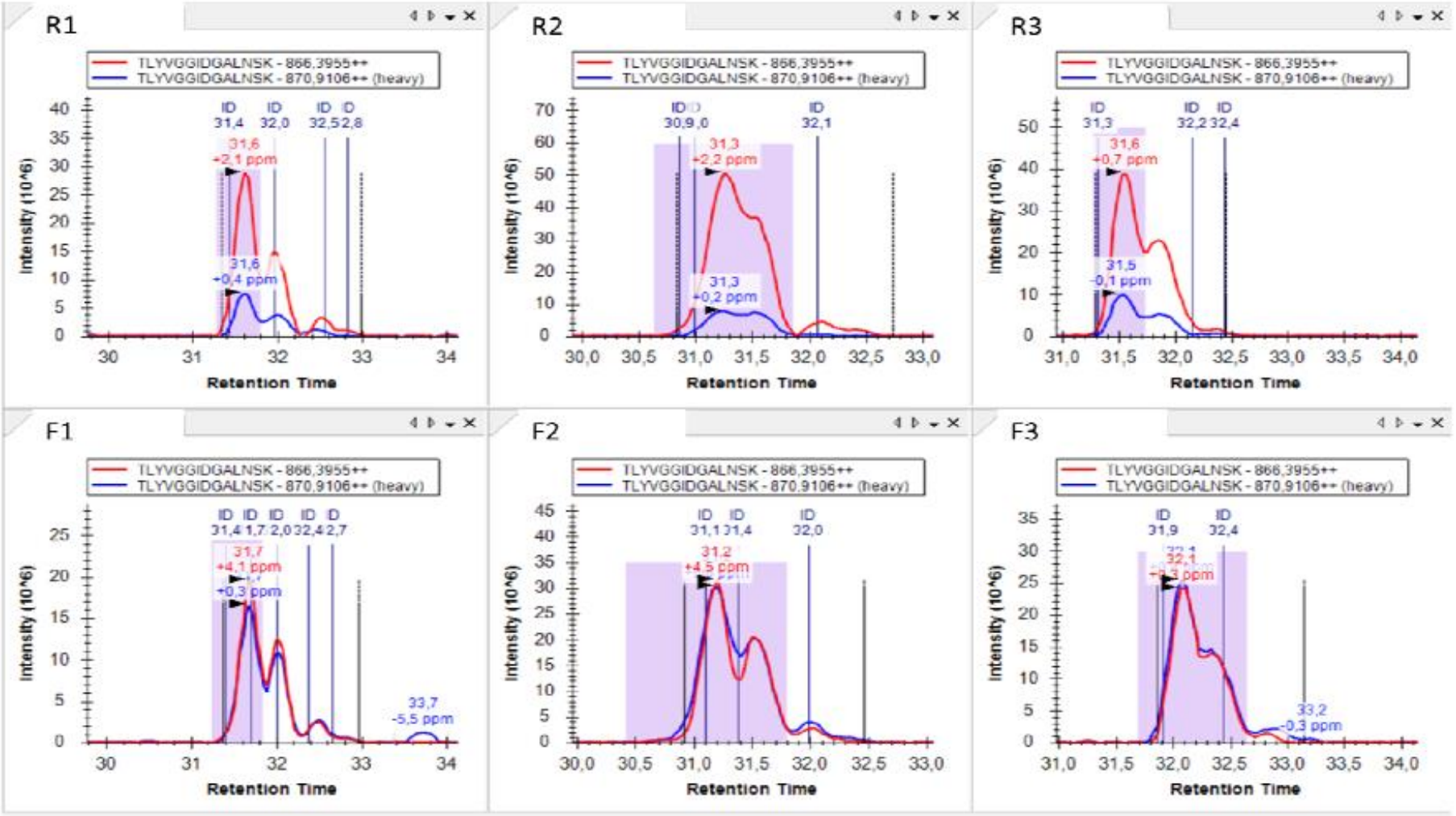
A.6.33: TEVLDCFGR+U



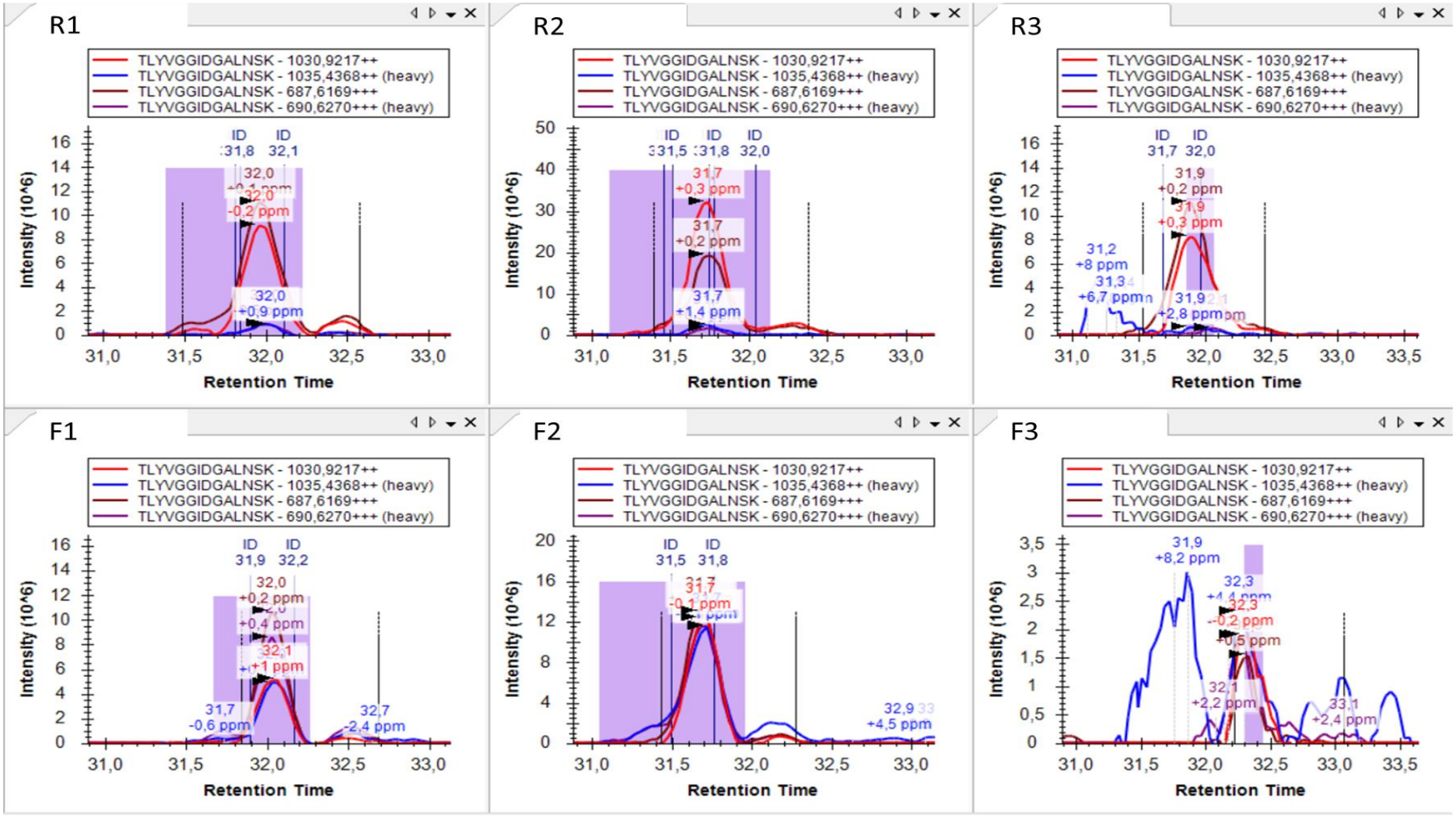
A.6.34: TEVLDCFGR+U+152



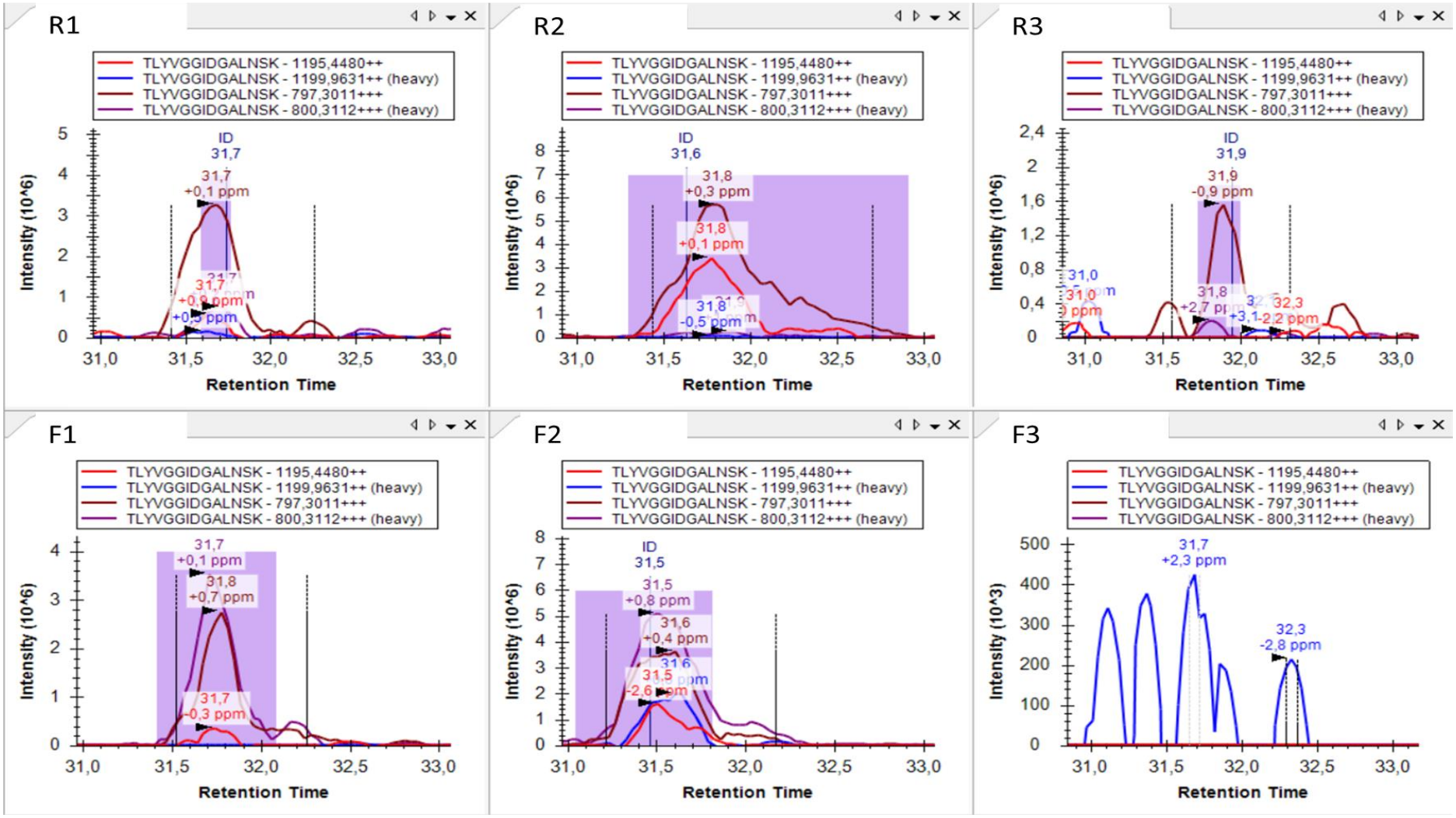
A.6.35: TLYVGGIDGALNSK+U



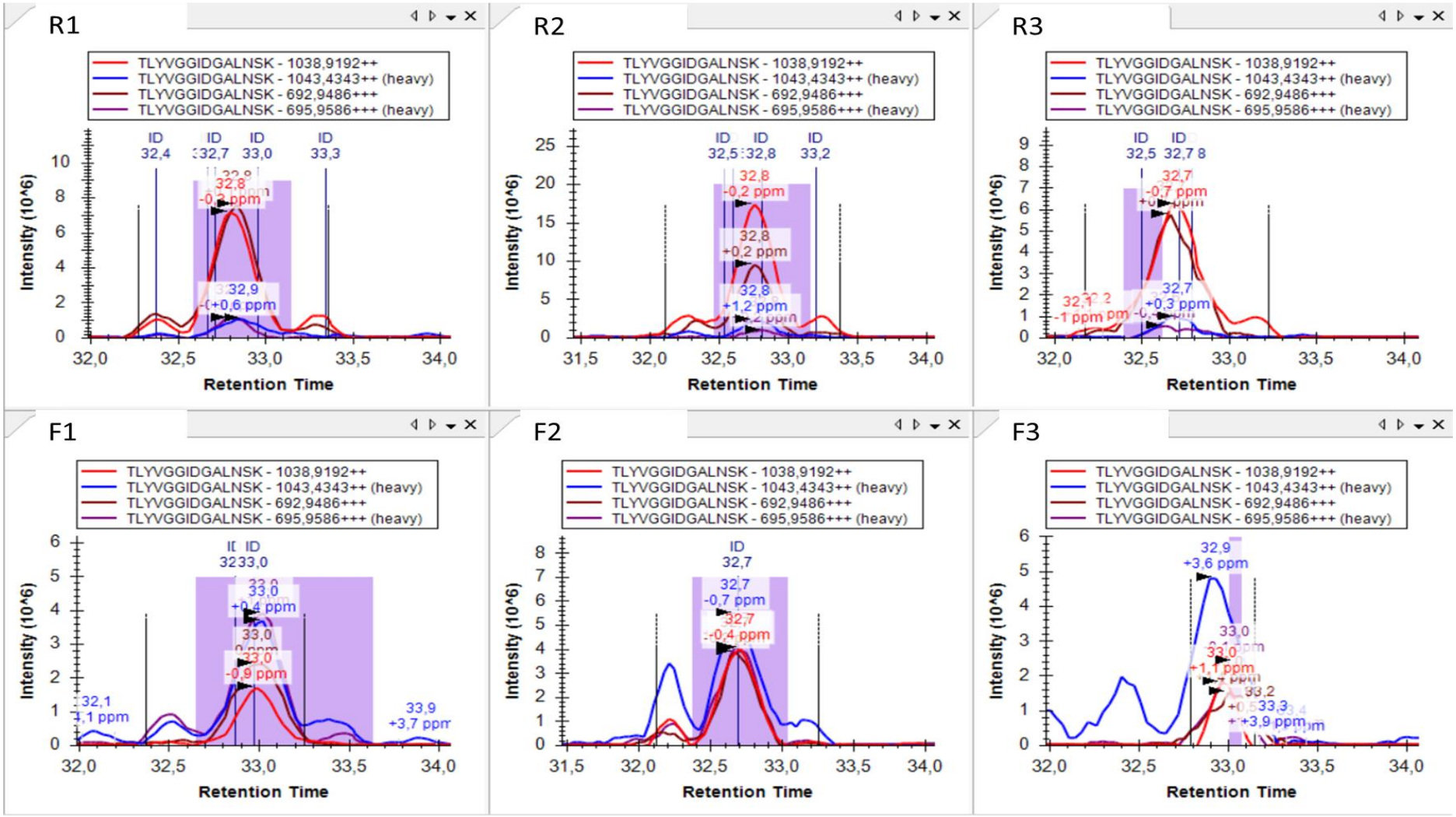
A.6.36: TLYVGGIDGALNSK+UA



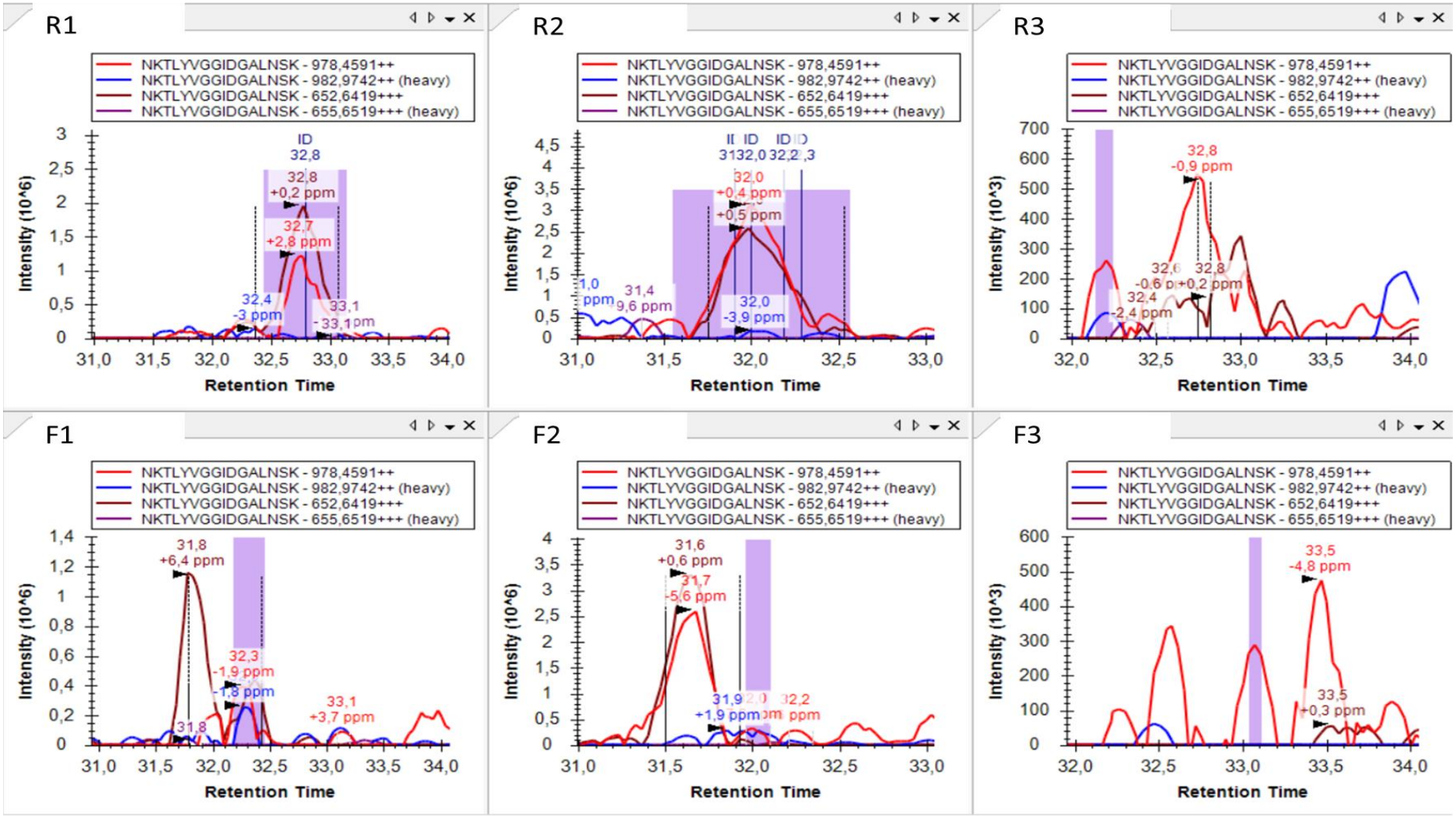
A.6.37: TLYVGGIDGALNSK+UAA



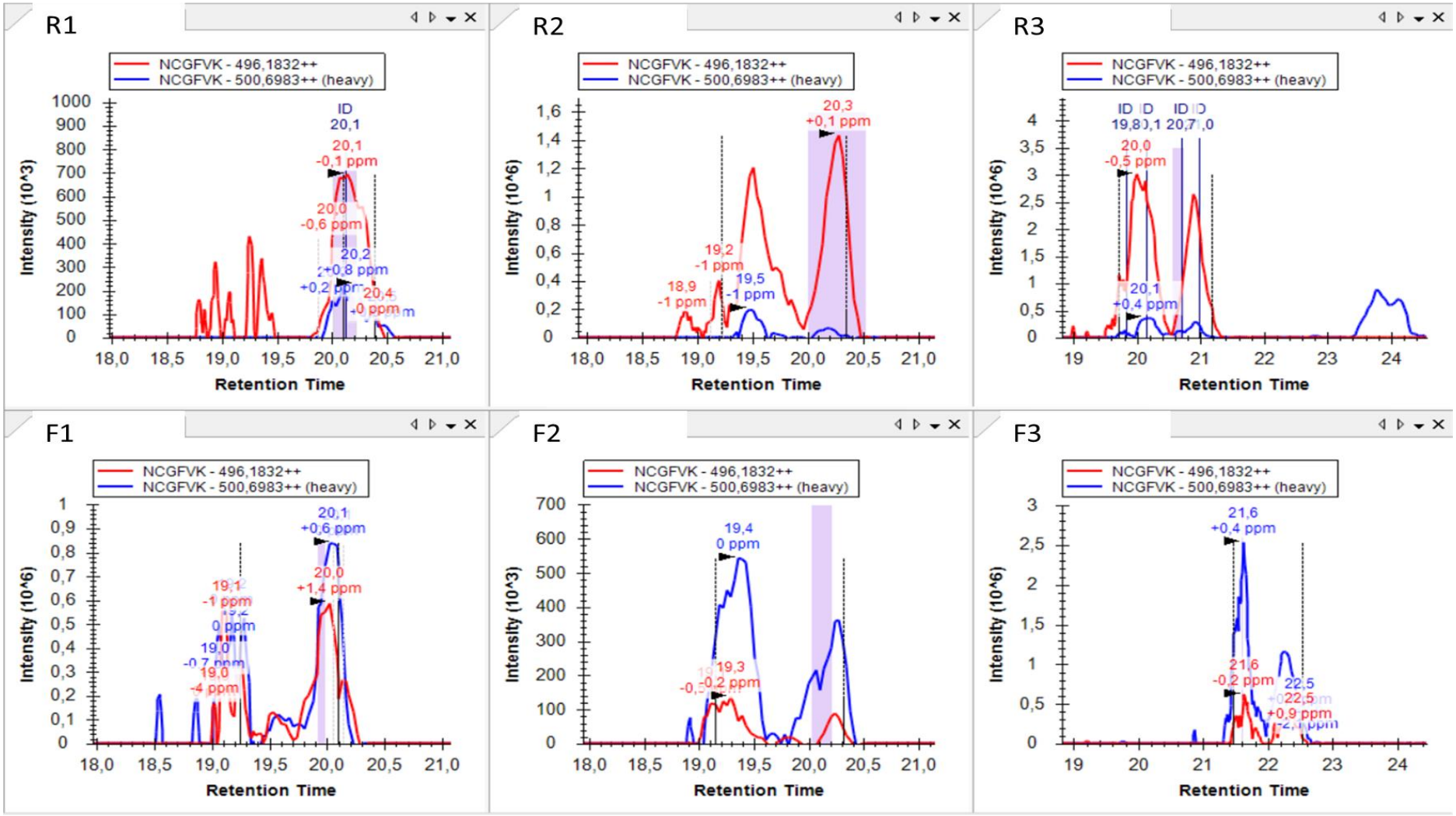
A.6.38: TLYVGGIDGALNSK+UG



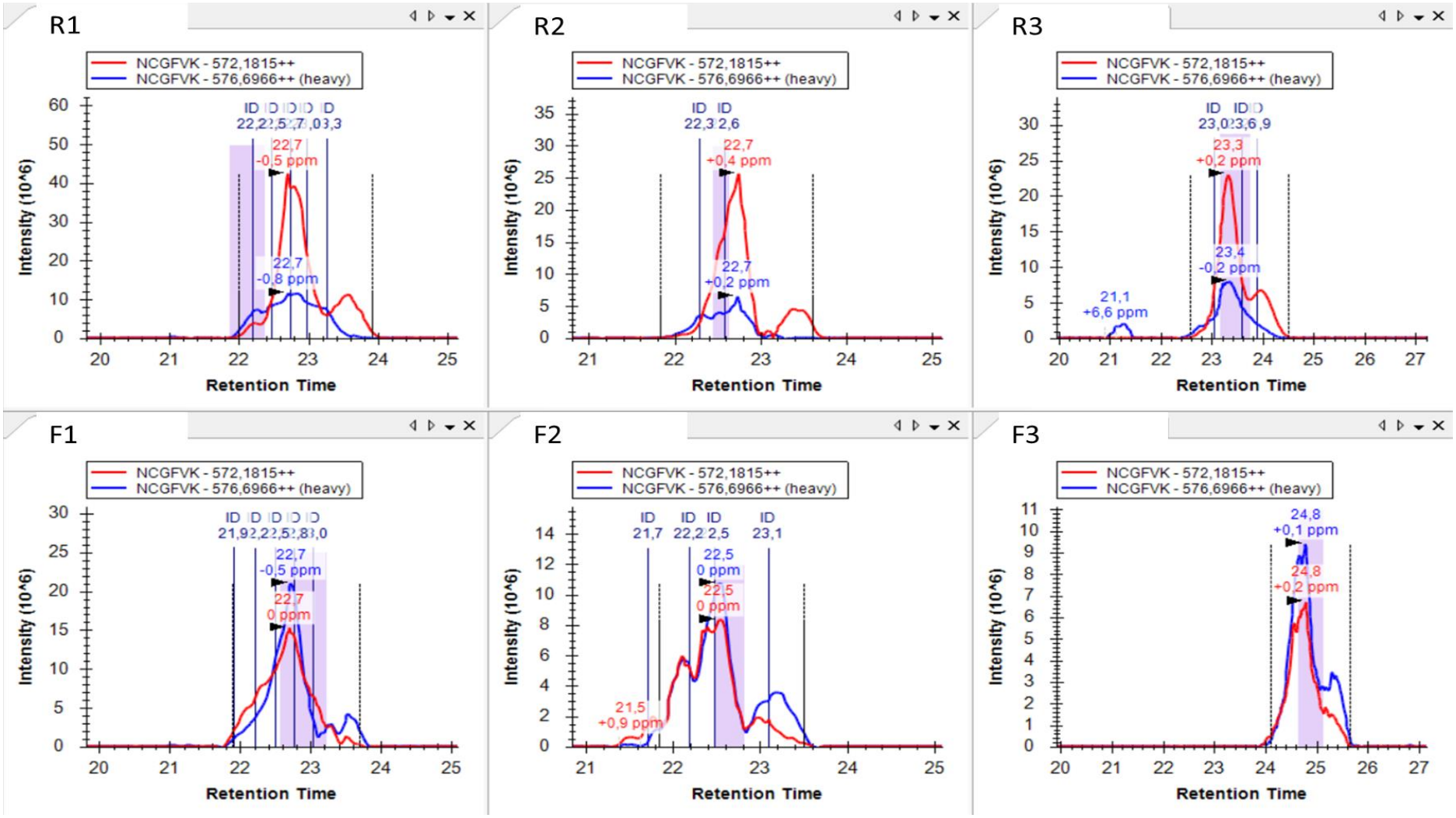
A.6.39: TLYVGGIDGALNSK+UG



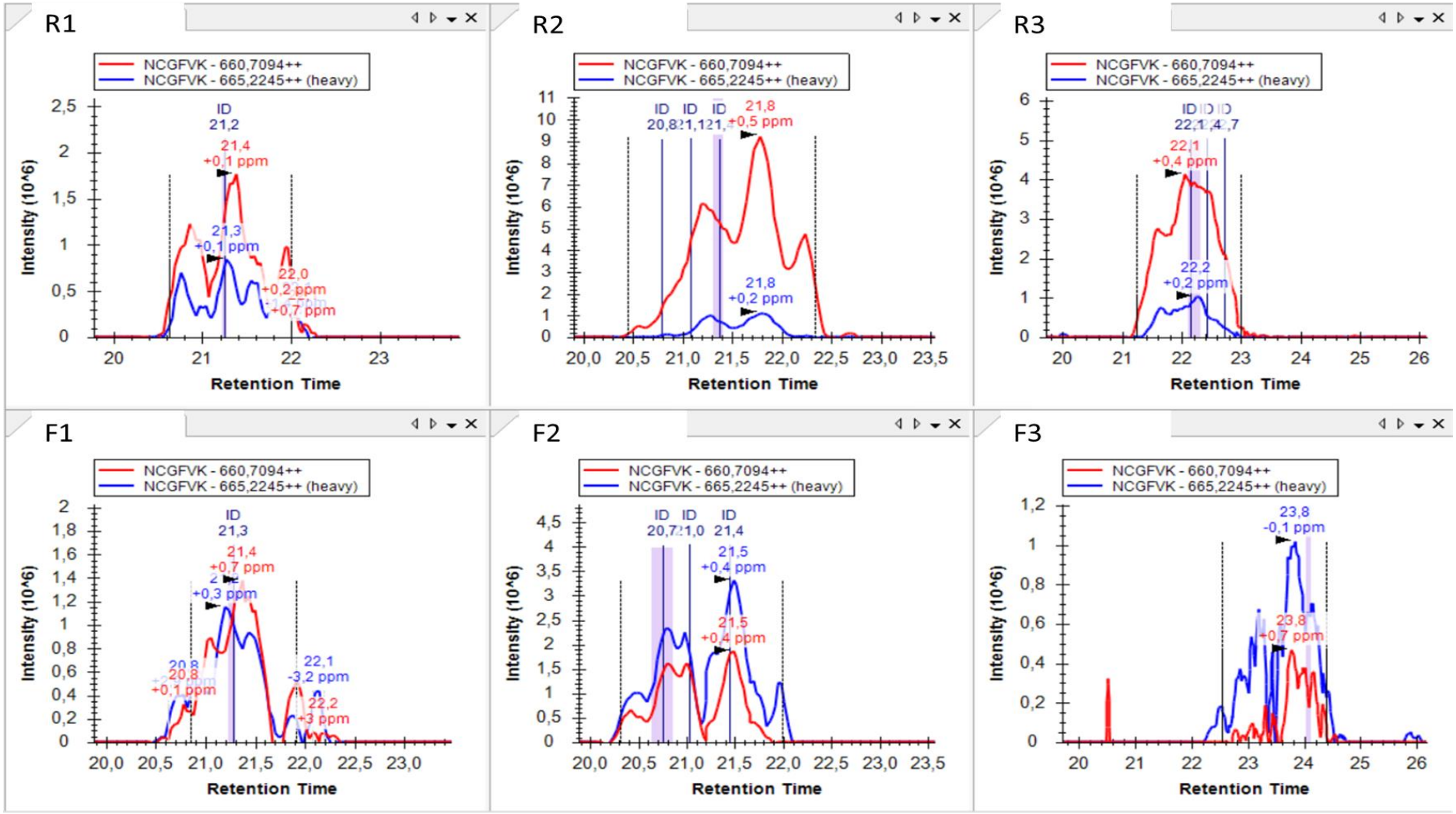
A.6.40: NCGFVK+U



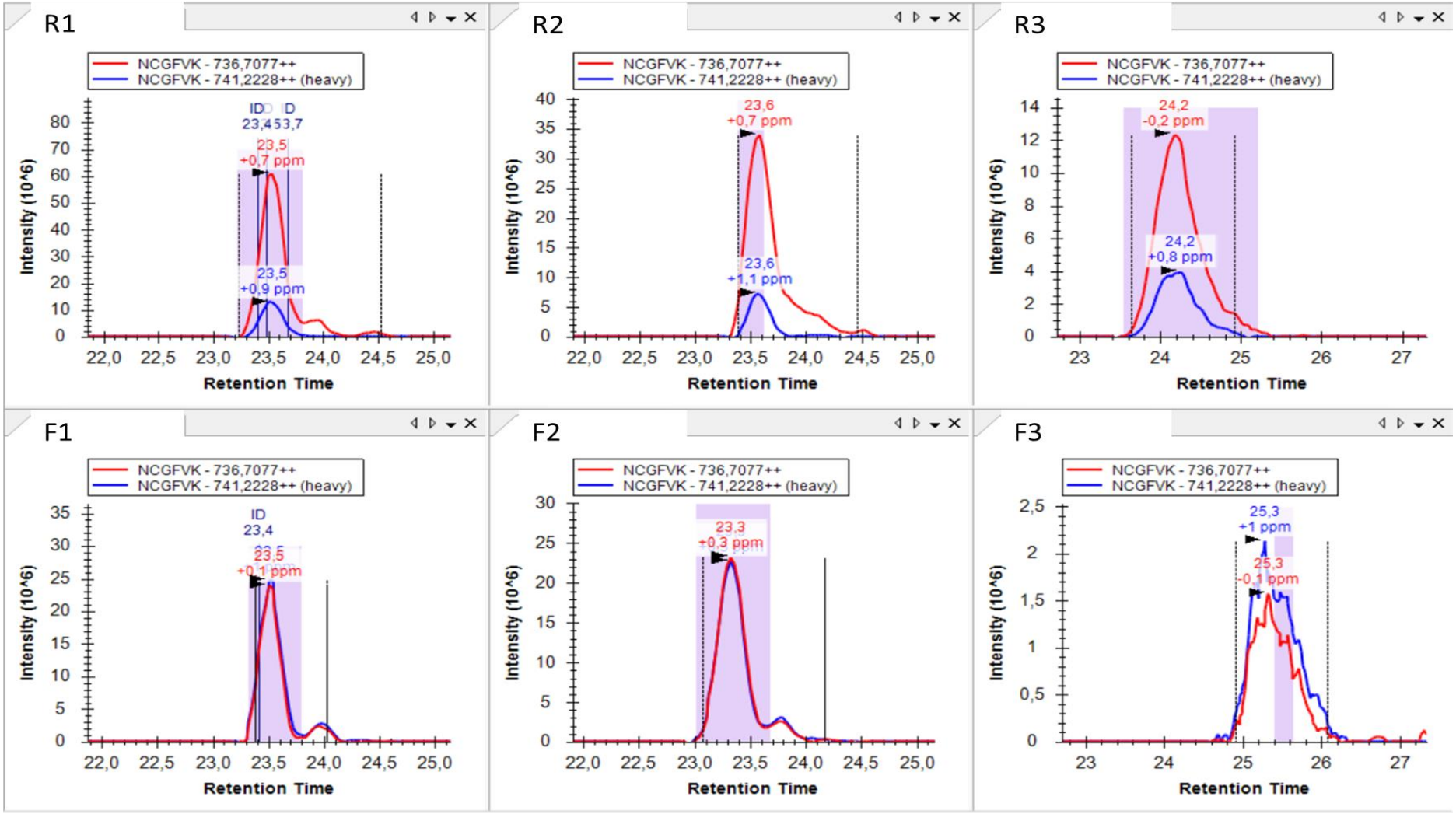
A.6.41: NCGFVK+U+152



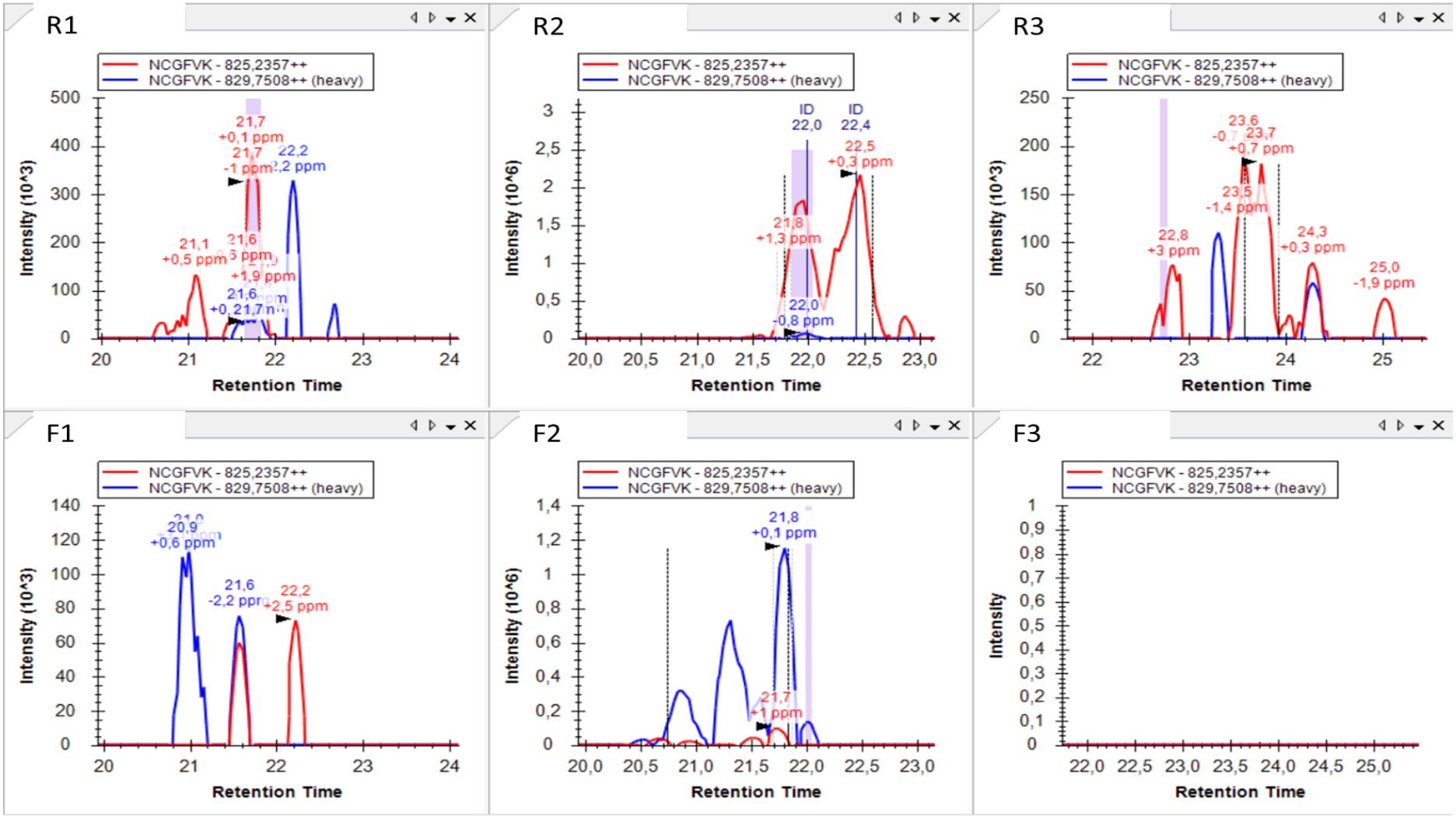
A.6.42: NCGFVK+UA



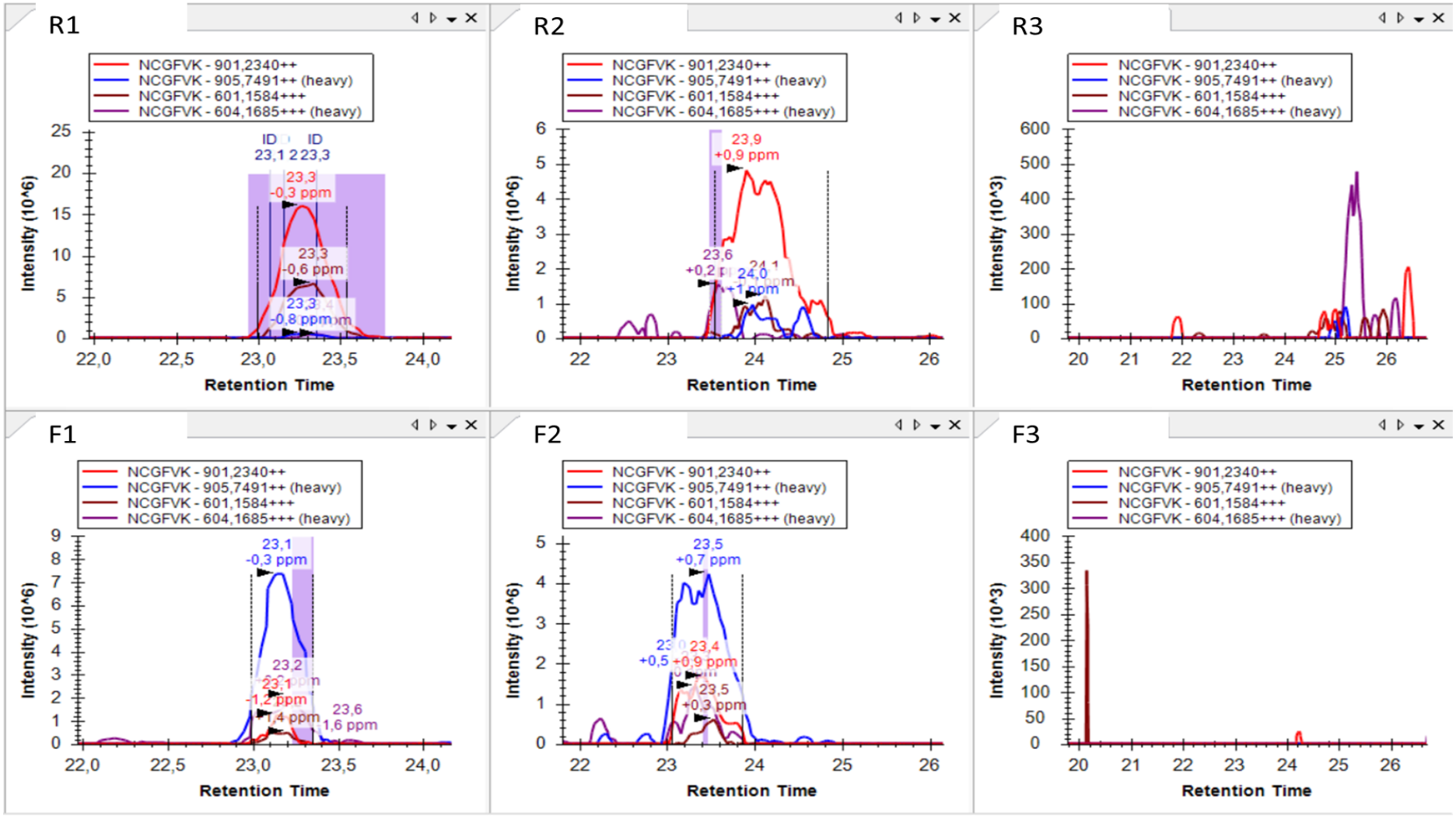
A.6.43: NCGFVK+UA+152



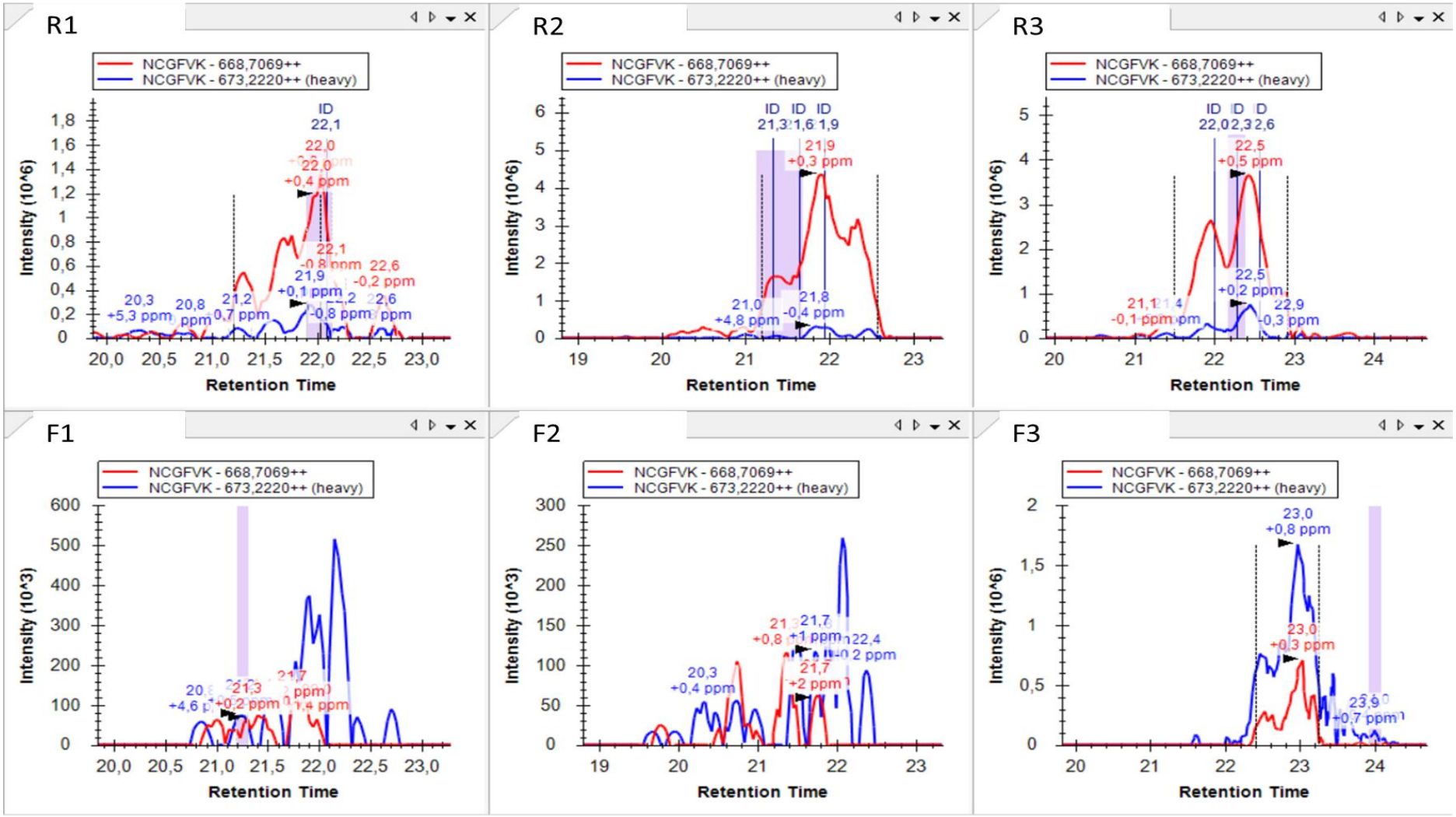
A.6.44: NCGFVK+UAA



A.6.45: NCGFVK+UAA+152



A.6.46: NCGFVK+UG



A.6.47: NCGFVK+UG+152

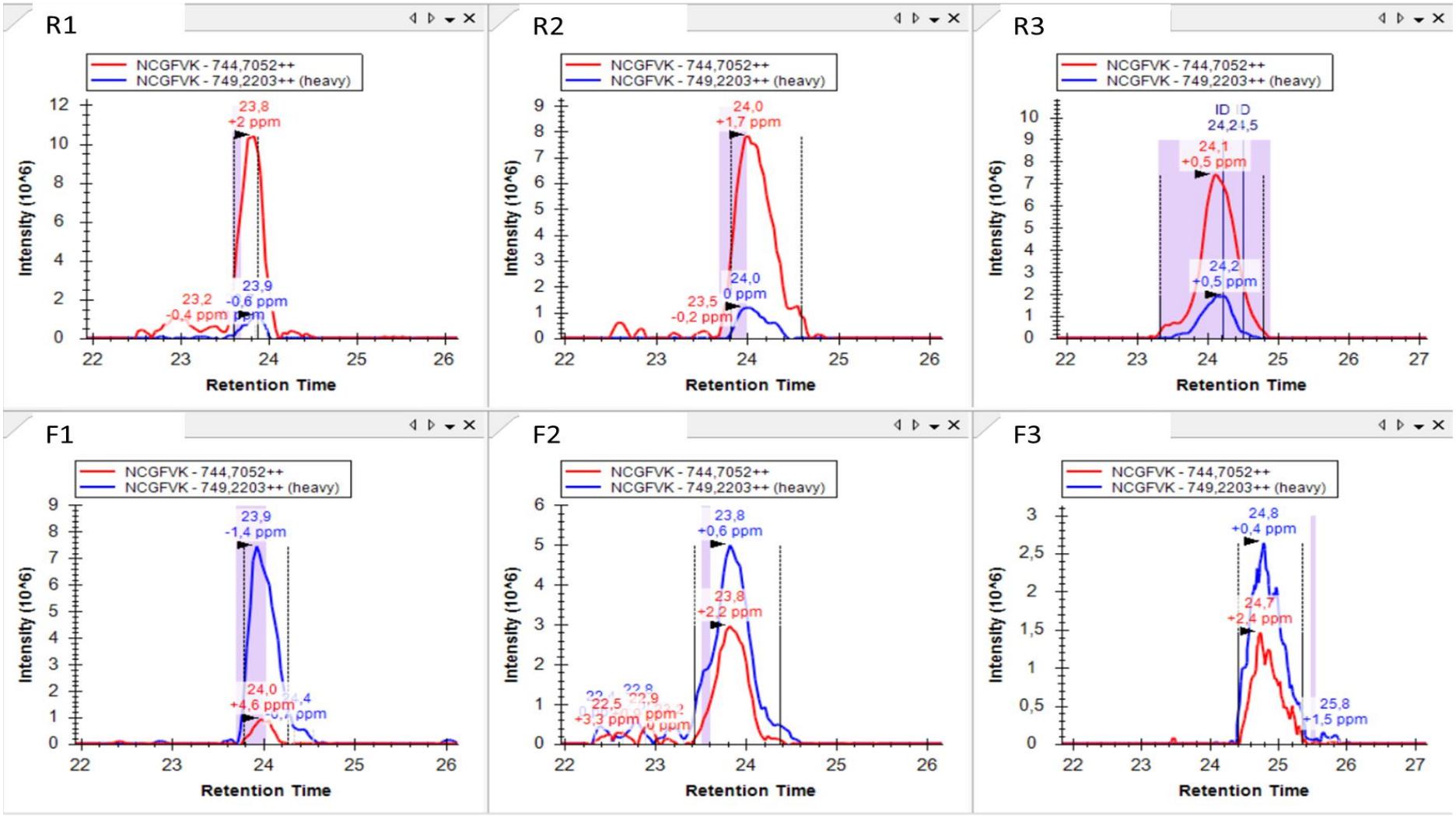
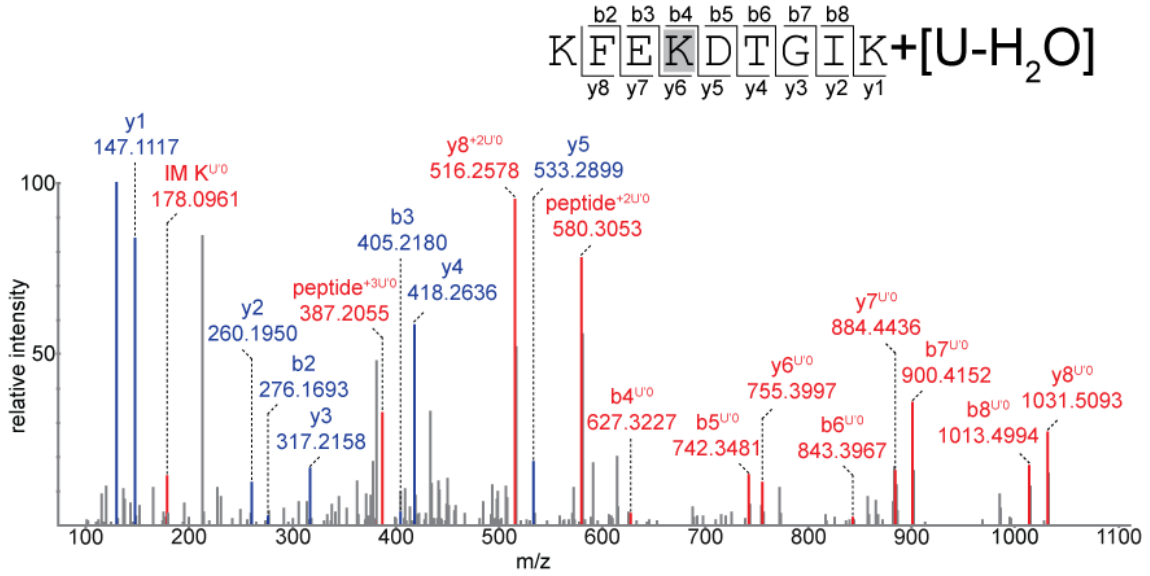


Table 6.9: RNA-protein cross-links identified from MS2-MBP protein complex followed by corresponding MS/MS spectra.

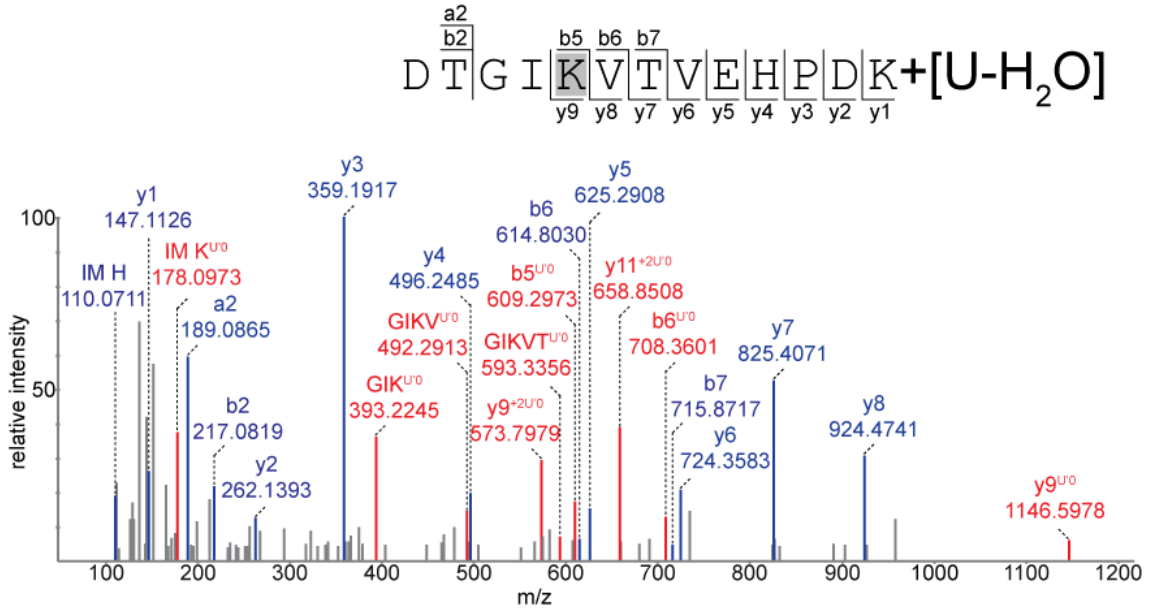
Protein (UniProt ID)	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
MBP (P0AEX9)	^{27}KFEKDTGIK^{35}	K^{30}	U-H_2O	1064.5865	306.0253	1370.6118	3	457.8784	457.8761	-5.02
	^{31}DTGIKVTVEHPDK^{43}	K^{35}	U-H_2O	1437.7463	306.0253	1743.7716	3	582.2650	582.2641	-1.54
	^{36}VTVEHPDKLEEK^{47}	K^{43}	UA-H_2O	1422.7354	635.078	2057.8134	3	686.9456	686.9440	-2.33
	^{68}FGGYAQSGLLAEITPDKAFQDK^{89}	K^{84}	U-H_2O	2355.1746	306.0253	2661.1999	3	888.0744	888.0703	-4.62
	^{121}DLLPNPPKTWEEIPALDK^{138}	^{124}P-T^{129}	U-H_2O	2075.0938	306.0253	2381.1191	3	794.7141	794.7142	0.12
	^{129}TWEEIPALDKELK^{141}	^{138}K/E^{139}	U-H_2O	1570.8242	306.0253	1876.8495	3	626.6243	626.6250	1.12
	^{275}ELAKEFLENYLLTDEGLEAVNK^{296}	^{275}E-L^{285}	U-H_2O	2537.2900	306.0253	2843.3153	3	948.7795	948.7779	-1.68
	^{279}EFLENYLLTDEGLEAVNKDKPLGAVALK^{306}	^{285}L-K^{296}	U-H_2O	3088.6331	306.0253	3394.6584	3	1132.5606	1132.5640	3.00
	^{297}DKPLGAVALK^{306}	^{297}D/K^{298}	U-H_2O	1010.6123	306.0253	1316.6376	2	659.3266	659.3304	5.76
	^{307}SYEEELAKDPR^{317}	^{314}K/D^{315}	U-H_2O	1335.6306	306.0253	1641.6559	3	548.2264	548.2258	-1.09
	^{356}QTVDEALKDAQTNSSSVPGR^{375}	K^{363}	U-H_2O	2102.0239	306.0253	2408.0492	3	803.6908	803.6930	2.73
	^{364}DAQTNSSSVPGRGSIEGR^{381}	^{370}S-R^{375}	UCA	1816.8663	958.13	2774.9963	3	926.0065	926.0037	-3.02
MS2 (P03612)	^{438}KYTIKVEVPK^{447}	K^{442}	U-H_2O	1203.7226	306.0253	1509.7479	3	504.2571	504.2557	-2.77
	^{443}VEVPKGAWR^{451}	K^{447}	U-H_2O	1040.5766	306.0253	1346.6019	2	674.3087	674.3092	0.74

Protein (UniProt ID): Protein name along with its UniProt ID; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotide found to be cross-linked to the peptide; m(calc) Peptide: Theoretical mass of the cross-linked peptide; m(calc) RNA: Theoretical mass of the cross-linked RNA; m(calc) Cross-link: Calculated mass of the RNA-protein cross-link; z: Charge state of the cross-link; *m/z*(calc): Calculated *m/z* of the cross-link by using formula m+z(mH)/z; *m/z*(exp): Observed *m/z* of the cross-link; Δm (ppm): Calculated mass error in ppm by using formula $[m/z(\text{exp}) - m/z(\text{calc})/m/z(\text{calc})] \times 10^6$.

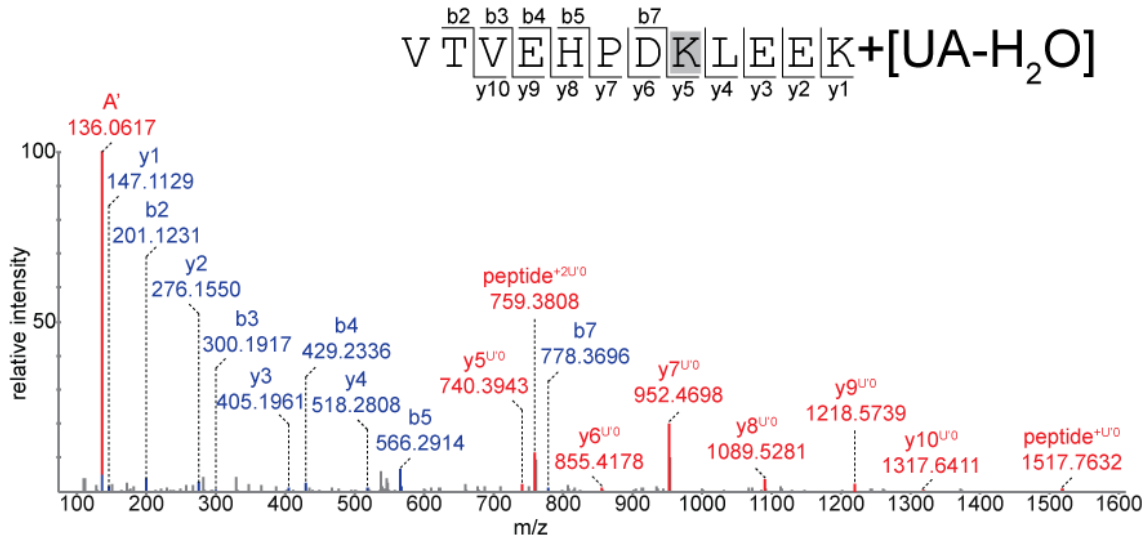
A.6.48


b2 b3 b4 b5 b6 b7 b8
K F E K D T G I K+[U-H2O]
y8 y7 y6 y5 y4 y3 y2 y1
y1 147.1117
IM K U'0 178.0961
y8 +2U'0 516.2578
y5 533.2899
peptide +2U'0 580.3053
b3 405.2180
y4 418.2636
peptide +3U'0 387.2055
y2 260.1950
b2 276.1693
y3 317.2158
b4 U'0 627.3227
b5 U'0 742.3481
y6 U'0 755.3997
y7 U'0 884.4436
b6 U'0 843.3967
b7 U'0 900.4152
b8 U'0 1013.4994
y8 U'0 1031.5093
relative intensity
m/z

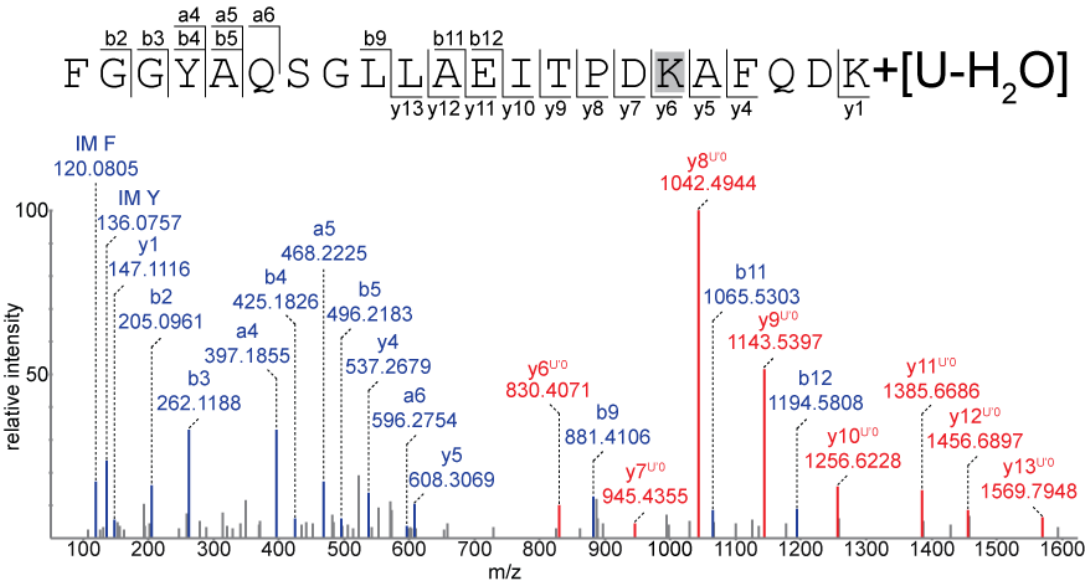
A.6.49


a2
b2
b5 b6 b7
D T G I K V T V E H P D K+[U-H2O]
y9 y8 y7 y6 y5 y4 y3 y2 y1
y3 359.1917
y1 147.1126
y5 625.2908
b6 614.8030
IM H 110.0711
IM K U'0 178.0973
y4 496.2485
b5 U'0 609.2973
y11 +2U'0 658.8508
a2 189.0865
GIKV U'0 492.2913
GIKVT U'0 593.3356
b6 U'0 708.3601
y7 825.4071
GIK U'0 393.2245
b7 715.8717
y9 +2U'0 573.7979
y6 724.3583
y8 924.4741
b2 217.0819
y2 262.1393
y9 U'0 1146.5978
relative intensity
m/z

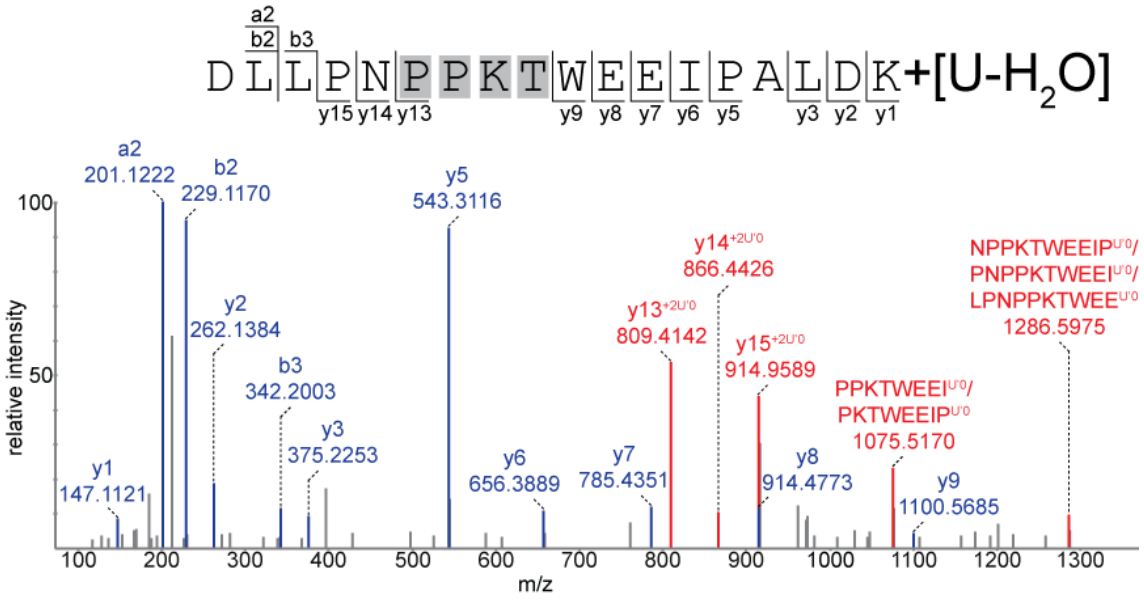
A.6.50


b2 b3 b4 b5 b7
V T V E H P D K L E E K+[UA-H2O]
y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
A'
136.0617
y1
147.1129
b2
201.1231
y2
276.1550
b3
300.1917
b4
429.2336
y3
405.1961
y4
518.2808
b5
566.2914
peptide+2U'0
759.3808
b7
778.3696
y5U'0
740.3943
y7U'0
952.4698
y6U'0
855.4178
y8U'0
1089.5281
y9U'0
1218.5739
y10U'0
1317.6411
peptide+U'0
1517.7632
relative intensity
m/z

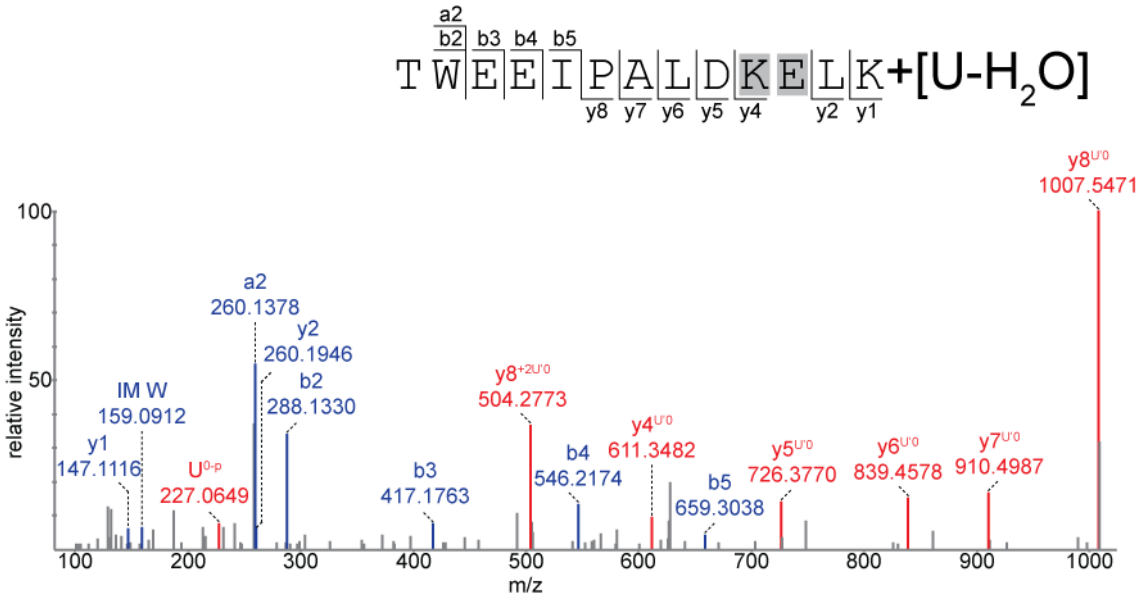
A.6.51


a4 a5 a6
b2 b3 b4 b5 b9 b11 b12
F G G Y A Q S G L L A E I T P D K A F Q D K+[U-H2O]
y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y1
IM F
120.0805
IM Y
136.0757
y1
147.1116
b2
205.0961
a4
397.1855
b3
262.1188
a5
468.2225
b4
425.1826
b5
496.2183
y4
537.2679
a6
596.2754
y5
608.3069
y6U'0
830.4071
b9
881.4106
y7U'0
945.4355
y8U'0
1042.4944
b11
1065.5303
y9U'0
1143.5397
b12
1194.5808
y10U'0
1256.6228
y11U'0
1385.6686
y12U'0
1456.6897
y13U'0
1569.7948
relative intensity
m/z

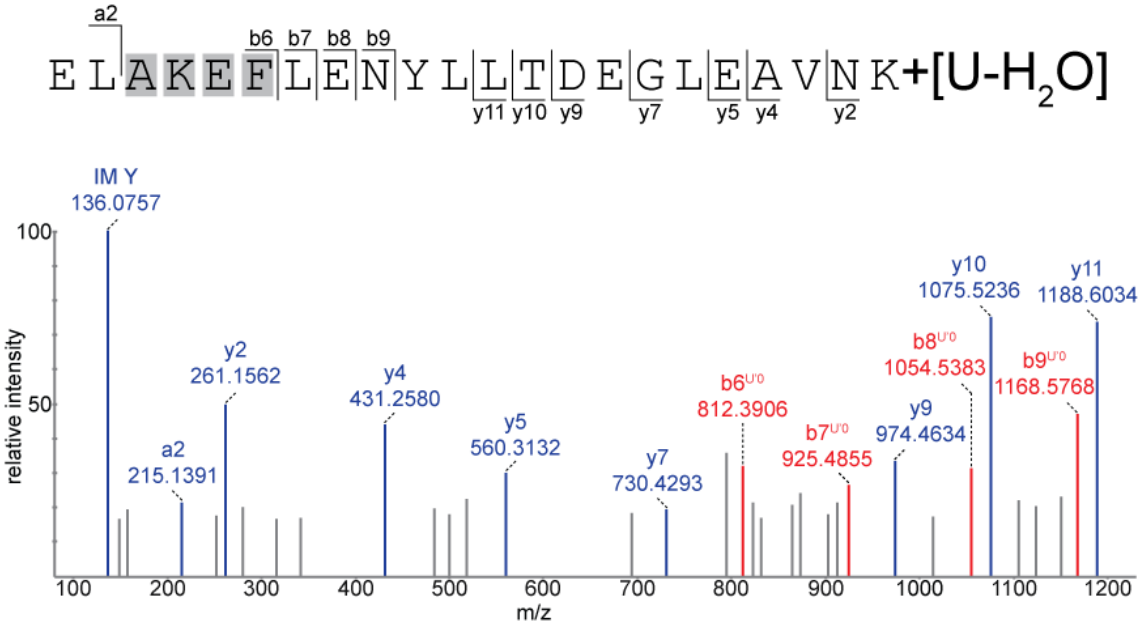
A.6.52


a2
b2 b3
D L L P N P P K T W E E I P A L D K+[U-H2O]
y15 y14 y13 y9 y8 y7 y6 y5 y3 y2 y1
a2 201.1222
b2 229.1170
y5 543.3116
y14+2U'0 866.4426
NPPKTWEEIPU'0/ PNPPKTWEEIU'0/ LPNPPKTWEEU'0 1286.5975
y13+2U'0 809.4142
y15+2U'0 914.9589
y2 262.1384
b3 342.2003
PPKTWEEIU'0/ PKTWEEIPU'0 1075.5170
y3 375.2253
y1 147.1121
y6 656.3889
y7 785.4351
y8 914.4773
y9 1100.5685
relative intensity
m/z

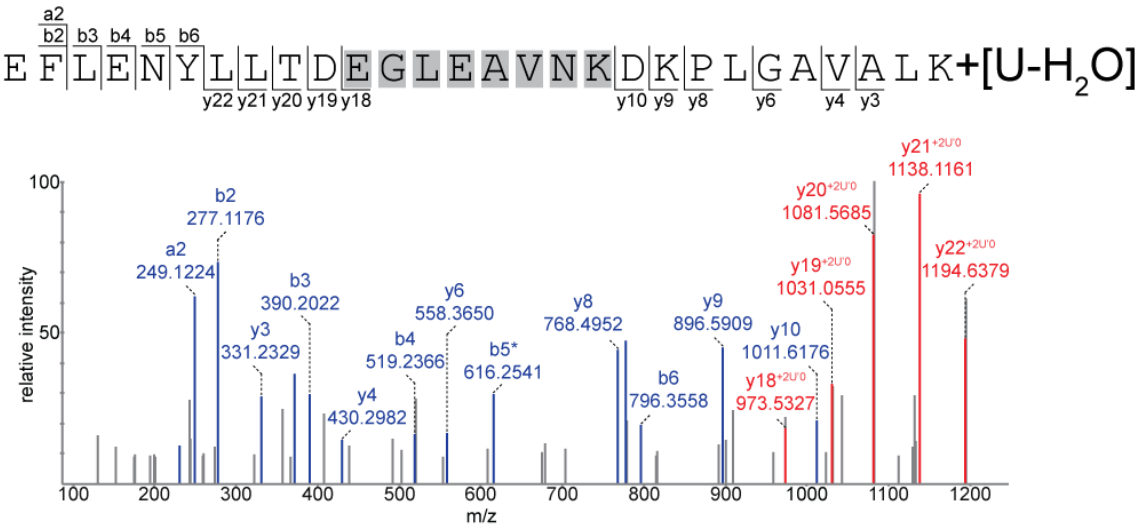
A.6.53


a2
b2 b3 b4 b5
T W E E I P A L D K E L K+[U-H2O]
y8 y7 y6 y5 y4 y2 y1
y8U'0 1007.5471
a2 260.1378
y2 260.1946
y8+2U'0 504.2773
IM W 159.0912
b2 288.1330
y4U'0 611.3482
y1 147.1116
U0-P 227.0649
b3 417.1763
b4 546.2174
y5U'0 726.3770
y6U'0 839.4578
y7U'0 910.4987
b5 659.3038
relative intensity
m/z

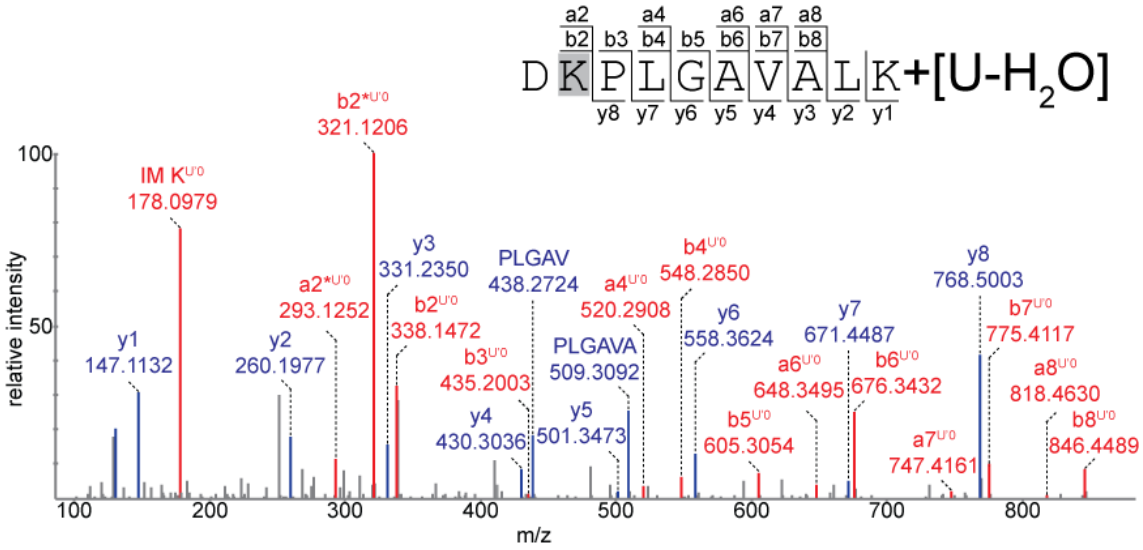
A.6.54


a2
b6 b7 b8 b9
E L A K E F L E N Y L L T D E G L E A V N K+[U-H2O]
y11 y10 y9 y7 y5 y4 y2
IM Y
136.0757
y2
261.1562
a2
215.1391
y4
431.2580
y5
560.3132
y7
730.4293
b6U'0
812.3906
b7U'0
925.4855
y9
974.4634
b8U'0
1054.5383
y10
1075.5236
b9U'0
1168.5768
y11
1188.6034
relative intensity
m/z

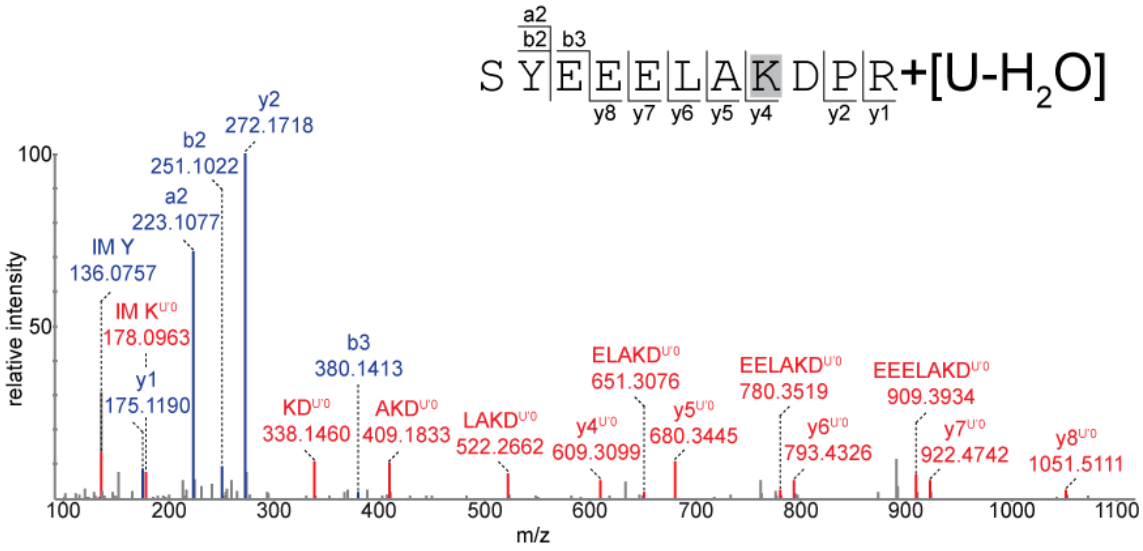
A.6.55


a2
b2 b3 b4 b5 b6
E F L E N Y L L T D E G L E A V N K D K P L G A V A L K+[U-H2O]
y22 y21 y20 y19 y18 y10 y9 y8 y6 y4 y3
b2
277.1176
a2
249.1224
b3
390.2022
y3
331.2329
y6
558.3650
b4
519.2366
b5*
616.2541
y4
430.2982
y8
768.4952
b6
796.3558
y9
896.5909
y10
1011.6176
y18+2U'0
973.5327
y19+2U'0
1031.0555
y20+2U'0
1081.5685
y21+2U'0
1138.1161
y22+2U'0
1194.6379
relative intensity
m/z

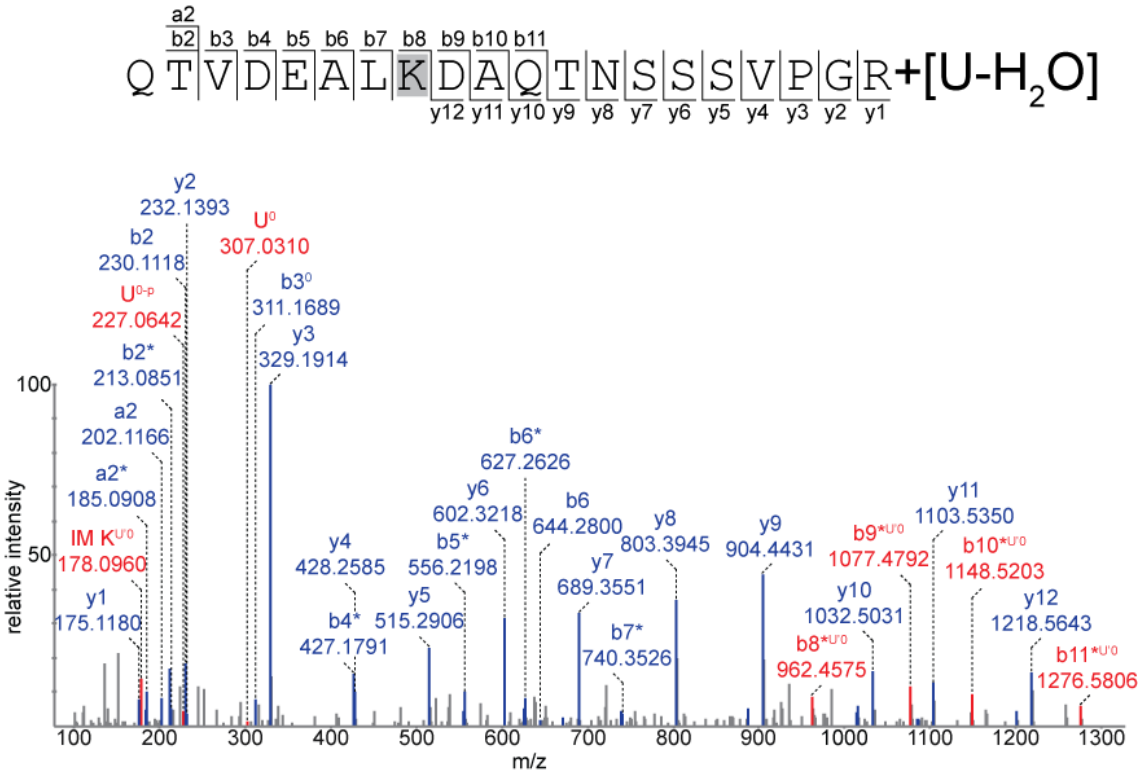
A.6.56


a2 b2 b3 a4 b4 b5 a6 b6 a7 b7 a8 b8
D K P L G A V A L K +[U-H2O]
y8 y7 y6 y5 y4 y3 y2 y1
relative intensity
m/z
b2*U'0 321.1206
IM K U'0 178.0979
y3 331.2350
PLGAV 438.2724
b4U'0 548.2850
a2*U'0 293.1252
a4U'0 520.2908
y6 558.3624
y8 768.5003
y7 671.4487
b7U'0 775.4117
b2U'0 338.1472
y1 147.1132
y2 260.1977
b3U'0 435.2003
PLGAVA 509.3092
a6U'0 648.3495
b6U'0 676.3432
a8U'0 818.4630
y4 430.3036
y5 501.3473
b5U'0 605.3054
a7U'0 747.4161
b8U'0 846.4489

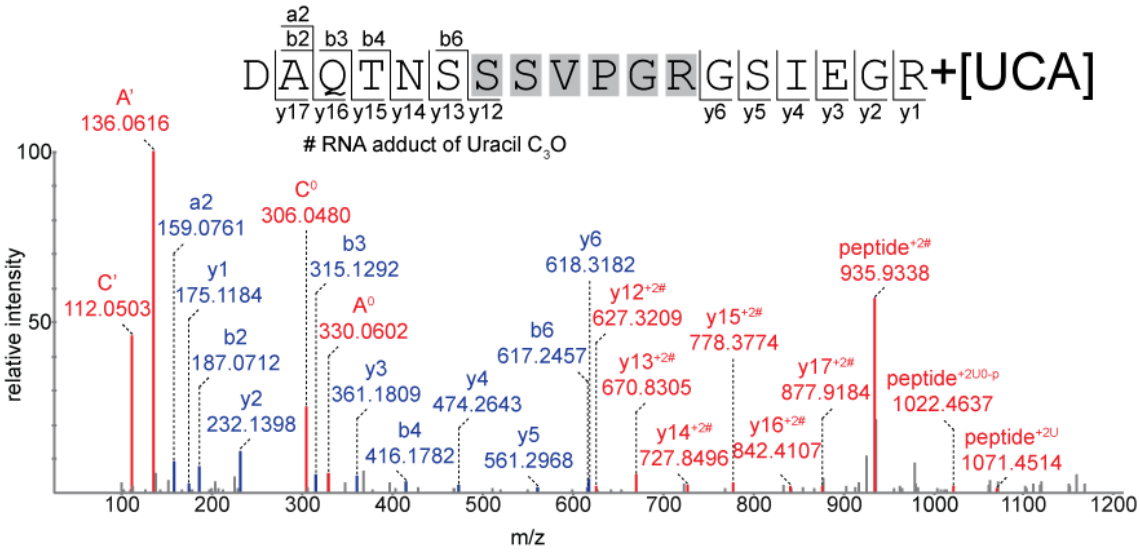
A.6.57


a2 b2 b3
S Y E E E L A K D P R +[U-H2O]
y8 y7 y6 y5 y4 y2 y1
relative intensity
m/z
y2 272.1718
b2 251.1022
a2 223.1077
IM Y 136.0757
IM K U'0 178.0963
y1 175.1190
b3 380.1413
KD U'0 338.1460
AKD U'0 409.1833
LAKD U'0 522.2662
y4U'0 609.3099
ELAKD U'0 651.3076
y5U'0 680.3445
EELAKD U'0 780.3519
y6U'0 793.4326
EEELAKD U'0 909.3934
y7U'0 922.4742
y8U'0 1051.5111

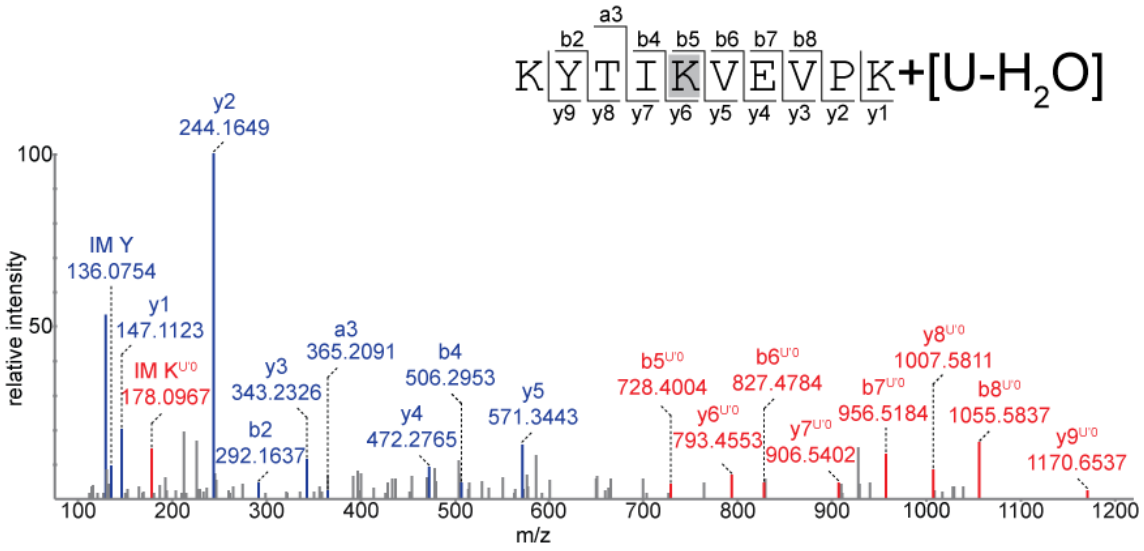
A.6.58


a2 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11
Q T V D E A L K D A Q T N S S S V P G R +[U-H2O]
y12 y11 y10 y9 y8 y7 y6 y5 y4 y3 y2 y1
y2 232.1393
U0 307.0310
b2 230.1118
U0-p 227.0642
b30 311.1689
y3 329.1914
b2* 213.0851
a2 202.1166
a2* 185.0908
IM K U'0 178.0960
y1 175.1180
b6* 627.2626
y6 602.3218
b6 644.2800
y8 803.3945
y9 904.4431
y4 428.2585
b5* 556.2198
y7 689.3551
y5 515.2906
b4* 427.1791
b7* 740.3526
b9*U'0 1077.4792
y11 1103.5350
b10*U'0 1148.5203
y10 1032.5031
y12 1218.5643
b8*U'0 962.4575
b11*U'0 1276.5806
relative intensity
m/z

A.6.59


a2 b2 b3 b4 b6
D A Q T N S S S V P G R G S I E G R +[UCA]
y17 y16 y15 y14 y13 y12 y6 y5 y4 y3 y2 y1
RNA adduct of Uracil C3O
A' 136.0616
C0 306.0480
a2 159.0761
b3 315.1292
y6 618.3182
peptide+2# 935.9338
C' 112.0503
y1 175.1184
A0 330.0602
y12+2# 627.3209
y15+2# 778.3774
b2 187.0712
b6 617.2457
y13+2# 670.8305
y17+2# 877.9184
peptide+2U0-p 1022.4637
y3 361.1809
y4 474.2643
y2 232.1398
b4 416.1782
y5 561.2968
y14+2# 727.8496
y16+2# 842.4107
peptide+2U 1071.4514
relative intensity
m/z

A.6.60


a3
b2 b4 b5 b6 b7 b8
K Y T I K V E V P K+[U-H2O]
y9 y8 y7 y6 y5 y4 y3 y2 y1
y2 244.1649
IM Y 136.0754
y1 147.1123
IM K U'0 178.0967
y3 343.2326
a3 365.2091
b2 292.1637
b4 506.2953
y4 472.2765
y5 571.3443
b5U'0 728.4004
b6U'0 827.4784
y6U'0 793.4553
y7U'0 906.5402
b7U'0 956.5184
y8U'0 1007.5811
b8U'0 1055.5837
y9U'0 1170.6537
relative intensity
m/z

A.6.61

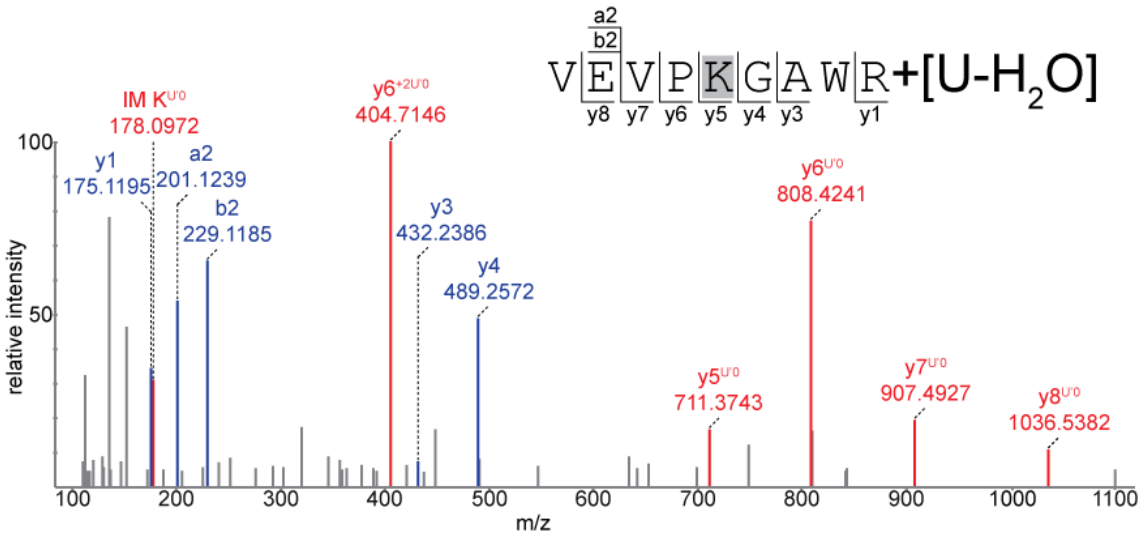

a2
b2
V E V P K G A W R+[U-H2O]
y8 y7 y6 y5 y4 y3 y1
IM KU'0 178.0972
y1 175.1195
a2 201.1239
b2 229.1185
y6+2U'0 404.7146
y3 432.2386
y4 489.2572
y6U'0 808.4241
y5U'0 711.3743
y7U'0 907.4927
y8U'0 1036.5382
relative intensity
m/z

Table 6.10: Selective RNA-protein cross-linking hits, identified from RNP complex from *HeLa* nuclear extract and their calculations.

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
ATP-dependent RNA helicase DDX3X/DDX3Y/Probable ATP-dependent RNA helicase DDX5/DDX17	M(Ox)LDMGFEPQIR	$M^{355}/M^{353}/M^{356}/M^{333}$	U	1351.6264	324.0359	1675.6623	2	838.8389	838.8421	3.81
ELAV-like protein 1	VAGHSLGYGFVNYVTAK	-	UU	1781.9100	630.0612	2411.9712	3	804.9982	804.9986	0.49
	DANLYISGLPR	Y^{109}	U	1217.6404	324.0359	1541.6763	2	771.8459	771.8444	-1.94
	GVAFIR	F^{151}	UU	661.3911	630.0612	1291.4523	2	646.7339	646.7335	-0.61
Endonuclease G, mitochondrial	YQVIGK	Y^{215}-Q^{216}	U-H_2O	706.4013	306.0253	1012.4266	2	507.2211	507.2232	4.14
Far upstream element-binding protein 2	M(Ox)ILIQDGSQNTNVDKPLR	G^{273}-S^{274}	U	2057.0575	324.0359	2381.0934	3	794.7056	794.7044	-1.50
	CGLVIGR	C^{436}	U	716.4003	324.0359	1040.4362	2	521.2259	521.2253	-1.15
Far upstream element-binding protein 3	CGLVIGK	C^{366}	U	688.3941	324.0359	1012.4300	2	507.2228	507.2238	1.97
H/ACA ribonucleoprotein complex subunit 4	IMLPGVLR	M^{316}	AU-H_2O	897.5469	635.0778	1532.6247	3	511.8827	511.8815	-2.34
Heterogeneous nuclear ribonucleoprotein A0	AVSREDSARPGAHAK	-	UGA	1550.7913	998.1358	2548.9271	4	638.2395	638.2401	0.94
Heterogeneous nuclear ribonucleoprotein A1	NQGGYGGSSSSSSYGSGR	Y^{366}	U	1693.6928	324.0359	2017.7287	2	1009.8721	1009.8710	-1.08
Heterogeneous nuclear ribonucleoprotein A1/A1-like 2	LFIGGLSFETTDESLR	F^{23}/F^{23}	AU-H_2O	1783.8991	635.0778	2418.9769	3	807.3334	807.3330	-0.49
	GFAFVTFDDHDSVDK	-	AU	1698.7525	653.0884	2351.8409	3	784.9548	784.9534	-1.78
	AVSREDSQRPGAHLTVK	-	U	1849.9758	324.0359	2174.0117	4	544.5107	544.5093	-2.57
	IEVIEIM(Ox)TDRGSGK	R^{140}/R^{140}	U	1562.7973	324.0359	1886.8332	2	944.4244	944.4310	6.98
	IEVIEIMTDR	M^{137}/M^{137}	U	1217.6325	324.0359	1541.6684	2	771.842	771.8428	1.03
Heterogeneous nuclear ribonucleoprotein A1/A1-like 2/A3	IFVGGIK	$F^{108}/F^{108}/F^{129}$	U	732.4533	324.0359	1056.4892	2	529.2524	529.2517	-1.32

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Heterogeneous nuclear ribonucleoprotein A3	IETIEVMEDR	M^{158}	U	1233.5910	324.0359	1557.6269	2	779.8212	779.8226	1.79
	AVSREDSVKPGAHLTVK	-	U	1792.9794	324.0359	2117.0153	4	530.2616	530.2609	-1.32
Heterogeneous nuclear ribonucleoprotein A/B	IFVGGLNPEATEEK	F^{157}	U	1502.7616	324.0359	1826.7975	2	914.4065	914.4047	-1.96
	EVYQQQQYGSGGR	Y^{240}	U	1498.6800	324.0359	1822.7159	2	912.3657	912.3638	-2.08
	GFGFILFK	-	AU	927.5217	653.0884	1580.6101	2	791.3128	791.314	1.51
	GFVFITFK	F^{197} or F^{199} or F^{202}	U	957.5323	324.0359	1281.5682	2	641.7919	641.7918	-0.15
	M(Ox)FVGGLSWDTSK	W^{78}	U	1342.6226	324.0359	1666.6585	2	834.3370	834.3350	-2.39
Heterogeneous nuclear ribonucleoproteins A2/B1	AVAREESGKPGAHVTVK	R^{99}	U	1734.9376	324.0359	2058.9735	4	515.7511	515.7504	-1.35
	EESGKPGAHVTVK	G^{103}-K^{104}	U-H_2O	1337.6938	306.0253	1643.7191	3	548.9142	548.9133	-1.63
	GFGFVTFDDHDPVDK	-	AU	1694.7576	653.0884	2347.8460	3	783.6231	783.6216	-1.91
	YHTINGHNAEVR	Y^{174}	U	1409.6799	324.0359	1733.7158	3	578.9131	578.9120	-1.90
	ALSRQEM(Ox)QEVQSSR	-	AU	1663.7947	653.0884	2316.8831	3	773.3022	773.3005	-2.19
	IDTIEIITDRQSGK	R^{147}	U	1587.8467	324.0359	1911.8826	3	638.302	638.3021	0.15
Heterogeneous nuclear ribonucleoproteins C1/C2	GFAFVQYVNER	F^{52} or F^{54}	UU	1328.6513	630.0612	1958.7125	2	980.3640	980.3620	-2.04
Heterogeneous nuclear ribonucleoproteins C1/C2/C-like 1	VFIGNLNTLVVK	F^{19}/F^{19}	UU	1315.7863	630.0612	1945.8475	2	973.9315	973.9293	-2.25
Heterogeneous nuclear ribonucleoprotein D0	IFVGGLSPDTPEEK	-	U	1487.7507	324.0359	1811.7866	3	604.9367	604.9354	-2.14
Heterogeneous nuclear ribonucleoprotein D0/D-like	GFGFVLFK	F^{142}/F^{193}	AU	913.5061	653.0884	1566.5945	3	523.2060	523.2049	-2.10
Heterogeneous nuclear ribonucleoprotein K	LLIHQSLAGGIIGVK	G^{157}	U	1517.9292	324.0359	1841.9651	3	614.9962	614.9948	-2.27
	DLAGSIIGK	G^{400}	U	872.4967	324.0359	1196.5326	2	599.2741	599.2725	-2.66
	NAGAVIGK	G^{55}	AU	728.4180	653.0884	1381.5064	2	691.761	691.7574	-5.20
Heterogeneous nuclear ribonucleoprotein L	YGPQYGHPPPPPPPPEYGPH-ADSPVLM(Ox)VYGLDQSK	Y^{386}	U	3799.7979	324.0359	4123.8338	4	1031.9662	1031.9630	-3.10
	NPNGPYPYTLK	Y^{574}	U	1262.6295	324.0359	1586.6654	2	794.3405	794.3460	6.92
Heterogeneous nuclear ribonucleoprotein L/L-like	LNVCVSK	-	U	761.4105	324.0359	1085.4464	2	543.731	543.7314	0.73

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Heterogeneous nuclear ribonucleoprotein M	VGEVTYVELLM(Ox)DAEGK	Y^{100}	U	1767.8600	324.0359	2091.8959	3	698.3064	698.3055	-1.28
Heterogeneous nuclear ribonucleoprotein Q	VTEGLTDVILYHQPDDK	Y^{276}	U	1941.9683	324.0359	2266.0042	3	756.3425	756.3414	-1.45
	DYAFIHFDER	-	AU	1311.5883	653.0884	1964.6767	3	655.9000	655.8990	-1.52
Heterogeneous nuclear ribonucleoprotein Q/R	IKALLER	K^{125}	U-H_2O	841.5385	306.0253	1147.5638	2	574.7897	574.7876	-3.65
	LFVGSIPK	S^{249}/S^{252}	ACU	859.5167	958.1297	1817.6464	2	909.8310	909.8294	-1.75
	VLFVR	F^{341}/F^{344}	U	632.4009	324.0359	956.4368	2	479.2262	479.2254	-1.66
	SAFLCGVM(Ox)K	C^{96}/C^{99}	U	970.4615	324.0359	1294.4974	2	648.2565	648.2547	-2.77
Heterogeneousnuclear ribonucleoprotein R	DYAFVHFEDR	-	U	1297.5727	324.0359	1621.6086	3	541.5440	541.5432	-1.47
	STAYEDYYYHPPPR	-	U	1757.7685	324.0359	2081.8044	3	694.9426	694.9410	-2.30
Heterogeneous nuclear ribonucleoprotein U	M(Ox)CLFAGFQR	C^{594}	U	1087.4942	324.0359	1411.5301	2	706.7728	706.7718	-1.41
	GYFEYIEENK	Y^{257}	U	1290.5768	324.0359	1614.6127	2	808.3141	808.3140	-0.12
Matrin-3	NYILM(Ox)R	Y^{526}	U	824.4214	324.0359	1148.4573	2	575.2364	575.2381	2.95
Nucleolin	VEGTEPTTAFNLFVGNLNFNK	-	AU	2311.1484	653.0884	2964.2368	3	989.0867	989.0855	-1.21
	FGYVDFESAEDLEK	Y^{351}	CU	1647.7303	629.0772	2276.8075	3	767.2710	767.2742	4.17
	NLPYKVTQDELK	Y^{402}	AU	1446.7718	653.0884	2099.8602	3	700.9612	700.9605	-0.99
	GIAYIEFK	Y^{433}	AU	939.5065	653.0884	1592.5949	3	531.8728	531.8720	-1.50
	GYAFIEFASFEDAK	F^{527}	U	1593.7350	324.0359	1917.7709	2	959.8932	959.8922	-1.04
	GFGFVDFNSEEDAK	F^{612} or F^{614} or F^{617}	U	1560.6732	324.0359	1884.7091	2	943.3623	943.3599	-2.54
	TLVLSNLSYSATEETLQEVFEK	Y^{495}	U	2500.2584	324.0359	2824.2943	3	942.4392	942.4390	-0.21
	VFGNEIKLEKPK	K^{377}	U-H_2O	1400.8027	306.0253	1706.828	3	569.9504	569.9516	2.10
	SISLYYTGEK	Y^{462}	U	1159.5760	324.0359	1483.6119	2	742.8137	742.8154	2.28
Nucleolin/Origin recognition complex subunit 1/DNA-binding protein RFX8/Titin	TLFVK	$F^{575}/F^{155}/F^{285}/F^{5783}$	UGG	606.3740	1014.129	1620.503	2	811.2593	811.2587	-0.73
Nucleolysin TIAR	DTSNHFHVFVGDLSPEITTEDIK	-	UU	2600.2394	630.0612	3230.3006	3	1077.7747	1077.7725	-2.04

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Nucleoporin GLE1/Terminal uridylyltransferase 4/Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1	EKFVK	$F^{484}/F^{567}/F^{686}$	U	649.3798	324.0359	973.4157	2	487.7156	487.7149	-1.43
Poly(rC)-binding protein 1/2/3	INEIRQM(Ox)SGAQIK	-	U-H_2O	1502.7874	306.0253	1808.8127	3	603.9454	603.9444	-1.65
Poly(rC)-binding protein 2/3	LVVPASQCGSLIGK	C^{109}/C^{141}	U	1370.7591	324.0359	1694.7950	3	565.9395	565.9387	-1.41
Poly(rC)-binding protein 1/2/3/4	EVGSIIGK	E^{24}-G^{26}/ E^{24}-G^{26}/ E^{56}-G^{58}/ E^{28}-G^{30}	U	801.4595	324.0359	1125.4954	2	563.7555	563.7556	0.177
Poly(U)-binding-splicing factor PUF60	VYVGSIYYELGEDTIR	Y^{138}	U	1875.9254	324.0359	2199.9613	3	734.3282	734.3284	0.27
Polypyrimidine tract-binding protein 1	HQNVQLPR	H^{411}	U	990.5359	324.0359	1314.5718	2	658.2937	658.2924	-1.97
	NFQNIFPPSATLHLSNIPPSVSEEDLK	-	U-H_2O	2993.5133	306.0253	3299.5386	3	1100.8540	1100.8506	-3.08
Probable ATP-dependent RNA helicase DDX5	LIDFLECGK	C^{234}	U	1036.5262	324.0359	1360.5621	2	681.2888	681.2887	-0.14
Probable E3 ubiquitin-protein ligase TRIML2	RLFLEK	-	UU	804.4857	630.0612	1434.5469	2	718.2812	718.2799	-1.80
Proteasome subunit beta type-3	LNLYELK	Y^{74}	U	891.5065	324.0359	1215.5424	2	608.779	608.7781	-1.47
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	YM(Ox)TDGM(Ox)LLREAM(Ox)NDPLLER	-	ACU	2315.0595	958.1297	3273.1892	4	819.3051	819.3037	-1.70
Putative RNA-binding protein 3	GFGFITFTNPEHASVAM(Ox)R	-	U	1996.9464	324.0359	2320.9823	3	774.6686	774.6671	-1.93

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
RNA-binding motif, single-stranded-interacting protein 1/2/3	TNLYIR	$Y^{65}/Y^{59}/Y^{64}$	U	778.4337	324.0359	1102.4696	2	552.2426	552.2431	0.90
RNA-binding protein 39	IESIQLMM(Ox)DSETGR	M^{282}	U	1624.7436	324.0359	1948.7795	3	650.6010	650.6003	-1.07
RNA-binding protein 39/Probable RNA-binding protein 23	LYVGSLHFNITEDM(Ox)LR	-	U	1922.9559	324.0359	2246.9918	3	750.0051	750.0032	-2.53
RNA-binding protein 47/ APOBEC1 complementation factor/Mitotic spindle assembly checkpoint protein MAD2B	ILYVR	$Y^{249}/Y^{234}/Y^{32}$	AU	662.4115	653.0884	1315.4999	2	658.7577	658.7585	1.21
RNA-binding protein FUS	TGQPMINLYTDR	-	AU	1407.6816	653.0884	2060.7700	3	687.9311	687.9281	-4.36
	APKPDGPGGGPGGSHM(Ox)GGNYGDDR	-	U	2267.9613	324.0359	2591.9972	4	649.0071	649.0054	-2.61
Serine/arginine-rich splicing factor 1	IYVGNLPPDIR	Y^{19}	AU	1255.6924	653.0884	1908.7808	3	637.2681	637.2670	-1.72
	YGPPSR	Y^{112}	U	675.3340	324.0359	999.3699	2	500.6927	500.6918	-1.79
Serine/arginine-rich splicing factor 2/8	GFAFVR	F^{59}	U	695.3754	324.0359	1019.4113	2	510.7134	510.7134	0
	VDNLTYR	Y^{23}	U	879.4450	324.0359	1203.4809	2	602.7482	602.7488	0.99
	VGDVYIPR	Y^{44}	U	917.4970	324.0359	1241.5329	2	621.7742	621.7734	-1.28
Serine/arginine-rich splicing factor 3	SVWVAR	W^{40}	UU	716.3969	630.0612	1346.4581	2	674.2368	674.2363	-0.74
	VYVGNLGNNGNK	Y^{13}	U	1247.6258	324.0359	1571.6617	2	786.8386	786.8327	-7.49
Serine/arginine-rich splicing factor 5	VFIGRLNPAAR	G^{9}-R^{10}	CU	1212.7090	629.0772	1841.7862	3	614.9365	614.9358	-1.13
Serine/arginine-rich splicing factor 6	NGYGFVEFEDSR	-	U	1418.6102	324.0359	1742.6461	2	872.3308	872.3293	-1.71
Serine/arginine-rich splicing factor 9	IYVGNLPTDVR	-	AU	1245.6717	653.0884	1898.7601	3	633.9278	633.9265	-2.05

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform/Zinc finger protein 585A/585B/420	RAFIR	$F^{224}/F^{587}/F^{587}/F^{344}$	UU	661.4023	630.0612	1291.4635	2	646.7395	646.7409	2.16
Signal recognition particle 9 kDa protein	VTDDLVCLVYK	C^{48}	U	1266.6529	324.0359	1590.6888	2	796.3522	796.3525	0.37
Small nuclear ribonucleoprotein E	VQKVM(Ox)VQPINLIFR	K^{12}	U-H_2O	1699.9806	306.0253	2006.0059	3	669.6764	669.6810	6.86
Splicing factor 3B subunit 4	NQDATVYVGGLDEK	Y^{16}	U	1507.7154	324.0359	1831.7513	2	916.8834	916.8826	-0.87
Splicing factor U2AF 65 kDa subunit	NFAFLEFR	F^{199}	CU	1042.5235	629.0772	1671.6007	3	558.2080	558.2070	-1.79
	GAKEEHGGLIR	K^{70}	U-H_2O	1165.6203	306.0253	1471.6456	3	491.5563	491.5535	-5.69
	LFIGGLPNYLNDDQVK	L^{261}-F^{262}	U	1804.9359	324.0359	2128.9718	3	710.6651	710.6640	-1.54
	AFNLVKDSATGLSK	D^{293}-S^{294}	UU	1449.7827	630.0612	2079.8439	3	694.2891	694.2908	2.44
Transcriptional activator protein Pur-alpha/beta/Purine-rich element-binding protein gamma	FYLDVK	$F^{74}/F^{47}/F^{73}$	AU	783.4166	653.0884	1436.505	3	479.8428	479.8420	-1.66
U1 small nuclear ribonucleoprotein 70 kDa	RVLVDVER	L^{175}	U	984.5716	324.0359	1308.6075	3	437.2103	437.2100	-0.68
	LGGGLGGTR	L^{196}	UCA	786.4347	958.1297	1744.5644	3	582.5292	582.5285	-1.20
	YDERPGPSPLPHR	Y^{219}	U	1519.7531	324.0359	1843.789	3	615.6041	615.5987	-8.77
Y-box-binding protein 1	RPQYSNPPVQGEVM(Ox)EGADN-QGAGEQGRPVR	Y^{208}-N^{210}	U	3238.5173	324.0359	3562.5532	4	891.6461	891.6450	-1.23
	GAEAANVTGPGGVPVQGSKYAADR	S^{136}-Y^{138}	U-H_2O	2271.1243	306.0253	2577.1496	3	860.0576	860.0518	-6.74
	NYQQNYQNSESGEK	Y^{158}	UA	1687.7074	653.0884	2340.7958	3	781.2730	781.2691	-4.99
Y-box-binding protein 1/2/3	NGYGFINR	$F^{74}/F^{109}/F^{106}$	U	939.4562	324.0359	1263.4921	2	632.7538	632.7525	-2.05
	EDVFVHQTAIK	-	ACU	1285.6666	958.1297	2243.7963	3	748.9399	748.9388	-1.46
Y-box-binding protein 1/Protein TASOR/Proteoglycan 3	NFNYR	$Y^{287}/Y^{145}/Y^{146}$	AU	712.3292	653.0884	1365.4176	2	683.7166	683.7149	-2.48

Continued.....

Protein Name	Peptide	Amino acid	RNA	m(calc) Peptide	m(calc) RNA	m(calc) Cross-link	z	*m/z*(calc)	*m/z*(exp)	Δm (ppm)
40S ribosomal protein S2	TYSYLTPDLWK	Y^{248} or Y^{250}	U	1385.6867	324.0359	1709.7226	2	855.8691	855.8680	-1.28
60S ribosomal protein L5	HIM(Ox)GQNVADYMR	Y^{207}	U	1465.6441	324.0359	1789.6800	3	597.5678	597.5669	-1.50
60S ribosomal protein L6	EKYEITEQR	E^{238}-K^{239}	U-H_2O	1194.5880	306.0253	1500.6133	3	501.2122	501.2118	-0.79
60S ribosomal protein L34	AFLIEEQK	I^{97}	GU	976.5229	669.0833	1645.6062	2	823.8109	823.8096	-1.57

Protein (UniProt ID): Protein name along with its UniProt ID; Peptide: Sequence of the cross-linked peptide along with its position within the protein sequence; Amino acid: One letter symbol of the cross-linked amino acid along with its position within the protein sequence. The amino acids highlighted in grey show the probable amino acids found to be cross-linked as the exact cross-linking site cannot be specified further; RNA: Nucleotide found to be cross-linked to the peptide; m(calc) Peptide: Theoretical mass of the cross-linked peptide; m(calc) RNA: Theoretical mass of the cross-linked RNA; m(calc) Cross-link: Calculated mass of the RNA-protein cross-link; z: Charge state of the cross-link; *m/z*(calc): Calculated *m/z* of the cross-link by using formula m+z(mH)/z; *m/z*(exp): Observed *m/z* of the cross-link; Δm (ppm): Calculated mass error in ppm by using formula $[m/z(\text{exp}) - m/z(\text{calc})/m/z(\text{calc})] \times 10^6$.

*MS2 spectra of the cross-linked peptide hits given in Table 6.10 can be provided in soft copy on the demand of the examiners. Refer to the supplementary material of the published paper entitled “Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins” by Kramer et al., 2014 containing several MS2 spectra.

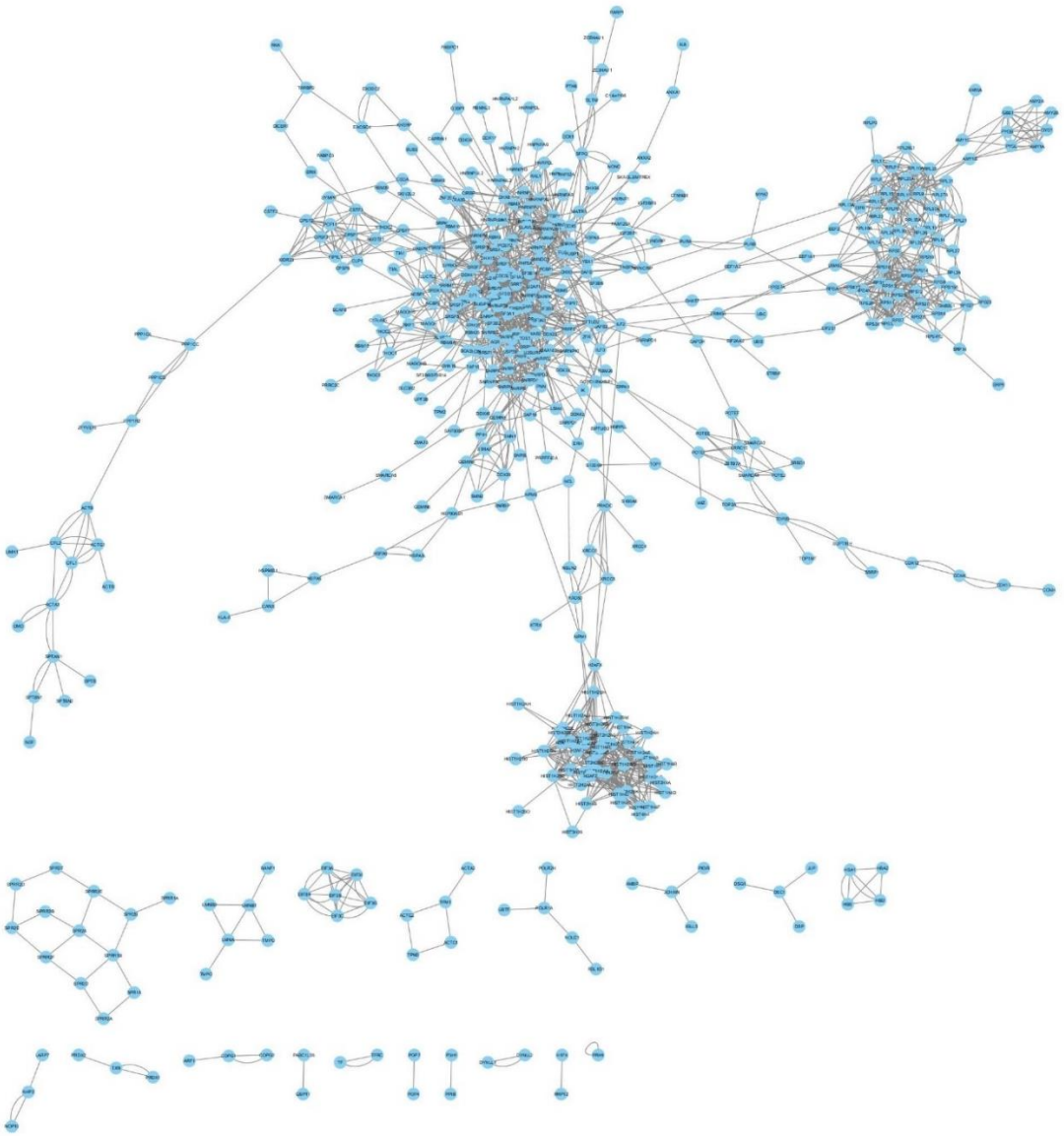


Figure 6.62: Interactome of isolated RNP complex (before cross-linking) from *HeLa* nuclear extract assembled on MINX pre-mRNA. Interactome network was constructed based on the information from the experiments. The structural and functional relationships are shown as edges that are connecting individual proteins (nodes). Nodes were labeled according to the gene names as reported by STRING database (Jensen et al., 2009). Mapping of Protein-Protein interactions revealed a network with 493 nodes and 1865 edges. Analysis of various attributes of the interaction map presented an average degree of 7.57 having a network diameter of 19. Average path length was calculated to be 6.29. Graph density was found to be 0.008. 299 components were found to be strongly connected whereas average clustering coefficient was found to be 0.162.

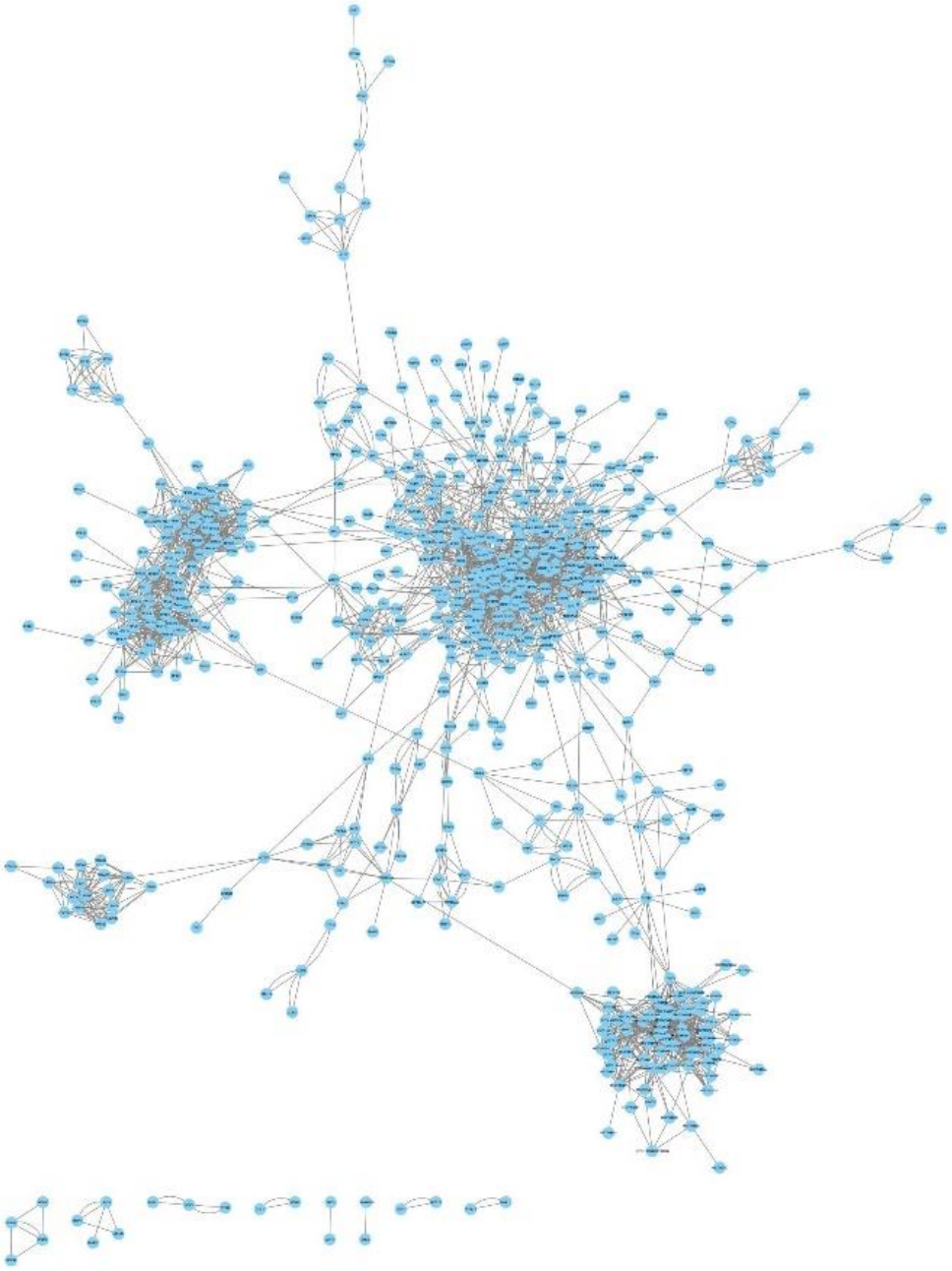


Figure 6.63: Interactome of isolated RNP complex (before cross-linking) from *HeLa* nuclear extract assembled on PM5 pre-mRNA. Interactome network was constructed based on the information from the experiments. The structural and functional relationships are shown as edges that are connecting individual proteins (nodes). Nodes were labeled according to the gene names as reported by STRING database (Jensen et al., 2009). Protein-Protein Interaction (PPI) analysis of complex assembled on PM5 pre-mRNA produced a network with 596 nodes and 2603 edges. Network over view rendered an average degree of 4.37 with the network diameter of 14. 364 components were found to be strongly connected in the network. Average path length presenting the average distance between pairs of nodes was calculated to be 4.98. Graph density was found to be 0.007. Average clustering coefficient is the overall indication of how the nodes are embedded in the neighborhood with in the network and was calculated to be 0.169.