DISS. ETH NO. 26775

Cross-linking methods to identify the mRNA targets of microRNAs

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

ANNA LUIZA MALINOWSKA

(Magister inżynier in Medicinal Chemistry, Lodz University of Technology)

born on 03.07.1988

citizen of Poland

accepted on the recommendation of

Prof. Dr Jonathan Hall, examiner

Prof. Dr Karl-Heinz Altmann, co-examiner

2020

Acknowledgements

Completion of the presented work would not be possible without the input and immense support of many people, to whom I am incredibly grateful.

First and foremost, I would like to thank Prof. Dr Jonathan Hall for giving me the opportunity to conduct my doctoral research in his group. I very much appreciate all the scientific discussions and valuable guidelines that he gave me throughout the last five years. I am also grateful that he allowed me to follow my curiosity in research.

I would also like to thank:

- Prof. Dr Karl-Heinz Altmann, for kindly acting as a co-examiner during my PhD defence and for devoting time to read through this work.
- Sylvia Peleg, for administrative support and daily optimism.
- Mauro Zimmermann, for all the help with the synthesizer, RP-HPLC, LC-MS and solving IT problems with great music in the background.
- Dr Pascal Röthlisberger, for long discussions about the science, all the help with the chemistry-related troubleshooting, sharing experimental tips and tricks and keeping me motivated. And for correcting a big part of this thesis.
- Dr Alice Ghidini, for being there to discuss all types of scientific and non-scientific problems, valuable suggestions and ideas, motivation, sarcasm and proofreading of parts of this thesis.
- Timo Hagen and Dr Helen Lightfoot, for the collaboration on the G-quadruplex project.
- Dr Yuluan Wang, for the collaboration on the miR-CLIP project and all the explanations related to the miR-CLIP technique.
- Dr Ugo Pradère, for the introduction to the CLICK chemistry and probe synthesis.
- Dr Paulina Ćwiek, for introducing me to the cell culture world, great sense of humour and endearing directness.
- Artur Łaski, for answering biological questions, teaching me how to run the luciferase assay and all the help with gel electrophoresis.
- Martina Bigatti, for equipping our lab with a solid stock of the modified phosphoramidites.
- Dr Mirjam Menzi, for the introduction to thermal melting studies.
- Anna Knörlein, for the help with the digestion assay and IE-HPLC.
- Alexandra Title, for testing of the anti-miRs.
- Dr François Halloy, for sharing the ups and downs of being a PhD student and a great company in the writing room.
- Dr Christian Berk, for the discussions on all kinds of topics and help with the IE-HPLC.
- Verena Schlösser, for the commitment to organize things in the lab and for talks over lunch.
- Dr Meiling Li, for being an inspiration during the first years of my PhD.
- Dr Elodie Decuypère, for the discussions on chemistry and not only.
- Dr Alok Behera, for being such a positive person. And for bringing the sunflowers!
- Dr Pavithra Iyer, for the company over lunch.
- Evangelia Lekka, for the exceptional kindness and amazing work attitude.
- Anastasiia Lepikhina, for being an optimistic and easy-going person.

- Dr Alyssa Hill, for help in solving English language dilemmas.
- Jonas Römer, for the contribution to the digestion assay.

I would like to thank Dr Bernhard Pfeiffer for all the help related to the NMR analysis and the Molecular and Biomolecular Analysis Service team for the mass analysis and helpful discussions.

I am very grateful to Paulina, Alice and François for all the Glühweins, dinners, trips, talks and fun. And for the support that I have received from you, especially during the last months of my PhD.

Special 'Thank You' goes to Katarzyna Karcz, for being there for all these years as a great and extremely supportive and understanding friend. I would also like to acknowledge Katarzyna Konior and Magdalena Wellicome, for keeping their fingers crossed and for raising my spirits.

Last but not least, I wish to express my deepest gratitude to my family, and especially to my Mom. For her faith in me, love and never-ending support.

Table of Contents

Abstract

microRNAs (miRNAs, miRs) constitute a class of small, endogenous, non-coding RNAs (ncRNAs) and have a great influence on various processes within the cell. By base-pairing selectively to partially-complementary sites located predominantly in the 3'-untranslated region (3'-UTR) of target messenger RNAs (mRNAs), miRNAs participate in the post-transcriptional regulation of gene expression. miRNAs are responsible for controlling the expression of the majority of human protein-coding genes and their dysregulation has been related to many pathological processes and diseases, including cancer. Thus, an in-depth understanding of the miRNA mechanisms of action is of great importance. One of the key challenges is the elucidation of the exact sites of the canonical and non-canonical interactions in a cell-specific context. An imperfect pairing between miRNAs and their target RNAs in animals, as well as high false-positive and false-negative rates for current prediction algorithms, generate a need for the experimental validation of the predicted binding sites.

For this purpose, various cross-linking-based methods for target identification have been established. The microRNA cross-linking and immunoprecipitation (miR-CLIP) approach developed previously in our group allows capturing predicted and unpredicted miRNA targets in cells. The miR-CLIP technique employs the pre-miR-CLIP probes site-specifically modified with trioxsalen and biotin moieties introduced pre- or post-synthetically at the 2'-O-position of the nucleoside by the copper(I)-catalysed azide-alkyne cycloaddition (CuAAC/'CLICK') reaction. It was shown that a miR-106a-5p pre-miR-CLIP probe was able to cross-link to the complementary regions present in the RNA targets. However, further tests suggested that the cross-linking with this probe design is more sequence-dependent than initially expected and therefore the outcome of the reaction is more difficult to predict.

The overall objective of this research was to design miR-CLIP probes for all miRNAs of a cell, for which an understanding of the cross-linking chemistry with respect to sequence preferences, linker composition and sites of conjugation was required.

The first project described in this thesis focused on exploring the nature of the cross-linking with the original miR-106a-5p miR-CLIP probe. *In vitro* cross-linking assays performed with a set of trioxsalen-labelled miR-106a-5p analogues and probes for other miRNAs allowed us to establish a working hypothesis about the sequence preferences of cross-linking. The results suggested that inter-strand cross-linking most likely took place with a uridine positioned 2-3 base pairs upstream from the desired (juxtaposed) cross-linking site. This allowed proposing an alternative strategy for the positioning of the trioxsalen modification within these probes.

In the second project, diazirine-based probes were synthesized and tested. The results showed that the attachment of the cross-linker at the 2'-O-position of the ribose or C^5-position of the pyrimidine was ineffective and resulted in quenching of reactive species by the solvent molecules.

Finally, a different approach aiming to achieve a versatile (sequence-independent) cross-linking was developed and tested. The new strategy was based on the utilization of probes in which the trioxsalen molecule was introduced at the N^4-position of cytidine through short poly(ethylene glycol) (PEG)-based linkers. Probes for miR-124-3p were functionalized post-synthetically by the application of the convertible nucleoside approach and showed more efficient cross-linking than analogous probes prepared with the use of the CLICK chemistry.

Résumé

Les microARNs (miARNs) sont une classe de petits ARNs endogènes et non codants (ARNnc). Ils exercent une grande influence sur plusieurs processus cellulaires. Les miARNs participent à la régulation post-transcriptionnelle de l'expression des gènes *via* leur hybridation à des sites spécifiques, partiellement complémentaires, généralement situés dans la région 3'-UTR (*three prime untranslated region* en anglais) des ARN messagers (ARNm) cibles. Les miARNs sont responsables du contrôle de l'expression de la majorité des gènes humains codants pour des protéines, et leur dérégulation a été liée à de nombreuses dégénérescences et évolutions pathologiques, dont le cancer. Une compréhension approfondie des miARNs et de leurs mécanismes d'action revêt ainsi une grande importance. Un des principaux défis de la recherche est d'identifier les sites exacts des interactions canoniques et non-canoniques des miARNs en milieu intra-cellulaire. En effet, un appariement imparfait entre les miARNs et leurs ARNs cibles a été noté chez divers organismes, et les algorithmes de prédiction actuels génèrent des taux élevés de faux-positifs et de faux-négatifs, tous ces éléments soulignant un besoin de validation expérimentale des sites prédits d'hybridation.

Dans ce but, plusieurs méthodes basées sur la réticulation ont été établies pour l'identification des cibles. L'approche de réticulation et d'immunoprécipitation des microARNs (abrégée « miR-CLIP » en anglais), développée précédemment dans notre groupe, a permis de capturer *in cellulo* des cibles de miARNs prédites et non-prédites. La technique miR-CLIP utilise des sondes modifiées avec des groupes trioxsalène et biotine à des positions spécifiques. Ces groupes sont ajoutés pré- ou post-synthèse en position 2'-*O* du nucléoside par cycloaddition azide-alcyne catalysée par le cuivre (I) (chimie « CLICK »). Au cours du miR-CLIP pour le miR-106a-5p, il a été démontré que la sonde se réticulait bien avec les régions complémentaires présentes dans les ARN cibles. Cependant, des investigations complémentaires ont aussi révélé que la réticulation de la sonde était plus dépendante de la séquence qu'anticipé initialement, ce qui pourrait rendre le résultat de la réaction plus aléatoire.

Le premier projet de cette thèse a été dédié à une exploration détaillée de la réticulation de la sonde miR-CLIP présentée précedemment. Les tests de réticulation *in vitro*, effectués avec un ensemble d'analogues de miR-106a-5p mono-marqués avec un trioxsalène, et également avec des sondes pour d'autres miARNs, nous ont permis d'établir des hypothèses mécanistiques sur le comportement de nos sondes miR-CLIP. Nos résultats indiquent que la réticulation a très probablement hors du site de réticulation souhaité avec une uridine positionnée 2-3 paires de bases en amont sur la séquence complémentaire. Ces résultats nous ont permis de proposer un design amélioré pour placer la modification trioxsalène sur les sondes miR-CLIP.

Dans un deuxième projet, des sondes comprenant des groupes réticulants diazirine ont été synthétisées et testées. Il a été démontré que le positionnement de l'agent de réticulation en 2'-*O* du ribose ou en C^5 des pyrimidines était inefficace, et entraînait une réaction non désirée avec les molécules de solvant.

Enfin, une dernière approche visant à obtenir une méthode polyvalente de réticulation a été développée et testée. Cette stratégie s'est basée sur l'utilisation d'un mélange de sondes, dans laquelle la molécule de trioxsalène a été introduite en position N^4 d'une cytidine *via* des linkers courts à base de motifs polyéthylène glycol (PEG). A cet effet, des sondes pour le miR-124-3p ont été fonctionnalisées de manière post-synthétique, en utilisant la méthodologie du nucléoside convertible. Ces sondes ont montré une réticulation plus efficace que leurs analogues préparés par chimie CLICK.

List of abbreviations

2'-F/ FNA – 2'-fluoro
2'-OMe – 2'-*O*-methyl
2-MOE – 2'-O-methoxyethyl
3'-UTR – 3'-untranslated region
4SU – 4-thiouridine
5'-UTR – 5'-untranslated region
8-MOP – 8-methylpsoralen

A – adenosine
Å – angstrom
A_{260} – absorbance at 260 nm
ACN – acetonitrile
ac-pre-miRNA – Argonaute-cleaved pre-miRNA
ADAR – adenosine deaminase acting on RNA
ADPKD – autosomal dominant polycystic kidney disease
AGO – Argonaute
AMT – 4'-aminomethyltrioxsalen
anti-miR/ AMO – anti-microRNA/ anti-microRNA oligonucleotide
ASGPR – asialoglycoprotein receptor
ATP – adenosine triphosphate

Boc – *tert*-butyloxycarbonyl

C – cytidine
C. elegans – Caenorhabditis elegans
CD – circular dichroism
cDNA – complementary DNA
ceRNA – competing endogenous RNA
cEt – constrained ethyl
circRNA – circular RNA
CL – cross-linking
CLIP – cross-linking and immunoprecipitation
CPG – controlled pore glass
CTCL – cutaneous T-cell lymphoma
CTX – chlorotoxin
CuAAC – copper(I)-catalysed azide-alkyne cycloaddition
Cy3 – cyanine 3
Cy5 – cyanine 5

D – duplex
DBU – 1,8-diazabicyclo[5.4.0]undec-7-ene
dC – deoxycytidine
DCM – dichloromethane
DGCR8 – DiGorge syndrome chromosomal region 8
DMAP – 4-dimethylaminopyridine
DMF – *N,N*-dimethylformamide
DMSO – dimethyl sulphoxide
DMTr – 4,4-dimethoxytrityl

DNA – deoxyribonucleic acid
dsRNA – double-stranded RNA
dT – deoxythymidine

EDC – 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA – ethylenediaminetetraacetic acid
EGFR – epidermal growth factor receptor
EMT – epithelial to mesenchymal
eq – equivalent
EtOAc – ethyl acetate
EtOH – ethanol

fc ctstr – fully complementary counter-strand
FDA – Food and Drug Administration
FRET – Förster resonance energy transfer
FSCN1 – Fascin Actin-Bundling Protein 1

g – gram (in the experimental part)
g – guide
G – guanosine
GalNAc – *N*-acetylgalactosamine
GTP – guanosine triphosphate

h – hour
HCl – hydrochloric acid
HCV – Hepatitis C virus
HITS-CLIP – high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation
HMDS – hexamethyldisilazane
HMT – 4'-hydroxymethyltrioxsalen
HOBt – hydroxybenzotriazole
HoxB10 – Homeobox D10
HoxB8 – Homeobox B8
Hsc70 – heat shock cognate 71 kilodalton protein
Hsp90 – heat shock protein 90

I – inosine
IE-HPLC – Ion-Exchange High-Performance Liquid Chromatography
IRAK3 – Interleukin 1 Receptor Associated Kinase 3

J – coupling constant

kb – kilobase
kDa – kilodalton
KLF4 – Kruppel-like factor 4

LC-MS – Liquid Chromatography Mass-Spectrometry
let-7 – lethal-7
LNA – locked nucleic acid
lncRNA – long non-coding RNA

M – molar
MeOH – methanol
MF – mycosis fungoides
MHz – megahertz
MID – middle
min – minute
miR-CLIP – microRNA cross-linking and immunoprecipitation
miRNA/miR – microRNA
mL – millilitre
mmol – milimol
MPM – malignant pleural mesothelioma
mRNA – messenger RNA
MsCl – methanesulphonyl chloride/ mesyl chloride

NASH – non-alcoholic steatohepatitis
n-BuLi – *n*-butyllithium
nc – non-coding
NCR – non-coding region
ncRNA – non-coding RNA
nm – nanometre
NMP – *N*-methyl-2-pyrrolidone
NMR – Nuclear Magnetic Resonance
NSCLC – non-small cell lung cancer
nt – nucleotide

ORN – oligoribonucleotides

PACT – protein activator of the interferon-induced protein kinase
PAL – photo-affinity labelling
PAZ – Piwi-Argonaute-Zwille
PBS – phosphate-buffered saline
PEG – poly(ethylene glycol)
pHLIP – pH (low) insertion peptide
PIWI – P-element-induced wimpy testes
PMO – phosphorodiamidate morpholino oligonucleotides
PNA – peptide nucleic acid
PO – phosphodiester
Pol II – polymerase II
Poly(A) – polyadenosine
ppm – parts per million
pre-miRNA – precursor microRNA
pri-miRNA – primary microRNA
PS – phosphorothioate
PTX – paclitaxel
PUVA – psoralen and ultraviolet A
Py – pyridine

r.t. – room temperature
Ran – Ras-related Nuclear protein
RhoB – Ras Homolog Family Member B
RISC – RNA-induced silencing complex

RNA – ribonucleic acid
RNAi – RNA interference
RNase – ribonuclease
RP-HPLC – Reversed-Phase High-Performance Liquid Chromatography
rpm – rounds per minute
R_t – retention time
RT-qPCR – Reverse Transcription quantitative Polymerase Chain Reaction

siRNA – small-interfering RNA
SMIR – small molecule inhibitor of miRNAs
SNALP – stable nucleic acid lipid particle
SOCS1 – Suppressor of cytokine signalling 1

t – target
T – thymidine
TAR – trans-activation response
TBAF – tetra-*n*-butylammonium fluoride
TBAI – tetra-*n*-butylammonium iodide
TBDMS – *tert*-butyldimethylsilyl
TBTA – tris((1-benzyl-4-triazolyl)methyl)amine
TDMD – target RNA-directed miRNA degradation
THF – tetrahydrofuran
TLC – thin-layer chromatography
T_M – melting temperature
TMP – trioxsalen, 4,5′,8-trimethylpsoralen
TMSOTf – trimethylsilyl trifluoromethanesulphonate
TRBP – human *trans*-activation response (TAR) RNA binding protein
Tris – tris(hydroxyethyl)aminomethane
TsCl – *p*-toluenesulphonyl chloride/ tosyl chloride

U – uridine
UV – ultraviolet

VEGF – vascular endothelial growth factor

XPO5 – Exportin 5
Xrn1/2 – 5′-3′ exoribonuclease 1/2

δ – chemical shift
λ – wavelength

I. Chapter I: General introduction

I.1. microRNAs – an overview

A section about the microRNAs was written based on the previously published book chapter on the related topic [1].

I.1.1. Biogenesis and function

microRNAs (abbreviated miRNAs or miRs) are a class of ~22 nucleotides (nt)-long, single-stranded endogenous non-coding (nc) ribonucleic acids (RNAs) that are highly conserved across all species. Discovered independently in *Caenorhabditis elegans* (*C. elegans*) by Ambros and Ruvkun labs in 1993 [2, 3], miRNAs have emerged as powerful regulators of a vast variety of physiological processes in different organisms. Recent reports show that miRNAs may be responsible for post-transcriptional regulation of the majority of human protein-coding genes [4].

Most of the human microRNAs are encoded within the intronic regions of the non-coding or protein-coding transcripts or, less often, also in the exonic regions of transcription units [4, 5]. In certain cases, miRNAs are transcribed as polycistronic clusters, which allows expressing many different miRNAs at the same time [6, 7].

I.1.1.1. miRNA biogenesis

The biogenesis of miRNAs consists of several sequential processing steps [5, 8] (**Figure 1**). In the so-called canonical pathway, the microRNA gene is transcribed by RNA Polymerase II (Pol II) [9] to give a long (usually more than 1 kb) primary microRNA (pri-miRNA) transcript [6] with an imperfect stem-loop structure flanked by single-stranded segments. This primary transcript is equipped with the 5'-cap and usually polyadenosine (poly(A)) tail, which are the attributes of Pol II transcription [9]. The pri-miRNA is recognized and trimmed by Microprocessor complex comprising ribonuclease (RNase) III-type endonuclease Drosha and two units of its partner protein, double-stranded RNA (dsRNA) binding protein DiGeorge syndrome critical region gene 8 (DGCR8) [10-14]. This processing event yields a ~70-nt-long precursor microRNA (pre-miR) in a form of the small hairpin with one- or two-nucleotide overhang at the 3'-end and monophosphate at the 5'-end. The 3'-overhangs, characteristic for the RNase III-mediated cleavage [10], play an important role in nuclear exporting and further processing of the pre-miRNA [15-17]. In the next step, protein Exportin-5 (XPO5) forms a complex with the guanosine triphosphate-binding nuclear protein Ran (Ran-GTP) [18] and shuttles the pre-miR through the nuclear pore complex into the cytoplasm, where it is released upon GTP hydrolysis [15, 16, 19, 20]. The final maturation step, in which the terminal loop of the pre-miRNA is excised, is performed by Dicer [21-23]. Dicer is another RNase III endonuclease which associates with its dsRNA binding partners: human trans-activation response (TAR) RNA-binding protein (TRBP) [24, 25] and/or protein activator of the interferon-induced protein kinase (PACT) [26, 27]. Dicer consists out of multiple domains and serves as a 'molecular ruler'; its Piwi-Argonaute-Zwille (PAZ) domain recognizes and binds 3'-overhang and the 5'-phosphate of the pre-miRNA [5, 28-30]. Depending on the species, the RNase III domain dimer makes a cut 21-25 nucleotides from the end of the helix [5, 28, 29].
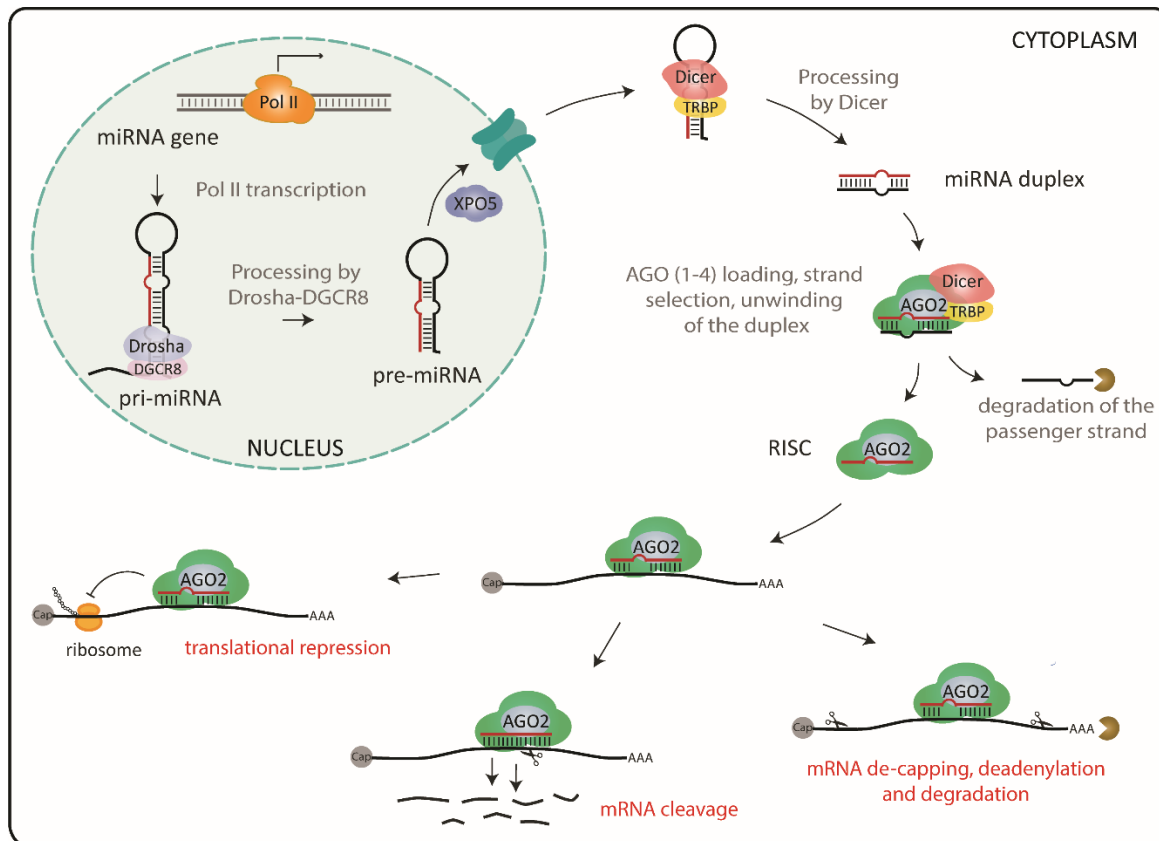
Figure 1: Overview of the canonical miRNA biogenesis pathway.

In alternative/non-canonical pathways, Drosha- or Dicer-mediated maturation steps can be bypassed. To give an example, mirtrons are generated by splicing of small introns [31, 32]. On the other hand, Argonaute 2-cleaved pre-miRNAs (ac-pre-miRNAs) are produced by Dicer-independent cleavage and trimming of the pre-miRNA executed by Argonaute 2 (AGO2) protein [33-35]. Nevertheless, Dicer and especially Drosha seems to be inevitable for proper processing of the miRNA within the canonical biogenesis pathway while Exportin-5 is important but not critical for the microRNA maturation [36].

I.1.1.2. RISC loading and strand selection

The Dicer-mediated cleavage step produces an imperfect miRNA duplex, consisting of two strands annotated as 5p strand (when it originates from the 5'-arm of the pri-/pre-miRNA hairpin) and 3p strand (when derived from the 3'-arm of the pri-/pre-miRNA hairpin) (**Figure 2**). This imperfect duplex is then loaded into an AGO, which associates with Dicer and TRBP to form a precursory version of the multicomponent effector complex, known as RNA-induced silencing complex (RISC) [37]. In humans, AGO family consists of four members, AGO1-AGO4. All of them can induce translational repression, but only AGO2 has 'slicing' properties, enabling cleavage of the target messenger RNA (mRNA) [38]. It has been suggested that TRBP plays a role in the recruitment of the AGO2 protein to the miRNA duplex bound by the Dicer-TRBP complex and is therefore responsible for the initiation of the RISC assembly [24, 27].

The duplex loading step is adenosine triphosphate (ATP)-dependent and assisted by Hsc70/Hsp90 (heat shock cognate 71 kDa protein/ heat shock protein 90) chaperone machinery [39]. It has been proposed that chaperons are responsible for keeping AGO in an energetically unfavourable conformation required to initiate the accommodation of miRNA duplex in AGO [40].
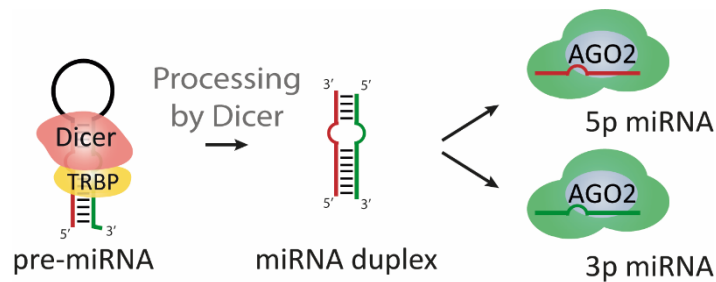
Figure 2: Generation of the 5p and 3p miRNA.

During the loading of the duplex, a process called 'strand selection' takes places, which is based on the relative thermodynamic stability of the ends of the duplex. The strand which is less stable at its 5'-terminus (usually the one with AU base pair at its 5'-end) is recognized by AGO as a guide strand (mature miRNA) and is in most cases more biologically active than its counter-strand [41, 42]. However, there are premises to believe that the strand selection process is much more complicated and influenced by multiple factors [26, 43].

AGO proteins, denominated as a functional core of the RISC, consist of four well-conserved domains: N-terminal domain, PAZ domain, middle (MID) domain and C-terminal P-element-induced wimpy testes (PIWI) domain connected by linkers (L1 and L2) [44-46] (**Figure 3**). The 5'-phosphate group of the strand recognized as a guide strand is anchored (through hydrogen bonds) in the basic 5'-phosphate-binding pocket present at the interface of the MID and PIWI domains of AGO. The 3'-end of the same strand is bound by the PAZ domain [45, 47]. Based on the crystal structure of human AGO2 in complex with four nucleoside monophosphates, it has been shown that the MID domain of AGO2 is biased towards recognition of the uridine (U) and adenosine (A) as the 5'-nucleotide of the guide strand [47]. This is due to the favourable interactions between those bases and a rigid loop ('nucleotide specificity loop') present in the MID domain, which are opposite for guanosine (G) and cytidine (C).

The other strand of the duplex, called the passenger strand, is unwound, dissociated and degraded resulting in the formation of a mature RISC, responsible for the miRNA-mediated post-transcriptional gene regulation. In principle, both strands can act as guides, depending on the cell type and biological context [43].
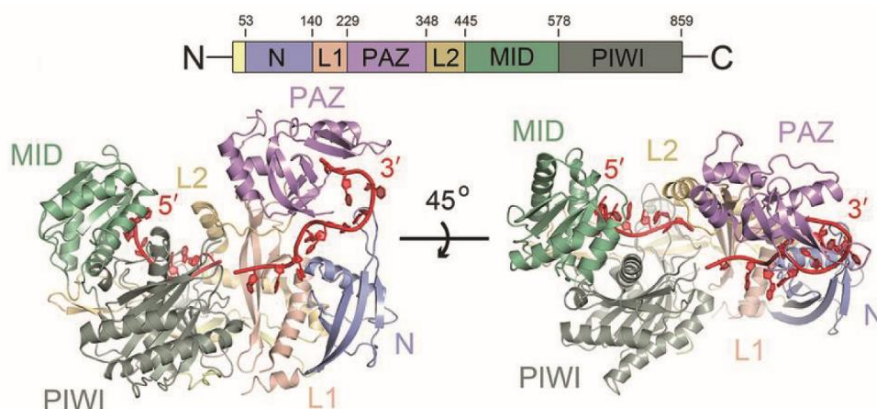
Figure 3: Structure of the AGO2-guide complex. Top: Schematic representation of the AGO2 primary sequence. Bottom: Crystal structure of the miRNA guide strand (marked in red) bound to the AGO2 protein. Taken from [46].

I.1.1.3. Target binding and miRNA functions

While working by a canonical mechanism, mature microRNAs bind to the complementary sites in the 3'-untranslated region (3'-UTR) of the target mRNAs and interfere with their translation. miRNA binding sites are often well conserved across species [48]. In most cases, target identification is based on a stepwise, fully complementary base-pairing between the so-called seed region (nucleotides 2 to 7 counting from the 5'-end of the microRNA guide strand) and target mRNA [49]. The seed region of the AGO-bound miRNA is pre-organized in the A-type helix, with nucleotides 2 to 5 (5' domain of the seed) exposed to the aqueous environment for target recognition. Base-pairing of these nucleotides with the target induces a conformational change of the α-helix-7 (helix-7) present in the linker L2 domain of AGO. This conformational change, in turn, enables engagement of the 3'-end of the seed (nucleotides 6 and 7 of the miRNA) in target binding and facilitates the formation of a stable miRNA-target duplex [45, 46, 50]. In the absence of the complementarity within the 3'-domain of the seed, helix-7 acts as a 'molecular wedge', responsible for the elimination of the non-target RNAs, and thus contributes to the fast and accurate target binding [50]. The seed binding can be further strengthened by a supplementary base-pairing to guide (g) nucleotides g_{13}-g_{16} (nomenclature adapted from Schirle et al. [46]). This supplementary interaction is feasible due to the presence of a 'supplementary chamber' created by AGO2, able to accommodate additional pairing [51].

The crystal structure of a human AGO2 revealed the existence of a narrow pocket, demonstrating an increased affinity for binding targets (t) with adenosine at the position t_1 [46, 52]. This finding is in accordance with the previously mentioned bias of the MID domain for binding guide strands with the uridine at the 5'-end [47].

miRNAs which share the same seed sequence are classified as a miRNA family [53]. Due to a common seed sequence, members of the same miRNA family can often regulate the same mRNA targets and therefore show redundant or partially redundant functions [54].

The seed-based ('canonical') target recognition is so far the most abundant and important one, as disruption of the base-pairing within this region can abolish the silencing abilities of the miRNA. The tolerance for mismatches or GU wobbles within the seed region is disputable and seems to highly depend on the cellular context of the silencing [54, 55]. Apart from the aforementioned canonical targeting, various non-canonical miRNA-mRNA interaction motifs have been reported (**Figure 4**) [54, 56]. The seed pairing can be supported by the additional base-pairing involving nucleotides g_{13}-g_{16} of the miRNA, referred to as the 3'-supplementary motif and enhancing target identification [46, 53]. In cases when the seed interaction is distorted or unusually short, additional binding at the 3'-end of the miRNA (the 3'-compensatory binding) can compensate for the impaired seed and enable regulation of target expression [54]. It has been hypothesized that supplementary binding in the 3'-end of the miRNA can be used by members of the same miRNA family to repress different targets [54]. Target recognition repertoire includes also perfect or imperfect centred interactions (lacking 5'- and 3'-interactions which are replaced by a contiguous base-pairing to nucleotides 4-15 of the miRNA) [57, 58], or binding to the bulged mRNAs [59]. Growing evidence shows that the seed-like interactions with targets having a G-bulge opposite to position g_5 and g_6 may be also tolerated and have an important biological role [60]. Last but not least, miRNAs can interact with mRNAs *via* fully complementary ('small-interfering RNA-like', siRNA-like) base-pairing, a phenomenon commonly occurring in plants [61, 62].
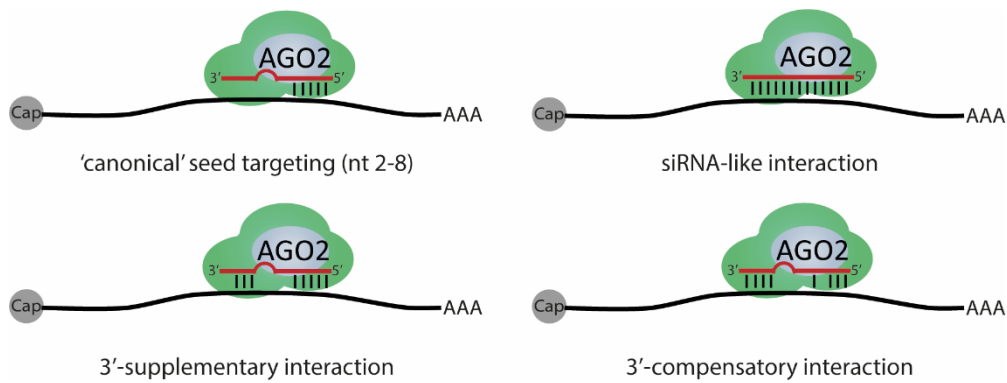
Figure 4: Illustrative examples of different types of miRNA-target interactions. Adapted from [56].

The degree of the complementarity between the miRNA and the mRNA transcript determines the mechanism by which protein expression is suppressed. Imperfect base-pairing usually leads to translational repression [63-65] and/or mRNA degradation [66], the latter being the dominant mechanism in cultured mammalian cells [67, 68]. On the other hand, fully or nearly fully complementary base-pairing may trigger mRNA cleavage [69], as it was shown on the example of miR-196 and Homeobox B8 (HOXB8) mRNA [70]. In such cases, the target is cut by the enzymatically active PIWI domain of the AGO2 [38] at the position juxtaposed to the phosphate bond (scissile phosphate) between nucleotides g_{10} and g_{11} [71, 72].

When bound to the AGO, miRNAs have been reported to have half-lives ranging from hours to days, depending on the cell type and general cell conditions [73-75]. The stability is ensured by the AGO protecting the ends of the guide strand from the enzymatic degradation [46, 74]. What happens to the guide miRNA after it fulfils its regulatory role has not been explained in detail. It has been reported that interactions with highly complementary targets may stimulate the release of the guide from the AGO [74] or trigger a non-templated tailing and trimming of the miRNAs leading to their degradation [76, 77]. This phenomenon, referred to as a target RNA-directed miRNA degradation (TDMD), seems to be particularly important in neuronal cells, as they need to dynamically regulate miRNA levels in response to quickly changing environmental conditions [78]. This type of modulation is possible due to the additional base-pairing within the 3'-region of the miRNA, which induces conformational changes and the release of the 3'-terminus of the guide strand from the PAZ domain [76].

Apart from the canonical interactions involving pairing to the binding sites located in the 3'-UTR, microRNAs have been reported to be involved in the non-canonical ones, such as binding to the 5'-untranslated region (5'-UTR) [79] or coding sequence of the mRNA [80, 81]. miRNAs can also target non-coding RNAs (ncRNAs), like long non-coding RNAs (lncRNAs) [82-84] and circular RNAs (circRNAs) [85, 86]. Additionally, it has been shown that in certain cases miRNA-mRNA interaction can result in an enhanced translation of the targeted mRNA [87, 88].

I.1.1.4. isomiRs – different variants of the same miRNA

As miRNAs are engaged in the regulation of crucial developmental and cellular processes, which in many cases are tissue- and/or stage-specific, miRNA biogenesis and function are also tightly controlled [89-91]. Stepwise processing and compartmentalization during the miRNA maturation are one of the means to fine-tune miRNA levels [6, 12]. In order to get properly processed and to fulfil their biological functions, miRNAs (or their precursors) require interactions with various co-factors [90]. These co-factors additionally provide protection from the degradation by nucleases omnipresent in the cellular

environment [15, 16, 20]. Association with distinct proteins can influence miRNA mode of action. For instance, the position at which Drosha and Dicer cleave the base of the pri-miRNA and the terminal loop of the pre-miRNA, respectively, determines the identity of the miRNA. Cleavage imprecision, together with the 3'-tailing and trimming ('nibbling') results in the formation of biologically relevant variants of the same miRNA, dubbed isomiRs [92], with mRNA targets partially overlapping with those regulated by the canonical miRNAs [93, 94]. It is speculated that the expression of the isomiRs may be cell- and stage-specific and therefore significantly contributes to the precise regulation of the post-transcriptional gene expression [95]. The 3'-isomiRs, differing in length and nucleotide composition at the 3'-end of the miRNA, are the most common population of isomiRs. The 3'-end of the miRNA, loosely bound by the PAZ domain of AGO, is prone to enzymatic modification [94]. Recent studies showed that 3'-isomiRs may have a more significant role than initially expected, as these isomiRs can stabilize supplementary binding, thus contributing to the specificity of the target recognition [51]. On the other hand, the seed sequence is known to play a key role in target identification. Changing of the seed register ('seed shifting'), taking place when the 5'-isomiRs are formed, can alter the identity of the targeted mRNAs, which can lead to significant functional consequences [95]. Apart from variations in the population of regulated transcripts, addition or deletion at the 5'-end of the miRNA can affect the process of strand selection through modulation of the thermodynamic stability at the ends of the duplex [94]. This is especially the case for a Dicer cleavage, which seems to be less strict with respect to the position at which it makes a cut. It has been reported that the kinetics and the exact cleavage site is often affected by Dicer's binding partners, TRBP and PACT, based on their ability to discriminate between different secondary structures of the pre-miRNA [26, 96].

A-to-I (inosine) editing carried out by the adenosine deaminase acting on RNA (ADAR) enzyme is another relevant process resulting in an isomiR formation. The occurrence of this type of editing within the seed region of the pri- or pre-miRNA may cause alterations in the targeting profile, as it has been shown for miR-376 [94, 97].

The highly intricate regulatory pathway of the miRNA maturation and function described above shed a light on how miRNAs can ensure spatiotemporal control of a wide variety of cellular processes. This highlights their importance as key players in the post-transcriptional regulation of gene expression.

I.1.1.5. Importance of the function of the miRNAs

There are several factors which add to the challenge of an in-depth understanding of the silencing process carried out by the miRNAs. First, one microRNA can regulate many different messenger transcripts and one mRNA is under coordinated control of multiple miRNAs. Secondly, due to the common seed sequence, members of the same miRNA family usually control the same set of target genes (unless their sequence differs significantly in the central domain and the 3'-end).

Finally, the discovery of competing endogenous RNA (ceRNA) network brought to light a completely new aspect of miRNA regulation [83, 98]. Basically, any ncRNA transcript (not only mRNA) having a binding site that is reversed complementary to the sequence of the miRNA of interest can operate as a ceRNA and compete with the mRNA targets for shared miRNAs [83, 98]. By functioning as a 'molecular sponge', ceRNA can alter the activity of the miRNA and change the expression patterns of its mRNA targets [83]. It has been shown that lncRNAs, ciRNAs or pseudogene mRNAs can function as ceRNAs [82, 83, 85]. All of these testify to a high level of redundancy and complexity in the miRNA-related regulatory pathways, additionally challenging their understanding.

Based on the initial research on nematodes, in which deletion mutations in the vast majority of the miRNA genes did not show severe phenotypes [99], it was concluded that the miRNAs are not indispensable for the development and survival of organisms. Instead, it was proposed that they act as micromanagers, responsible for a fine-tuning of the mRNA levels according to the cell type and changeable environmental conditions [100]. However, further loss-of-function studies in mammals showed that impaired function of some miRNAs can be related with more severe phenotypes [4, 101]. As an example, point mutations within the seed of the human miR-96 result in a progressive hearing loss [102].

I.1.2. miRNAs and disease

As miRNAs influence many physiological processes within the cell, it could be expected that substantial changes/anomalies in microRNA performance may lead to the cell malfunction. Indeed, alterations to the miRNA expression patterns have been related to many pathological activities and diseases such as diabetes (miR-103 and miR-107) [103], cardiovascular diseases (miR-208, miR-133a, miR-1, miR-15) [104-107], atherosclerosis (miR-33) [108-110], Hepatitis C (miR-122) [111] and cancer [112, 113].

Numerous studies have demonstrated a strong correlation between miRNAs and cancer (summarized in: [114-116]). Significant alterations (in most of the cases: downregulation [117]) of the microRNA levels have been associated with cancer onset and progression [114], tumour environment [118], angiogenesis [119], epithelial-mesenchymal transition (EMT) [120] and other metastasis-related events [121, 122]. A lot of effort is taken to unravel the intricate mechanisms driving carcinogenic processes and establish the molecular roles that miRNAs play in them.

Depending on the type of targeted mRNA, most of the cancer-related miRNAs can be classified as tumour suppressor microRNAs or tumour-promoting miRNAs (dubbed oncomiRs) [121, 123]. The group of tumour suppressor miRNAs comprise miRNAs targeting transcripts coding for known oncoproteins (for example miR-34a or *lethal*-7 (let-7)). Their expression is usually downregulated in diseased cells. Tumour-promoting miRNAs, such as miR-21 or miR-17~92 cluster, repress mRNAs coding for the tumour suppressor proteins. The observation that many cancer-associated miRNAs or miRNA clusters are often located in chromosomal regions being either amplified or deleted in cancer cells [116, 124, 125] provided one of the first unambiguous links between the miRNA and cancer. However, a detailed understanding of the correlation between the two is challenging for numerous reasons. miRNAs are engaged in many different regulatory networks which intersperse with each other. The role of some of the miRNAs seems to be redundant. miRNAs can act either as oncogenes or tumour suppressors in a cell/tissue-dependent manner [121, 126]. This underlines the complexity of the task not only in cancer but also in all the other diseases involving dysregulation of the miRNAs.

In the last few years, a new class of microRNAs related to the development and acquisition of metastasis has been identified. Members of this class, collectively called metastamiRs [127], can control numerous signalling pathways of invasion-metastasis cascade, with the ability to either promote or suppress the metastasis process [121]. Despite significant improvements in cancer therapy, metastasis is still considered to be the most common cause of cancer-related mortality [122], emphasizing the need for precise identification of key players in the metastasis process and the development of new strategies against it.

For certain cancers, distinct expression patterns of the miRNAs can be used to specify subtype of the tumour and help to match the best treatment [117, 128]. Moreover, the discovery of cancer-specific

miRNAs present in the body fluids (referred to as circulating miRNAs) holds great potential as they can be possibly applied as non-invasive diagnostic markers [114, 129]. Therefore, microRNAs are emerging not only as the therapeutic targets but also as potential diagnostic and prognostic biomarkers for a wide range of diseases [106, 115].

Although a change in the expression profile of one miRNA might not be a direct cause of a disease, accurate levels of the miRNAs are crucial for the proper functioning of organisms.

I.1.3. miRNAs as therapeutics

Various approaches have been developed to target miRNAs (**Figure 5**). These include microRNA sponges and small molecule inhibitors of miRNAs (SMIRs). microRNA sponges are RNA transcripts produced from transgenes in cells [130]. These competitive inhibitors with multiple tandem binding sites to a miRNA or miRNAs of interest serve as decoys to 'fish out' the endogenous microRNAs. Stable sponges might be used to mimic the conditions in cancer cells, in order to study the biological effects of downregulation of miRNAs of choice [131, 132]. SMIRs, characterized by favourable pharmacokinetic properties, can interact with the potential binding sites such as internal/ terminal loops or bulges within the pri- or pre-miRNAs, thus preventing the processing of miRNA precursors [133]. Unfortunately, one of the most important drawbacks of the small molecule-based approach is that SMIRs have low specificity, which can significantly increase the chance for serious side-effects [115, 134]. Roos et al. reported *N*-methyl-*N*-[3-(3-methyl[1,2,4]triazolo[4,3-*b*]-pyridazin-6-yl)-phenyl]acetamide as the first small molecule antagonist of Lin28, a protein known to control the maturation of the let-7 microRNA family. The molecule, selected in a novel Förster resonance energy transfer (FRET)-based screening assay, was shown to prevent binding of Lin28 to the let-7 precursors resulting in the rescue of the let-7 processing and functioning in cells [135]. Although appealing, most of these approaches are used for research purpose only, being still far from the therapeutic application.

The most advanced strategy to influence miRNA levels developed so far is based on the application of the oligonucleotides. There are two main concepts to target disease-related microRNAs: microRNA replacement/ mimicking therapy, used to substitute downregulated miRNAs, and microRNA inhibition therapy, employed to reduce the levels of the microRNAs that are overexpressed in a particular disease. Hence, miRNAs can be considered as either therapeutic agents or therapeutic targets. miRNA mimics are synthetic duplexes imitating the function of their endogenous counterparts [136]. On the other hand, suppression of the miRNA levels is achieved with the use of miRNA inhibitors, called anti-microRNAs. Anti-microRNAs (also known as anti-miRs or anti-microRNA oligonucleotides, AMOs) are chemically modified single-stranded synthetic molecules of various length with a sequence fully complementary to the targeted mature miRNA. The concept of targeting the microRNA with anti-miRs and the milestones in this topic will be briefly summarized in the next sections of this chapter (based on [1]).
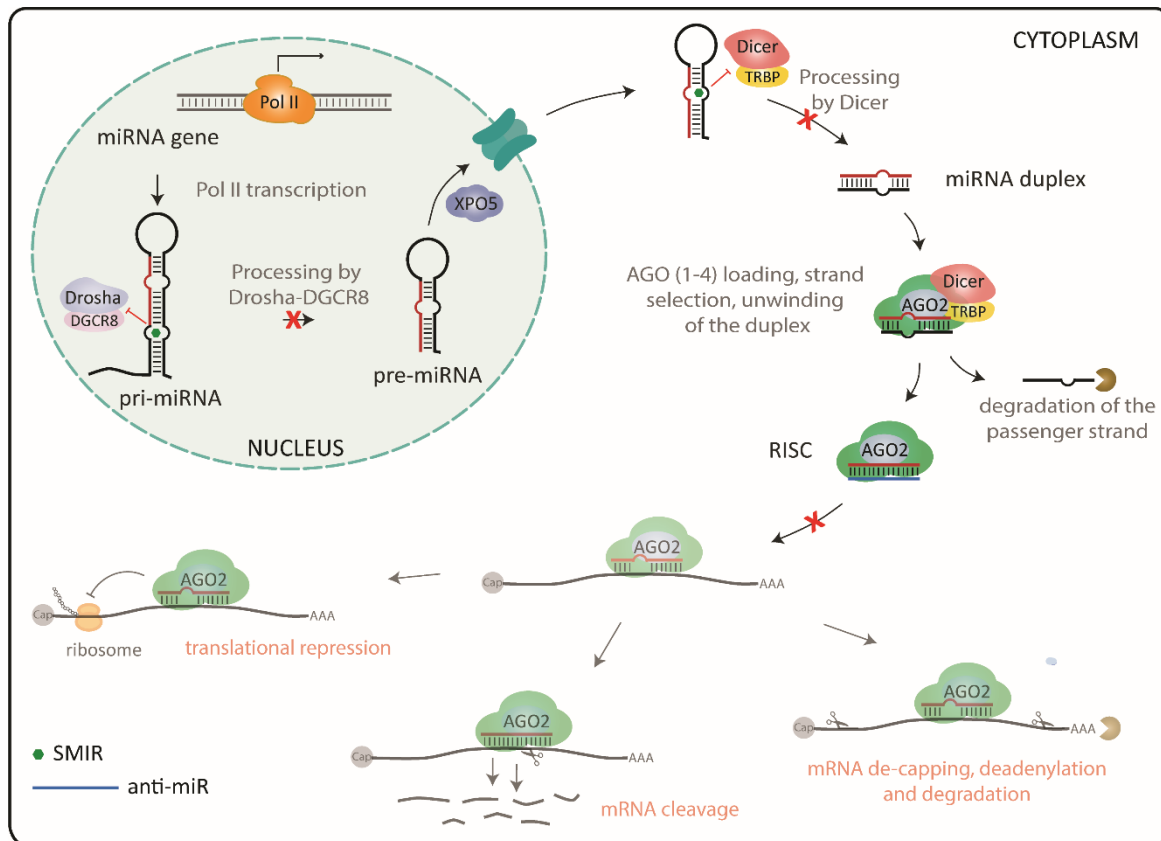
Figure 5: Targeting of the miRNA biogenesis and function with SMIRs and anti-miRs. Binding of the anti-miRs to the pri- and pre-miRNAs has been also reported [137].

I.1.3.1. Application of the anti-miRs for miRNA targeting - challenges

There are numerous hurdles that need to be overcome by the anti-miRs so that they could fulfil their molecular function [1, 136, 138]. The first challenge is related to the stability of these molecules. When delivered systemically, anti-miRs are exposed to degradation by endo- and exonucleases omnipresent in serum, readily recognising and cleaving phosphodiester (PO) bonds. Secondly, they need to evade premature clearance *via* renal excretion and circulate in the bloodstream long enough to have a chance to get into the cells. When circulating in the body, anti-miRs should avoid the triggering of the innate or adaptive immune response causing toxic effects. Subsequently, these rather large, hydrophilic, and negatively charged molecules need to traverse the also negatively charged cellular membranes. Upon delivery to the cells, anti-miRs incorporated through different mechanisms such as endocytosis pathway have to escape from the endosomes to the cytoplasm. Finally, anti-miRs need to efficiently and specifically bind to the designated microRNA, winning the competition with the endogenous microRNA targets without blocking the RISC machinery.

Different strategies have been employed to overcome these obstacles. Stability, affinity, and specificity (ability to discriminate mismatches) matters can be addressed by the incorporation of chemical modifications into the structure of an anti-miR. Besides, various conjugates can be attached to either inhibitor itself or to the carrier system with an aim to support the delivery to the tissue of interest, facilitate entry to the cell or even boost endosomal escape.

The field of oligonucleotide-based microRNA inhibitors is under constant development and is following the path similar to that for antisense oligonucleotides targeting mRNAs (reviewed in [1, 139-

142]). Generally, chemical modifications are introduced into the phosphodiester backbone, ribose moiety or both (**Figure 6**). There are two main purposes of backbone modification. The first one is to improve the stability of the oligonucleotides against exo- and endonucleases. Incorporation of the 'unnatural' elements into the backbone makes oligonucleotides less recognizable for the nucleases and therefore extend their half-life in serum. The second goal is to prolong the systemic circulation of the oligonucleotides by increasing their binding to the serum proteins. Modifications to the sugar ring, on the other hand, are used to boost the binding affinity for the RNA targets and, usually, also to improve the stability towards the degradation by nucleases.

I.1.3.2. Medicinal chemistry of anti-miRs

One of the very first trials to deplete microRNA for loss-of-function study in *Drosophila* embryos has been performed with an unmodified DNA molecule [143]. It has been pointed out that the observed developmental defects were most likely achieved due to direct microinjection preventing degradation of the deoxyribonucleotide (DNA) molecules [140]. This was followed by experiments with anti-miRs in which 2'-*O*-position of the ribose was modified with a methyl group (2'-*O*-methyl; 2'-OMe). The main advantage of 2'-OMe-modified oligonucleotides is that they hamper the degradation process and increase binding affinity to RNA targets [141]. Hutvagner et al. employed 31-nt, 2'-OMe-per-modified anti-miR to downregulate let-7 in HeLa cells and *C. elegans* [144], whereas Meister et al. used a shorter (24-nt) oligonucleotide against miR-21 in HeLa cells [145]. Although in both cases anti-miRs were able to inhibit target microRNAs, it became clear that further refinement in the anti-miR design will be needed, especially for *in vivo* tests.
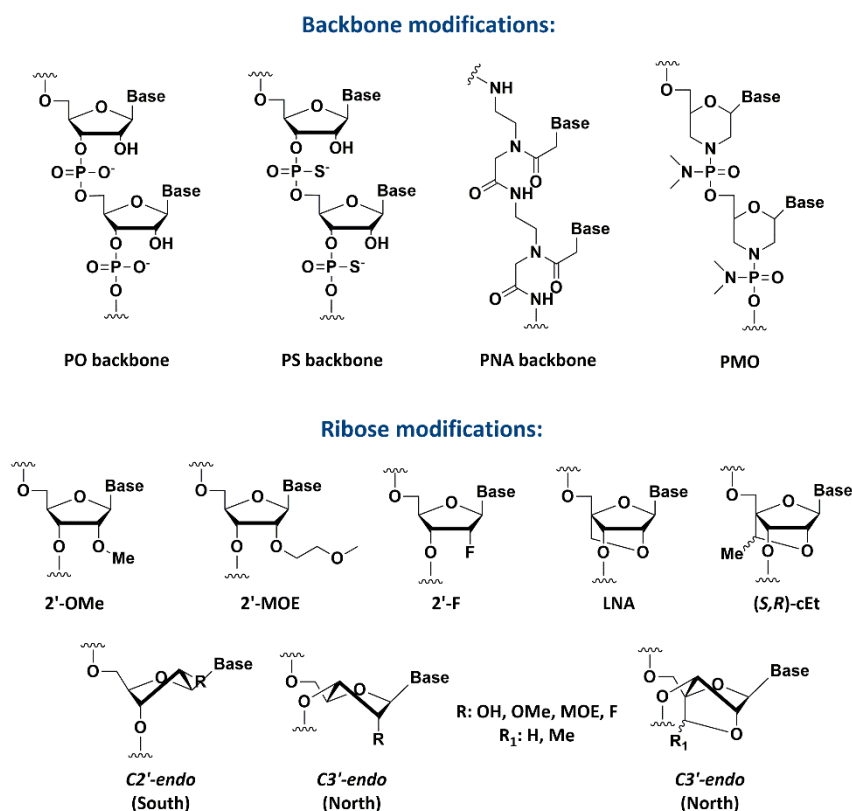
Figure 6: Selected backbone and ribose modifications commonly applied in anti-miR chemistry. Adapted from [1].

The first significant improvement was the application of the phosphorothioate (PS) linkages, in which the non-bridging oxygen of the naturally occurring phosphodiester backbone was replaced by a sulphur atom [146]. The main benefit of using PS modification is that it protects oligonucleotides (including anti-miRs) from degradation by nucleases [147]. Additionally, the presence of sulphur atoms increases the hydrophobicity of the oligonucleotide, thus promoting a non-specific binding to proteins in serum, slowing down renal excretion and enhancing cellular uptake of oligonucleotides administered systemically *in vivo* [146, 148, 149]. On the other hand, the main drawback of the PS linkages is that they lower binding affinity for the complementary target strands when compared to the corresponding phosphodiester molecules [146, 150]. This can be compensated by mixing PS modification with the affinity-increasing ribose modifications. Currently, anti-miRs with the PS backbone are widely used, and this modification is necessary for the enhanced *in vivo* delivery, especially if no formulation is applied.

PS modification was critical for the function of antagomirs, the first oligoribonucleotide-based microRNA inhibitors used in mice [151]. Antagomirs constituted a new class of chemically modified and cholesterol-conjugated oligoribonucleotides, able to efficiently and specifically downregulate endogenous miRNAs *in vivo*. Tha original antagomir, reported by Krützfeldt et al., was a 2'-OMe-per-modified oligoribonucleotide fully complementary to the miR-122 sequence, with two phosphorothioate linkages at the 5'-terminus, four phosphorothioate linkages at the 3'-end and with a cholesterol moiety conjugated at the 3'-terminus of the molecule to support cellular uptake [151]. Administration of miR-122 antagomir in mice resulted in a specific and powerful reduction in the miR-122 levels caused by target microRNA degradation and shed a light on the role of miR-122 in the regulation of the cholesterol biosynthetic pathway. Due to the high dosage (3 injections of 80 mg/kg) needed to influence miRNA levels, antagomirs have found an application mainly as experimental tools for the *in vitro* and *in vivo* loss-of-function studies [152].

Incorporation of the 2'-*O*-methoxyethyl (2'-MOE) modification into anti-miR sequences resulted in a greater nuclease resistance and binding affinity in comparison with the 2'-OMe modification [141, 153]. With the use of 2'-MOE/PS-modified AMOs, Esau et al. studied the role of the microRNAs in adipocyte differentiation and discovered that targeting of miR-143 in primary human pre-adipocytes hampered their differentiation and prevented adipocyte-specific gene expression [154]. The same modification pattern was then employed to antagonize miR-122 *in vivo*, to gain more insight into its function [155]. The results showed that downregulation of miR-122 lead to a remarkable reduction in cholesterol levels in plasma and upregulation of miR-122 target genes. This proved that unconjugated MOE/PS anti-miRs can efficiently decrease microRNA levels at doses lower than those reported for the aforementioned miR-122 antagomir and highlighted their potential in improving AMOs properties.

The influence of 2'-fluoro (2'-F or FNA) modification on anti-miR properties has been evaluated as well. Introduction of a fluorine atom at the 2'-position of the ribose constrains the sugar ring in an RNA-like, 3'-endo conformation, which gives rise to remarkably enhanced binding affinity towards RNA targets [156, 157]. Since 2'-F-modified PO oligonucleotides do not display sufficient resistance against nucleases [153, 158, 159], they usually have also a PS backbone. Indeed, the target-binding affinity of the 2'-F/PS-modified oligonucleotide was higher than this for its 2'-OMe/PS and 2'-MOE/PS counterparts, as shown in a head-to-head comparison of PS anti-miRs with various chemistries at the 2'-*O*-position directed against miR-21 [153]. Furthermore, capping of the 2'-F/PS inhibitor of miR-122 with two 2'-MOE modifications at both ends resulted in a significant de-repression of the miR-122 target gene *in vivo* [160].

Locked nucleic acids (LNAs) are to date one of the most powerful modifications applied in the antisense field. The characteristic feature of LNAs is the presence of at least one 2'-*O*,4'-*C*-methylene-linked bicyclic ribonucleoside unit which, similarly to the 2'-F modification, impart 3'-endo sugar puckering [161]. LNAs manifest outstanding binding affinity to the complementary RNA, with an increase in the melting temperature (T_M) as high as 4-6°C/LNA moiety [162]. Moreover, the unusual structure of LNA significantly enhances its stability towards nucleases.

The first study describing the use of LNA-modified AMOs employed an LNA/DNA chimaera (an LNA central core flanked from both sides by seven DNA bases) directed against miR-21 in glioblastoma cells [163]. This compound showed rather moderate activity, most likely due to the degradation of the DNA wings by nucleases. On the other hand, all-LNA anti-miR was able to efficiently and specifically inhibit *bantam* microRNA in HEK293 or *Drosophila* cells [164]. Nonetheless, LNA oligonucleotides have been reported to form LNA-LNA hybrids with extremely high T_Ms which can impair their function [140]. To overcome this problem, careful design of the oligonucleotide in terms of a number of LNA modifications and their positioning within a sequence is needed. For example, the standard approach for the design of the LNA-based probes used for the detection of the miRNAs in the *in situ* hybridization experiments assumes the use of a repeating motif, in which two DNA bases are followed by one LNA modification [140, 165, 166]. Going further, DNA bases can be replaced with other 2'-*O* chemistries, giving rise to various LNA mixmers with improved activity [167, 168]. Elmen et al. utilized 16-nucleotide-long PS LNA/DNA mixmer to target miR-122 in mice [169]. Upon further optimization of the chemistry design this anti-miR, known as miravirsen, became the first oligonucleotide-based microRNA inhibitor to enter the clinical trials (the development of miravirsen will be described in detail in section I.1.4.1). Obad et al. took advantage of an exceptional binding affinity of LNAs to design so-called tiny LNAs, with an aim to simultaneously antagonise all the members of a given miRNA family [170]. Tiny LNAs, fully LNA-modified 8-mers with a PS backbone, bound to the seed region of the miRNA and suppressed the activity of all of the family members sharing the same seed sequence in cultured cells and *in vivo*. Moreover, although shortening of the sequence increased the number of potential binding sites in target messages, tiny LNAs seemed to have fewer off-target effects, which added to their uniqueness. LNAs are state-of-the-art chemistry in the anti-miR field and many of the AMOs in pre-clinical or clinical trials are based on this modification.

Constrained ethyl (cEt, 2',4'-constrained 2'-*O*-ethyl) unit is a 2',4'-bridged analogue (locked in *N*-type conformation), in which position 6 of the 2,5-dioxabicyclo ring system is substituted with a methyl group, resulting in either *R*- or *S*-stereoisomer. cEts are close analogues of LNAs, with a comparable binding affinity, mismatch discrimination and remarkably higher than LNAs resistance towards the degradation by nucleases [171]. AMOs modified with cEts were used in the *in vitro* and *in vivo* studies focusing on the anti-miRs mechanism of action [172]. Some anti-miRs modified with cEt units are currently evaluated in clinical trials.

Peptide nucleic acids (PNAs) and phosphorodiamidate morpholino oligonucleotides (PMOs or Morpholinos) are two other backbone modifications studied in the search for the design of anti-miRs with improved properties. PNAs are electrostatically uncharged, synthetic nucleic acid analogues, with a backbone composed of the *N*-(2-aminoethyl)-glycine units and nucleobases attached to the scaffold through a methylene carbonyl linker [173]. Lack of the negatively charged phosphate groups within the PNA backbone eliminates electrostatic repulsion and ensures strong binding to the target RNA. The unique, peptide-like scaffold provides superior stability of PNAs under physiological conditions, as they are not recognized and degraded by nucleases. The main limitation associated with the use of

PNAs is their decreased solubility in an aqueous environment and the tendency to aggregate. Besides, PNA molecules are also not easily taken up by the cells and relatively fast removed from the bloodstream. These issues can be addressed by conjugation of the PNA-based anti-miRs to cationic amino acids (to increase solubility) or cell-penetrating peptides (to support cellular delivery *in vitro* and *in vivo*) [167, 174-176]. Unfortunately, efficient tissue uptake is still the main concern regarding PNA-based anti-miRs and the main reason why these molecules are not yet in clinical trials.

PMO backbone consists of methylenemorpholine rings connected by the phosphorodiamidate linkers. Due to the unusual, uncharged scaffold, morpholinos are resistant to digestion by nucleolytic enzymes [177, 178]. Diminished interactions with plasma proteins result in low toxicity of morpholinos but also reduce their ability to efficiently pass through the plasma membrane. PMOs show relatively good solubility in water and high specificity towards complementary RNAs, nevertheless they bind their targets with a lower affinity when compared to the unmodified RNAs. Flynt et al. used a PMO-modified anti-miR to target miR-214 in *Zebrafish*, which uncovered its function in muscle cell development [179]. It has been demonstrated that in addition to targeting mature microRNA, morpholinos are also able to bind to the pri-microRNA and pre-microRNA, thus interfering with their processing by Drosha and Dicer, respectively [180]. Despite some favourable properties, PMO scaffold is rather rarely used in the anti-miR design.

I.1.3.3. The power of targeted delivery and combined treatments

With LNAs being an undisputed leader among chemical modifications in the anti-miR field, the attention of scientists has been focused on the delivery issue or, to be more precise, targeted delivery. This is typically addressed by conjugation to the anti-miR or to the delivery systems peptides or small molecules having high affinity for a particular tissue type or cellular environment.

To give an example, PNA-based anti-miR was conjugated to the pH (low) insertion peptide (pHLIP) to target acidic tumour environment in a mouse model of lymphoma [181]. A unique feature of pHLIP, a water-soluble polypeptide derived from the bacteriorhodopsin C helix, is that it can insert across the lipid bilayer (membrane translocation peptide) and form a stable transmembrane α-helix [182, 183]. In the acidic conditions being a hallmark of solid tumours [184], negatively charged residues of pHLIP get protonated, thus increasing the hydrophobicity of the peptide and enhancing the affinity between the peptide and a membrane [182]. This triggers the peptide folding and its insertion into the cell membrane, with N-terminus of the pHLIP pointing into the extracellular space and the C-terminus placed in the cytosol [173]. The energy from the protein folding is used to translocate cell-impermeable cargo molecules into the cytoplasm. Based on this concept, the pHLIP peptide was attached at its C-terminus to the anti-miR-155 molecule *via* a disulphide bond, cleavable in the reducing environment of the cytosol [185]. miR-155 is an oncomiR known to be upregulated in lymphomas [186]. While administered intravenously, pHLIP-anti-miR significantly impeded tumour growth and held back metastasis of neoplastic lymphocytes to other organs in mice with miR-155-addicted lymphoma subcutaneous tumour model [181].

miR-21 is a well-known tumour-promoting miRNA playing a key role in cell function and survival. It has been reported to be significantly upregulated in a great number of cancers such as glioblastoma, neuroblastoma, lymphoma or breast, pancreatic, colorectal and lung cancer [163, 187]. Therefore, a lot of effort is put into trials to effectively reduce levels of this microRNA in diseased tissues. For instance, encapsulation of the LNA-based anti-miR-21 into stable nucleic acid lipid particles (SNALP) and its systemic administration to mice with induced brain tumours resulted in preferential

accumulation of the AMO in the tumour associated with the inhibition of miR-21 [188]. To enhance glioblastoma tumour-specific targeting of miR-21 *in vivo*, SNALPs used for encapsulation of the anti-miR were conjugated with chlorotoxin (CTX), a peptide known for its specific binding to glioma cells [189]. Indeed, application of the CTX-SNALP-formulated anti-miR-21 resulted in a higher expression of RhoB (Ras Homolog Family Member B), a direct target of miR-21. Additionally, when intravenous injections were followed by oral administration of the tyrosine kinase inhibitor sunitinib, a decline in tumour size was observed, together with enhanced apoptosis and slightly prolonged survival of the animals. As shown above, targeted delivery is a very important matter as it can not only increase the potency of the anti-miR but also lower the risk of the off-target effects (as the same microRNA might have a different function in different tissues).

Recent studies highlight the possible therapeutic potential of the combined strategies, such as mixing of the anti-miR treatment with a miRNA replacement therapy or co-administration of anti-miR/miRNA mimic together with the anti-cancer small-molecule drug, to name few [114].

Following this concept, liposome-based co-delivery of miR-10b antagomir and paclitaxel (PTX) was tested for simultaneous inhibition of tumour growth and metastasis in murine metastatic mammary tumour model [190]. miRNA-10b, highly overexpressed in metastatic breast cancer cells, is known to be responsible for the initiation of invasion and metastasis [191]. Although it has been shown previously that high dose of antagomir-10b was able to hamper metastasis, it did not hold back the growth of primary tumours [192]. Paclitaxel is a cytotoxic agent widely used in chemotherapy against various types of cancer and it was envisioned that the co-treatment of the tumour cells with both, antagomir and PTX might bring more promising results. To improve endosomal escape, liposomes were decorated with a pH-responsive, anti-microbial peptide [D]-H_6L_9 (D-Lip) [193]. In acidic conditions, prevailing in endosomes or lysosomes, histidines of D-Lip get protonated, which in turn triggers lysis of the endosomal membrane and release of its content [193, 194]. D-Lip-supported delivery of the cholesterol-conjugated antagomir-10b, either with or without PTX, de-repressed expression of a downstream target of miR-10b, Hoxd10 (Homeobox D10) both, *in vivo* and *in vitro* [190]. Furthermore, *in vivo* tests showed that co-delivery of D-Lip/antagomir-10b and PTX induced a more significant reduction of tumour volume and decreased metastasis to lungs (thanks to the PTX and antagomir activity, respectively).

I.1.4. From bench to bedside: miRNAs in clinical trials

Eventually, improvements in the design and delivery of anti-miRs allowed some of them to be transferred from the pre-clinical to clinical stage, giving hope for their application as therapeutics (reviewed in [112, 195, 196], **Table 1**)

### I.1.4.1. Anti-miR-based treatment of HCV

Miravirsen, a 15-nt LNA/DNA chimaera with a fully phosphorothioated backbone, was the first oligonucleotide-based miRNA inhibitor to enter clinical trials as a drug candidate for the treatment of Hepatitis C virus (HCV). Infection with HCV is usually followed by the development of chronic hepatitis C leading to even more severe liver diseases which highlights the need for effective HCV treatments [197, 198].

The origin of the development of miravirsen (summarized in detail by Baek [198]) goes back to the discovery that miR-122 can bind to three highly conserved binding sites at the 5'- (two binding sites) and 3'-end (one binding site) of the non-coding region (NCR) in hepatitis C virus RNA, thus facilitating

replication of the virus in host liver cells [111, 199]. This finding was surprising, as the observed mechanism was in contrary to the canonical role of microRNAs. miR-122 binds to the 5′-end of the viral RNA in an unusual way, in which 3′ region of the miRNA serves as a camouflage hiding 5′-terminal sequence of the HCV genome. Thus, miR-122 protects viral RNA from degradation by cellular 5′-3′ exoribonuclease 1 and 2 (Xrn1 and Xrn2) [200, 201], prevents stimulation of innate immune responses and increases the stability of the viral RNA [202]. miR-122 attracted the attention of researchers much earlier, as it is highly and specifically expressed in liver hepatocytes, having an important role in cholesterol and lipid metabolism [151, 155].

Initially, 16-nt-long PS LNA/DNA anti-miR fully complementary to the 5′-terminus of miR-122 has been shown to antagonize its function in a dose-dependent way when administered systemically in mice [169]. This was accompanied by a decline in cholesterol level and de-repression of the miR-122 targets. The inhibitory effect was achieved by the formation of a stable heteroduplex between the anti-miR and target miR-122, though follow-up studies showed that pri- and pre-miR-122 can be also targeted [137]. It is worth to note that prolonged silencing of the miR-122 was achieved with much lower doses than those used for cholesterol-conjugated antagomirs [151] or MOE-based anti-miRs [155]. Upon further fine-tuning of the chemistry, the new candidate (15-nt-long DNA phosphorothioate oligonucleotide highly substituted with LNA modifications; **Figure 7a**) was tested in African green monkeys showing convergent results [203]. The same compound, henceforth called SPC3649 or miravirsen, was subsequently tested as a potential anti-HCV therapy in chronically infected chimpanzees [204]. The treatment resulted in a long-lasting suppression of HCV viremia and no rebound in viremia or adaptive mutations in microRNA-122 target sites have been observed. In 2010 miravirsen entered a phase 2a proof-of-concept clinical trial. The monotherapy with saline-formulated miravirsen caused a dose-dependent and prolonged reduction in HCV levels in patients with chronic HCV genotype 1 infection and a decrease of miR-122 plasma levels [205, 206]. No prominent adverse effects or resistance-associated mutations have been reported. However, C3U nucleotide change in the 5′-UTR of viral RNA has been observed more recently in patients with a viral relapse occurring after treatment with miravirsen and seemed to be responsible for a resistance towards this compound [207]. Further clinical development of miravirsen was discontinued due to the approval of the small molecule drug sofosbuvir as the treatment for the hepatitis C.
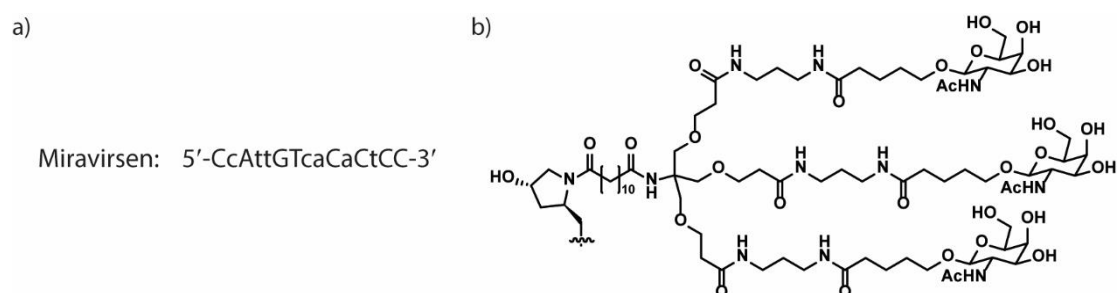
Figure 7: a) Sequence of miravirsen (SPC3649), an LNA/DNA anti-miR with a full PS backbone. Upper case: LNA; lower case: DNA; capital C: LNA C^5-methylcytidine; b) structure of the triantennary *N*-acetylgalactosamine (GalNAc). Adapted from [198, 204].

In late 2014, clinical trial with another anti-miR molecule directed against miR-122, RG-101 (Regulus Therapeutics), has been launched [208]. RG-101 is a *'mixed chemistry phosphorothioate oligonucleotide inhibitor of miR-122'*, conjugated to *N*-acetylgalactosamine (GalNAc) [208] for an enhanced uptake by hepatocytes [209, 210]. GalNAc (**Figure 7b**) is a targeting ligand for the

asialoglycoprotein receptor (ASGPR) abundantly expressed on hepatocytes [211]. When delivered to hepatocytes, RG-101 is metabolized to its main active metabolite (RG1694) [208].

Based on a measured modulation of direct miR-122 target genes, RG-101 was announced to have nearly 20-fold better activity in mice when compared to the unconjugated active oligonucleotide [212]. Treatment with RG-101 resulted in a substantial reduction of viral levels in mice and was reported to be well tolerated in preliminary toxicology studies in mice, rats, and non-human primates. Results from the 1b phase of a clinical trial showed that single subcutaneous injection (2 mg/kg or 4 mg/kg) led to a significant decrease in HCV RNA levels and a decline in total plasma cholesterol levels, without causing serious adverse events [208]. Nevertheless, viral relapse was observed in most of the treated patients. In the patients with HCV genotype 1, this was mainly associated with C3U mutation, which was observed in miravirsen treatment before [207]. Due to a report of a second serious adverse event of jaundice in a patient treated with a single dose of RG-101, the American Food and Drug Administration (FDA) initiated the clinical hold of RG-101 (Regulus Therapeutics press release from June 2016 [213]). Impaired bilirubin transport was pointed out as a most probable cause of the observed hyperbilirubinemia (Regulus Therapeutics press release from June 2017 [214]). In June 2017 Regulus announced that upon completion of the remaining clinical study, clinical development of RG-101 will be terminated [214].

Although none of the two compounds was accepted as a drug, both studies showed the therapeutic potential of the oligonucleotide-based miRNA inhibitors. An important lesson from these clinical trials was that a careful and long-term evaluation of possible side effects through extensive follow-up studies is a must.

Table 1: Summary of anti-miRs and miRNA mimics in clinical trials. *Information collected from the clinicaltrials.gov database [219] and the European Union Clinical Trials Register [220].

Drug candidate	miRNA	Target disease	Company	Clinical trial phase*	Clinical trial identifier*
anti-miRs					
Miravirsen (SPC3649)	miR-122	Hepatitis C Virus	Santaris Pharma/Roche	2 (completed)	NCT01200420
RG-101	miR-122	Hepatitis C Virus	Regulus Therapeutics	2 (completed)	2015-004702-42
SAR339375 (RG-012)	miR-21	Alport Syndrome	Genzyme/ Sanofi	2 (recruiting)	NCT02855268
RGLS4326	miR-17	Autosomal dominant polycystic kidney disease	Regulus Therapeutics	1 (ongoing)	-
AZD4076 (RG-125)	miR-103/ miR-107	Non-alcoholic steatohepatitis	AstraZeneca	1 (active, not recruiting)	NCT02612662
Cobomarsen (MRG-106)	miR-155	Cutaneous T-Cell Lymphoma/Mycosis Fungoides	miRagen Therapeutics	1/2 (recruiting)	NCT02580552 NCT03713320
miRNA mimics					
MRX34	miR-34	Primary liver cancer, small-cell lung cancer, melanoma, lymphoma	Mirna Therapeutics	1 (terminated)	NCT01829971
Remlarsen (MRG-201)	miR-29	Keloids	miRagen Therapeutics	2 (active, not recruiting)	NCT03601052
TargomiR	miR-16	Malignant Pleural Mesothelioma, Non-Small Cell Lung Cancer	EnGeneIC	1 (completed)	NCT02369198

I.1.4.2. Other miRNA-related drug candidates in clinical trials

When administered systemically, single-stranded oligonucleotides with phosphorothioate backbone tend to preferentially accumulate in the liver and kidneys [215]. This is reflected in clinical trials, as most of the drug candidates which were (miravirsen, RG-101) or are currently investigated are directed against miRNAs dysregulated in these two organs. Some types of kidney diseases have been connected with higher levels of miR-21 in this organ and silencing of miR-21 has been indicated as a promising therapeutic strategy for chronic kidney diseases [216].

SAR339375 (RG-012) is an anti-miR directed against miR-21 for the treatment of Alport Syndrome [217]. The recruitment of participants for Phase 2 clinical development to evaluate the safety and efficacy of RG-012 is currently ongoing [218].

To give another example, data from the previous studies suggested that the miRNA-17 family, and especially miR-17, is implicated in the modulation of mitochondrial metabolism and supports the progression of polycystic kidney disease [74]. Regulus Therapeutics proposed RGLS4326 (anti-miR-17 with mixed chemistry and PS backbone) as a potential remedy for autosomal dominant polycystic kidney disease (ADPKD) [221]. Phase 1 investigation was in partial clinical hold due to the unexpected discoveries associated with chronic toxicity study in mice (Regulus Therapeutics press release from July 2018 [222]). Upon running of the additional chronic toxicity studies and fulfilling all the FDA data requirements, the partial clinical hold for multiple ascending doses was announced to be lifted (Regulus Therapeutics press release from December 2019 [223]).

AZD4076 (RG-125) is a GalNAc-conjugated anti-miR against miR-103/107 investigated as a potential treatment for the non-alcoholic steatohepatitis (NASH). AZD4076 is officially still in Phase 1 of the clinical development (active, not recruiting) but it has been announced that AstraZeneca is planning to terminate this clinical program (Regulus Therapeutics press release from June 2017 [214]).

miR-155 is a microRNA with an important role in differentiation, function, and proliferation of blood and lymph cells. Thus, it has been proposed as a therapeutic target in lymphoma cells showing aberrant proliferation. Cobomarsen (MRG-106) from miRagen Therapeutics is an LNA-based anti-miR-155 being in Phase 2 clinical trials (recruitment of patients [224]) for the treatment of mycosis fungoides (MF) form of the cutaneous T-cell lymphoma (CTCL). Although data obtained from the pre-clinical studies look promising [225-227] therapeutic approach based on targeting of miR-155 requires a careful validation [228], as it has been suggested that in some cancer cells miR-155 may also act as a tumour suppressor [229].

Apart from the oligonucleotide-based inhibitors of the miRNA function, miRNA mimics have reached the clinical trial stage as well. Downregulation of the miR-34, being a well-known tumour suppressor, has been reported in many solid and haematological cancer cells [230]. MRX34 from Mirna Therapeutics is a liposomal miR-34 mimic proposed for the treatment of different types of advanced solid tumours [231]. Despite promising results from the pre-clinical studies, Phase 1 clinical trials with this drug candidate were terminated due to several severe adverse events [232]. Remlarsen (MRG-201), a microRNA mimic of miR-29 for the treatment of keloids (an excessive scar tissue formation) [233] is currently in Phase 2 of the clinical investigation, initiated to assess its safety, efficacy and tolerability [234]. EnGeneIC presented TargomiRs as possible replacement therapy for the treatment of recurring malignant pleural mesothelioma (MPM) and advanced non-small cell lung cancer (NSCLC). TargomiRs are minicells loaded with the synthetic miR-16 mimic targeted to Epidermal Growth Factor

Receptor (EGFR) -expressing cells recognized by a specific antibody [235, 236]. Although results from phase 1 investigation showed an acceptable safety profile, they need to be confirmed by studies on a larger population of patients [236].

I.2. RNA-RNA cross-linking

Although first miRNA-related treatments entered clinical trials, the process of the uncovering of miRNA roles in diverse biological pathways is still on-going. A recent discovery of new, non-canonical functions and binding sites [79-87] emphasized that the miRNA world still hides many secrets.

I.2.1. Identification of the miRNA targetome – an overview

Elucidation of the targetome of each miRNA is crucial for the understanding of miRNA mode of action and its function in a biological context [4].

Computational approaches are based on algorithms designed to predict potential binding sites according to established targeting rules and patterns, such as binding to the mRNA target *via* the seed region. Though highly useful, computer-based prediction methods are not without flaws, as they have troubles with the detection of the non-canonical interactions [56]. Moreover, currently available prediction algorithms are characterized by high false positive (contain seed-matching sequence but are not true targets) and false negative (targets which are not predicted by the algorithms but in fact do interact with miRNA) detection rates [237]. This generated a need for experimental validation of theoretically predicted binding sites, which triggered a search for reliable high-throughput methods, able to fulfil this task (summarized in [238]). This gave rise to the development of so-called capture methods, based on the cross-linking and immunoprecipitation (CLIP) protocols. The first CLIP methodology applied for the identification of miRNA targets was the high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation (HITS-CLIP) [239]. HITS-CLIP consists of several core steps, including RNA-protein cross-linking, lysis of the cell and RNA fragmentation, purification of the protein-RNA complex and preparation of the complementary DNA (cDNA) library for sequencing [240]. The cross-linking between RNAs (miRNA or RNA target) and AGO takes place upon irradiation with the ultraviolet (UV) light (λ=254 nm). The covalent bond is preferentially formed between nucleobases (mostly pyrimidines) and amino acids residues (predominantly phenylalanine, tyrosine and cysteine) of the surrounding protein [241-244]. The main advantage coming from the application of the cross-linking step is that it allows to capture binding partners being in the direct contact at the moment of irradiation. Additionally, the generation of a covalent bond between RNA and protein allows more stringent purification steps which helps to reduce the number of false-positive rates.

In the photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) methodology, cells are pre-treated with photo-reactive nucleosides such as 4-thiouridine (4SU) [245]. Their incorporation into endogenous transcripts increases the efficiency of the cross-linking induced by λ=365 nm light. In addition, cross-linking reaction with 4SU is frequently related to thymidine into cytidine mutations observed in the generated cDNA sequence, which facilitates mapping of the interaction sites.

Although these sequencing-based experimental approaches provide an overview of genome-wide interactions with RISC, they do not enable to elucidate binding sites for the particular miRNA. In order to be able to capture the targetome for a miRNA of interest, a miRNA cross-linking and immunoprecipitation (miR-CLIP) technique was developed in our laboratory [84]. The miR-CLIP

approach is based on the synthetic pre-miR-CLIP probes (capture probes), modified with trioxsalen and biotin. In short, cells are first transfected with the bis-labelled pre-miR-CLIP probe (**Figure 8**). In the next step, cells are irradiated, first with λ=254 nm light to cross-link RNAs to proteins in RISC, and then with λ=365 nm light to covalently attach miRNA to its targets. Subsequent lysis of the cells, followed by AGO2 immunoprecipitation (IP) step results in isolation of AGO2 proteins together with the cross-linked RNAs. The biotin-streptavidin affinity purification, preceded by digestion of AGO2 proteins, allows to catch only the probe-linked RNAs. cDNAs libraries are then prepared and sequenced to obtain a set of enriched potential targets.
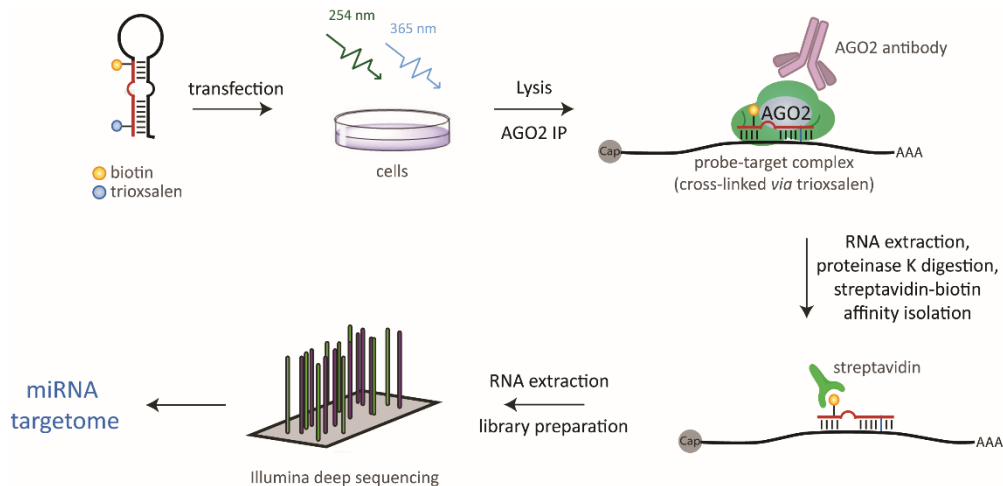
Figure 8: Overview of the miR-CLIP protocol. IP: immunoprecipitation. Adapted from [84].

Generation of the covalent linkage between the miRNA and its target is a key step of the miR-CLIP protocol and is carried out by the trioxsalen molecule attached to the pre-miR-CLIP probe. The application of trioxsalens (psoralen derivatives) for nucleic acids cross-linking has a rich history, which will be briefly presented in the next sections of this chapter.

I.2.2. The photochemistry of psoralens

Psoralens are a family of naturally occurring and synthetic bifunctional, tricyclic furanocoumarins originally discovered in plants [246, 247]. The characteristic feature of psoralens is their ability to covalently link (cross-link) double-stranded nucleic acids upon irradiation with long-wavelength UV light (320-380 nm). The cross-linking occurs due to the formation of covalent adducts between psoralen and pyrimidine bases of nucleic acids.

The process of cross-linking with free (unbound) psoralens can be divided into three stages [247, 248]. In the first step, psoralen intercalates into a double-helical region of the nucleic acid. Upon the absorption of the first photon (λ=360 nm), and provided that the pyrimidine base is present in the proximity of the psoralen molecule, a cyclobutane mono-adduct is formed *via* [2+2] cycloaddition reaction (**Figure 9**). If the covalent bond is generated between the 5,6 double bond of pyrimidine and the 3,4 double bond of the psoralen, the product is referred to as a pyrone-side mono-adduct. When the reaction involves the 4',5' double bond of the psoralen, the product is a furan-side mono-adduct. Side at which the first mono-addition takes place is important as it determines whether the second mono-addition can occur or not. If the furan-side mono-adduct is formed first, and if the unreacted pyrone double bond is located at a cross-linkable site (is in proximity to pyrimidine base from the other strand) [249], the second photon can be absorbed resulting in the formation of a pyrimidine-psoralen-

pyrimidine di-adduct (cross-linked product) [246, 247]. On the other hand, cross-linking does not happen when pyrone-side mono-adduct is generated first since it cannot absorb at 360 nm. Previous research has established that conformational change within the DNA structure is required for the cross-linking to occur [250]. The photoreaction may be reversed upon irradiating with short-wavelength UV light (λ=254 nm) [247]. Psoralens have been shown to react mostly with thymidines in DNA and uridines in RNA, although adduct formation with other bases has also been reported [246].
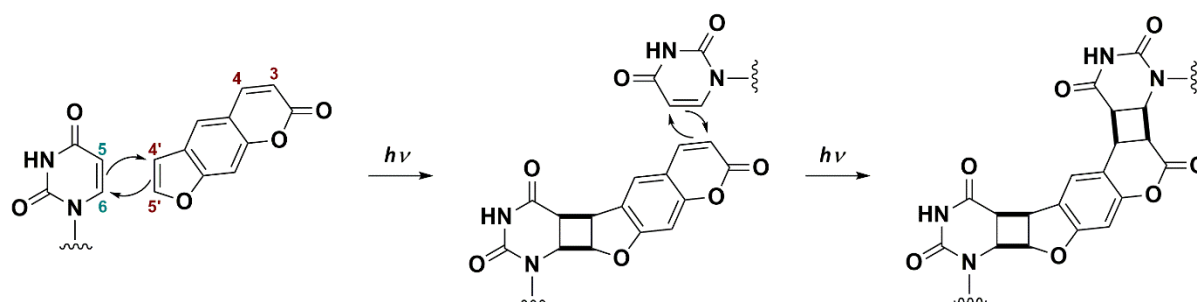
Figure 9: Mechanism of the psoralen-based RNA-RNA cross-linking. Adapted from [251].

Diverse psoralen analogues have been synthesized to modulate its chemical and physical properties such as reactivity or solubility [246, 252]. Among the synthesized compounds, trioxsalen (4,5',8-trimethylpsoralen or TMP), 4'-aminomethyltrioxsalen (AMT) and 8-methoxypsoralen (8-MOP) are the most known and widely used ones (**Figure 10**) [246]. A lot of effort has been made to explore rules governing the photochemistry of psoralens and most of the acquired knowledge comes from the fundamental experiments with free psoralen or its derivatives and DNA or RNA duplexes. Studies on mono-addition of 4'-hydroxymethyltrioxsalen (HMT) to tRNA showed that the reactivity of the RNA is correlated with its secondary structure [253]. HMT addition seems to occur preferentially with uridines located at the ends of loose helices or with uridine in GU pairs within the helix (so at the locations facilitating intercalation) and is influenced by the salt concentration and temperature [253, 254]. Psoralen forms mono-adducts preferentially with pyrimidines at the 5'-TpA-3' sites in DNA [255], while in RNA sites containing adjacent uridines (5'-UpU-3') seems to be preferred [253, 254].
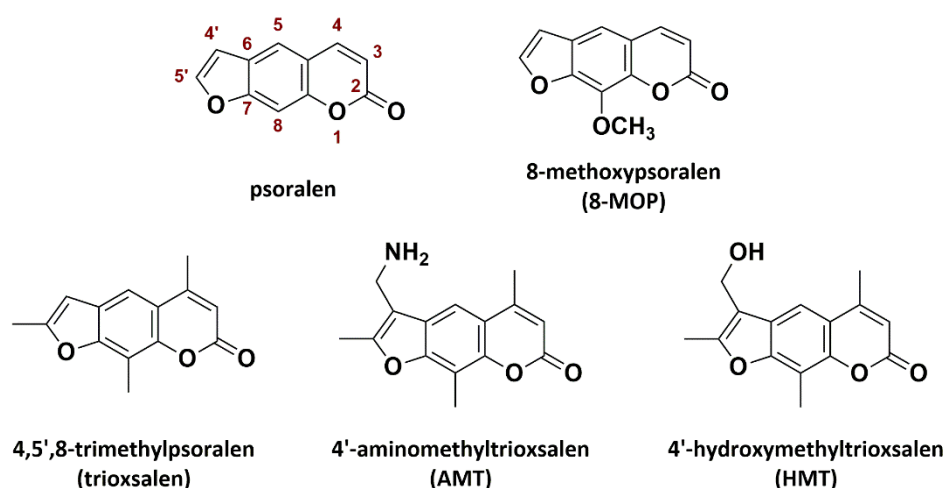
Figure 10: Chemical structures of psoralen and its selected analogues.

Psoralens have a clinical application in psoralen and ultraviolet A (PUVA) therapy for the treatment of skin disorders such as psoriasis or vitiligo [248, 256]. Moreover, due to their unique ability to 'freeze'

long-range interactions, psoralens are considered as one of the most important classes of photochemical reagents that can be used to explore nucleic acid structure and function [248]. The main arguments supporting the use of this class of compounds are the high stability of psoralen and its adducts [257] and the fact that the cross-linking can be performed under various conditions (buffer composition, temperature) and with different psoralen analogues, allowing fine-tuning of the experimental settings according to the investigated system [248]. Furthermore, psoralens absorption band (320-380 nm) is higher than the absorption band of nucleic acids and proteins which makes the irradiation less detrimental for the cells. Finally, their photochemistry and the protocols for RNA cross-linking are well established [258].

Free psoralens were used in many biological inquires (summarized in [248]), including investigation of the chromatin structure [259], secondary structure of various single-stranded DNAs or RNAs and the interactions between them [260, 261]. As psoralens have been reported to cross-link to proteins (though to a much lower extent than with nucleic acids), they can be also applied to study protein-nucleic acid interactions [257, 262].

I.2.3. Site-specific incorporation of trioxsalen into oligonucleotides

The growing need for a site-specific cross-linking with various RNA and DNA structures resulted in the synthesis of multiple psoralen-conjugated oligo(deoxy)ribonucleotide analogues. Psoralens (mainly trioxsalen) were attached at different sites of the oligonucleotides, such as the 5'-end phosphate group, nucleobase or sugar ring (**Table 2**) [263]. Many of these studies were initiated with a purpose to improve the properties of antisense oligonucleotides by their cross-linking to the target. The term 'cross-linking' used here (and later on in this dissertation) should be understood as the generation of the covalent bond between the two strands of the DNA and/or RNA of interest. When the cross-linking is achieved with the application of the trioxsalen molecule that is already attached to one of the strands, it is, according to the photochemistry of psoralens, actually a mono-addition reaction.

Initial efforts focused on the synthesis of compounds in which trioxsalen was appended at the 5'-end of oligodeoxyribonucleotides *via* linkers of various lengths and with a use of different types of chemistry (structures I-III, **Table 2**) [264-268]. It has been observed that the length of the linker has a significant influence on the outcome of the cross-linking process [264]. In reported cases, derivatization with aminoethyl linker gave the best results. Even though longer linkers could also place the trioxsalen in the proximity of the targeted uridine, the enhanced freedom of the furanocoumarin reduced the efficiency of the cross-linking and increased off-target reactions [265]. The effectiveness of cross-linking was also affected by the sequence complementarity and the structure of the target strand [264, 269].

Conclusions concerning properties of the linkers appended at the 5'-end hold true for the linkers used to conjugate trioxsalen at the 2'-*O*-position of the ribose moiety. Higuchi et al. claimed that oligodeoxynucleotide with trioxsalen at the 2'-*O*-position of designated adenosine (attachment *via* methylene linker; 2'-Ps-met) (structure IV, **Table 2**) seems to manifest higher sequence-specificity than these with trioxsalen introduced at the 5'-end, as 2'-Ps-met was able to distinguish a single base mutation in the target RNA strand [270]. The main drawback of this approach was that the yield of the cross-linking with the 2'-Ps-met was relatively low and required long irradiation time (25% yield after 60 min of irradiation). To overcome this shortcoming, the influence of length and flexibility of the applied linker on the efficiency of the cross-linking process was explored. A set of alkoxymethylene linkers was synthesized and employed to attach trioxsalen at the 2'-*O*-position of the adenosine

(structure V, **Table 2**) [271]. The design of linkers was based on conformational calculations for the model DNA-RNA duplex, which showed that trioxsalen introduced *via* ethoxymethylene (2'-Ps-eom) or propoxymethylene (2'-Ps-pom) linkers should be placed nearby the target uridine, while methylene linker (2'-Ps-met) seemed to be too short to enable efficient cross-linking. These predictions were confirmed experimentally. Both, 2'-Ps-pom and especially 2'-Ps-eom were able to efficiently crosslink to the target sequence in a sequence-specific manner upon shorter irradiation time (respectively 50% and 75% conversion to the cross-linked product after 1 min of treatment with the UV light). Reaction with butoxymethylene linker did not result in the formation of the inter-strand cross-linking product.
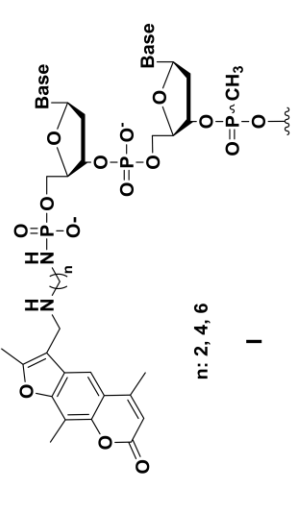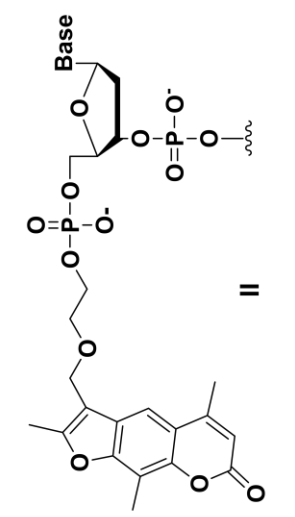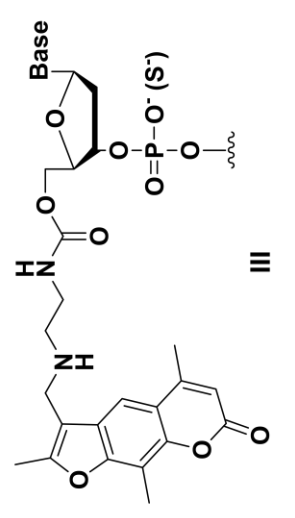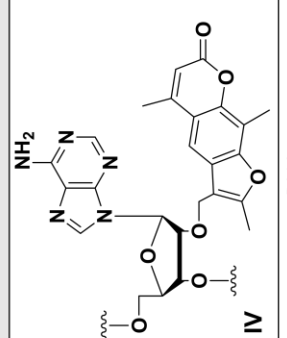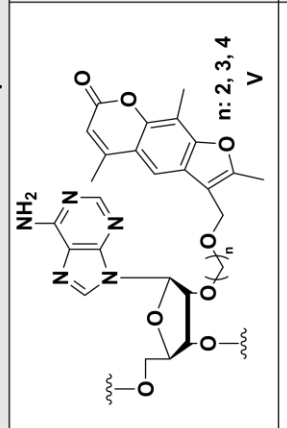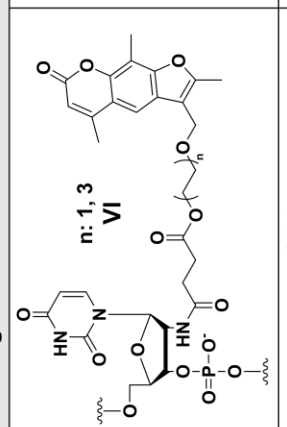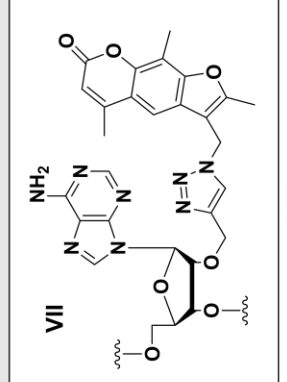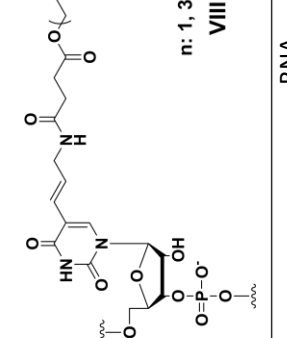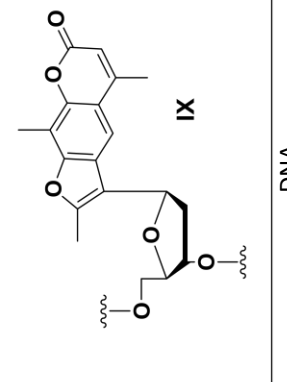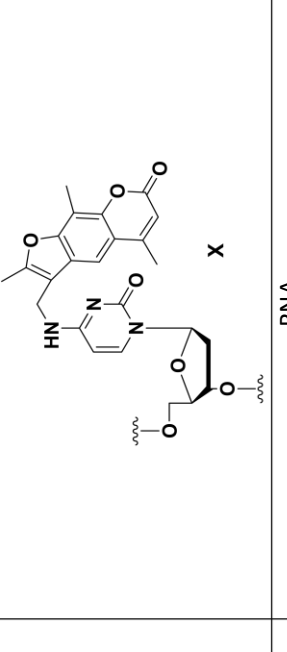
What can be learnt from those experiments is that the site of attachment and linker length have a great impact on the outcome of the cross-linking process with trioxsalen. Too short or too rigid linkers may not be able to deliver the cross-linker close enough for the reaction to occur. On the other hand, excessive length and flexibility of the linker may enhance cyclobutane formation with off-target uridines present in the vicinity. Perfect positioning can perhaps only be found by empirical experimentation.

Additionally, with a view to facilitate target identification, trioxsalen was also incorporated into miRNA sequence, which upon irradiation cross-linked to the functional mRNA targets in cells [272]. Photo-reactive moiety was appended either at the 2'- or C^5-position of the uridine (structure VI and VIII, **Table 2**) *via* two types of linkers (a short one and a long one); the shorter linkers were reported to be more efficient. Photo-reactive antisense oligonucleotides containing adenosine modified with trioxsalen attached through ethoxymethylene linker (structure V, n=2, **Table 2**) were used to impede RISC function in cells [273]. Last but not least, trioxsalen was also used in the aforementioned miR-CLIP approach, where it was pre-synthetically conjugated at the 2'-*O*-position of adenosine (structure VII, **Table 2**) with use of copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) reaction [84]. Strengths and weaknesses of this strategy will be discussed in the next chapter. Finally, oligonucleotides in which trioxsalen replaces the nucleobase (structure IX, **Table 2**) [274] or is attached at the N^4 position of deoxycytidine (structure X, **Table 2**) [275] have been also described.

Cross-linking of the nucleic acids has been also achieved with the use of other classes of photo-reactive compounds, some of which were transferred from the RNA-protein cross-linking field. A DNA-DNA or DNA-RNA cross-linking with coumarin analogues [276], 5-cyanovinyldeoxyuridine [277] and diazirines (which will be described more in detail in Chapter III) [278] have been reported.

Table 2: Selected examples of the site-specific incorporation of trioxsalen into oligonucleotide sequence for DNA-DNA, DNA-RNA or RNA-RNA cross-linking.

Trioxsalen at the 5´-end of the oligonucleotide

Structure		
I n: 2, 4, 6	**II**	**III**
Target		
DNA, RNA	DNA	RNA
Reference		
[264, 265]	[266]	[267, 268]

Trioxsalen at the 2´-O-position of the oligonucleotide

Structure		
IV	**V** n: 2, 3, 4	**VI** n: 1, 3
Target		
RNA	RNA	RNA
Reference		
[270]	[271, 273]	[272]

Trioxsalen attached to the nucleobase/ mimicking nucleobase

Structure		
VIII n: 1, 3	**IX**	**X**
Target		
RNA	DNA	RNA
Reference		
[272]	[274]	[275]

VII

RNA

[84]

23

II. Chapter II: Exploring the nature of the cross-linking mechanism with the trioxsalen-labelled microRNA probes

II.1. Introduction

In the original miR-CLIP protocol, the use of bis-labelled probes for pre-miR-106a was described [84]. miR-106a-5p is a member of the miR-17-5p family having a potentially oncogenic function [279]. Capturing of the miR-106a-5p targetome in HeLa cells revealed its interaction with canonical but also non-canonical target sites, such as binding to the lncRNA H19.

As mentioned in section I.2.1, pre-miR-CLIP probes are synthetic analogues of the pre-miRNA of interest. The probes are bis-labelled, with trioxsalen and biotin employed respectively for the miRNA-target cross-linking and biotin-streptavidin pulldown (second purification step). Both modifications are attached by a highly selective copper(I)-catalysed azide-alkyne 1,3-dipolar cycloaddition (CuAAC) reaction (also known as 'CLICK' chemistry). This variant of the CLICK chemistry, developed independently by Meldal and Sharpless groups [280, 281], is a versatile and straightforward tool widely used for the efficient post-synthetic functionalization of oligonucleotide sequences [282].

Upon irradiation with the UV light (λ=365 nm), trioxsalen preferentially reacts with uridines (in RNAs) [248]. Thus, its attachment to adenosine so that the cross-linking to the base-paired uridine in the target strand could occur appeared as a reasonable approach. Hence, in the published miR-CLIP account, the trioxsalen was appended at the 2'-*O*-position of one of the adenosines present in the seed sequence, as the pairing between a miRNA and an mRNA in the seed region is known to be complementary. The synthesis of pre-miR-CLIP probes requires site-specific attachment of two different functional moieties. Therefore, orthogonal chemistries need to be applied to avoid cross-reactivity at the activated sites during the functionalization [283].
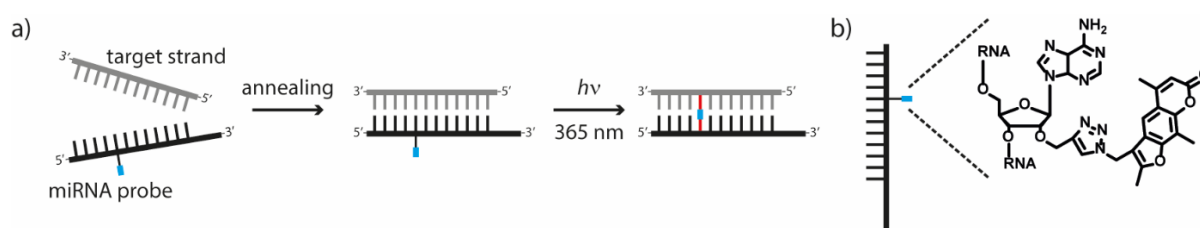
Figure 11: a) Schematic representation of the *in vitro* photo-cross-linking between the trioxsalen-labelled miRNA and the counter-strand (target strand); b) structure of adenosine modified with trioxsalen at the 2'-*O*-position of the ribose.

For this reason, the 4'-azidomethyltrioxsalen was pre-synthetically CLICKed with the 2'-*O*-propargyl adenosine and introduced during the solid phase oligonucleotide synthesis as a phosphoramidite. Biotin, on the other hand, was attached post-synthetically, with use of the CLICK chemistry on the solid support (reaction between triethylene glycol-biotin azide and 2'-*O*-propargyl adenosine).

An *in vitro* photo-cross-linking assay (schematically depicted on **Figure 11a**) with a mono-labelled miR-106a-5p analogue bearing only trioxsalen and its short (15-nt), fully complementary counter-strand (fc ctstr) mimicking the target confirmed the formation of a cross-linked product. However, the miR-106a-5p sequence is unusual, in that it has four contiguous adenosines at its 5'-end (nucleotides g_1-g_4,

Figure 12), that will base-pair with four uridines in its target mRNAs (nucleotides t_1-t_4, **Figure 12**). Each of these uridines can potentially react with trioxsalen and form a cyclobutane mono-adduct.

fc ctstr (ORN-2): $t_1 t_2 t_3 t_4$
3'-UUUUCACGAAUGUCA-5'

miR-106a-5p (ORN-1): 5'-AAAXGUGCUUACAGUGCAGGUAG-3'
                        $g_1 g_2 g_3 g_4$

                        -3 -2 -1
3'-UUUUCACGAAUGUCA-5'
                          ?
5'-AAAXGUGCUUACAGUGCAGGUAG-3'
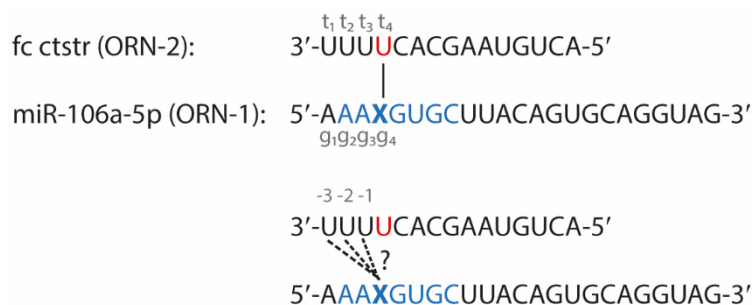
Figure 12: The desired (top) and possible (bottom) cross-linking sites for the D1 duplex indicated by constant or dashed lines, respectively. The seed region of the miRNA is shown in blue and the originally expected target site of the cross-linking (uridine t_4) is marked in red. Fc ctstr: fully complementary counter-strand.

The CLICK reaction between the azide and propargyl functional groups generates a 1,2,3-triazole moiety in the linker (**Figure 11b**). The presence of the triazole increases the rigidity of the molecule, affecting the overall "reach" of the trioxsalen. As mentioned in section I.2.3, the length and flexibility of the linker greatly influence the efficiency of the cross-linking process. Thus, the linker used in the original miR-CLIP protocol may be unable to place the trioxsalen so that it could cross-link to the juxtaposed uridine when applied for other miRNAs with less A-rich sequences/U-rich targets. In fact, the actual site of the cross-linking for the reported miR-106a-5p probe has not been determined and unpublished observations suggested that the covalent linkage is not formed with the uridine at position t_4. It seemed likely that the efficient cross-linking of this miR-106a-5p probe was favoured by the presence of four contiguous uridines in the counter-strands. Bis-labelled miR-CLIP probes in a form of duplexes were prepared for let-7g-5p and miR-34a-5p. Although they were reported to be active in cellular assays, they were not tested in the *in vitro* photo-cross-linking experiments [84].

II.2. Aim of the project

The ultimate goal of this part of the project was to characterize the trioxsalen-modified miR-106a-5p probe used in the published miR-CLIP study and to determine which of the four uridines (t_1 to t_4, **Figure 12**) in the miR-106a-5p counter-strand participated in cross-linking. Simultaneously, trioxsalen-modified probes for other miRNAs of interest were prepared and investigated. Observations from these experiments were expected to provide guidelines on how to improve the design of a miR-CLIP probe for any miRNA.

II.3. Results and discussion

II.3.1. *In vitro* photo-cross-linking – pilot experiments

The first experiments aimed to reproduce conditions from the *in vitro* photo-cross-linking assay established by Dr Andreas Brunschweiger and verify if the cross-linking can be achieved for other miRNAs. For this purpose, trioxsalen-labelled analogues of miR-124-3p and miR-132-3p (sequences in **Figure 13**), two miRNAs of ongoing interest in the laboratory, were prepared. miR-132-3p and miR-124-3p are highly enriched in brain tissues. What makes these miRNAs particularly interesting, is that they were reported to be subjected to the TDMD process in rodent primary neurons [78].

miR-124-3p: 5'-UAAGGCACGCGGUGAAUGCC-3'

miR-132-3p: 5'-UAACAGUCUACAGCCAUGGUCG-3'

Figure 13: Native sequences of miR-124-3p and miR-132-3p.

The initial cross-linking assays were performed under various conditions. Samples (containing either single-strands or duplexes) were analysed by direct injection on the Liquid-Chromatography Mass-Spectrometry (LC-MS) system. Apart from a putative intra-strand cross-linking detected for some of the probes (observed as peaks with shifted retention time (R_t)), no inter-strand cross-linking within the duplex was observed for any of the tested conditions (data not shown). This provided the first experimental evidence that the probe design reported by Imig et al. might not be universal. To exclude the possibility that lack of the observed cross-linking was related to the technical issues, the original trioxsalen-modified miR-106a-5p probe was re-synthesized and used as a positive control. When this probe was tested (D1), the cross-linked product was detected (**Figure 14**).
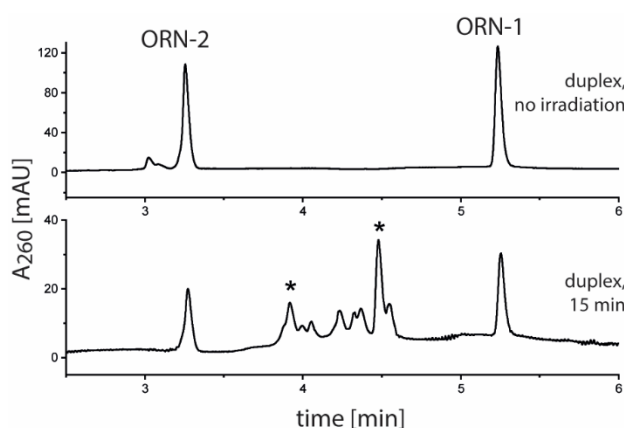
Figure 14: Representative RP-HPLC chromatograms from the *in vitro* photo-cross-linking assay for duplex D1 between ORN-1 (trioxsalen-modified miR-106a-5p) and ORN-2 (fully complementary counter-strand). Chromatograms present samples before (top) and after 15 min (bottom) of the irradiation at λ=365 nm. Peaks marked with an asterisk (*) correspond to the cross-linked duplex (mass calc.: 12452.0, mass found: 12452.4 and 12473.8 [M+Na]).

II.3.2. Synthesis of the small library of the trioxsalen-modified miR-106a-5p analogues

To explore the selectivity of the cross-linking for the miR-106a-5p probe, a small library of trioxsalen-labelled miR-106a-5p analogues was synthesized (**Table 3**). The library was assembled by the introduction of systematic AU to CG mutations at chosen positions within the duplex. The trioxsalen was introduced post-synthetically by the CLICK reaction between the propargyl-bearing oligonucleotide and 4'-azidomethyltrioxsalen, following the protocol developed in the laboratory [284]. Counter-strands used in these experiments were 15-nt-long and were for the most part fully complementary to the 5'-end of the miRNA analogue (ORN-17 and ORN-18 had one mismatch) to ensure stable duplex formation under cross-linking conditions and easy separation by LC-MS.

Table 3: Synthesized trioxsalen-labelled miR-106a-5p analogues and their counter-strands. Trioxsalen was attached post-synthetically (CuAAC reaction on the solid support) at the 2′-*O*-position of adenosine (**X**), guanosine (**Y**), uridine (**Z**) or cytidine (**W**). In red: AU to CG base pair mutations with regard to the original sequence of miR-106a-5p.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found
1	ORN-1	AAA**X**GUGCUUACAGUGCAGGUAG	23	7754.0	7754.9
2	ORN-2	ACUGUAAGCACUUUU	15	4698.0	4697.1
3	ORN-3	CAA**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7733.1
4	ORN-4	ACUGUAAGCACUUUG	15	4736.9	4736.9
5	ORN-5	ACA**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7730.2
6	ORN-6	ACUGUAAGCACUUGU	15	4736.9	4736.1
7	ORN-7	AAC**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7730.9
8	ORN-8	ACUGUAAGCACUGUU	15	4736.9	4736.0
9	ORN-9	CCA**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7706.8
10	ORN-10	ACUGUAAGCACUUGG	15	4776.0	4775.1
11	ORN-11	CAC**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7707.0
12	ORN-12	ACUGUAAGCACUGUG	15	4776.0	4774.9
13	ORN-13	ACC**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7706.8
14	ORN-14	ACUGUAAGCACUGGU	15	4776.0	4775.1
15	ORN-15	CCC**X**GUGCUUACAGUGCAGGUAG	23	7683.8	7682.5
16	ORN-16	ACUGUAAGCACUGGG	15	4815.0	4814.2
17	ORN-17	ACUGUAAGCACCGGU	15	4775.0	4774.0
18	ORN-18	ACUGUAAGCACCGUG	15	4775.0	4771.1
19	ORN-19	AAAA**Y**UGCUUACAGUGCAGGUAG	23	7756.0	7755.4
20	ORN-20	CACC**Y**UGCUUACAGUGCAGGUAG	23	7683.9	7682.8
21	ORN-21	ACUGUAAGCACGGUG	15	4815.0	4813.9
22	ORN-22	AAAAG**Z**GCUUACAGUGCAGGUAG	23	7756.0	7755.0
23	ORN-23	ACACG**Z**GCUUACAGUGCAGGUAG	23	7707.9	7706.5
24	ORN-24	ACUGUAAGCACGUGU	15	4775.9	4775.1
25	ORN-25	AAAAGU**Y**CUUACAGUGCAGGUAG	23	7756.0	7755.1
26	ORN-26	AAAAGUG**W**UUACAGUGCAGGUAG	23	7756.0	7755.1
27	ORN-27	GCCAGUGACC**X**CUGUGCAGGUAG	23	7722.9	7722.0
28	ORN-28	ACAGUGGUCACUGGC	15	4791.0	4790.4

Initially, single mutations were introduced to disturb the four contiguous uridines (t_1-t_4) in the target strand (**Table 4** and **Figure 15**). Uridines were systematically replaced with guanosines to eliminate cross-linking with a nucleotide at this position (guanosines were chosen as cytidines may potentially react with trioxsalen [248]). Samples were prepared by mixing equivalent amounts of the two strands which were then lyophilized and re-dissolved in 200 µL of phosphate buffer. Samples were annealed by incubation at 95°C for 5 min and then cooled to room temperature over 2 h. Subsequently, samples were transferred to an open 24-well plate which was placed on ice and irradiated for 15 min. Non-irradiated samples were used as negative controls.

Table 4: Summary of the *in vitro* photo-cross-linking experiments performed with miR-106a-5p analogues with one mutation within the sequence. In red: AU to CG base pair mutations with regard to the original sequence of miR-106a-5p. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 1 µM. **X** designates adenosine with the trioxsalen attached at the 2′-*O*-position of the ribose. Fc ctstr: fully complementary counter-strand. The classification 'detected' or 'not detected' used here refers to the ability to identify the cross-linking product in the *in vitro* assay. Masses of the isolated peaks are presented in section VII.1.4.1.

Duplex	ORN	Sequence	Cross-linked product
		the original miR-106a-5p probe	
D1	ORN-1 (probe)	5′–AAA**X**GUGCUUACAGUGCAGGUAG–3′	detected
	ORN-2 (fc ctstr)	3′–UUUUCACGAAUGUCA–5′	
		miR-106a-5p probe analogues with a single mutation	
D2	ORN-3 (probe)	5′–CAA**X**GUGCUUACAGUGCAGGUAG–3′	not detected
	ORN-4 (fc ctstr)	3′–GUUUCACGAAUGUCA–5′	
D3	ORN-5 (probe)	5′–ACA**X**GUGCUUACAGUGCAGGUAG–3′	detected
	ORN-6 (fc ctstr)	3′–UGUUCACGAAUGUCA–5′	
D4	ORN-7 (probe)	5′–AAC**X**GUGCUUACAGUGCAGGUAG–3′	detected
	ORN-8 (fc ctstr)	3′–UUGUCACGAAUGUCA–5′	

Surprisingly, cross-linking was abolished only when the first of the four uridines at the 3′-end of the counter-strand (t_1 base-pairing with g_1; duplex D2, **Table 4**) was mutated, even though the uridine t_4 was present. This suggested that uridine t_1 was engaged in the cross-linking with the original miR-106a-5p probe (duplex D1). For duplexes D3 and D4, having the second (t_2) and the third (t_3) uridine mutated, respectively, masses corresponding to the cross-linked products were detected.
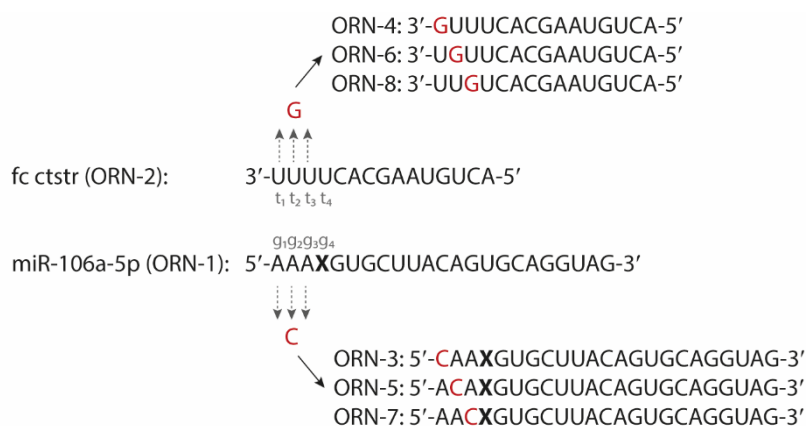
Figure 15: The design of the first set of mutations within the miR-106a-5p sequence. **X**: adenosine with trioxsalen at the 2′-*O*-position of the ribose. In red: AU to CG base pair mutations with regard to the original sequence of miR-106a-5p. Fc ctstr: fully complementary counter-strand.

Single mutations were followed by double and triple mutations (duplexes D5-D10, **Table 5**) in the same region of the duplex. Results from these experiments were in line with previous observations. The presence of uridine at the very 3′-end of the counter-strand appeared to be necessary, but not sufficient for the cross-linking to occur with this probe, as the cross-linking product was not detected for duplex D9 (**Table 5**). Interestingly, these findings were partially in line with conclusions from RNA-RNA cross-linking with free trioxsalen, which suggested that the trioxsalen mono-adducts are preferentially

formed at 5'-UpU-3' sites and/or at the ends of helical regions of a duplex [253, 254]. Favoured intercalation at the end of the duplex could be explained by the more relaxed structure of the duplex in this region, a phenomenon referred to as 'breathing' of the duplex [285, 286]. Cross-linking was observed for duplex D7 but not for duplex D9, which differed by the presence of a mismatched base at position t_4, juxtaposed to the modified adenosine (**Table 5**). This suggested that the change in the local topology of the duplex, driven by mismatched base-pairing of the trioxsalen-bearing adenosine, resulted in abolished cross-linking. Eventually, it is also possible that for some of the tested duplexes the cross-linking did occur but was not detected.

Table 5: Summary of the *in vitro* photo-cross-linking experiments with miR-106a-5p analogues with two or more mutations in the sequence. In red: AU to CG base pair mutations with regard to the original sequence of the miR-106a-5p. Underlined: mismatched base pairs. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 1 µM. **X** designates adenosine with the trioxsalen attached at the 2'-*O*-position of the ribose. (Fc) ctstr: (fully complementary) counter-strand. Masses of the isolated peaks are presented in section VII.1.4.1.

Duplex	ORN	Sequence	Cross-linked product
miR-106a-5p probe analogues with two or more mutations		
D5	ORN-9 (probe)	5'–CCA**X**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-10 (fc ctstr)	3'–GGUUCACGAAUGUCA–5'	
D6	ORN-11 (probe)	5'–CAC**X**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-12 (fc ctstr)	3'–GUGUCACGAAUGUCA–5'	
D7	ORN-13 (probe)	5'–ACC**X**GUGCUUACAGUGCAGGUAG–3'	detected
	ORN-14 (fc ctstr)	3'–UGGUCACGAAUGUCA–5'	
D8	ORN-15 (probe)	5'–CCC**X**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-16 (fc ctstr)	3'–GGGUCACGAAUGUCA–5'	
D9	ORN-13 (probe)	5'–ACC**X**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-17 (ctstr)	3'–UGGCCACGAAUGUCA–5'	
D10	ORN-11 (probe)	5'–CAC**X**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-18 (ctstr)	3'–GUGCCACGAAUGUCA–5'	

The third set of experiments was based on probes in which trioxsalen was systematically attached at the 2'-*O*-position of proximal bases in the miR-106a-5p sequence (g_5-g_8 and g_{11}), towards the central part of the duplex (D11-D17, **Table 6**).

Due to the difficulties with an unambiguous interpretation of the chromatograms obtained by direct injection of the samples on the LC-MS system, the protocol for the sample analysis was modified. After irradiation, samples were injected on the Reverse-Phase High-Performance Liquid Chromatography (RP-HPLC). Peaks with an intensity (A_{260} absorbance) above the set threshold (10-25 mAU, depending on the experiment and the current performance of the instrument) were collected, dried, re-dissolved in ultrapure water and subjected to the analysis by the LC-MS. This produced cleaner chromatograms, although cross-linked products formed in low yields might not have been collected. Illustrative chromatograms from the *in vitro* cross-linking with duplexes for which the cross-linking was (D14) or was not (D13) detected are presented in **Figure 16**.

Table 6: Summary of the *in vitro* photo-cross-linking experiments performed with miR-106a-5p analogues with trioxsalen attached to different bases. In red: AU to CG base pair mutations with regard to the original sequence of miR-106a-5p. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 1 µM. **X**, **Y**, **Z**, **W** designate adenosine, guanosine, uridine or cytidine with the trioxsalen attached at the 2'-*O*-position of the ribose, respectively. Fc ctstr: fully complementary counter-strand. Masses of the isolated peaks are presented in section VII.1.4.1.

Duplex	ORN	Sequence	Cross-linked product
miR-106a-5p probe analogues with trioxsalen attached to different bases		
D11	ORN-19 (probe)	5'–AAAA**Y**UGCUUACAGUGCAGGUAG–3'	detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D12	ORN-20 (probe)	5'–CACC**Y**UGCUUACAGUGCAGGUAG–3'	not detected
	ORN-21 (fc ctstr)	3'–GUGGCACGAAUGUCA–5'	
D13	ORN-22 (probe)	5'–AAAAG**Z**GCUUACAGUGCAGGUAG–3'	not detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D14	ORN-23 (probe)	5'–A**C**ACG**Z**GCUUACAGUGCAGGUAG–3'	detected
	ORN-24 (fc ctstr)	3'–U**G**UGCACGAAUGUCA–5'	
D15	ORN-25 (probe)	5'–AAAAGU**Y**CUUACAGUGCAGGUAG–3'	detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D16	ORN-26 (probe)	5'–AAAAGUG**W**UUACAGUGCAGGUAG–3'	not detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D17	ORN-27 (probe)	5'–GCCAGUGACC**X**CUGUGCAGGUAG–3'	not detected
	ORN-28 (fc ctstr)	3'–CGGUCACUGGUGACA–5'	

Figure 16: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with the duplexes for which the cross-linked product a) was (D14) and b) was not (D13) detected. Top panels present the non-irradiated samples (references). Bottom panels represent samples irradiated for 15 min (λ=365 nm). Peaks marked with an asterisk (*) correspond to the cross-linked duplex (mass calc.: 12483.8, mass found: 12553.2 and 12531.1 for D14). The shoulders visible on the chromatograms are related to the issues with the used RP-HPLC column.

The top panels represent the non-irradiated samples which served as references. Peaks with the lowest and the highest R_ts correspond to the unmodified counter-strands and trioxsalen-labelled probes, respectively. The hydrophobic nature of the trioxsalen causes a significant shift in the R_t of the probes, which simplifies separation and analysis. Cross-linked products appear on the chromatograms between the parent compounds, concomitant with the decrease in the intensity of both single-strands. A drop in the intensity of the probe strand only usually indicates a change of its secondary structure or formation of an intra-strand cross-linking product upon irradiation.

The observed cross-linking pattern was inconsistent, suggesting its dependence on various factors. This includes the sequence and the region of the duplex (terminal or central part) within which the modification was attached. In a study published by Matsuyama et al., trioxsalen appended at the 2'-O-position of the adenosine through a short flexible ethoxymethylene linker (structure V, n=2; **Table 2**) cross-linked to the base-paired uridine in the target strand [273]. However, although this strategy is appealing, it was demonstrated for only one sequence and its general applicability was not assessed.

A time-course cross-linking experiment was conducted for duplexes, for which the most prominent cross-linking had been observed. Samples were irradiated for 5, 10 or 15 minutes. In miR-CLIP experiments, it is important that irradiation conditions are mild. As presented in **Figure 17a**, the intensity of the newly-formed cross-linked product increased with the time of irradiation, and was accompanied by the disappearance of signals from both of the single-strands. For D3, the maximum cross-linking was achieved after 10 min of irradiation (**Figure 17b**).
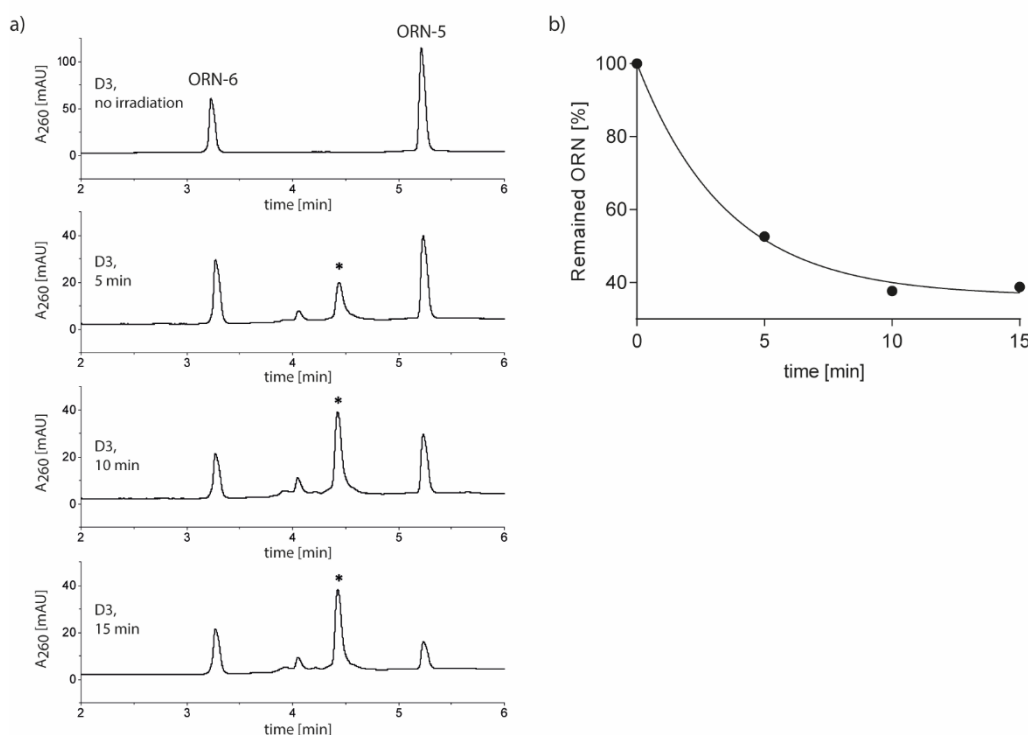
Figure 17: a) RP-HPLC chromatograms from the *in vitro* photo-cross-linking time-course experiment with D3; peaks indicated with an asterisk (*) correspond to the cross-linked duplex (mass calc.: 12468.9, mass found: 12469.6); b) graph showing remaining ORN-6 [%] after 0, 5, 10 and 15 min of irradiation (λ=365 nm).

II.3.3. Cross-linking with miR-124-3p and miR-132-3p probes

A large part of our research was focused on probes for miR-124-3p and miR-132-3p. miR-CLIP experiments with pre-miR-CLIP probes for these miRNAs gave promising results, as described in the doctoral thesis of Dr Yuluan Wang entitled 'Tailoring CLIP-based methods for exploring the miRNA targetome' [287]. The pre-miR-CLIP probes for miR-124-3p and miR-132-3p were designed following the strategy from the original miR-CLIP, i.e. to cross-link to the juxtaposed uridine.

II.3.3.1. miR-124-3p

The properties of two pre-miR-CLIP-124 probes were compared in cellular experiments to investigate whether the location of the trioxsalen has an influence on the capture of the miRNA targetome [287]. Thus, their cross-linking properties *in vitro* were also compared. In miR-124-3p-1 probe (ORN-29), trioxsalen was placed in the 3'-region of the miRNA, while miR-124-3p-2 probe had trioxsalen attached within the seed region (ORN-30) (**Table 7**).

Table 7: Synthesized trioxsalen-labelled miR-124-3p and miR-132-3p analogues and their counter-strands. Trioxsalen was attached post-synthetically (CuAAC reaction on the solid support) at the 2'-*O*-position of adenosine (**X**).

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found
		miR-124-3p analogues			
1	ORN-29	UAAGGCACGCGGUG**X**AUGCC	20	6766.4	6765.5
2	ORN-30	UA**X**GGCACGCGGUGAAUGCC	20	6766.4	6765.4
3	ORN-31	UCACCGCGUGCCUUA	15	4687.9	4687.1
4	ORN-32	ACAGCGGACCUUGAU	15	4775.0	4774.0
5	ORN-33	AGUGUUAGCAUCGUGUGUCUUG	22	7013.2	7012.4
6	ORN-34	UAUUCACUCUGUGCCUUAA	19	5920.6	5919.8
7	ORN-35	CAAUGUAGUGGCCUUG	16	5098.1	5097.2
		miR-132-3p analogues			
8	ORN-36	UAAC**X**GUCUACAGCCAUGGUCG	22	7322.7	7321.7
9	ORN-37	GGCUGUAGACUGUUA	15	4792.9	4792.2
10	ORN-38	CUUUCGAUUGUUACU	15	4651.8	4651.1
11	ORN-39	GCCAUUGGCUGUGGACUGUU	20	4785.9	4785.1

Cross-linking was initially performed with a short (15-nt), fully complementary counter-strand (ORN-31, D18 and D23) as usual. However, we also used a partially complementary 'native' counter-strand corresponding to the passenger strand of miR-124 (ORN-32, D19 and D24) as a generic mimic for the natural mRNA targets of miR-124-3p (**Table 8**). Although for most of the tested duplexes cross-linking was detected (an illustrative example is presented in **Figure 18**), the yield of the cross-linking reaction was much lower compared to that of the miR-106a-5p probe.

As a precautionary measure, the individual strands of the duplexes (probes and counter-strands) were also checked for their stability to the irradiation conditions. While irradiation of miR-124-3p-2 left the probe unaffected, irradiation of miR-124-3p-1 probe gave rise to new peaks in the chromatogram of identical mass to the starting material. These most likely corresponded to intra-strand cross-linking products (**Figure 18**).

Table 8: Summary of the *in vitro* photo-cross-linking experiments performed with miR-124-3p probes with a fully complementary counter-strand (fc ctstr) (D18 and D23) or a native counter-strand (native ctstr) with a sequence corresponding to the passenger strand of miR-124 (D19 and D24). Underlined: mismatches and wobble base pairs. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 5 μM. **X**: adenosine with trioxsalen attached at the 2'-*O*-position of the ribose. Masses of the isolated peaks are presented in section VII.1.4.2.

Duplex	ORN	Sequence	Cross-linked product
		miR-124-3p-1 probe	
D18	ORN-29 (probe)	5'–UAAGGCACGCGGUG**X**AUGCC–3'	detected
	ORN-31 (fc ctstr)	3'–AUUCCGUGCGCCACU–5'	
D19	ORN-29 (probe)	5'–UAAGGCACGCGGUG**X**AUGCC–3'	not detected
	ORN-32 (native ctstr)	3'–UAGUUCCAGGCGACA–5'	
		miR-124-3p-2 probe	
D23	ORN-30 (probe)	5'–UA**X**GGCACGCGGUGAAUGCC–3'	detected
	ORN-31 (fc ctstr)	3'–AUUCCGUGCGCCACU–5'	
D24	ORN-30 (probe)	5'–UA**X**GGCACGCGGUGAAUGCC–3'	detected
	ORN-32 (native ctstr)	3'–UAGUUCCAGGCGACA–5'	
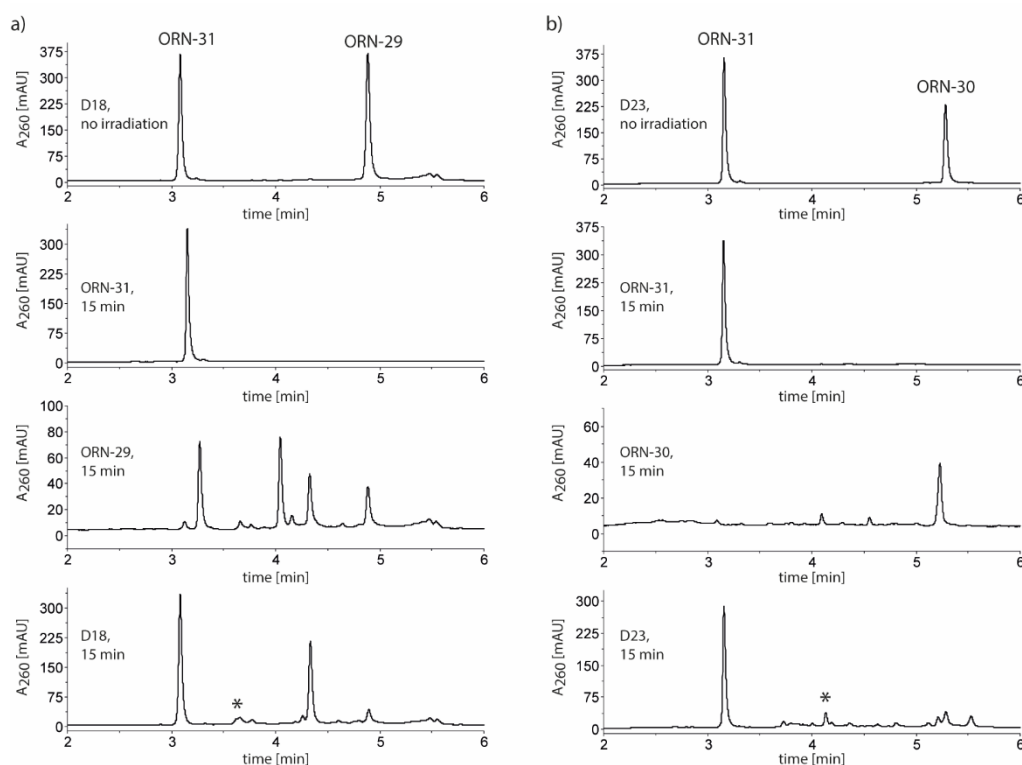
Figure 18: Illustrative RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with a) D18 and b) D23. Four panels represent respectively chromatograms for a) non-irradiated D18 (reference), ORN-31 single-strand, ORN-29 single-strand and D18 irradiated for 15 min (λ=365 nm); b) non-irradiated D23 (reference), ORN-31 single-strand, ORN-30 single-strand and D23 irradiated for 15 min (λ=365 nm). Peaks marked with an asterisk (*) correspond to the cross-linked duplex; a) mass calc. [M+K]: 11493.3, mass found: 11491.4 and b) mass calc.: 11452.3, mass found: 11452.8.

The predicted secondary structure of miR-124-3p (RNA folding performed with freely-available mfold software [288]) showed, that the location of the trioxsalen in the miR-124-3p-1 probe may indeed result in the intra-strand cross-linking (**Figure 19**).
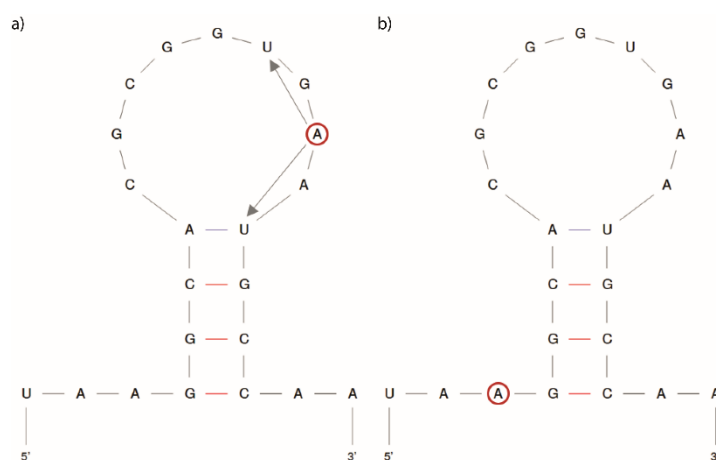
Figure 19: The predicted folding (Mfold [288]) of the trioxsalen-modified a) miR-124-3p-1 probe (ORN-29) and b) miR-124-3p-2 probe (ORN-30). Red circles mark the position of the trioxsalen within the sequences. Grey arrows point possible sites of the intra-strand cross-linking.

The fact that all of the tested counter-strands remained unaffected by irradiation confirmed that the cross-linking yield can be calculated based on the consumption of this strand.

To mimic the cross-linking with native target sites in mRNAs, a series of counter-strands (ORN-33 to ORN-35, **Table 7**) with sequences corresponding to the selected miR-124-3p targets identified by Dr Wang in miR-CLIP experiments was prepared. Target counter-strands corresponding to LRCH1 (Leucine-Rich Repeats And Calponin Homology Domain Containing 1) (ORN-33, **Table 7**) and MDC1 (Mediator Of DNA Damage Checkpoint 1) (ORN-34, **Table 7**) were chosen since they were highly enriched in the miR-CLIP experiments [287]. These partially complementary counter-strands were longer (22- and 19-nt-long for LRCH1 and MDC1, respectively) than the model 15-mer counter-strands to ensure the formation of the extended duplex structure for both of the probes.

Cross-linked products were detected for duplexes containing both of these native counter-strands (D20, D21, D25 and D26, **Table 9**, **Figure 20**). This finding was particularly striking in case of cross-linking with the miR-124-3p-1 probe (ORN-29), as it has trioxsalen in the 3'-part of the miRNA, which is often deprived of complementarity to the target mRNA. On the other hand, the cross-linking was not observed for the duplex formed with a counter-strand mimicking other mRNA target site, MINK1 (Misshapen-like kinase 1) (D22 and D27, **Table 9**).

Table 9: Summary of the *in vitro* photo-cross-linking experiments performed with miR-124-3p probes with counter-strands mimicking target sites in mRNA. Underlined: mismatches and wobble base pairs. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 5 µM. **X**: adenosine with trioxsalen attached at the 2'-*O*-position of the ribose. Masses of the isolated peaks are presented in section VII.1.4.2.

Duplex	ORN	Sequence	Cross-linked product
miR-124-3p-1 probe		
D20	ORN-29 (probe)	5'–U̲AAGGCACG̲CGGUG**X**AUGCC–3'	detected
	ORN-33 (LRCH1)	3'–G̲UUC̲UGUGUGCUACGAUUGUGA–5'	
D21	ORN-29 (probe)	5'–UAAGGCACG̲CGGUG**X**AU̲GCC–3'	detected
	ORN-34 (MDC1)	3'–AAUUCCGUGUC̲UCACUUAU̲–5'	
D22	ORN-29 (probe)	5'–UAAGGC–ACG̲CGGUG**X**AUGCC–3'	not detected
	ORN-35 (MINK)	3'–GUUCC̲GGUGA̲UGUAAC–5'	
miR-124-3p-2 probe		
D25	ORN-30 (probe)	5'–U̲A**X**GGCACG̲CGGUGAAUGCC–3'	detected
	ORN-33 (LRCH1)	3'–G̲UUC̲UGUGUGCUACGAUUGUGA–5'	
D26	ORN-30 (probe)	5'–UA**X**GGCACG̲CGGUGAAUGCC–3'	detected
	ORN-34 (MDC1)	3'–AAUUCCGUGUC̲UCACUUAU̲–5'	
D27	ORN-30 (probe)	5'–UA**X**GGC–ACG̲CGGUGAAUGCC–3'	not detected
	ORN-35 (MINK)	3'–G̲UUCC̲GGUGA̲UGUAAC–5'	
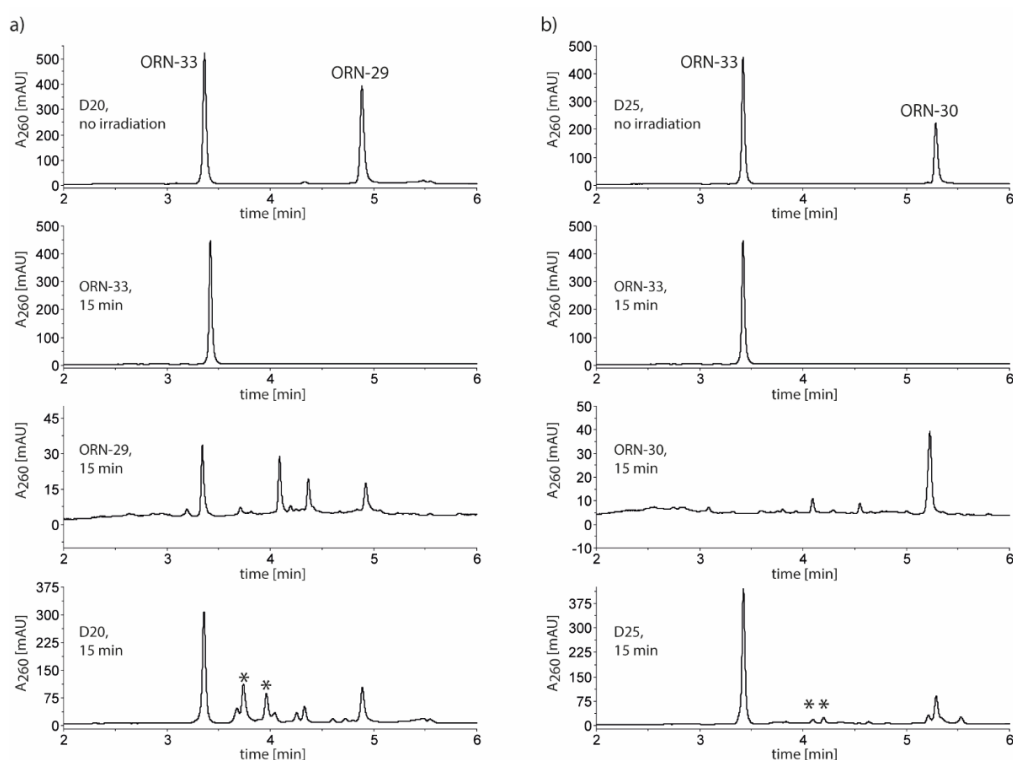
Figure 20: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with a) D20 and b) D25. Four panels represent respectively chromatograms for a) non-irradiated D20 (reference), ORN-33 single-strand, ORN-29 single-strand and D20 irradiated for 15 min (λ=365 nm); b) non-irradiated D25 (reference), ORN-33 single-strand, ORN-30 single-strand and D25 irradiated for 15 min (λ=365 nm). Peaks indicated with an asterisk (*) correspond to the cross-linked duplex; a) mass calc.: 13779.6, mass found: 13778.0; b) mass calc.: 13779.6, mass found: 13775.9 and 13799.6 ([M+Na]).

It should be emphasized that the conditions for *in vitro* cross-linking and intracellular cross-linking during miR-CLIP are very different. In particular, the amplification step in the miR-CLIP protocol may identify targets that are captured even with a poor cross-linking efficiency. This is not the case for the *in vitro* cross-linking assay, where the identification of the covalently linked product is possible only if the cross-linking yield exceeds a certain threshold defined by the technical capability of the analytics apparatus. Thus, the fact that the cross-linking product for the investigated probe was not detected *in vitro* does not mean that the same probe will not capture the targets in cells.

II.3.3.2. miR-132-3p

In contrast to the pre-miR-CLIP-124 probes, both pre-miR-CLIP-132 probes had trioxsalen attached within the seed region. A mono-labelled version (ORN-36) of one of them was prepared and tested in the cross-linking assays with a variety of counter-strands (**Table 10**).

Table 10: Summary of the *in vitro* photo-cross-linking experiments performed with miR-132-3p probe and a fully complementary counter-strand (fc ctstr) (D28), a native counter-strand (native ctstr) with a sequence corresponding to the passenger strand of miR-132 (D29) or a counter-strand with a sequence corresponding to the binding site in GNL3L (D30). Underlined: mismatches and wobble base pairs. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 1 µM. **X**: adenosine with trioxsalen attached at the 2'-*O*-position of the ribose. Masses of the isolated peaks are presented in section VII.1.4.3.

Duplex	ORN	Sequence	Cross-linked product
		miR-132-3p	
D28	ORN-36 (probe)	5'–UAAC**X**GUCUACAGCCAUGGUCG–3'	detected
	ORN-37 (fc ctstr)	3'–AUUGUCAGAUGUCGG–5'	
D29	ORN-36 (probe)	5'–UAAC**X**GUCUACAGCCAUGGUCG–3'	not detected
	ORN-38 (native ctstr)	3'–UCAUUGUUAGCUUUC–5'	
D30	ORN-36 (probe)	5'–UAAC**X**GUCUACAGCCAUGGUCG–3'	detected
	ORN-39 (GNL3L)	3'–UUGUCAGGUGUCGGU–5'	

Among performed experiments, cross-linking with the counter-strand mimicking the endogenous target of miR-132-3p, GNL3L (ORN-39), gave the most compelling results. GNL3L (guanine nucleotide-binding protein-like 3-like protein) was captured by the pre-miR-132 probe in miR-CLIP experiments performed in HEK293T cells [287]. A potential, highly complementary binding site for miR-132-3p was identified in its 3'-UTR and cross-linking to this site was investigated *in vitro*. Irradiation of the miR-132-3p probe with a GNL3L mimic as a counter-strand (D30) resulted in a clean and efficient cross-linking (**Figure 21a**). The time-course cross-linking experiment confirmed the formation of the cross-linked product after 5 min of irradiation and almost full conversion of the starting material after 30 min of irradiation (**Figure 21b**).
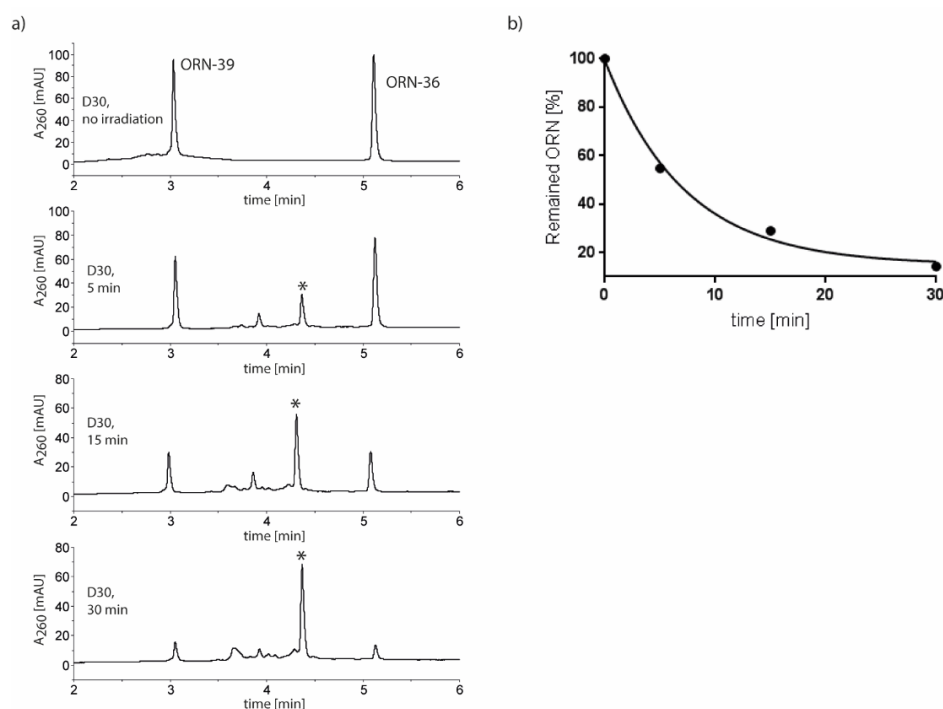
Figure 21: a) RP-HPLC chromatograms from the *in vitro* photo-cross-linking time-course experiment with D30; peaks marked with an asterisk (*) correspond to the cross-linked duplex (mass calc.: 12107.6, mass found: 12107.2); b) graph showing remaining ORN-39 [%] after 0, 5, 15 and 30 min of irradiation (λ=365 nm).

II.3.4. Miscellaneous miRNAs

Based on experiments with miR-106a-5p analogues, a 'shifted cross-linking' hypothesis emerged, in which trioxsalen cross-linking was assumed to occur in the target strand two or three base pairs upstream (towards the 5'-end of the miRNA) from the juxtaposed uridine. To further explore this assumption, probes for some other biologically relevant miRNAs were synthesized and tested (**Table 11**).

Table 11: Synthesized trioxsalen-labelled miRNA analogues and their counter-strands. Trioxsalen was attached post-synthetically (CuAAC reaction on the solid support) at the 2'-*O*-position of adenosine (**X**) or cytidine (**Y**).

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found
1	ORN-40	UGAGGUAGU**X**GUUUGUACAGUU	22	7382.7	7381.2
2	ORN-41	AACUGUACAAACUACUACCUCA	22	6913.3	6913.5
3	ORN-42	UAAGGCACGCGGAGA**X**UGCC	20	6789.4	6788.5
4	ORN-43	GGCAUUCUCCGCGUG	15	4743.9	4743.1
5	ORN-44	UAC**Y**CUGUAGAACCGAAUUUGUG	23	7629.8	7628.9
6	ORN-45	CGGUUCUACAGGGUA	15	4791.9	4791.2
7	ORN-46	UAG**Y**UUAUCAGACUGAUGUUGA	22	7325.7	7324.7
8	ORN-47	CAGUCUGAUAAGCUA	15	4759.9	4759.2

First, the concept of the cross-linking to a "-3" nucleotide in the counter-strand was examined. Trioxsalen was attached to the adenosines located in the central or 3'-region of let-7g-5p (ORN-40, **Figure 22a**) and miR-124-3p mutant (ORN-42, **Figure 22b**), respectively. In both cases, the fully complementary counter-strands used for the cross-linking had uridines juxtaposed and located three base pairs upstream from the modification site (**Figure 22a** and **Figure 22b**). Unfortunately, no cross-

linking was detected for these duplexes, strongly suggesting that the "rules" for miR-106a-5p could not be extrapolated to other sequences (**Table 12**). The fact that no formation of the covalently bound product was observed suggests that trioxsalen might have not been able to intercalate into such a compact duplex structure.
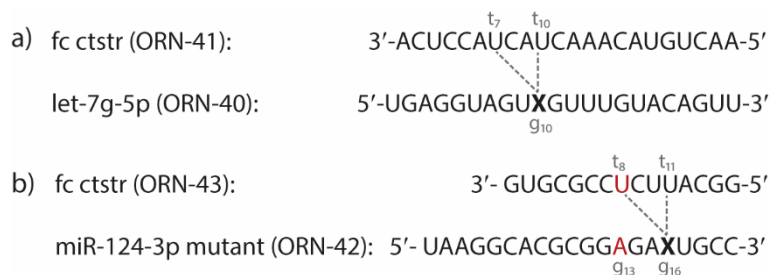
Figure 22: Sequences of the trioxsalen-modified a) let-7g-5p (ORN-40) and b) miR-124-3p mutant (ORN-42) and their fully complementary counter-strands (fc ctstr; ORN-41 and ORN-43, respectively) used in the *in vitro* photo-cross-linking study. In red: UA to AU mutation with regard to the original sequence of miR-124-3p. Putative cross-linking sites are indicated by dashed lines. **X**: adenosine with trioxsalen at the 2'-*O*-position of the ribose.

Table 12: Summary of the *in vitro* photo-cross-linking experiments performed with let-7g-5p ad miR-124-3p mutant analogues and their fully complementary counter-strands (fc ctstr). In red: UA to AU mutation with regard to the original miR-124-3p sequence. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 1 µM. **X** designates adenosine with the trioxsalen attached at the 2'-*O*-position of the ribose. Masses of the isolated peaks are presented in section VII.1.4.4.

Duplex	ORN	Sequence	Cross-linked product
let-7g-5p			
D31	ORN-40 (probe)	5'-UGAGGUAGU**X**GUUUGUACAGUU-3'	not detected
	ORN-41 (fc ctstr)	3'-ACUCCAUCAUCAAACAUGUCAA-5'	
miR-124-3p mutant			
D32	ORN-42 (probe)	5'-UAAGGCACGCGG**A**GA**X**UGCC-3'	not detected
	ORN-43 (fc ctstr)	3'-GUGCGCC**U**CUUACGG-5'	

Cross-linking to a '-2' base pair target site was then investigated. To evaluate this concept, probes for two other microRNAs, miR-10b-5p and miR-21-5p, were designed (**Table 11**). Trioxsalen was appended to the base situated two base pairs downstream (towards the 3'-end of the miRNA) from the adenosine defining the possible reactive uridine (t_2). Therefore, for both miRNAs, the trioxsalen-bearing base was cytidine (g_4) (**Figure 23a**, **Figure 23b** and **Table 13**).
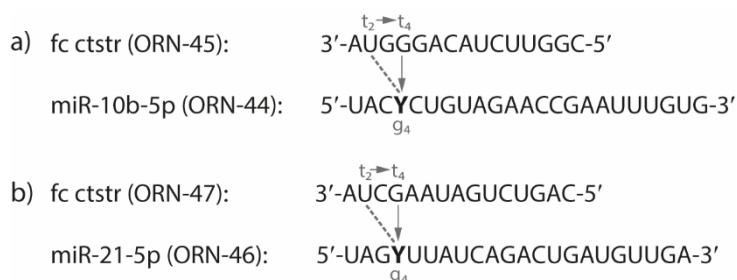
Figure 23: Sequences of the trioxsalen-modified a) miR-10b-5p (ORN-44) and b) miR-21-5p (ORN-46) and their fully complementary counter-strands (fc ctstr; ORN-45 and ORN-47, respectively) used in the *in vitro* photo-cross-linking study. The putative cross-linking sites are indicated by dashed lines. **Y**: cytidine with trioxsalen attached at the 2'-*O*-position of the ribose.

An extended cross-linking experiment was performed for these miRNAs. The creation of the covalently linked product was observed for D33 as well as for D34 (**Figure 24a** and **Figure 24b**, respectively).

Table 13: Summary of the *in vitro* photo-cross-linking experiments performed with miR-10b-5p ad miR-21-5p analogues. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 5 µM. **Y** designates cytidine with the trioxsalen attached at the 2'-*O*-position of the ribose. Fc ctstr: fully complementary counter-strand. Masses of the isolated peaks are presented in section VII.1.4.4.

Duplex	ORN	Sequence	Cross-linked product
		miR-10b-5p	
D33	ORN-44 (probe)	5'-UAC**Y**CUGUAGAACCGAAUUUGUG-3'	detected
	ORN-45 (fc ctstr)	3'-AUGGGACAUCUUGGC-5'	
		miR-21-5p	
D34	ORN-46 (probe)	5'-UAG**Y**UUAUCAGACUGAUGUUGA-3'	detected
	ORN-47 (fc ctstr)	3'-AUCGAAUAGUCUGAC-5'	
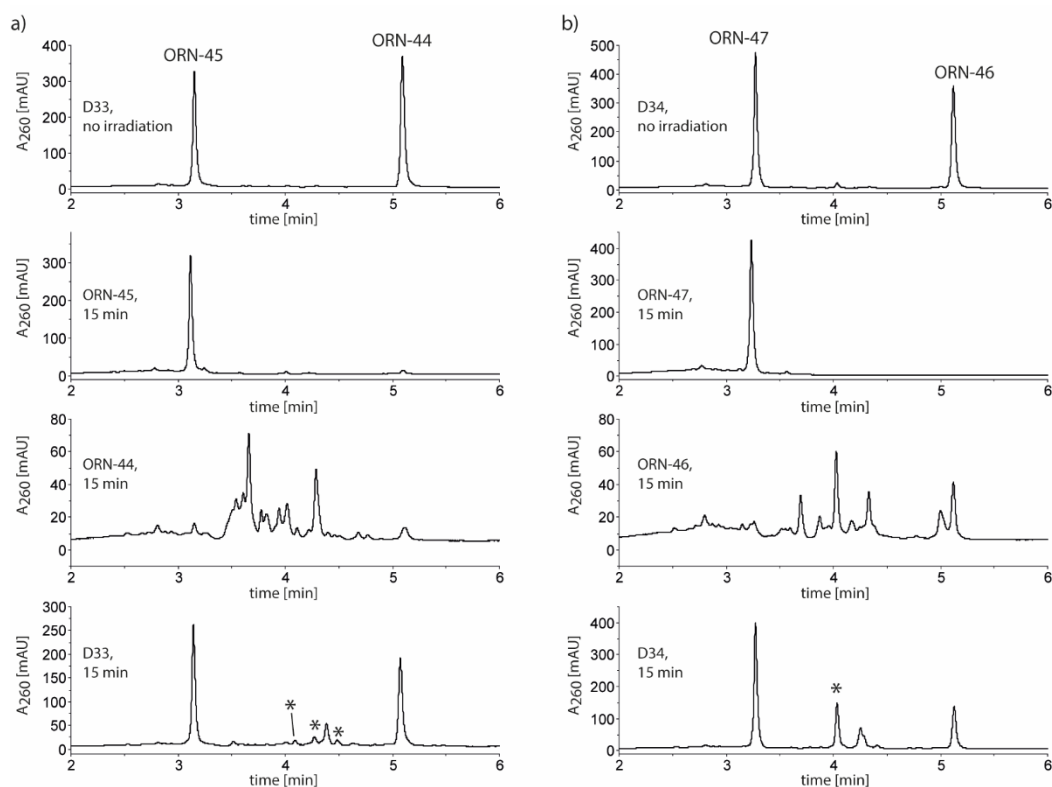
Figure 24: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with a) miR-10b-5p (D33) and b) miR-21-5p (D34) analogues having the potential cross-linking site (uridine) two base pairs upstream from the position at which the trioxsalen was incorporated. Four panels represent respectively chromatograms for a) non-irradiated D33 (reference), ORN-45 single-strand, ORN-44 single-strand and D33 irradiated for 15 min (λ=365 nm); b) non-irradiated D34 (reference), ORN-47 single-strand, ORN-46 single-strand and D34 irradiated for 15 min (λ=365 nm). Peaks marked with an asterisk (*) correspond to the cross-linked duplex; a) mass calc. [M+K]: 12460.7, mass found: 12458.7; b) mass calc.: 12085.6, mass found: 12084.1.

Similarly to miR-124-3p-1 probe, irradiation of the single-stranded miR-10b-5p and miR-21-5p probes resulted in the appearance of new faster eluting peaks in the chromatograms (ORN-44 and ORN-46, 15 min panel in **Figure 24a** and **Figure 24b**). Also in these cases, folding with the mfold predicted various possible secondary structures for both of these miRNAs (data not shown), supporting the concept of the intra-strand cross-linking. Interestingly, the mass of the cross-linked dimer was also detected for miR-21-5p (the single-strand probe irradiation). Indeed, a prediction of self-complementarity run with the freely-available software tool Oligo Calc forecasted a formation of stable homo-dimers, especially for miR-21-5p (**Figure 25**).

```
3'-AGUUGUAGUCAGACUAUUYGAU-5'
5'-UAGYUUAUCAGACUGAUGUUGA-3'
```

Figure 25: A miR-21-5p homo-dimer structure predicted with the Oligo Calc. The self-complementary region is marked in blue. **Y**: cytidine modified with trioxsalen.

As an alternative to RP-HPLC analysis of samples, a denaturing polyacrylamide gel electrophoresis assay was also tested as a potential method for the detection of RNA-RNA cross-linking (**Figure 26**). The cross-linking assay was performed according to the standard procedure described before. Four samples were prepared for each miRNA: a single-strand probe (**Figure 26**, lanes 1 and 6, ORN-46 and ORN-44, respectively), a single-strand counter-strands (**Figure 26**, lanes 4 and 9, ORN-47 and ORN-45, respectively) and non-irradiated and irradiated duplexes (**Figure 26**, lanes 2-3, D34 and lanes 7-8, D33). Samples were analysed after irradiation through denaturing polyacrylamide gel electrophoresis. The gels showed the presence of the new bands with lower mobility. As the cross-linked products were detected for these duplexes by the RP-HPLC/LC-MS analysis, therefore the new bands (indicated with arrows) most likely corresponded to the cross-linked product. This demonstrated that the denaturing gel electrophoresis technique can be applied to support the analysis of the cross-linking experiment. However, if this method were to be employed as the sole method of analysis, it would require further evaluation of the identity of the new bands. The addition of a control sample with an irradiated single-strand probe will be necessary to verify the migration of the band corresponding to the possible product of the intra-strand cross-linking.
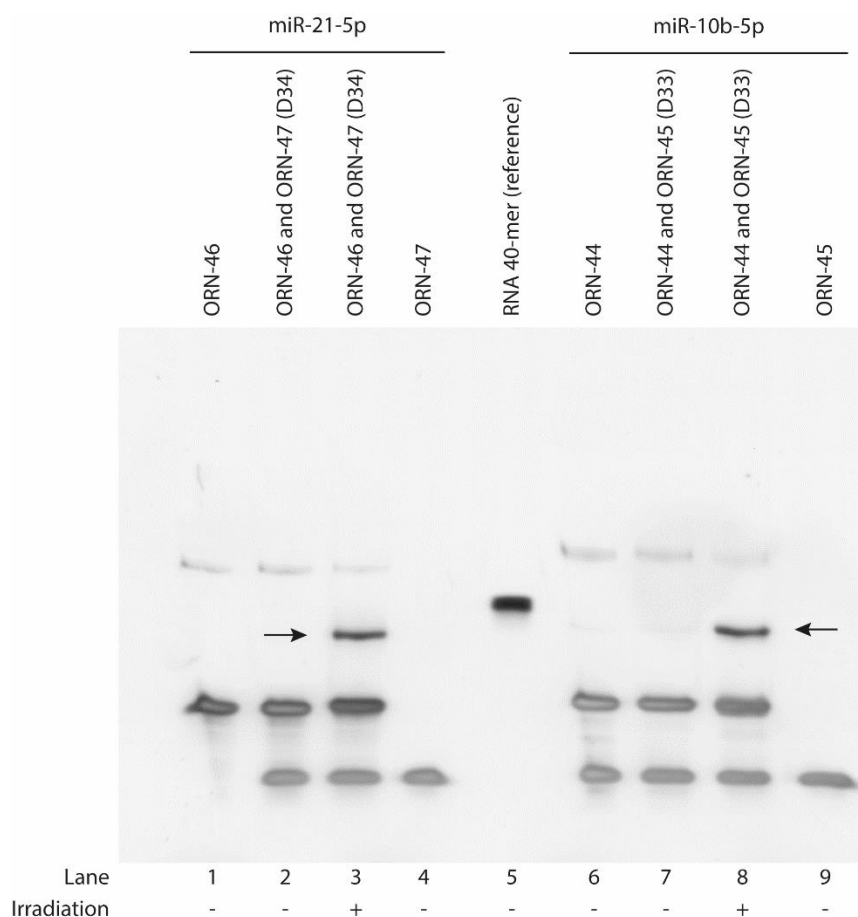
Figure 26: 20% denaturing polyacrylamide gel electrophoresis used to analyse the results from the *in vitro* photo-cross-linking experiment with D33 and D34. The '-' sign under the lane number refers to the fact that the sample was not subjected to the irradiation. The '+' sign designates that the sample was irradiated for 15 min (λ=365 nm). RNA 40-mer was used as a ladder (lane 5). The gel was visualized with SYBR® Gold. N = 1. The experiment was run together with Artur Łaski.

II.4. Conclusions and outlook

The cross-linking experiments described in this chapter were performed with the goal of a better understanding of the nature of the covalent bond formation between the trioxsalen-bearing miRNA analogues and their counter-strands in terms of sequence and position dependency. In all cases, the same type of chemistry, CuAAC reaction, was used to attach the cross-linker at the 2'-*O*-position of the ribose, following the strategy implemented in the original miR-CLIP publication [84]. This investigation was necessary to verify if the same probe design can be used to achieve the cross-linking with other miRNAs, having less adenosine-rich sequences.

The results from the tests with mutated miR-106a-5p analogues provided an overview of the cross-linking process and helped to develop the hypothesis about the possible cross-linking site in the target strand. Experiments showed that the cross-linking does not detectably occur with the uridine that is base-paired with the trioxsalen-modified adenosine. Instead, the cross-linking site is likely shifted two or three base pairs aside from the envisioned (juxtaposed) site of interaction. The cross-linking distance appears to be shortened with an increase of the compactness of the duplex.

Tests with miR-124-3p and miR-132-3p validated the ability of the probes to cross-link to the non-fully complementary, native-like targets. Furthermore, as demonstrated by the miR-124-3p-1 probe, cross-

linking could have been also achieved when the trioxsalen was placed outside of the seed region. The presence of bulges and mismatches can facilitate the accommodation of the trioxsalen within the duplex, as it was shown by the tests with free trioxsalen [253]. Nevertheless, the presence of potential reactive sites (i.e. uridines or cytidines) accessible by the trioxsalen with our current linker is difficult to predict outside of the seed region.

Finally, the small set of experiments with probes for various miRNAs allowed to further fine-tune the working hypothesis about the cross-linking. Tests with let-7g-5p and miR-124-3p mutant confirmed that interaction with uridine placed as far as three base pairs away from the attachment point of the cross-linker was rather exceptional for the miR-106a-5p sequence. The most likely general scenario which emerged is that the trioxsalen is reacting with the uridine placed two ('-2') base pairs upstream from the base juxtaposed to the trioxsalen-bearing nucleotide. Nevertheless, whether the cross-linking will happen appears to be dependent on the local structure and the region of the duplex.

Conducted experiments allowed to gain an insight into the cross-linking process with the investigated probe and highlighted its complexity. Based on the obtained results, an alternative approach for the probe design is proposed. When preparing the probe for a miRNA of interest, it would be suggested to first search for adenosines within the seed region. This would imply the presence of the uridine in the target sequence and the recommendation will be to attach the trioxsalen at the 2'-*O*-position of the guide nucleotide placed two base pairs downstream from the identified adenosine, as it was done for miR-10b-5p and miR-21-5p. This strategy can be also applied outside of the seed region, though the outcome may be compromised by the lack of complementarity.

It should be mentioned that these observations and conclusions should be considered as general guidelines rather than strict patterns. The fact that the cross-linking product with an expected mass is detected does not provide a piece of information about the exact cross-linking site and can be misleading. It is possible that in some cases the covalent bond was generated with cytidine and not with uridine. To get more concrete information on the behaviour of the probes, the precise identification of the cross-linking site in the target strand would be necessary.

II.5. Contributions

2'-*O*-propargyl phosphoramidites and 4'-azidomethyltrioxsalen were taken from the common laboratory stock synthesized by Martina Bigatti as a part of the NCCR platform. The first batch of the mono-labelled miR-124-3p and miR-132-3p probes was prepared by Dr Ugo Pradère and then re-synthesized by Anna L. Malinowska. The counter-strands mimicking the endogenous target sites for miR-124-3p and miR-132-3p were designed by Anna L. Malinowska based on the results from the miR-CLIP experiment with the pre-miR-124- and pre-miR-132-miR-CLIP probes conducted and shared by Dr Yuluan Wang. Synthesis of all the other oligoribonucleotides and cross-linking experiments were designed, performed and analysed by Anna L. Malinowska. The denaturing polyacrylamide gel electrophoresis was run by Anna L. Malinowska and Artur Łaski according to the protocol provided by Artur Łaski.

III. Chapter III: Cross-linking properties of miRNA mimics modified with a diazirine moiety

III.1. Introduction

As briefly mentioned in Chapter I, not only psoralens have been used to achieve an RNA-RNA cross-linking. The cross-linkers utilized to create a covalent bond between nucleic acids have been also adapted from the field of RNA-protein cross-linking. The photo-affinity labelling (PAL) technique is widely used to study protein-protein, protein-ligand, protein-nucleic acid and many other interactions taking place in cells [289]. In the PAL approach, a ligand modified with a photo-reactive group (and usually also a detection tag) is employed. Upon the irradiation with an appropriate wavelength, the covalent bond is generated between the ligand and its macromolecular binding partner present in the proximity. Benzophenones, aryl azides and diazirines are one of the most popular types of photophores employed in PAL (**Figure 27**). All of them work through slightly different mechanisms and their application is associated with certain limitations and benefits [289-291].
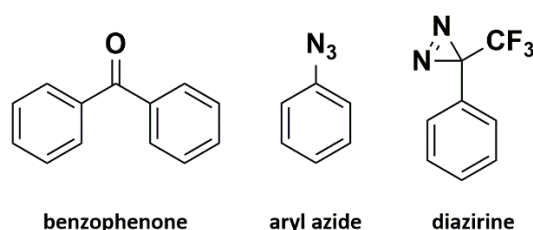
Figure 27: Examples of the most popular photo-reactive groups used in the PAL approach. Adapted from [290].

The commercially available benzophenones are activated with a long-wave UV light (λ>350 nm) but generally require long irradiation times in order to react. This promotes undesired non-specific labelling [290, 292]. Additionally, the need to accommodate a large moiety can interfere with the natural activity of the investigated system. In contrast to benzophenones, aryl azides have increased light-sensitivity and are excited by irradiation at shorter wavelengths (λ<300 nm), although irradiation at this wavelength can be detrimental for cellular components [293, 294], limiting the value of these compounds. On the positive side, the synthesis of the aryl azides is straightforward.

Diazirines constitute a class of compounds characterized by the presence of a three-member ring, in which two nitrogen atoms, connected by a double-bond, are bound to a single carbon atom. Diazirines are one of the smallest functional groups with photolabile properties. The irradiation of diazirines with a long-wavelength light (λ>320 nm) triggers their photolysis and generation of highly reactive carbenes [289]. The divalent carbon atom of the carbene is then able to insert into single bonds in proximity, including C-H, N-H and O-H bonds, forming stable adducts [289].

Photolysis of diazirines can also induce their isomerization with concomitant formation of less reactive and long-living linear diazo side products, thereby lowering the efficiency of the photo-labelling (**Figure 28**) [295-298]. These side products can be converted to the carbene by irradiation at shorter wavelength (λ=305 nm) or by prolonged irradiation at longer (λ=350 nm) wavelength [296, 299]. Attachment of the trifluoromethyl group to the carbon within the diazirine ring stabilizes the diazo structure and renders it much less reactive. Additionally, incorporation of such an electron-

withdrawing group contributes to the stability of the generated carbenes and prevents them from undergoing rearrangements [296]. Previous studies showed that during the photolysis process, 35% of the 3-trifluoromethyl-3-phenyldiazirine is converted to its linear diazoisomer [296].
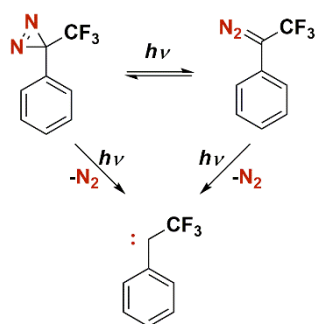
Figure 28: Mechanism of the carbene formation upon irradiation of the diazirine with the λ=365 nm light (adapted from: [297]).

Carbenes can exist in a singlet or a triplet form, which differ in reactivity and photochemical properties [300]. Which of the two carbene states will be generated is influenced by substituents on the diazirine ring. Electron-donating groups, such as phenyl, favour the formation of the singlet state by stabilization of its structure [301, 302]. In the context of the PAL, the generation of singlet states is usually preferred, as they reportedly react faster with a broad range of functional groups.

Diazirines can indiscriminately react with any molecules in their vicinity, including solvent molecules [296, 303-305]. As a consequence, the yields of photo-labelling reactions with diazirines are usually low, especially when aliphatic diazirines are used (yields for the aliphatic diazirines usually do not exceed 10%) [290]. Despite their demanding chemical synthesis, their chemical stability and high reactivity make 3-trifluoromethyl-3-phenyldiazirine one of the most widely used cross-linkers for PAL [295].

Diazirines have been incorporated at various positions (nucleobase, ribose, phosphate backbone) in oligonucleotide sequences for the investigation of DNA-protein [292, 306-308], RNA-protein [309] and DNA-DNA [310] interactions. In 2014, the Ueno group reported the synthesis of two photo-reactive nucleoside analogues, in which the nucleobase was replaced by an aromatic group substituted with a trifluoromethyldiazirine at either meta (Analogue A) or para (Analogue B) positions (**Figure 29**) [278]. The compounds were incorporated into the let-7 miRNA sequence and thermal stabilities and cross-linking abilities of the new probes were measured.
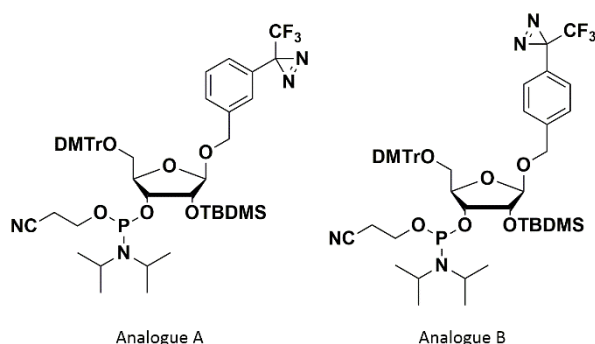
Figure 29: Two photo-reactive nucleoside analogues bearing an aryl trifluoromethyldiazirine reported by Nakamoto et al. [278].

Melting studies of the duplexes showed that the introduction of the two nucleoside analogues into the oligoribonucleotides resulted in a measurable destabilization of the duplexes. The effect was sequence-dependent and could be explained by the loss of the base-pairing at the position at which the analogue was introduced; generally, incorporation of Analogue B caused a smaller destabilization. Subsequently, probes containing Analogue A or Analogue B were tested in the *in vitro* cross-linking assay with 5'-fluorescein-labelled counter-strands. Duplexes were irradiated for 30 min with ultraviolet A light (λ=365 nm), followed by 10 min irradiation with ultraviolet B light (λ=302 nm). Samples were then analysed with a 20% polyacrylamide gel electrophoresis performed under denaturing conditions. Probes with both analogues were able to cross-link to their targets having different bases at the juxtaposed position, as new bands with lower mobility were detected for all of the samples. The cross-linking yields (calculated based on the intensity of the bands) varied from 6 to 32%, depending on the sequence and analogue (cross-linking with the Analogue B was more efficient). Furthermore, miR-199 mimics containing Analogue B in either 5p or 3p strand were shown to be active in a dual-luciferase reporter assay, used to assess the ability of the miRNA to repress gene expression (the principle of the luciferase assay will be described more in detail in section V.2.1).

As Analogue B displayed superior properties, it was used in further studies in cells [311]. The photo-reactive monomer was introduced at every position within the miR-145 sequence in a stepwise manner, to assess the tolerability for the modification at each location. Probes were additionally functionalized with a biotin at the 3'-end to facilitate the isolation and purification. Western blot analysis showed that incorporation of the Analogue B within the seed region reduced the miRNA activity, while its positioning outside of the seed region had a lower impact on miRNA performance. The bis-labelled probe with Analogue B at the position nine (g_9) and biotin at the 3'-end of the miRNA was subsequently tested in DLD-1 cells, where it was shown to bind to FSCN1 (Fascin Actin-Bundling Protein 1) and KLF4 (Kruppel-like factor 4) mRNAs, two reported targets of miR-145.

In the most recent report from the same group, the aryl trifluoromethyldiazirine scaffold was additionally modified with an ethynyl moiety [312]. The presence of the ethynyl group enables further functionalization of the probe which can be achieved with CLICK reaction performed after the cross-linking step. This would ensure that the biotin moiety does not interfere with RISC formation.

III.2. Aim of the project

The goal of this project was to develop a simplified diazirine-based probe for miRNA-mRNA cross-linking. A sequence-independent photo-reactivity of the aforementioned Analogue B appeared promising in terms of its application for the formation of the covalent bond between the miRNA and its target. However, this approach had some significant drawbacks. Preparation of the presented photo-reactive phosphoramidite required a multistep synthesis with some of the steps (especially on the nucleoside level) giving low yields. Moreover, the incorporation of this photophore as a base surrogate decreased the thermal stability of the duplex, as the aryl trifluoromethyldiazirine cannot base-pair with the opposite base in the target. Therefore, it was envisioned to incorporate the aryl trifluoromethyldiazirine into the miRNA sequence using a less demanding synthetic pathway, without destroying base-pairing. This was accomplished by its incorporation at the 2'-*O*-position of the ribose or the C^5-position of the pyrimidine within the miRNA sequence with the use of the CuAAC chemistry. The ability of such miRNA probes to cross-link to their counter-strands was then tested.

III.3. Results and discussion

III.3.1. Synthesis of the 3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3H-diazirine

The synthesis of the 3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3H-diazirine (**8**, in this chapter abbreviated as azidodiazirine) was performed following the published protocols with minor changes (**Scheme 1**) [278, 313].
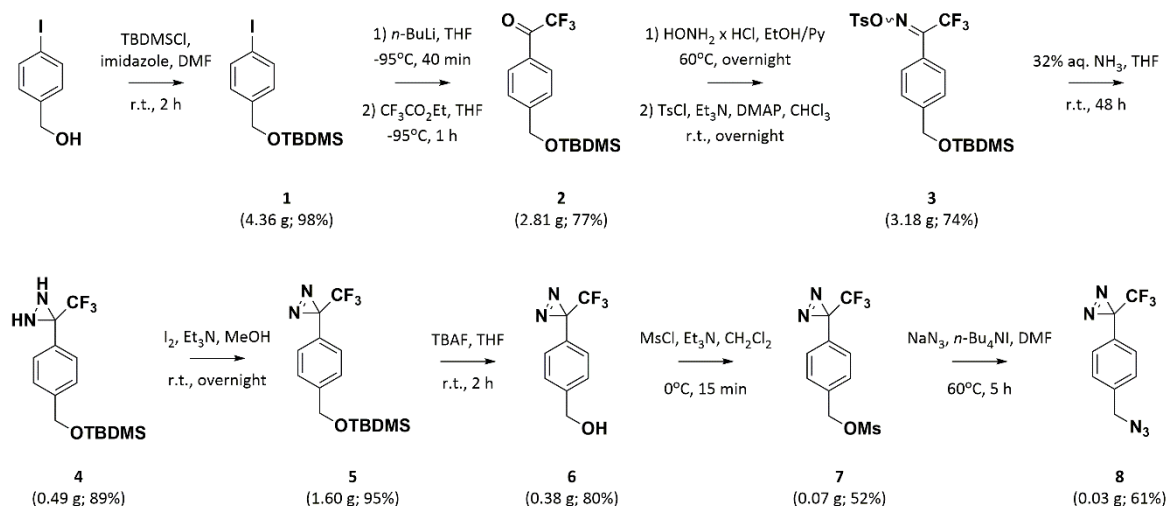
Scheme 1: Schematic representation of the synthesis of the 3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3H-diazirine (**8**) [278, 313].

First, 4-iodobenzyl alcohol was quantitatively protected with *tert*-butyldimethylsilyl chloride (TBDMSCl). The product (**1**) was converted to the trifluoroacetophenone (**2**) by the consecutive treatment with *n*-BuLi and ethyl trifluoroacetate. In the initial experiments run at -78°C, a substantial formation of by-products (20-40%) was observed. Lowering the reaction temperature below -90°C and quenching of the reaction mixture at the same temperature significantly reduced the generation of the undesired side products. Next, the conversion of the trifluoroacetophenone (**2**) to the tosyl-oxime (**3**) was performed. Compound **2** was reacted with hydroxylamine hydrochloride at 60°C overnight. The produced stereoisomeric (*E/Z*) mixture of oximes was subsequently tosylated to give the desired compound **3**. The product of this two-step procedure was contaminated with an excess of the tosylic acid and required multiple purification steps which decreased the yield reaction. Additionally, it had been noticed that some of the product was lost during the aqueous workup after the first step (formation of the oxime). Omitting the first work-up slightly improved the yield, though this was connected with extensive salt precipitation (most likely pyridinium and triethylammonium chlorides) while running the tosylation reaction. The formation of a diaziridine (**4**) was achieved by the treatment of compound **3** with aqueous ammonia. The reaction was performed in a sealed 20 mL microwave reaction vial. In cases when the scale of the reaction exceeded the volume of the vial, the starting material was dissolved in the solvent and distributed equally into several vials to which ammonia was subsequently added. The diaziridine (**4**) was oxidized to diazirine (**5**) with I$_2$, which upon treatment with tetra-*n*-butylammonium fluoride (TBAF) gave the deprotected diazirine (**6**). To introduce azide functionality, the hydroxyl group of diazirine (**6**) was mesylated and then treated with sodium azide to generate the slightly impure final product (**8**) in moderate yields.

III.3.2. Cross-linking tests with miR-106a-5p analogues

The synthesized azidodiazirine **8** was then incorporated into the miR-106a-5p sequence (**Table 14**). This particular sequence was chosen as it exhibited the most efficient cross-linking with the trioxsalen modification. Therefore, it could be used as a reference to assess the cross-linking abilities of the new diazirine-modified analogues.

Table 14: Synthesized diazirine-labelled miR-106a-5p analogues and counter-strands. Diazirine was attached post-synthetically (CuAAC reaction on the solid support) at the 2'-O-position of adenosine (**Y**), uridine (**Z**) or at the C^5-position of uridine (**W**).

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found
		miR-106a-5p analogues			
1	ORN-48	AAA**Y**GUGCUUACAGUGCAGGUAG	23	7713.8	7713.0
2	ORN-49	AAAAG**Z**GCUUACAGUGCAGGUAG	23	7713.8	7713.1
3	ORN-50	AAAAG**W**GCUUACAGUGCAGGUAG	23	7699.8	7698.9
4	ORN-51	GUAAGCACUUCUUAC	15	4696.9	4696.2
5	ORN-52	AAAAGUGCUUACAGUGCAGGUAG	23	7434.6	7433.6

The probes were designed as depicted below (**Figure 30**). The azidodiazirine (**8**) was post-synthetically CLICKed at three different locations within the miR-106a-5p sequence. The modification was introduced either at the 2'-O-position of adenosine g_4, to compare its behaviour with the original trioxsalen-modified probe, or at the 2'-O- or C^5-position of the uridine g_9, to directly compare the effect of placement of the modification in the minor and major groove of the duplex, respectively (ORN-48 to ORN-50, **Table 14** and **Figure 30**).
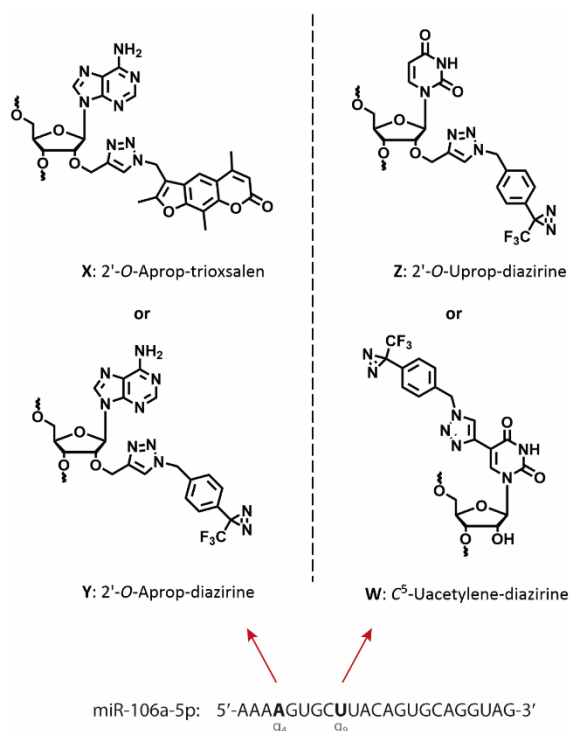
Figure 30: Design of the cross-linking study with diazirine-modified miR-106a-5p probes. Adenosine (**A**) was replaced by 2'-O-propargyl adenosine which was post-synthetically modified with the trioxsalen (**X**) or diazirine (**Y**). Uridine (**U**) was replaced by 2'-O-propargyl uridine (**Z**) or C^5-acetylene uridine (**W**) which were post-synthetically CLICKed with a diazirine.

Obtained compounds were then tested in the *in vitro* photo-cross-linking assay. Fully complementary (ORN-2) or 'native' (corresponding to the passenger strand of miR-106a, ORN-51) oligoribonucleotides were used as counter-strands to form duplexes (D35-D37 and D39-D41, **Table 15**). Duplex D1 bearing the trioxsalen modification at g_4 served as a positive control. Additionally, the cross-linking properties of the ORN-1 with a native counter-strand (D38) were tested. During the synthesis and preparation of the photo-cross-linking assays, the diazirine-bearing oligoribonucleotides were protected from ambient light to prevent a possible decomposition of the diazirine moiety.

Table 15: Summary of the *in vitro* photo-cross-linking experiments performed with miR-106a-5p probes and a fully complementary (fc ctstr) or a native counter-strand (native ctstr) with a sequence corresponding to the passenger strand of miR-106a. Samples were irradiated for 5, 15 or 30 min with the UV light λ=365 nm (non-irradiated samples served as reference). The concentration of each strand: 1 µM. Adenosine was modified with the trioxsalen (**X**) or diazirine (**Y**) attached at the 2'-*O*-position of the ribose; uridine was modified with a diazirine attached at the 2'-*O*-position of the ribose (**Z**) or the C^5-position of the nucleobase (**W**). Masses of the isolated peaks are presented in section VII.2.5.

Duplex	ORN	Sequence	Cross-linked product
miR-106a-5p analogues			
D1	ORN-1 (probe)	5'–AAA**X**GUGCUUACAGUGCAGGUAG–3'	detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D35	ORN-48 (probe)	5'–AAA**Y**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D36	ORN-49 (probe)	5'–AAAAG**Z**GCUUACAGUGCAGGUAG–3'	not detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D37	ORN-50 (probe)	5'–AAAAG**W**GCUUACAGUGCAGGUAG–3'	not detected
	ORN-2 (fc ctstr)	3'–UUUUCACGAAUGUCA–5'	
D38	ORN-1 (probe)	5'–AA–A**X**GUGCUUACAGUGCAGGUAG–3'	detected
	ORN-51 (native ctstr)	3'–CAUUCUUCACGAAUG–5'	
D39	ORN-48 (probe)	5'–AA–A**Y**GUGCUUACAGUGCAGGUAG–3'	not detected
	ORN-51 (native ctstr)	3'–CAUUCUUCACGAAUG–5'	
D40	ORN-49 (probe)	5'–AA–AAG**Z**GCUUACAGUGCAGGUAG–3'	not detected
	ORN-51 (native ctstr)	3'–CAUUCUUCACGAAUG–5'	
D41	ORN-50 (probe)	5'–AA–AAG**W**GCUUACAGUGCAGGUAG–3'	not detected
	ORN-51 (native ctstr)	3'–CAUUCUUCACGAAUG–5'	

The cross-linking experiments were performed according to the standard procedure. Calculated volumes of the two strands were mixed, dried and re-dissolved in 200 µL of the phosphate buffer so that the final concentration of each strand was 1 µM. After annealing, samples were subjected to irradiation for 5, 15 or 30 min (diazirine-containing probes). The 30-minute irradiation time point was adapted from the cross-linking experiment with a diazirine-modified miRNA reported by the Ueno group [278]. As in previous experiments described in section II.3, annealed but non-irradiated samples were used as references. Apart from duplex D1 serving as a positive control, the cross-linking between trioxsalen-modified miR-106a-5p (ORN-1) and a native counter-strand (ORN-51) was observed (**Figure 31**).
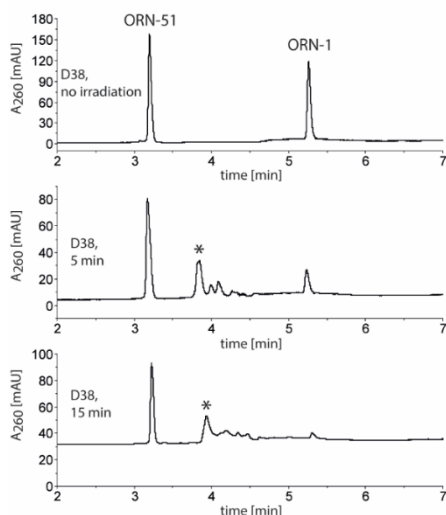
Figure 31: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment performed with duplex D38. Samples were irradiated for 5 or 15 min (λ=365 nm). The top panel represents a non-irradiated sample (reference). Peaks indicated with an asterisk (*) correspond to the cross-linked duplex (mass calc. [M+K]: 12491.9, mass found: 12490.1 and 12491.0).

On the other hand, the cross-linking was not detected for any of the probes modified with a diazirine (**Table 15**). **Figure 32** represents chromatograms from the *in vitro* photo-cross-linking assay performed for duplexes D36 and D37, in which the diazirine was clicked to the g_9 uridine in the minor or the major groove, respectively.
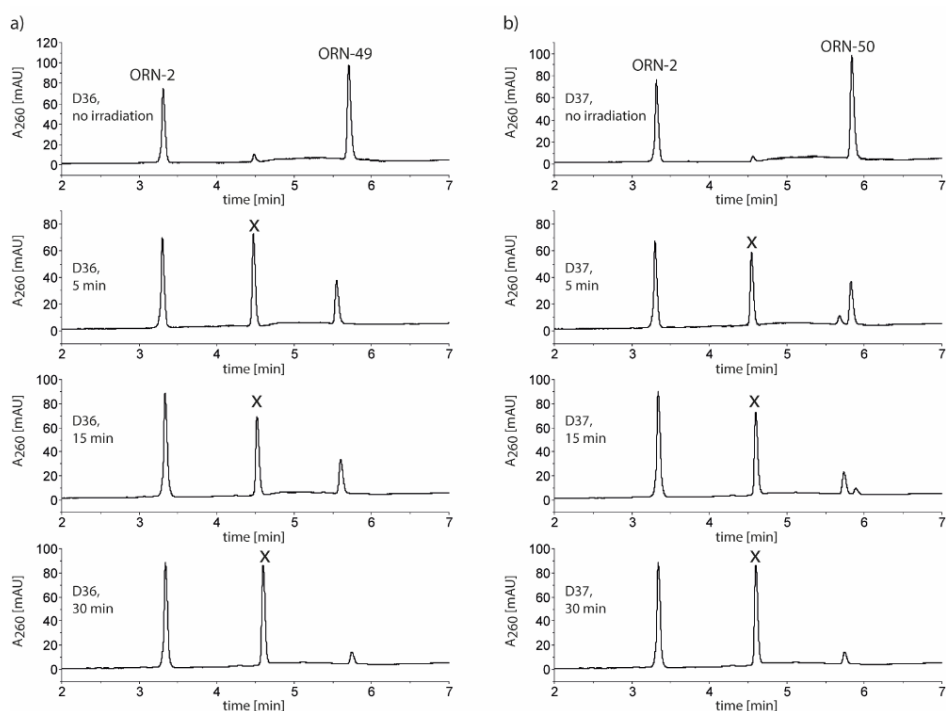
Figure 32: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment for duplexes a) D36 and b) D37. Samples were irradiated for 5, 15 or 30 min (λ=365 nm). Peaks marked with an 'x' correspond to the product of the insertion into water O-H bond; a) mass calc.: 7703.8, mass found: 7703.1; b) mass calc.: 7689.8, mass found: 7689.1.

LC-MS analysis of the newly-formed peaks suggested that for all of the tested probes, the generated carbene reacted with water, as masses corresponding to the [M-N$_2$+H$_2$O] products were detected (peaks marked with an 'x', **Figure 32**). This was not unexpected and can be explained by the 'extrahelical' location of the diazirine when attached at the 2'-*O*-position of the ribose or *C*^{5}-position of the nucleobase. Lack of the inter-strand DNA cross-linking for probes, in which the diazirine moiety was placed in the minor and major groove, has been reported previously [308]. Diazirines do not exhibit intrinsic intercalation abilities, thus there is no driving force that would direct them towards the duplex structure. As carbenes generated during the irradiation are very reactive and not selective, they could easily react with water (or any other solvent) molecules when positioned outside of the duplex structure [303, 304].

The presence of potassium cations associated with the duplex was often observed. Interestingly, as soon as the sample was subjected to the irradiation, the mass corresponding to the mass of the probe was not detected anymore (even for the peaks with a retention time corresponding to the parent compound), highlighting the reactivity of diazirines. The fact that the mass of the probe was detected in the reference sample confirmed that the annealing step did not compromise the diazirine structure, which was in agreement with the previous reports about the stability of the 3-trifluoromethyl-3-phenyldiazirine [296].

A similar set of cross-linking experiments was performed with the diazirine-modified miR-132-3p analogues. The diazirine was introduced at the 2'-*O*-position of adenosine g$_3$ and the 2'-*O*- or *C*^{5}-position of the uridine g$_7$ (data not shown). Samples were irradiated for 5 minutes in the phosphate (pH 7.0) or sodium chloride-based buffer (pH 7.2) used by Nakamoto et al. [278]. As it was the case for miR-106a-5p probes with diazirine, no cross-linking was observed also for miR-132-3p sequences. This further confirmed that extrahelical attachment of this diazirine derivative is not suitable for the RNA-RNA cross-linking.

III.3.3. Thermal stability studies

The thermal stabilities of the trioxsalen- and diazirine-modified duplexes were investigated by analysing their UV melting profiles. The melting temperature (T_M) is a temperature at which 50% of the duplex is hybridized. Denaturation of the duplex results in an increased absorbance of the UV light, a phenomenon called the hyperchromic shift. T_M provides information about the stability of the duplex and its affinity for the hybridization. The UV absorption profiles measured at 260 nm (the wavelength of maximum absorption of the UV light by RNA and DNA) are presented in **Graph 1**. A duplex containing an unmodified miR-106a-5p (ORN-52), served as a reference. In all duplexes, a fully complementary 15-nt counter-strand (ORN-2) was used as a second strand of the duplex.

The T_M values (calculated as the maximum of the first derivative of the melting curve) for all of the measured samples were convergent (**Table 16**). A slightly higher T_M value obtained for a duplex containing ORN-1 (61.1 +/- 0.34°C) might be due to an increased stabilization caused by the intercalation of the trioxsalen into the double helix. However, the effect was too small to draw any conclusions. In the same fashion, the introduction of a diazirine seemed to not interfere with a duplex formation.
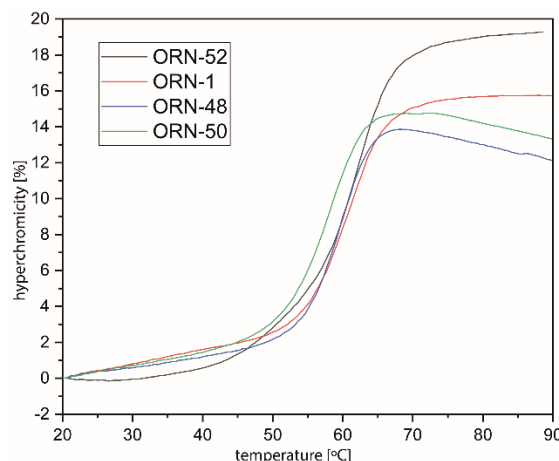
Graph 1: Melting profiles of the miR-106a-5p and its analogues modified with the trioxsalen or diazirine moieties incorporated at different positions within the sequence against the fully complementary counter-strand (ORN-2). Conditions: 2 µM concentration of each strand of the duplex in 100 mM NaCl, 10 mM Na-phosphate buffer and 0.1 mM Na$_2$EDTA, pH 7.0.

Table 16: Melting temperatures calculated for the investigated duplexes.

Sample	Modification	T_M [°C]	ΔT_M [°C]
ORN-52	none	59.7 ± 1.2	-
ORN-1 (D1)	2'-*O*-Aprop-trioxsalen	61.1 ± 0.3	1.4
ORN-48 (D35)	2'-*O*-Aprop-diazirine	60.0 ± 0.2	0.3
ORN-50 (D37)	C^5-Uacetylene-diazirine	58.5 ± 0.1	-1.2

III.3.4. Synthesis of the aryl trifluoromethyldiazirine nucleoside analogue

Attempts to synthesize the fully protected aryl trifluoromethyldiazirine nucleoside Analogue B were undertaken. This would enable to reproduce the cross-linking conditions reported by Nakamoto et al. and apply them for other miRNAs, for which the cross-linking with trioxsalen was not achieved.

The reported synthetic pathway was followed (**Scheme 2**) [278]. The glycosylation of compound **6** with commercially available 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-*β*-D-ribofuranose, followed by the removal of the benzoyl protecting groups gave the expected products with good yields (compounds **9** and **10**, respectively). The 4,4'-dimethoxytrityl (DMTr)-protection of the 5'-hydroxyl group was, in general, less efficient and cleavage of the acetal bond was occasionally observed. Treatment of the compound **11** with the TBDMSCl gave a mixture of 2'- and 3'-*O*-protected isomers formed in almost 1:1 ratio, that could be separated by preparative thin-layer chromatography (TLC). The isolated 2'-isomer (**12**) was subjected to the phosphitylation reaction but the formation of the product was not detected.
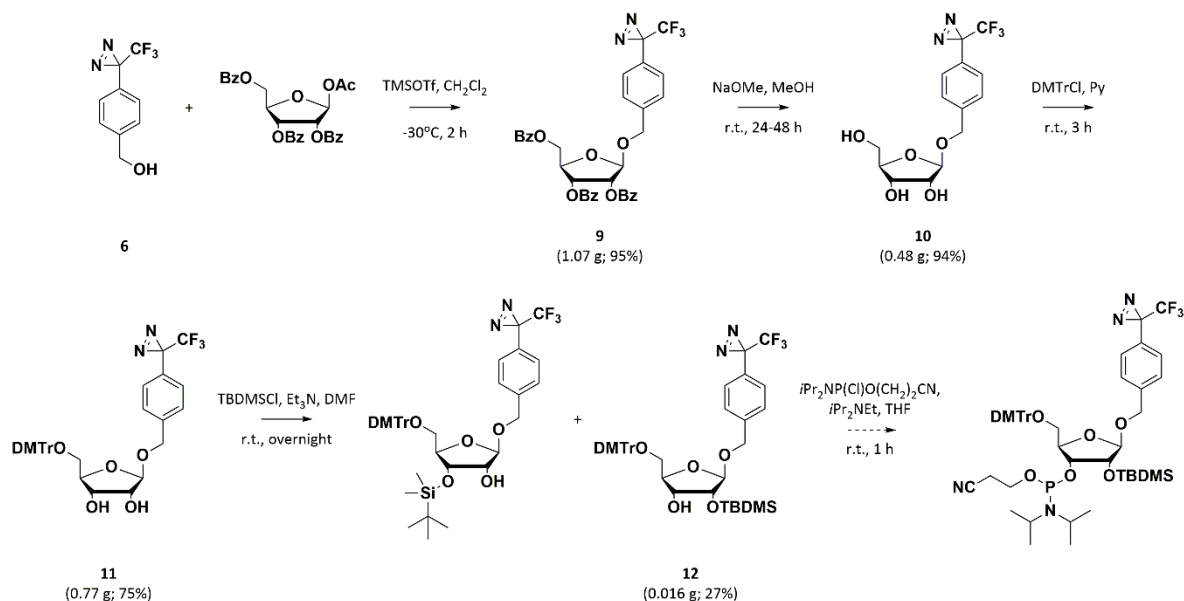
Scheme 2: Schematic representation of the synthesis of the fully protected nucleoside Analogue B bearing the aryl trifluoromethyldiazirine as a base [278].

Further optimization of the reaction conditions would most likely allow to obtain the final phosphoramidite. However, low efficiency of the 2'-OH protection step and laborious separation of the resulting mono- and bis-protected products contributed to discontinuation of this synthetic pathway. As an alternative, synthesis of the DNA analogue with the same photo-reactive scaffold was considered. This would allow omitting the yield-limiting step (protection of the 2'-OH group). Heck conditions were chosen to perform the glycosylation reaction [310]. For this purpose, 3-(4-iodophenyl)-3-(trifluoromethyl)-3H-diazirine (**14**) and furanoid glycal (**16**) building blocks were synthesized (**Scheme 3**).
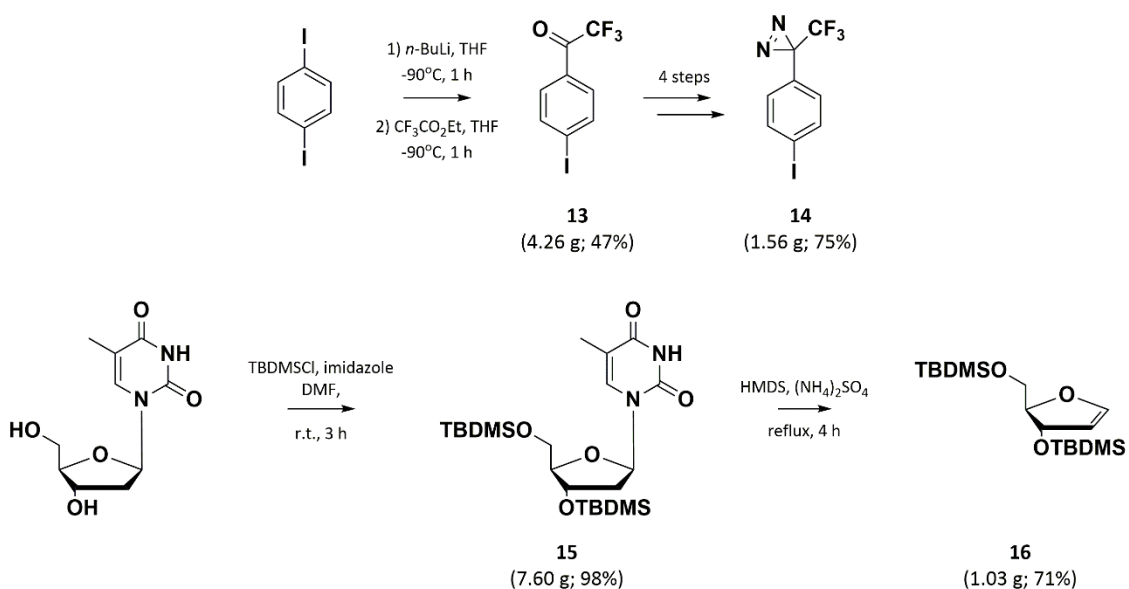
Scheme 3: Synthetic approach to obtain the 3-(4-iodophenyl)-3-(trifluoromethyl)-3H-diazirine (**14**) and furanoid glycal (**16**) building blocks.

3-(4-iodophenyl)-3-(trifluoromethyl)-3*H*-diazirine (**14**) was obtained from 1,4-diiodobenzene, adapting the five-step procedure reported by Savechenkov et al. [314]. The synthetic pathway was analogous to the synthesis of the (4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)phenyl)methanol (**6**). Furanoid glycal (**16**) was synthesized from thymidine in two steps, following the published protocol [315]. Heck reaction with palladium (II) acetate was run according to the published conditions [310] but only small amounts of the product were generated (data not shown). Test reactions with other palladium catalysts were performed but they did not improve the yield. Taking into consideration all the advantages and disadvantages of this strategy and the time limitations, the synthesis of the diazirine-based probes was not pursued further.

III.4. Conclusions and outlook

In this chapter, the efforts taken to achieve the miRNA-RNA cross-linking with use of the diazirine-modified probes have been summarized. In contrary to the synthesis of the fully protected nucleoside analogue bearing the trifluoromethyl aryldiazirine moiety (Analogue B), the synthesis of the 3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3*H*-diazirine (**8**) was straightforward. This is why the concept of post-synthetic functionalization of the oligoribonucleotides with the CLICK chemistry seemed worth testing.

An application of the phosphoramidite with an aryl trifluoromethyldiazirine as a nucleobase ensured the placement of the photo-reactive moiety inside the duplex structure, thus increasing the chances for the cross-linking with a target. Simultaneously, this most likely abolished the base-pairing at this position within the sequence, which resulted in a decreased target affinity. Although the diazirine-based probes proposed by the Ueno group presented some advantages, their ability to label known miRNA targets in cells was demonstrated only for miR-145. The usefulness of such probes for the identification of the unknown miRNA targets has not been assessed so far.

The utilization of the CuAAC reaction allows to readily modify different sites within the duplex. The presented results implied that CLICK-based attachment (either at the 2'-*O*- or C^5-position) positioned the diazirine functional group far from the other strand of the duplex. Thus, instead of the desired inter-strand cross-linking, insertion into the O-H bonds of the ubiquitous water molecules was observed. The molecular modelling studies would be valuable to further investigate this issue.

Performed experiments provided additional knowledge about the cross-linking process. Although the synthesized probes with the diazirine as a photophore seemed to be of no value in the context of the miRNA-RNA cross-linking, they might be useful to study miRNA-protein interactions on the molecular level. The fact that the diazirine moiety is protruding outside of the duplex structure would be beneficial in this case, as it increases the probability for capturing interactions with dsRNA-binding proteins.

III.5. Contributions

2'-*O*-propargyl phosphoramidites and 4'-azidomethyltrioxsalen were taken from the common laboratory stock synthesized by Martina Bigatti as a part of the NCCR platform. Synthesis of the azidodiazirine, the aryl trifluoromethyldiazirine nucleoside analogue and oligoribonucleotides, as well as the *in vitro* photo-cross-linking assays, melting study and data analysis was performed by Anna L. Malinowska.

IV. Chapter IV: Synthesis and application of PEG-based linkers for a more versatile miRNA-mRNA cross-linking

IV.1. Introduction

The potential application of the oligonucleotides bearing non-natural modifications as biological tools and probes created a need for their site-specific functionalization. In a 'traditional' fashion, this would require the step-wise synthesis of the fully-protected modified monomer (modified phosphoramidite) and its incorporation into the sequence during the solid-phase synthesis. However, this demands preparation of each building block and oligonucleotide separately (a 'dedicated monomer strategy' [316]). Such an approach is tedious and time-consuming, especially when a library of compounds differing by the linker length or substituents is investigated. This stimulated a search for new chemical strategies enabling more versatile functionalization of modified oligonucleotides.

In 1990, MacMillan and Verdine introduced the 'convertible nucleoside' approach [317]. This concept was based on the application of the building block (the title convertible nucleoside) equipped with a good leaving group attached to the nucleobase. The modified monomer (which should be easy to prepare) was incorporated into the oligonucleotide sequence during standard oligonucleotide synthesis. Subsequently, the leaving group was replaced by an amine- or thiol-bearing substituent in a post-synthetic nucleophilic displacement reaction, performed when the oligonucleotide was still attached to the solid support [316, 317]. In this manner, a set of oligonucleotides having the same sequence but carrying different tethers can be prepared from a single precursor oligonucleotide containing a convertible nucleoside. In the original example, the trimethylphenyl residue attached to the O^4 of deoxyuridine (TMP-dU phosphoramidite) (**Figure 33**) was substituted with various aqueous amines to give deoxycytidine with different moieties appended at its N^4 position [317].

4-O-(2,4,6-trimethylphenyl)-2'-deoxyuridine phosphoramidite, TMP-dU phosphoramidite

O^4-triazolyluridine phosphoramidite
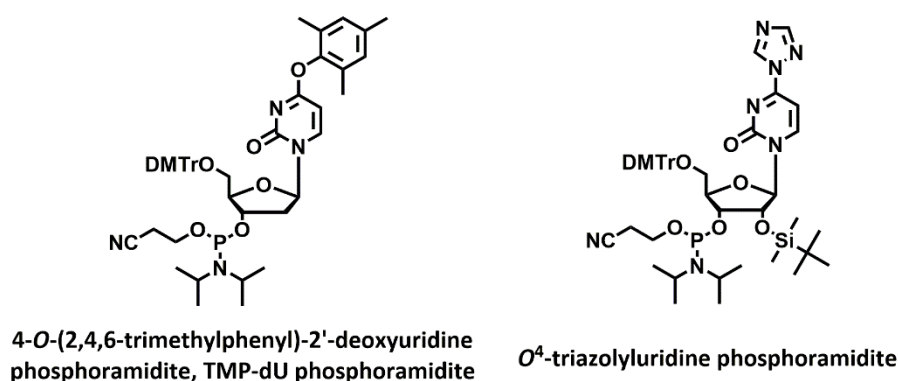
Figure 33: Examples of the convertible nucleosides used for the post-synthetic nucleophilic displacement with amines and thiols.

Substituents incorporated at this position are located within the major groove of the DNA duplex [317, 318]. Previous studies reported that mono-substitution on the exocyclic amine of deoxycytidine did not disturb the Watson-Crick base-pairing and the formation of helical DNA and RNA structures [317, 319-322]. However, this required that the residue appended to N^4 was protruding away from the Watson-Crick face (i.e. was in *trans* conformation with the N^3 atom of (deoxy)cytidine) (**Figure 34a**). It was postulated that mono-substituents attached to the exocyclic amine of (deoxy)cytidine may

preferentially adopt a *cis* conformation on the nucleoside [320, 323]. Nonetheless, the situation seemed to be different in the oligonucleotide context, where the rotation of the tether towards the 'CH'-edge was forced to enable favourable hybridization into a duplex structure [320]. On the other hand, the presence of a methyl group at the C^5 or dC^5 was shown to prevent rotation of the substituent due to a steric clash between the two residues (**Figure 34b**) [320]. This lead to the loss of Watson-Crick base-pairing at this position, manifested by a significant drop in the observed T_M.
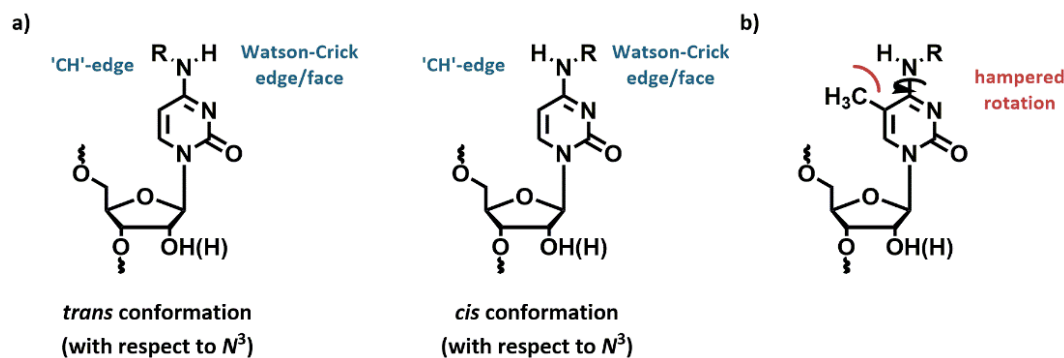
Figure 34: a) Possible conformations of the single substituents (R) attached to the exocyclic amine of cytidine or deoxycytidine. *Trans* conformation enables canonical base-pairing with the juxtaposed base in the complementary strand; b) the presence of a methyl group at C^5 can hamper the rotation around C^4-N^4 bond and thus disable Watson-Crick base-pairing. Adapted from: [320].

Shah et al. described the synthesis of the fully protected uridine phosphoramidite with a triazole moiety attached at the N^4 (O^4-triazolyluridine phosphoramidite in **Figure 33**), which upon incorporation into the RNA sequence served as a leaving group in the nucleophilic substitution reaction with the thiolacetic acid [324]. Nowadays, the triazole substitution is one of the most commonly used versions of the convertible nucleoside approach.

In the group of prof. Hall, the triazole-based convertible nucleoside approach was employed to synthesize miRNA mimics with mono- and bis-methylated cytidines to study their potential application as a new type of negative controls for RNA interference (RNAi) experiments [325]. This strategy was also utilized for the site-specific incorporation of the spermine into oligonucleotides and siRNA sequences to investigate the influence of this polycationic tether on the oligonucleotide's properties [319, 326].

Although not as popular as the CLICK chemistry, the convertible nucleoside approach allowed to significantly expand the repertoire of chemical modifications available for the functionalization of synthetic nucleic acids. In contrast to the CuAAC reaction, nucleophilic displacement of the triazole (or other leaving group employed) with amines does not create any additional scaffold between the reacting molecules. The cross-linking experiments with trioxsalen described in Chapter II suggested that the triazole moiety may have an influence on the cross-linking process. The presence of such a rigid structure within the linker can interfere with the intercalation of the trioxsalen in the neighbourhood of the base to which it is attached. Although the inter-strand cross-linking may be still achieved, the outcome of the reaction is highly dependent on the sequence and is difficult to predict.

IV.2. Aim of the project

The objective of this study was to develop an alternative cross-linking strategy that would enable to overcome the sequence-dependence of trioxsalen-mediated cross-linking, namely its bias towards reacting with pyrimidines.

We hypothesized that instead of using one defined modification, a mixture of probe homologues varying in linker length could be employed. Application of linkers with different lengths would deliver the cross-linker at a variety of positions with respect to the duplex structure and increase the range of the cross-linking process. To prepare such probes, the convertible nucleoside can be introduced at one (or more) positions within the miRNA sequence, and post-synthetically substituted with trioxsalen analogues bearing amino-linkers of various length (**Figure 35**).
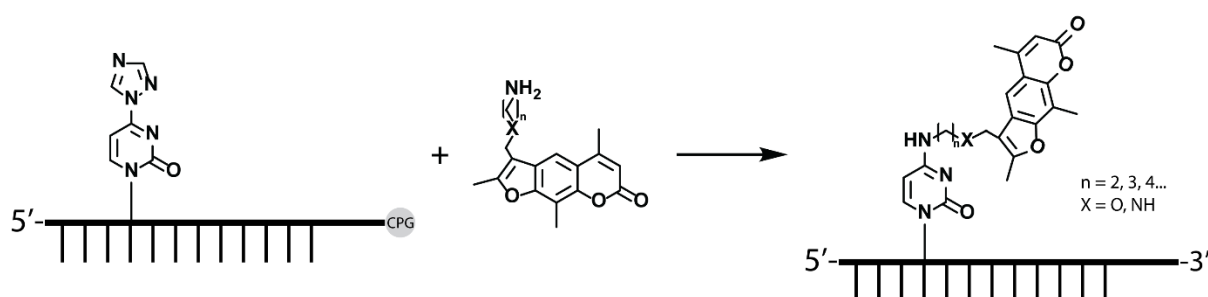
Figure 35: An overview of the application of the convertible nucleoside approach for the functionalization of the oligoribonucleotides with the trioxsalen attached through linkers of different lengths.

This would enable to obtain a small set of probes quickly and conveniently. Purified probes could be mixed stoichiometrically and simultaneously used in the cross-linking experiments (**Figure 36**). Much greater versatility would be possible for the trioxsalen to react with nearby uridines in a given target strand. This would ensure an increase in the probability of cross-linking events, as different putative cross-linking sites (uridines) could be simultaneously targeted. Therefore, this strategy could be considered as a potential way to overcome the issue of sequence-dependence connected to the use of trioxsalen.
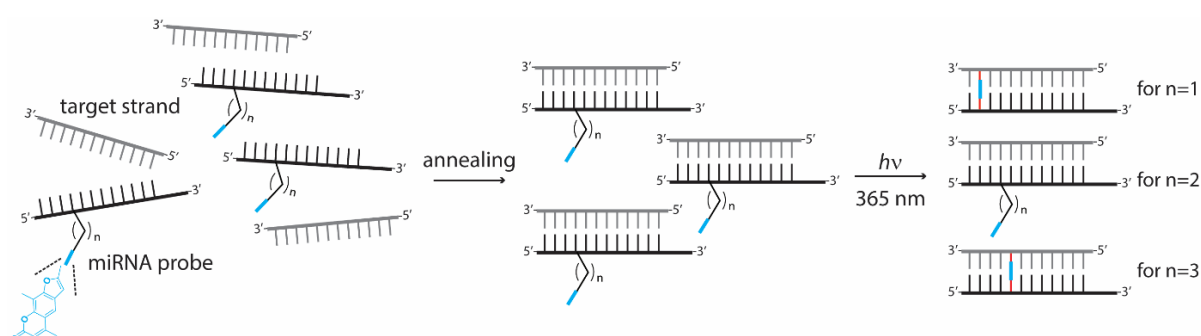
Figure 36: Schematic representation of the *in vitro* photo-cross-linking with trioxsalen-modified probes prepared with the use of the triazole-based convertible nucleoside approach. A mixture consisting of probes in which the trioxsalen is attached at the N^4 position of the same base (cytidine) *via* different linkers would be used. 'n' designates a number of atoms within the linker.

IV.3. Results and discussion

IV.3.1. Synthesis of the trioxsalen analogues with the amino-linkers of different lengths

To incorporate the trioxsalen into the oligoribonucleotide sequence with the convertible nucleoside approach, trioxsalen analogues with linkers having primary amines at their ends had to be synthesized. Initially, three possible approaches were considered to achieve this (**Scheme 4**).
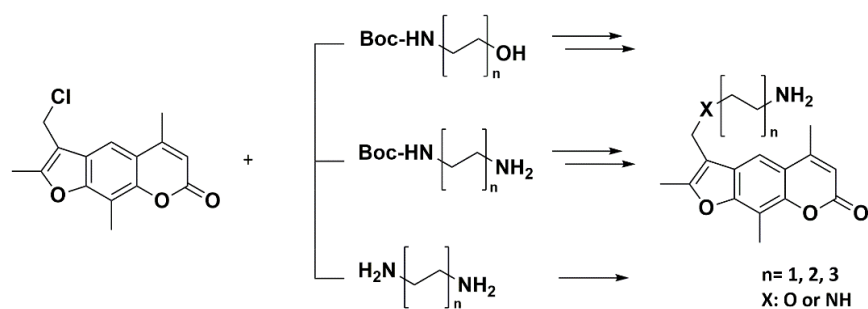
Scheme 4: Synthetic strategies to obtain trioxsalen analogues with C_2, C_4 and C_6 linkers having a terminal amine functionality.

All of them were based on the substitution of the 4'-chloromethyltrioxsalen with either *N-tert*-butyloxycarbonyl (Boc)-protected amino alcohol, mono-*N*-Boc-protected diamine or unprotected diamine with a C_2, C_4 or C_6 scaffold. Although for some of the linkers the substitution was achieved, reaction times were relatively long (approx. 60 h) and gave moderate yields (40-60%, depending on the strategy and the length of the linker; data not shown). Moreover, the Boc-deprotection step was troublesome and in some cases a partial cleavage of the linker was observed. On the other hand, isolation of the products of the substitution reaction with free diamines was tedious (time-consuming separation of the phases during the extraction), and small changes to the protocols did not improve the outcome of the tested reactions.

The nucleophilic displacement of the convertible nucleoside can be also performed pre-synthetically, yielding a substitution product which can be then incorporated into the oligonucleotide sequence as a phosphoramidite. However, pre-synthetic substitution with the trioxsalen analogue having a diamine linker would yield a product requiring additional protection of the secondary alkyl amine, in order to avoid side reactions of this group during oligonucleotide synthesis. Thus, the utilization of the diamine-based linkers was unfavourable. This prompted a search for an alternative pathway to obtain trioxsalen analogues.

In 1995, Hei patented the synthesis of the trioxsalen analogues with short poly(ethylene glycol) (PEG)-based linkers [327]. The main advantages for use of PEG linkers are that they are uncharged, highly flexible and more hydrophilic than the saturated aliphatic chains. The attachment of linear or branched PEG chains to oligonucleotides has been widely applied to enhance their pharmacokinetic properties, i.e. biodistribution, cellular uptake and lowering of the side effects [328]. Pegaptanib (pegaptanib sodium, trade name: Macugen) was the first RNA-based aptamer approved by the FDA for the treatment of the neovascular (wet) age-related macular degeneration [329, 330]. This aptamer selectively binds to the vascular endothelial growth factor (VEGF)-165 isoform involved in the development and progress of the disease. Incorporation of the 40 kDa Y-shaped PEG moiety at the 5'-end of the pegaptanib was associated with the reduced renal clearance and extended tissue residence

of the drug [329, 331]. However, much shorter ethylene glycol-based chains have also been employed to improve the pharmacokinetics of the oligonucleotides. The most prominent and simplest example is the 2'-MOE modification [332, 333], described in section I.1.3.2.

The reported synthetic pathway consisted of four relatively straightforward steps (from 4'-chloromethyltrioxsalen (**17**) as the starting material), which provided the desired products in moderate to good yields (**Scheme 5**).
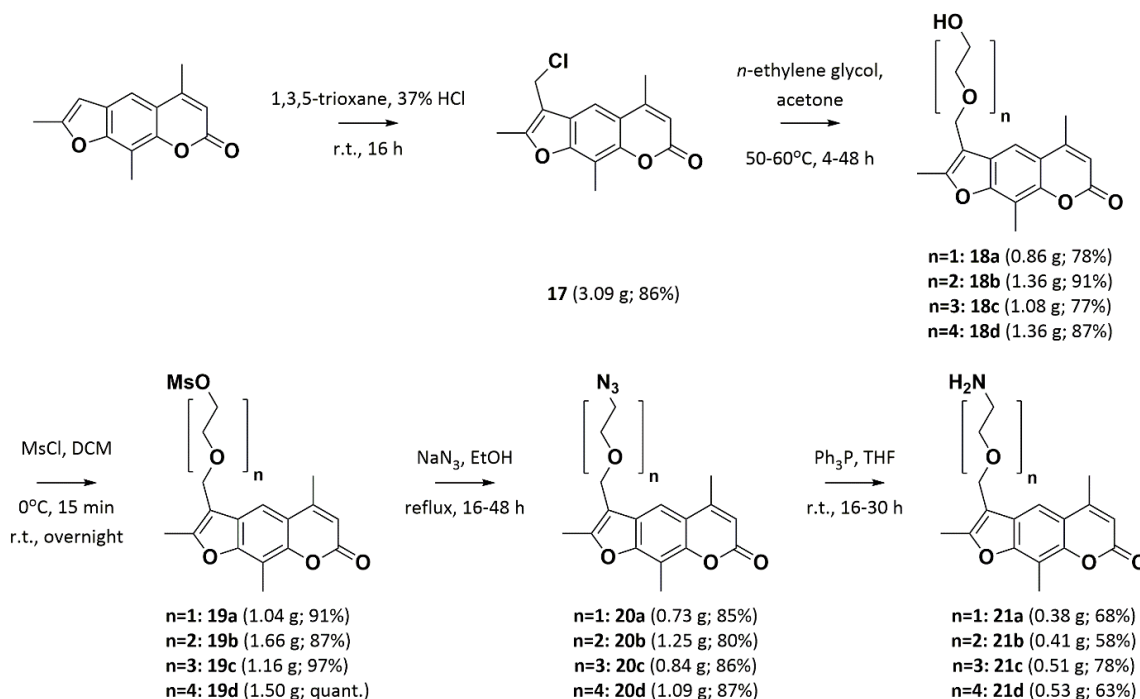
Scheme 5: Synthesis of the trioxsalen analogues with amino-PEG linkers of different lengths. Synthetic pathway adapted from [327].

Examples of two synthetic routes were described in the patent, starting from either 4'-bromomethyltrioxsalen or 4'-chloromethyltrioxsalen [327]. The synthetic approach employing 4'-chloromethyltrioxsalen (**17**) was chosen, as **17** can be easily obtained from the commercially available trioxsalen in one step (**Scheme 5**). The chloromethylation reaction gave the product (**17**) with a good yield. The product however was often contaminated with approx. 10-20% of the starting material (trioxsalen), despite the increase of the solvent volume used to dilute the thick suspension formed during the reaction. Purification by column chromatography did not improve the purity of the product. In fact, minor decreases in isolated yields were observed. Therefore, the slightly impure crude product was used in further steps. The synthesis of the 4'-[(4-amino-2-oxa)butyl]-4,5',8-trimethylpsoralen (**21a**, trioxsalen with a 'PEG1' linker) and 4'-[(7-amino-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (**21b**, trioxsalen with a 'PEG2' linker) was performed following the published protocol with minor changes [327]. 4'-chloromethyltrioxsalen (**17**) was substituted with ethylene glycol or diethylene glycol to give **18a** or **18b**, respectively. Purification by column chromatography performed at this step usually allowed to separate the residual starting material from the chloromethylation reaction. The hydroxyl groups were then mesylated and substituted with sodium azide to give **20a** and **20b**. The amounts of the sodium azide used and the reaction times were increased, as only partial conversion of the starting material was observed. The azides were then reduced to amines (**21a** and **21b**) under mild conditions with a Staudinger reaction. The purification of the final amines was performed by acid-base extraction.

This caused a small loss of the products due to the troubles with re-extraction of the free amines from the water phase.

The same strategy was adapted to obtain the trioxsalen with the 'PEG3' (**21c**) and 'PEG4' (**21d**) linkers. In general, reactions with extended linkers required longer reaction times. Furthermore, the number of equivalents of tri- and tetraethylene glycol was proportionately increased for the pegylation steps. Although this allowed to achieve quantitative conversions, the removal of the surplus glycols was tedious. Despite extensive washing with water and double purification by flash column chromatography, a small excess of the glycols was still present. These slightly impure compounds were used in the next steps. In total, four trioxsalen analogues with amino-PEG linkers were synthesized following this synthetic strategy.

The last compound, 4'-aminomethyltrioxsalen (**22**, trioxsalen with a 'PEG0' linker), was obtained performing the Gabriel synthesis (**Scheme 6**) [252]. In the first step, the potassium phthalimide was *N*-alkylated with 4'-chloromethyltrioxsalen (**17**). Upon reaction with hydrazine, the phthalimide was hydrolyzed, yielding the desired amino-product (**22**).
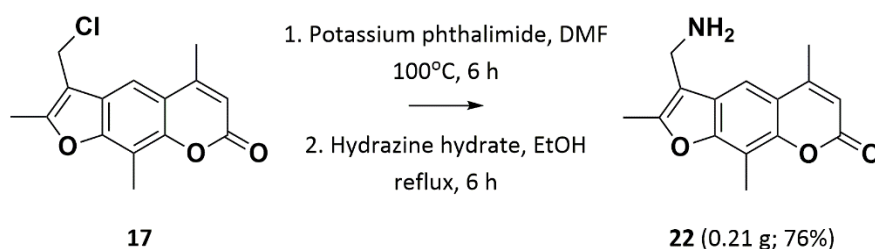
Scheme 6: Synthesis of the 4'-aminomethyltrioxsalen (**22**) [252].

IV.3.2. Synthesis of the probes

Prepared trioxsalen analogues with PEG0 to PEG4 linkers were subsequently employed to modify oligoribonucleotides. The miR-124-3p sequence was chosen as a test sequence. The cross-linking for miR-124-3p probes, with the trioxsalen incorporated through the CLICK chemistry as previously described, was rather inefficient (**Figure 37**). This left room for improvement of the efficiency of the cross-linking process for this miRNA. Moreover, the results from the experiments with the probes of the new design could be compared with those obtained for the miR-124-3p-1 (ORN-29) and miR-124-3p-2 (ORN-30) probes.
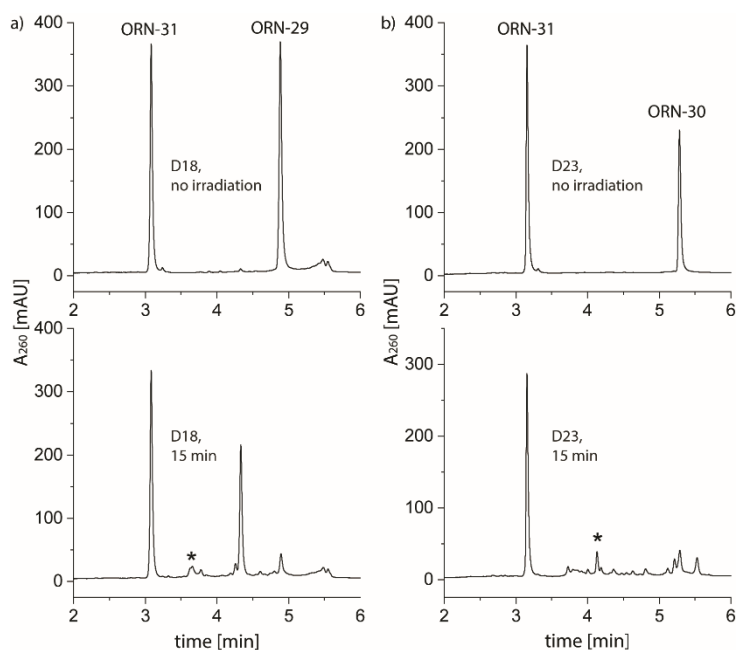
Figure 37: Selected RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment for a) D18 and b) D23. Top panels present the non-irradiated samples (references). Bottom panels represent samples irradiated for 15 min (λ=365 nm). Peaks designated with an asterisk (*) correspond to the cross-linked duplex; a) mass calc. [M+K]: 11493.3, mass found: 11491.4 and b) mass calc.: 11452.3, mass found: 11452.8. The full chromatograms from this experiment are presented in **Figure 18** in section II.3.3.1.

The main purpose of this project was to investigate the usefulness of the convertible nucleoside approach for the preparation of the trioxsalen-modified probes with diversified linker lengths and then to test their performance in the photo-cross-linking assays. An additional benefit from using the selected synthetic pathway was that it generated analogues with an azide functional group as intermediates (**20a-20d**). Thus, oligoribonucleotides with CLICKed modifications were also prepared. Therefore, two synthetic strategies, presented in **Figure 38a**, were employed to prepare the trioxsalen-bearing oligoribonucleotides. This yielded analogues which had the modification attached either in the major (when introduced *via* the convertible nucleoside approach) or in the minor groove (when appended through the CLICK chemistry at the 2'-*O*-position of the ribose).
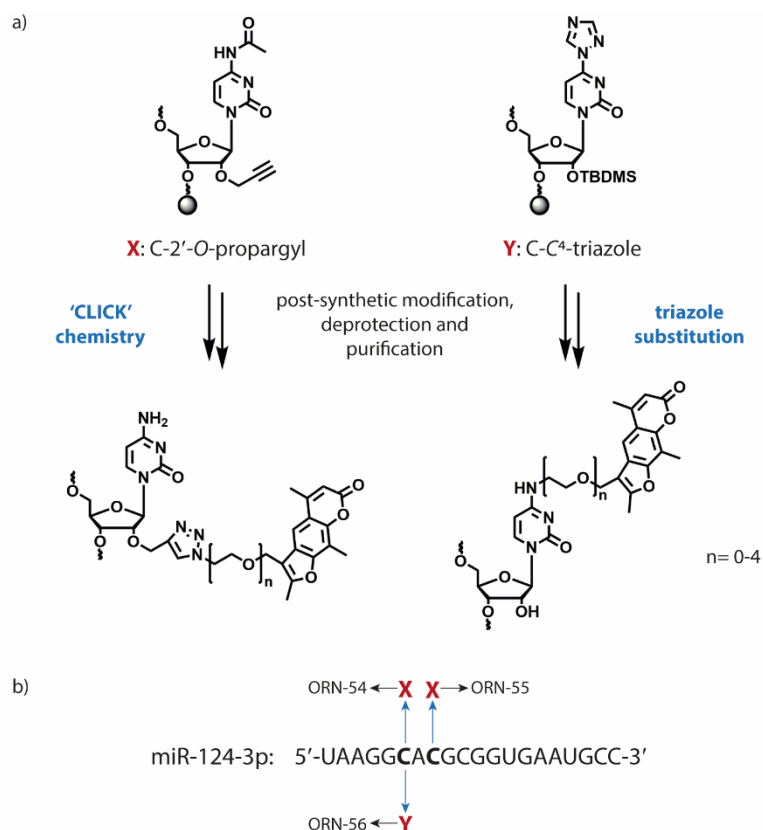
Figure 38: Design of the new probes in which trioxsalen was attached *via* PEG-based linkers of different lengths; a) strategies for the post-synthetic functionalization of the oligoribonucleotides; b) an indication of the positions within the miR-124-3p sequence at which the modifications were introduced. Cytidine (**C**) was replaced by either 2'-*O*-propargyl cytidine (**X**) or O^4-triazolyluridine (**Y**), which were then post-synthetically modified with the trioxsalen analogues bearing linkers terminated with an azide or an amine functionality, respectively.

Three oligoribonucleotides with miR-124-3p sequence were synthesized on 1 μmol scale (**Figure 38b**). ORN-54 and ORN-55 had a 2'-*O*-propargyl cytidine at positions 6 and 8 within the sequence, respectively. Five portions of approx. 15 mg of the solid support with either ORN-54 or ORN-55 were distributed into screw cap vials and treated with one of the trioxsalen azides (**20a-20d** or 4'-azidomethyltrioxsalen), according to the standard CLICK procedure which was performed on a slightly bigger scale (usually 5-10 mg of the solid support with oligoribonucleotide is used). The scale-up of the protocol was performed by doubling the amounts of the reagents. This gave rise to a set of miR-124-3p-C6- and miR-124-3p-C8-2'-*O*-modified oligoribonucleotides (ORN-54a to ORN-54e and ORN-55a to ORN-55e).

In the case of ORN-56, O^4-triazolyluridine was incorporated at position 6 of the miR-124-3p. Five 20 mg portions of the solid support were transferred into screw cap vials and treated with amines (**21a-21d** or **22**) dissolved in acetonitrile in the presence of DBU (1,8-diazabicyclo[5.4.0]undec-7-ene) as a non-nucleophilic base. Reaction mixtures were shaken for 6 h at 60°C, as these conditions were optimal. Shorter reaction times and lower temperatures showed reduced conversion, while longer reaction times did not increase the yields of the desired products. Afterwards, suspensions were concentrated and deprotected according to the standard procedure. During the DMTr-on purification of these compounds, all of the fractions with the retention times higher than 2 min were usually collected, combined and deprotected (**Figure 39a**). The RP-HPLC purification in the DMTr-off mode resulted in the clean separation of the products, although for some of the compounds more than one purification

round was needed. Three peaks were usually isolated. The first two were characteristic for all of the samples subjected to the nucleophilic displacement reaction and subsequent workup protocol. The mass of the fastest-eluting peak corresponded to the product of the substitution with ammonia, while the second peak seemed to be related to the product of the substitution with DBU or its derivative (the exact structure of the adduct was not elucidated). The mass of the third peak matched the mass of the desired product of the triazole displacement with the trioxsalen analogue (**Figure 39b**).
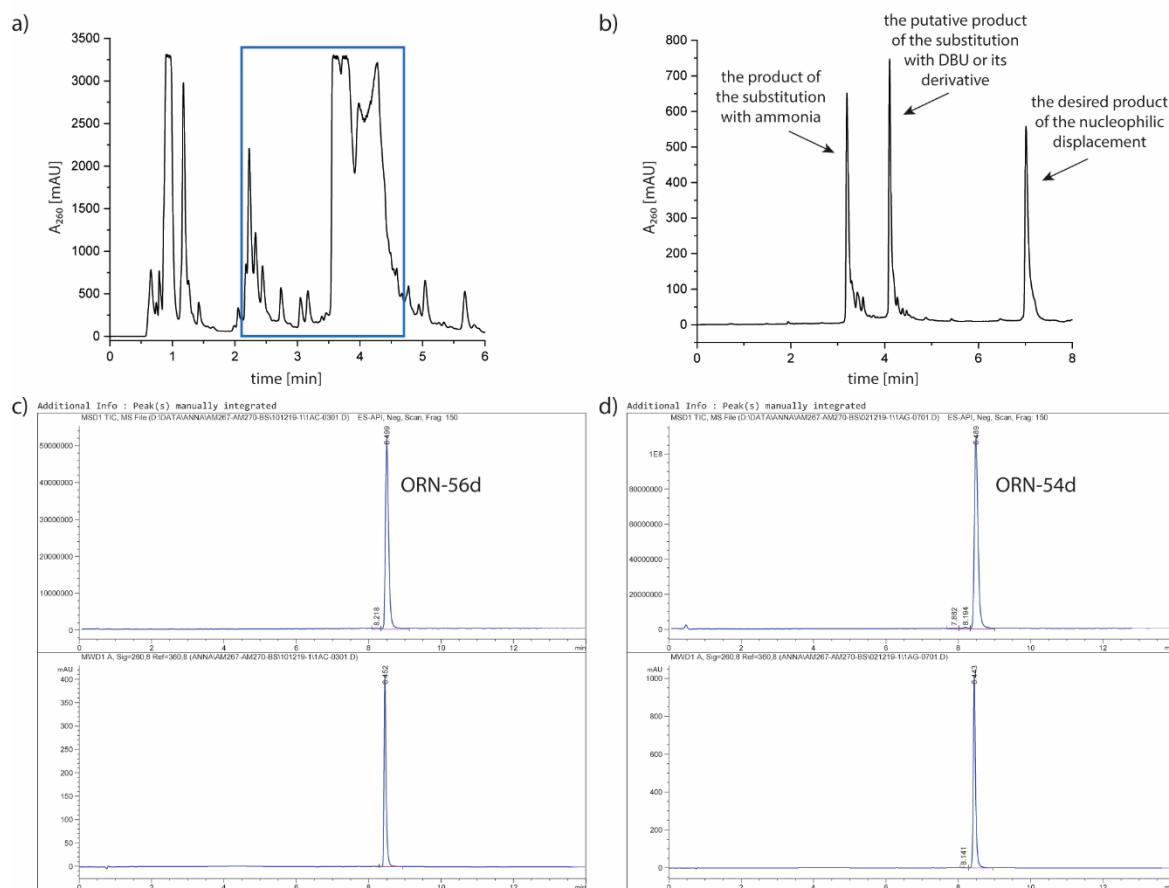
Figure 39: Illustrative examples of the chromatograms from the DMTr-on (a) and DMTr-off (b) purification (RP-HPLC) of the oligoribonucleotides obtained with the convertible nucleoside approach. The blue box on the chromatogram from the DMTr-on purification (a) indicates fractions which were combined and deprotected; c) and d): LC-MS chromatograms of the final products: ORN-56d and ORN-54d, respectively.

In total, a small library of 15 oligoribonucleotides modified with a trioxsalen attached through PEG0-PEG4 linkers was obtained (**Table 17**). The 'prop' in the miR-124-3p-C6-2'-*O*-prop-PEG and miR-124-3p-C8-2'-*O*-prop-PEG refers to the modifications that were appended by the CLICK chemistry with the 2'-*O*-propargyl moiety. Illustrative LC-MS chromatograms of the final oligoribonucleotides obtained with a convertible nucleoside approach (ORN-56d) or CLICK chemistry (ORN-54d) are presented in **Figure 39c** and **Figure 39d**, respectively.

Table 17: A small library of the synthesized miR-124-3p analogues modified with a trioxsalen attached through PEG-based linkers of different lengths. **X**: cytidine with the trioxsalen at the 2'-*O*-position of the ribose; **Y**: cytidine with the trioxsalen at the *N*⁴-position of the nucleobase. PEG0-PEG4 indicates how many ethylene glycol units are present within the linker.

Entry	ORN	Sequence (5' to 3')	Mass calc.	Mass found
1	ORN-53	UAAGGCACGCGGUGAAUGCC	6445.0	6444.7
miR-124-3p-C6-2'-*O*-prop-PEG				
2	ORN-54a	UAAGGX_{PEG0}ACGCGGUGAAUGCC	6766.4	6765.2
3	ORN-54b	UAAGGX_{PEG1}ACGCGGUGAAUGCC	6810.4	6809.3
4	ORN-54c	UAAGGX_{PEG2}ACGCGGUGAAUGCC	6854.5	6853.4
5	ORN-54d	UAAGGX_{PEG3}ACGCGGUGAAUGCC	6898.5	6897.4
6	ORN-54e	UAAGGX_{PEG4}ACGCGGUGAAUGCC	6942.6	6941.5
miR-124-3p-C8-2'-*O*-prop-PEG				
7	ORN-55a	UAAGGCAX_{PEG0}GCGGUGAAUGCC	6766.4	6765.3
8	ORN-55b	UAAGGCAX_{PEG1}GCGGUGAAUGCC	6810.4	6809.4
9	ORN-55c	UAAGGCAX_{PEG2}GCGGUGAAUGCC	6854.5	6853.4
10	ORN-55d	UAAGGCAX_{PEG3}GCGGUGAAUGCC	6898.5	6897.4
11	ORN-55e	UAAGGCAX_{PEG4}GCGGUGAAUGCC	6942.6	6941.6
miR-124-3p-C6-*N*⁴-PEG				
12	ORN-56a	UAAGGY_{PEG0}ACGCGGUGAAUGCC	6685.3	6684.2
13	ORN-56b	UAAGGY_{PEG1}ACGCGGUGAAUGCC	6729.3	6728.3
14	ORN-56c	UAAGGY_{PEG2}ACGCGGUGAAUGCC	6773.4	6772.3
15	ORN-56d	UAAGGY_{PEG3}ACGCGGUGAAUGCC	6817.4	6816.4
16	ORN-56e	UAAGGY_{PEG4}ACGCGGUGAAUGCC	6861.5	6860.4

IV.3.3. *In vitro* photo-cross-linking experiments

Table 18: Summary of the *in vitro* photo-cross-linking experiment performed with the miR-124-3p-C6-2'-*O*-prop-PEG probes. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 3 µM. **X**: cytidine with the trioxsalen attached at the 2'-*O*-position of the ribose. PEG0-PEG4 indicates how many ethylene glycol units are present within the linker. The estimated cross-linking yield is given in brackets. Masses of the isolated peaks are presented in section VII.3.5.1.

Duplex	ORN	Sequence	Cross-linked product
miR-124-3p-C6-2'-*O*-prop-PEG			
D42	ORN-53	5'-UAAGGCACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D43	ORN-54a (probe)	5'-UAAGGX_{PEG0}ACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D44	ORN-54b (probe)	5'-UAAGGX_{PEG1}ACGCGGUGAAUGCC-3'	detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	(~18%)
D45	ORN-54c (probe)	5'-UAAGGX_{PEG2}ACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D46	ORN-54d (probe)	5'-UAAGGX_{PEG3}ACGCGGUGAAUGCC-3'	detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	(~25%)
D47	ORN-54e (probe)	5'-UAAGGX_{PEG4}ACGCGGUGAAUGCC-3'	detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	(~21%)

The *in vitro* photo-cross-linking assays were conducted according to the standard procedure described in the previous sections (II.3.2 and III.3.2). Initial experiments were performed with single probes. In all of the samples, the 15-mer with a sequence fully complementary to the 5'-end of the probe was used as a counter-strand (ORN-31). The irradiation of the duplex containing native miR-124-3p strand served as a negative control (D42), to ensure that no radical-based cross-linking occurred upon irradiation. First, miR-124-3p-C6-2'-*O*-prop-PEG analogues were tested (**Table 18**).

Samples containing duplexes D42-D47 were irradiated for 15 minutes. The chromatograms from the *in vitro* photo-cross-linking experiment with these compounds are presented in **Figure 40**.
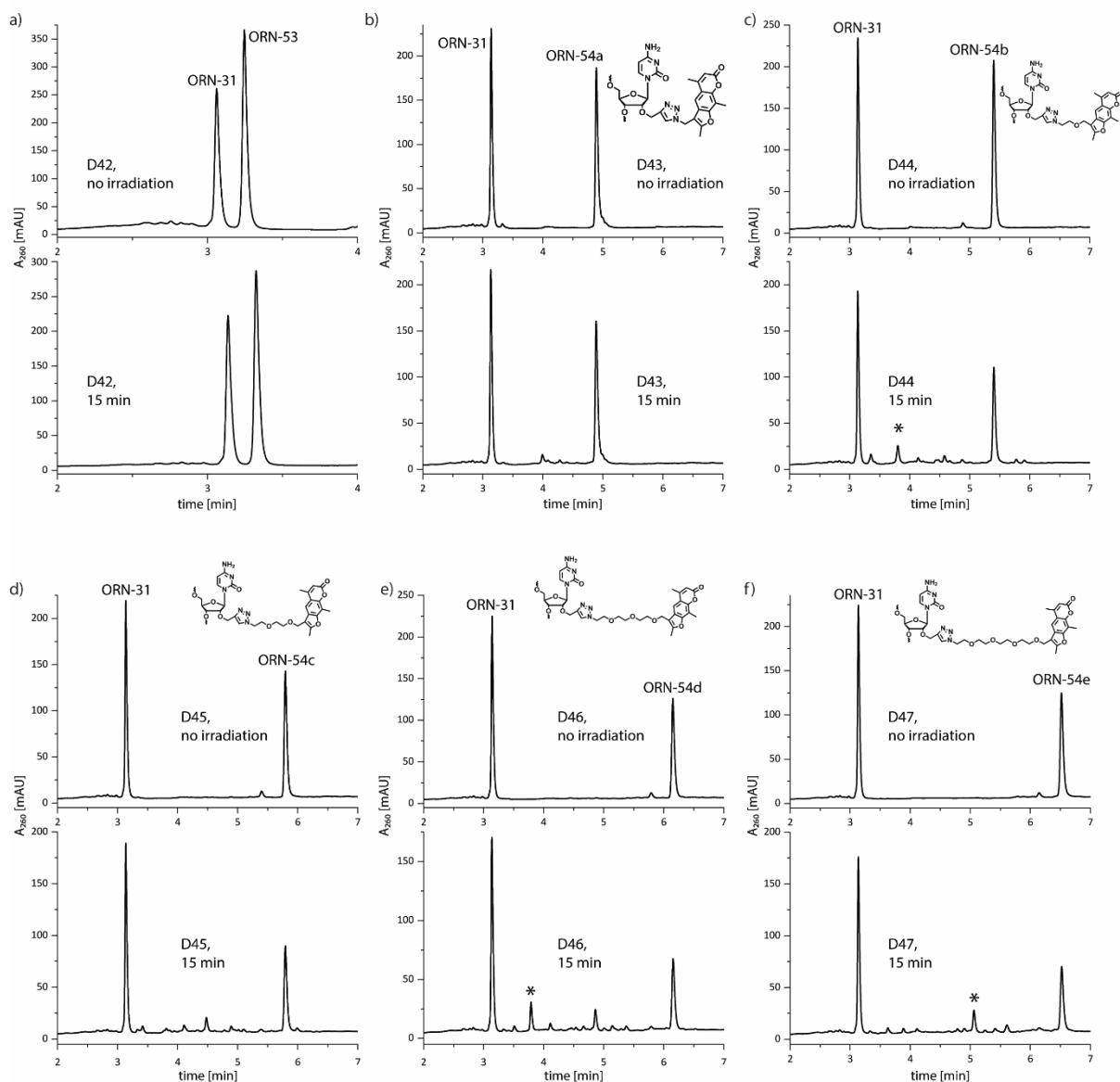
Figure 40: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with the miR-124-3p-C6-2'-*O*-prop-PEG probes containing linkers of different lengths [0-4 ethylene glycol units, b) to f), respectively]; a) an unmodified duplex D42 serving as a negative control. Top panels present the non-irradiated samples (references). Bottom panels represent samples irradiated for 15 min (λ=365 nm). Peaks marked with an asterisk (*) correspond to the cross-linked duplexes; c) mass calc.: 11498.3, mass found: 11587.1; e) mass calc. [M+K]: 11625.4, mass found: 11623.0; f) mass calc.: 11630.5, mass found: 11689.0.

Chromatograms are ordered according to the increasing linker length, starting from the non-modified duplex D42 (**Figure 40a**), followed by the D43 with PEG0 linker (**Figure 40b**) up to the duplex D47, in which the trioxsalen was attached *via* the PEG4 linker (**Figure 40f**). The top panels correspond to the non-irradiated samples (references), while the bottom panels represent the irradiated samples. The overall hydrophobicity of the probes was increased with increased length of the linker, which was reflected in the retention times. The cross-linking was detected for three duplexes, in which trioxsalen was appended with PEG1 (D44), PEG3 (D46) and PEG4 (D47) linkers. Most likely, the longer linkers allowed the cross-linking to uridine t_2 or t_3 (**Figure 41**). However, the cross-linking to cytidines t_4 and t_5 cannot be ruled out.
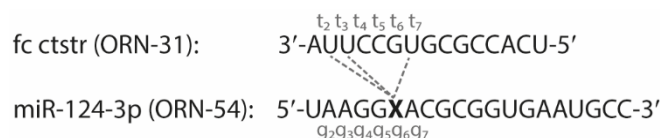
Figure 41: Sequence of the miR-124-3p analogue with the trioxsalen introduced at the C6-2'-*O*-position *via* PEG linkers of different length (ORN-54) and its fully complementary counter-strand (fc ctstr; ORN-31). The putative cross-linking sites are indicated by dashed lines. **X**: cytidine with the trioxsalen at the 2'-*O*-position of the ribose.

In the second round of experiments, the cross-linking properties of the miR-124-3p-C8-2'-*O*-prop-PEG analogues were investigated (**Table 19**).

Table 19: Summary of the *in vitro* photo-cross-linking experiment performed with the miR-124-3p-C8-2'-*O*-prop-PEG probes. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 3 μM. **X**: cytidine with the trioxsalen at the 2'-*O*-position of the ribose. PEG0-PEG4 indicates how many ethylene glycol units are present within the linker. The approximated cross-linking yield (calculated based on the consumption of the unmodified counter-strand, ORN-31) is given in bracket. Fc ctstr: fully complementary counter-strand. Masses of the isolated peaks are presented in section VII.3.5.2.

Duplex	ORN	Sequence	Cross-linked product
miR-124-3p-C8-2'-*O*-prop-PEG			
D42	ORN-53	5'-UAAGGCACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D48	ORN-55a (probe)	5'-UAAGGCA**X**_{PEG0}GCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D49	ORN-55b (probe)	5'-UAAGGCA**X**_{PEG1}GCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D50	ORN-55c (probe)	5'-UAAGGCA**X**_{PEG2}GCGGUGAAUGCC-3'	detected (~14%)
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D51	ORN-55d (probe)	5'-UAAGGCA**X**_{PEG3}GCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D52	ORN-55e (probe)	5'-UAAGGCA**X**_{PEG4}GCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	

For this set of compounds, the formation of the cross-linked product was observed for duplex D50 only (**Figure 42**). It suggested that solely PEG2 linker used in this probe was able to position the trioxsalen so that it could intercalate into the duplex structure near to a uridine in the target strand,

in this case most likely uridine t_7 (**Figure 43**). Therefore, the inter-strand covalent bond could have been generated upon irradiation.
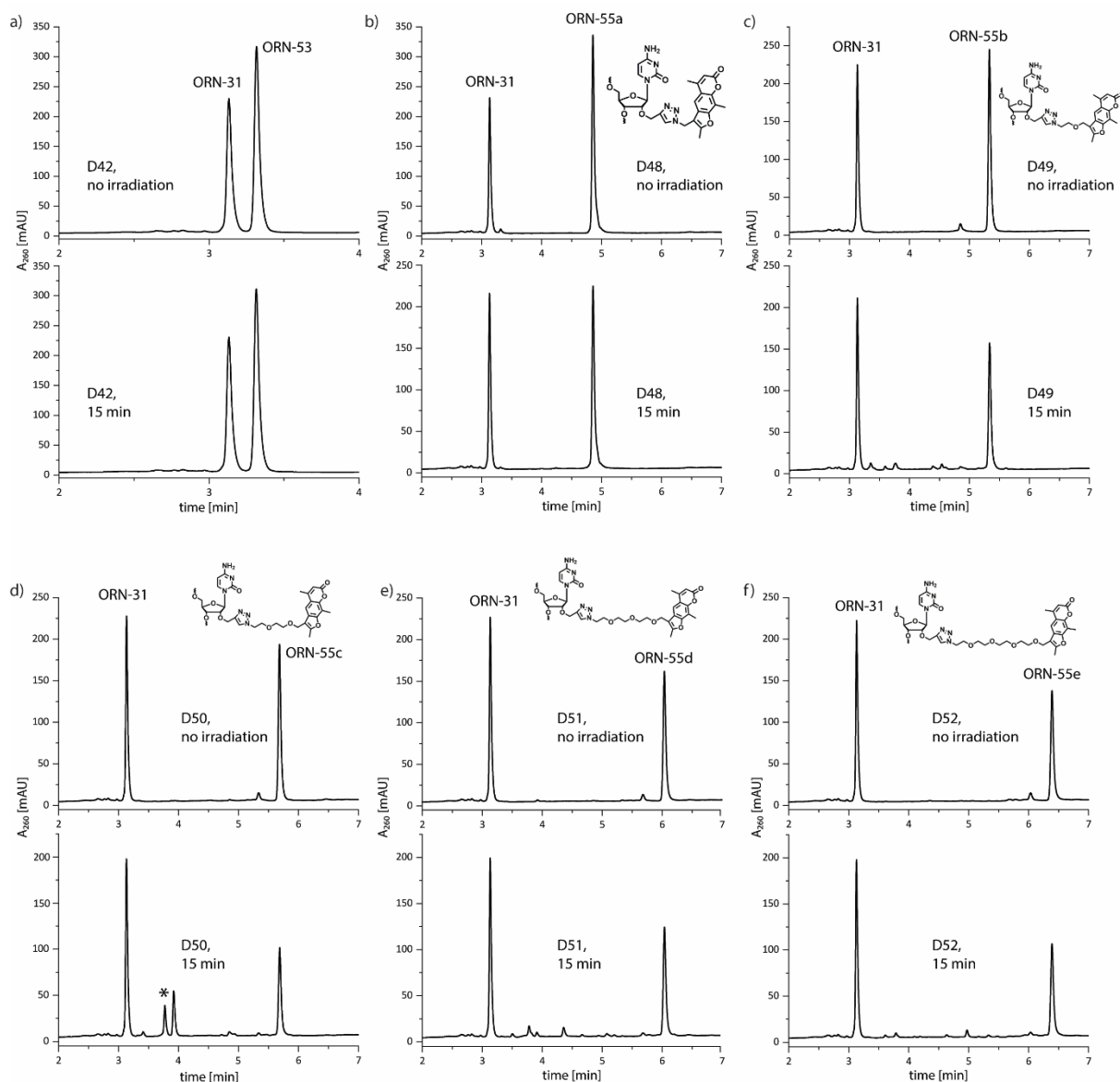
Figure 42: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with the miR-124-3p-C8-2′-*O*-prop-PEG probes with linkers of different lengths [0-4 ethylene glycol units, b) to f), respectively]; a) an unmodified duplex D42 serving as a negative control. Top panels present the non-irradiated samples (references). Bottom panels represent samples irradiated for 15 min (λ=365 nm). Peak indicated with an asterisk (*) corresponds to the cross-linked duplex; d) mass calc. [M+K]: 11581.4, mass found: 11579.2.
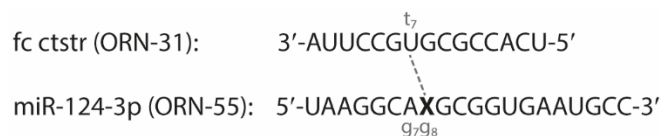
Figure 43: Sequence of the miR-124-3p analogue with the trioxsalen introduced at the C8-2′-*O*-position *via* PEG linkers of different length (ORN-55) and its fully complementary counter-strand (fc ctstr; ORN-31). The putative cross-linking site is indicated by a dashed line. **X**: cytidine with the trioxsalen at the 2′-*O*-position of the ribose.

Finally, the cross-linking assay with the miR-124-3p-C6-N^4-PEG probes (**ORN-56a** to **ORN-56e**) was performed (**Table 20**), with the same settings as for the previous tests.

Table 20: Summary of the *in vitro* photo-cross-linking experiment performed with the miR-124-3p-C6-N^4-PEG probes. Samples were irradiated for 15 min with the UV light λ=365 nm (reference samples were not irradiated). The concentration of each strand: 3 μM. **Y**: cytidine with the trioxsalen at the N^4-position of the nucleobase. PEG0-PEG4 indicates how many ethylene glycol units are present within the linker. The approximated cross-linking yield (calculated based on the consumption of the unmodified counter-strand, ORN-31) is given in brackets. Fc ctstr: fully complementary counter-strand. Masses of the isolated peaks are presented in section VII.3.5.3.

Duplex	ORN	Sequence	Cross-linked product
miR-124-3p-C6-N^4-PEG			
D42	ORN-53	5'-UAAGGCACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D53	ORN-56a (probe)	5'-UAAGG**Y**$_{PEG0}$ACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D54	ORN-56b (probe)	5'-UAAGG**Y**$_{PEG1}$ACGCGGUGAAUGCC-3'	not detected
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D55	ORN-56c (probe)	5'-UAAGG**Y**$_{PEG2}$ACGCGGUGAAUGCC-3'	detected (∼76%)
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D56	ORN-56d (probe)	5'-UAAGG**Y**$_{PEG3}$ACGCGGUGAAUGCC-3'	detected (∼34%)
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	
D57	ORN-56e (probe)	5'-UAAGG**Y**$_{PEG4}$ACGCGGUGAAUGCC-3'	detected (∼24%)
	ORN-31 (fc ctstr)	3'-AUUCCGUGCGCCACU-5'	

Pleasingly, a distinct and clean cross-linking was observed for duplexes D55, D56 and D57 (**Figure 44d** to **Figure 44f**). The cross-linking was the most efficient for the probe with a PEG2 linker (**Figure 44d**) and reached almost full conversion to the cross-linked duplex (∼76%). The yield of the cross-linking was calculated based on the consumption of the unmodified counter-strand (ORN-31) in the irradiated samples in comparison with the non-irradiated reference sample. The cross-linking efficiency gradually decreased with the increased length of the linker, achieving ∼34% and ∼24% for the oligoribonucleotides with PEG3 and PEG4 linkers, respectively.
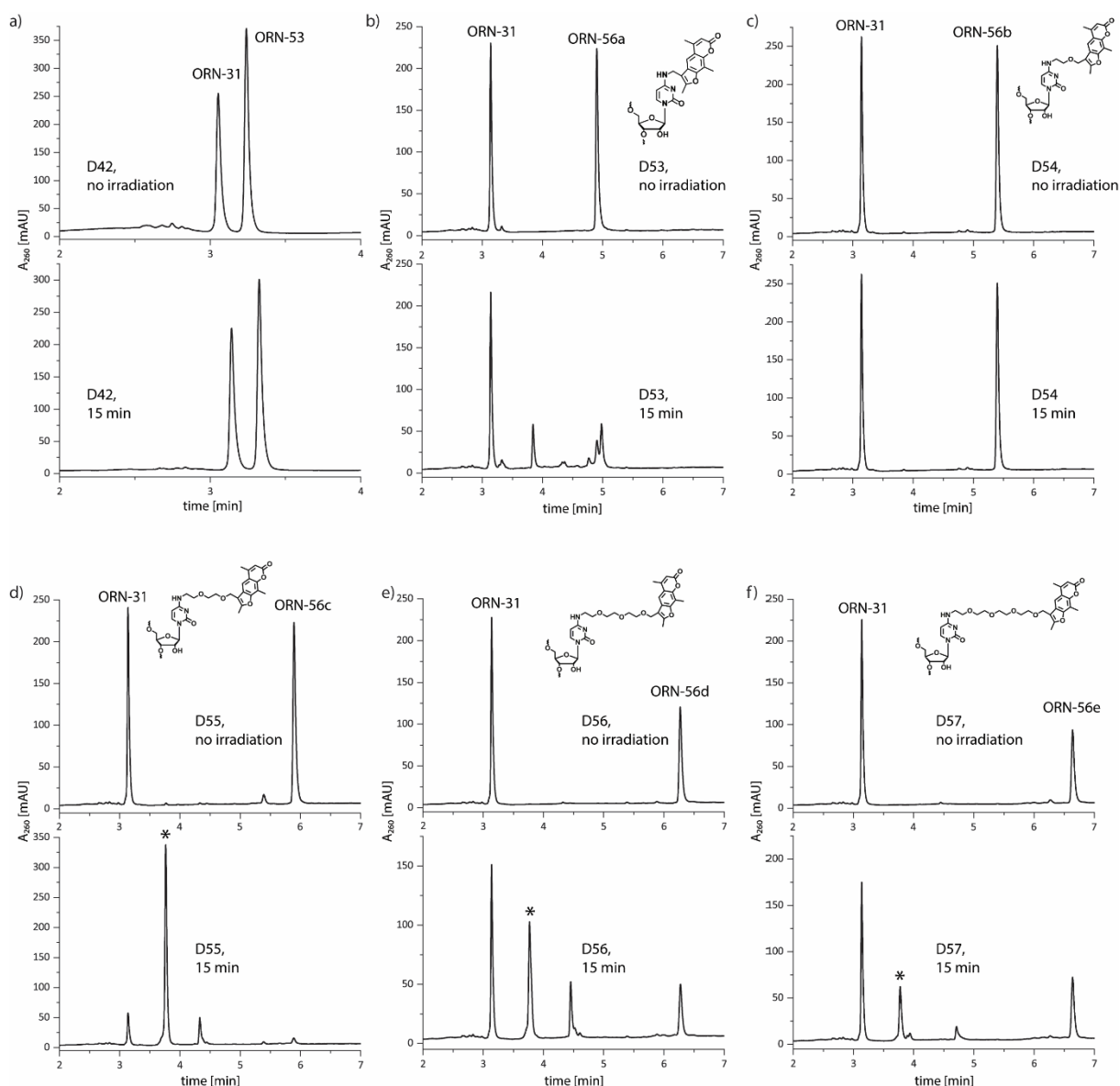
Figure 44: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with the miR-124-3p-C6-N^4-PEG probes with linkers of different lengths [0-4 ethylene glycol units, b) to f), respectively]; a) an unmodified duplex D42 serving as a negative control. Top panels present the non-irradiated samples (references). Bottom panels represent samples irradiated for 15 min (λ=365 nm). Peaks marked with an asterisk (*) correspond to the cross-linked duplexes; d) mass calc.: 11461.3, mass found: 11459.6; e) mass calc.: 11505.3, mass found: 11503.6; f) mass calc.: 11549.4, mass found: 11547.8.

A significantly improved cross-linking efficiency was observed for the probes with the cross-linker at the N^4-position when compared to the probes with the trioxsalen introduced at the 2'-*O*-position of the same cytidine (C6). This can be attributed to two factors. First, the shift of the modification from the minor to the major groove. Second, a higher flexibility of the linkers introduced *via* convertible nucleoside approach, as they do not contain a rigid triazole ring (which is the case when CuAAC chemistry is applied). Both of these factors can contribute to optimal positioning of the trioxsalen with respect to the duplex structure, thus facilitating intercalation and subsequent formation of the mono-adducts upon irradiation. In the investigated miR-124-3p sequence, uridines t_2 and t_3 were most likely engaged in the formation of the covalent bond with the trioxsalen (**Figure 45**). Interestingly, this corresponds to the 5'-UpU-3' site suggested as a preferred intercalation site in tests with the free

trioxsalen [253, 254]. Whether ORN-56c, ORN-56d and ORN-56e cross-link to the same or different uridines remains an open question. The digestion assay on the cross-linked product could be used to identify the exact target site (the concept of the digestion assay will be explained in section V.1).

fc ctstr (ORN-31):   3'-AUUCCGUGCGCCACU-5'

miR-124-3p (ORN-56):   5'-UAAGG**Y**ACGCGGUGAAUGCC-3'
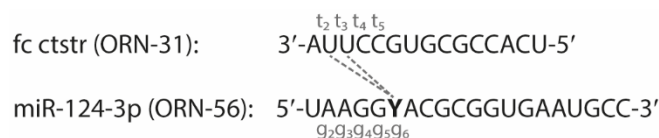
Figure 45: Sequence of the miR-124-3p analogue with the trioxsalen introduced at the C6-N^4-position *via* PEG linkers of different length (ORN-56) and its fully complementary counter-strand (fc ctstr; ORN-31). Putative cross-linking sites are indicated by dashed lines. **Y**: cytidine with the trioxsalen at the N^4-position of the nucleobase.

A long term goal would be to apply a stoichiometric mixture of probes differing by the length of the linker (here: number of the ethylene glycol units) to achieve a sequence-independent formation of the mono-adducts without focusing on a particular cross-linking site. For this purpose, one oligoribonucleotide sequence containing convertible nucleoside would be prepared and functionalized by displacement with trioxsalen analogues with various amino-linkers. Simultaneous use of a probe mixture could increase the probability of the cross-linking.

To verify if products of the cross-linking with such a mixture of probes could be separated and identified, equimolar (2 µM) amounts of ORN-56a to ORN-56e were mixed with the counter-strand (ORN-31, 10 µM) and tested in the *in vitro* photo-cross-linking assay. Chromatograms for both, non-irradiated and irradiated samples (**Figure 46**, the top and the bottom panel, respectively) showed a clean separation of the peaks. The deconvoluted mass of the main newly-formed peak (marked with an asterisk on the chromatogram) matched the expected masses of the cross-linked duplexes D55 and D56. Simultaneously, the intensity of the peak corresponding to ORN-56c significantly decreased. The intensities of ORN-56d and ORN-56e were moderately reduced, which reflected the lower efficiency of the cross-linking with these probes. The results were in agreement with the data obtained from the assay performed with single duplexes. This demonstrated that the application of the mixture of the probes to achieve the cross-linking is feasible.
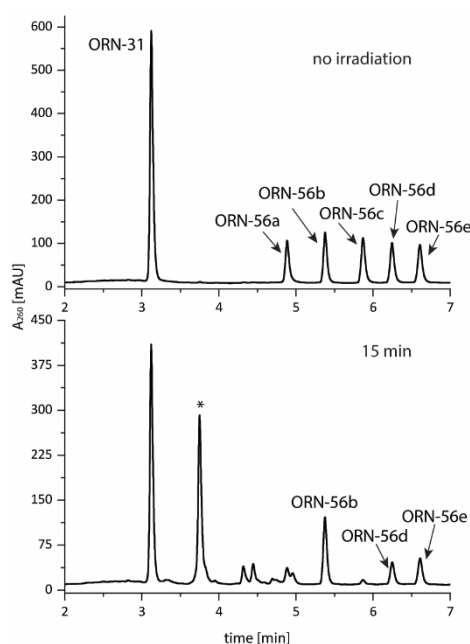
Figure 46: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with an equimolar mixture of the miR-124-3p-C6-N^4-PEG probes with linkers of different lengths. The top panel presents the non-irradiated sample (reference). The bottom panel represents sample irradiated for 15 min (λ=365 nm). Peak indicated with an asterisk (*) corresponds to the cross-linked duplexes; mass calc. [D55]: 11461.3, mass found: 11459.6; mass calc. [D56+K]: 11544.3, mass found: 11541.6. Masses of the isolated peaks are presented in section VII.3.5.4.

IV.3.4. Thermal stability studies

The UV melting profiles of the newly synthesized compounds were recorded and are plotted on **Graph 2**. A relatively high T_M (76.9 +/- 0.3°C) observed for the unmodified duplex D42 was attributed to the high content of the GC base pairs (60%) present within this sequence.
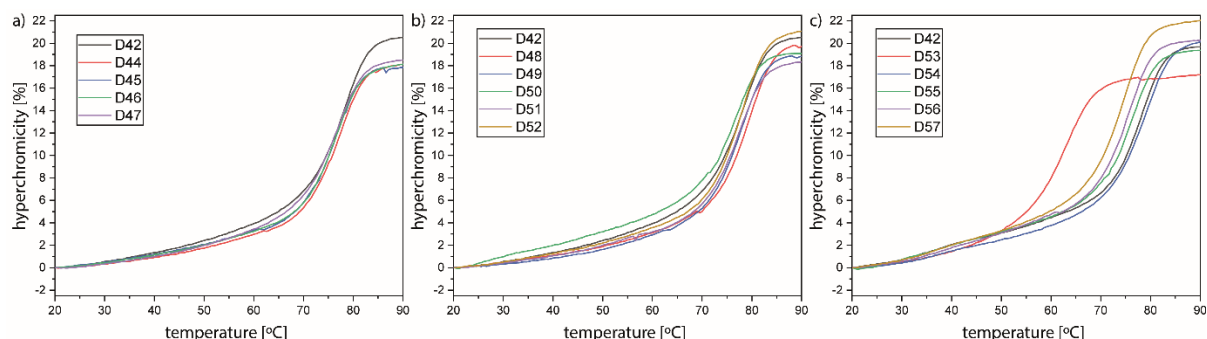
Graph 2: Melting profiles of the miR-124-3p analogues modified with the trioxsalen at C6-2'-*O*-position (a), C8-2'-*O*-position (b) or C6-N^4-position (c) of the cytidine through linkers of different lengths (number of ethylene glycol units: 0-4). Fully complementary counter-strand (ORN-31) was used as a second strand of the duplex. Conditions: 2 µM concentration of each strand of the duplex in 100 mM NaCl, 10 mM phosphate buffer and 0.1 mM Na_2EDTA, pH 7.0.

The calculated T_Ms for the miR-124-3p-C6-2'-*O*-prop-PEG analogues were similar (~75.7°C; D44-D47; **Table 21**) and slightly reduced when compared to the unmodified duplex D42. The T_Ms for duplexes containing miR-124-3p-C8-2'-*O*-prop-PEG and miR-124-3p-C6-N^4-PEG analogues were more heterogeneous. The highest T_Ms were observed for D48 (78.3 +/- 0.2°C) and D54 (78.1 +/- 0.3°C). The lowest T_Ms were in general observed for the analogues in which the modification was appended at the N^4 position of the nucleobase. Nevertheless, the destabilization was still rather mild (up to -3.6°C for PEG4 linker; D57) and in agreement with findings from the study on mono-substituted N^4-propargyl cytidines [320].

Table 21: Melting temperatures calculated for the investigated duplexes.

Duplex	Modification	T_M [°C]	ΔT_M [°C]
D42	no modification	76.9 ± 0.3	-
D44	C6-2'-*O*-prop-PEG1	75.8 ± 0.3	-1.1
D45	C6-2'-*O*-prop-PEG2	75.6 ± 0.4	-1.4
D46	C6-2'-*O*-prop-PEG3	75.5 ± 0.2	-1.5
D47	C6-2'-*O*-prop-PEG4	75.7 ± 0.1	-1.2
D48	C8-2'-*O*-prop-PEG0	78.3 ± 0.2	1.4
D49	C8-2'-*O*-prop-PEG1	77.2 ± 0.6	0.2
D50	C8-2'-*O*-prop-PEG2	75.9 ± 0.6	-1.0
D51	C8-2'-*O*-prop-PEG3	76.5 ± 0.1	-0.5
D52	C8-2'-*O*-prop-PEG4	76.9 ± 0.1	0.0
D53	C6-N^4-PEG0	62.1 ± 0.3	-14.8
D54	C6-N^4-PEG1	78.1 ± 0.3	1.2
D55	C6-N^4-PEG2	75.6 ± 0.3	-1.3
D56	C6-N^4-PEG3	74.5 ± 0.2	-2.4
D57	C6-N^4-PEG4	73.3 ± 0.2	-3.6

For duplex D53 with the shortest linker (PEG0), the calculated T_M (62.1 +/- 0.3°C) was lower than the T_M of the unmodified duplex by almost 15°C. Such a significant drop in the T_M value implied that the base-pairing at this position was abolished [320]. This could be explained by the nature of the PEG0 linker. Most likely, the linker was too short to position the trioxsalen so that it would not interfere with the duplex structure. This suggested that PEG0 linker should be excluded from a pool of linkers attached to the N^4 position of cytidine.

It is difficult to understand why the T_Ms for oligoribonucleotides from ORN-54 series were almost equally reduced while the same modifications incorporated two bases downstream within the sequence (ORN-55) influenced the T_M values to a different extent. Interestingly, the observed decrease in the T_M values might have been caused by the actual intercalation of the trioxsalen into the duplex structure. This would explain why for all of the duplexes for which the cross-linking was detected the calculated T_Ms were reduced. However, in order to further validate these results, the T_M measurements should be repeated.

IV.3.5. Preparation of the pre-synthetically substituted phosphoramidite

Simultaneously, efforts to pre-synthetically displace the triazole moiety with a trioxsalen carrying an amino-linker were undertaken. Such phosphoramidites could be useful when the oligonucleotide needs to be functionalized with more than one or two modifications. The synthesis, presented in **Scheme 7**, started with the standard protection of the 5'- and 2'-hydroxyl groups of the ribose. The TBDMS-protection step yielded the desired 2'-*O*-protected isomer **24** with moderate yield, due to the concomitant formation of the 3'-isomer. Afterwards, the 3'-hydroxyl group was acetylated, to prevent the side-reactions of this group during the next step. Subsequently, reaction with POCl$_3$ and triazole gave the protected O^4-triazolyluridine (**26**).
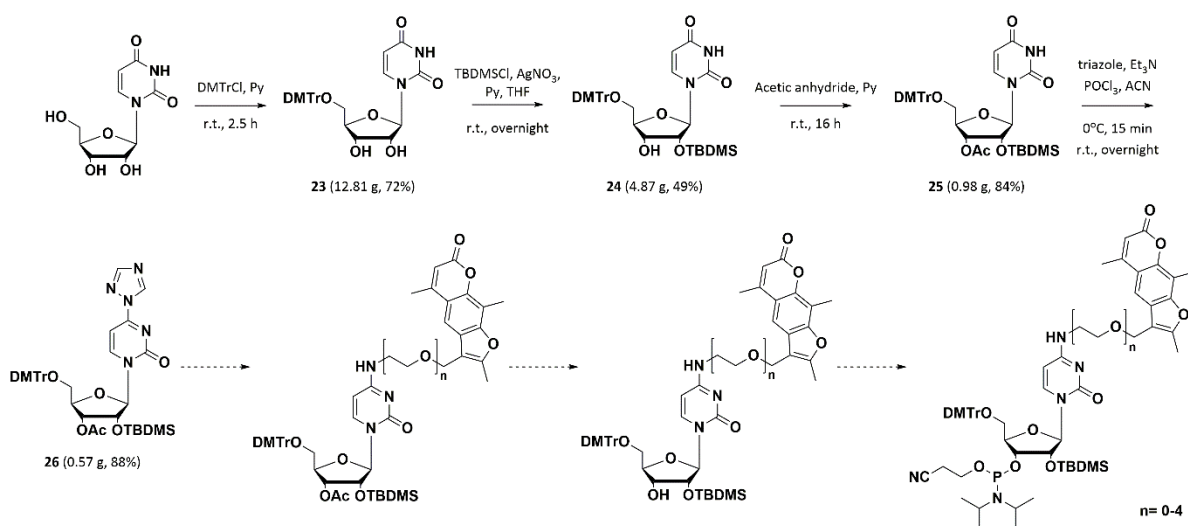
Scheme 7: General scheme of the synthesis of the fully protected cytidine phosphoramidite modified at the N^4 position with the trioxsalen attached through the PEG-based linkers.

The displacement reaction was performed adopting the published procedure [334]. A mixture of the O^4-triazolyluridine (**26**), 4'-aminomethyltrioxsalen (**22**) and triethylamine in dry acetonitrile was stirred overnight at room temperature. As no conversion was observed, DBU was added to the reaction but this did not result in the formation of the product. Due to strategic reasons, the synthesis was stopped at this point. For the future, it would be reasonable to run the reaction with the same conditions that

were applied for the post-synthetic displacement, so with DBU and at a higher temperature. Additionally, other trioxsalen analogues should be tested as nucleophiles, as lack of the substitution may be related to the lower reactivity of **22**, caused by the steric hindrance.

IV.4. Conclusions and outlook

The main objective of the study described in this chapter was to prepare and test the new type of photo-reactive miRNA probes prepared with the use of the convertible nucleoside approach. This required the synthesis of trioxsalen analogues bearing the amino-PEG linkers of various lengths. The synthesis was performed following a published procedure and provided the final products with moderate to good yields. For the future, it would be beneficial to further optimize the reaction conditions and purification after the pegylation step, especially when the longer linkers are introduced. The post-synthetic displacement with trioxsalen analogues generated the desired substitution products. To make the convertible nucleoside approach more economical, it would be worth to investigate whether reaction side products could be avoided. Moreover, direct substitution with a mixture of the amines could be performed (first trial with such simultaneous substitution reaction to obtain the multi-functionalized probe is currently on-going). This would avoid a separate preparation of each probe. It would be also interesting to verify (from a chemistry point of view) what would be the ratio of the generated substitution products when the solid support would be treated with an equimolar mixture of amines. However, based on the previous observations, the purification of such a mixture is expected to be challenging.

The results from the *in vitro* photo-cross-linking experiment with the miR-124-3p probes with the trioxsalen attached at the N^4 position of the nucleobase were promising. The fact that the formation of the inter-strand covalent bond was observed for three from five tested linkers suggests that use of a probe mixture will increase the chances to achieve the cross-linking in a sequence-independent fashion. The cross-linking efficiencies observed for the new probes were in all cases superior to the cross-linking yields observed with the miR-124-3p-1 and miR-124-3p-2 probes described in section II.3.3.1. On the other hand, no cross-linking was detected for the 2'-*O*-modified probes with PEG0 linker, corresponding to the modification used in the miR-124-3p-1 and miR-124-3p-2 probes. This highlighted once again the sequence-dependency of the cross-linking with the trioxsalen. Although this study was based on the application of the substituents with 0-4 ethylene glycol units, the preparation of the trioxsalen analogues with even longer linkers would be interesting. The use of such analogues could increase the chance for cross-linking with uridines located in more distant regions of the target sequence.

These preliminary experiments imply that the convertible nucleoside approach may be particularly useful for the preparation of the cross-linking probes as it does not result in the generation of any additional rigid scaffolds between the reacting moieties. The high flexibility of the linker seems to enhance the favourable positioning of the trioxsalen, resulting in the more efficient cross-linking. However, further characterization of these new compounds is required, especially in terms of the influence of the modifications on the stability and integrity of the duplex. To assess the general applicability of the described approach, the N^4-modified probes for other miRNAs should be synthesized and examined in the *in vitro* photo-cross-linking assay to verify if the efficient cross-linking can be achieved also for these sequences. Ideally, the most efficient cross-linking would be observed

for probes with linker lengths different that in case of ORN-56. The preparation of the probes for other miRNAs is currently on-going.

Additionally, the bis-labelled pre-miR-CLIP probes bearing the trioxsalen and the biotin should be prepared and tested in cells. This would demonstrate if the new probe design is tolerated by cells and whether the pre-miR-CLIP probe can be processed and incorporated into RISC to fulfil its biological function. Alternatively, proper processing can be also evaluated in the *in vitro* Dicer assay [335].

IV.5. Contributions

2'-*O*-propargyl cytidine phosphoramidite and 4'-azidomethyltrioxsalen were taken from the common laboratory stock synthesized by Martina Bigatti as a part of the NCCR platform. The O^4-triazolyluridine phosphoramidite was synthesized by Gwendolin Lieb. The synthesis of the trioxsalen analogues with different linkers as well as the synthesis of all of the oligoribonucleotides used in this study was performed by Anna L. Malinowska. The *in vitro* photo-cross-linking experiments and melting studies were designed, performed and analyzed by Anna L. Malinowska.

V. Chapter V: Side projects and collaborations

V.1. Elucidation of the cross-linking site for the miR-106a-5p probe – pilot studies

The cross-linking experiments with the miR-106a-5p analogues were the first important step in extending the knowledge about the behaviour of the probes with the trioxsalen attached at the 2'-O-position of the ribose by the CuAAC chemistry. However, further validation of the postulated hypothesis regarding the reactivity of the probes would require precise identification of the cross-linking site. It was envisioned that ^{13}C-labelled oligoribonucleotides could be used for this purpose.

The assay workflow would see the original miR-106a-5p probe (ORN-1) subjected to the irradiation with four fully complementary counter-strands. In each counter-strand, one out of four consecutive uridines at the positions t_1-t_4 would be replaced with the uridine phosphoramidite in which the carbon atoms of the ribose would be ^{13}C-labelled (**Figure 47a**). The isolated cross-linked products would be then digested to obtain the dimer structure, corresponding to the two nucleotides covalently bond by the trioxsalen. Digestion products would be analysed by the mass spectrometry. The cross-linking to the ^{13}C-labelled uridine would result in the detection of the product with a mass higher by 5 units (corresponding to the five ^{13}C atoms instead of five ^{12}C atoms in the ribose, **Figure 47b**) than in case of the cross-linking with unmodified uridine. Based on the knowledge which uridine was labelled in which counter-strand, the cross-linking site could be identified.
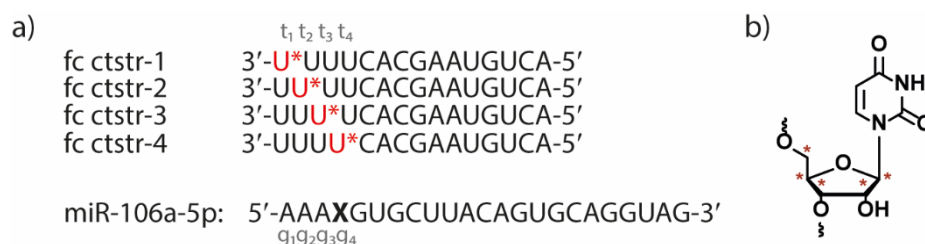
Figure 47: a) A schematic representation of the cross-linking with the miR-106a-5p probe for the identification of the cross-linking site. The cross-linking experiment would be performed with four fully complementary counter-strands (fc ctstr) having ^{13}C-labelled uridine (marked in red) inserted systematically at positions t_1 to t_4; b) the structure of the modified uridine. ^{13}C-labelled atoms are indicated by the red stars.

The presented strategy would require optimization on many different steps. Initial efforts focused on the elucidation of the digestion conditions. Ideally, the digestion process should allow to produce the intact cross-linked dimer, without concomitant depurination or degradation of the trioxsalen moiety. Such side reactions could complicate the identification of the product based on its mass. As the presence of the trioxsalen at the 2'-O-position of the adenosine may abolish the cleavage of the phosphate bond between the modified adenosine and neighbouring guanosine, the digestion product in a form of a trimer would be also expected.

To meet these requirements, basic hydrolysis conditions were chosen to cleave (digest) the oligoribonucleotide [336]. For the screening of the conditions, the unmodified, fully complementary miR-106a-5p counter-strand (ORN-2) was used. Calculated volumes of ORN-2 were transferred to screw cap tubes and concentrated. Then, 20 μL of 25% ammonia, 0.25 M NaOH or 0.5 M NaOH solutions were added and samples were subjected to the basic hydrolysis under various conditions, summarized in **Table 22**.

The main benefit of using ammonia solution was that samples could be concentrated, re-dissolved in water and directly subjected to analysis. The use of NaOH solutions required quenching of the reaction with acid (HCl aqueous solution) and subsequent solid-phase extraction (performed with Strata-X Polymeric Reversed-Phase Columns) to remove the excess salts. The extraction process resulted in an overall more demanding protocol, especially when many samples were analysed simultaneously. Moreover, the additional purification step increased the risk of losing a fraction of the material during the elution. In contrary, upon treatment with ammonia, the sample does not need any transfer to other tubes.

Table 22: Summary of the conditions tested for the base-catalysed digestion of ORN-2.

Entry	Reagent	Temperature	Time
1		room temperature	16 h
2		50°C	4 h
3		50°C	17 h
4		60°C	4 h
5	25% ammonia solution	60°C	6 h
6		60°C	17 h
7		70°C	4 h
8		70°C	6 h
9		70°C	17 h
10		room temperature	1 h
11	0.25 M NaOH solution	room temperature	4 h
12		room temperature	17 h
13		room temperature	17 h
14	0.5 M NaOH solution	50°C	1 h
15		90°C	20 min

The obtained samples were dissolved in ultrapure water and injected on the ion-exchange (IE)-HPLC to assess the level of digestion. Adenosine 3'-monophosphate (Ap), guanosine 3'-monophosphate (Gp) and AGp dimer were injected as references (**Figure 48a**). Hydrolysis with NaOH solutions (both, 0.25 M and 0.5 M) ensured complete digestion of the oligoribonucleotide, as only the peaks corresponding to the mononucleotides were detected (**Figure 48b** and **Figure 48c**). Nevertheless, it is possible that the full-length product was retained on the column during de-salting. In the case of the samples treated with 25% ammonia solution, different levels of digestion were observed, depending on the conditions applied (**Figure 48d**, **Figure 48e** and **Figure 48f**). This gave more information on how to adjust the hydrolysis conditions. Further use of NaOH would require the application of less concentrated solutions. However, the purification step would be still necessary. Based on this, it was proposed that the digestion with 25% ammonia solution at 60°C for 6 h would be the optimal conditions to perform the basic hydrolysis of the oligoribonucleotide.

To assess if the products of the digestion can be detected and identified, the digested samples were injected on the LC-MS. Although some peaks were present in the UV spectrum, it was not possible to determine their mass value. The peaks had the retention times below 1 min which suggested that they did not have retention on the column. Column dedicated for the separation of the polar, aromatic compounds (Synergi™ 4 μm Polar-RP 80 Å, Phenomenex) and various gradients were tested but it did not allow to detect the products (data not shown). It is possible that the mass was not detected due

to the quadrupole detection limit of the LC-MS apparatus. Subsequently, the trioxsalen-modified miR-106a-5p probe (ORN-1) was subjected to digestion to verify if the mononucleotide/dinucleotide with the trioxsalen attached (and therefore having a higher mass) could be detected but it was also not effective.
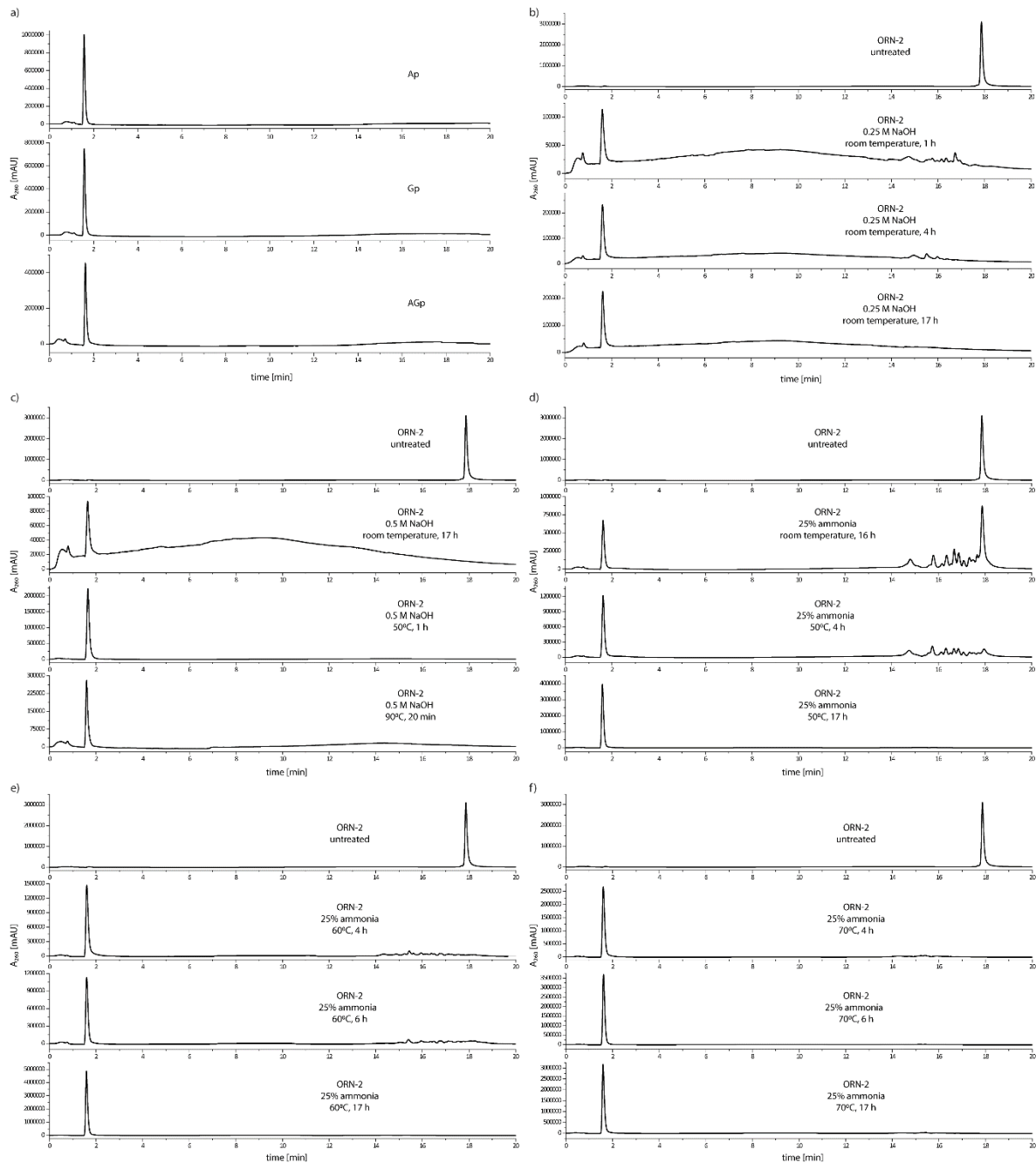
Figure 48: IE-HPLC chromatograms from the screening of the conditions for the digestion assay; a) reference chromatograms of Ap (adenosine 3'-monophosphate), Gp (guanosine 3'-monophosphate) and AGp dimer; b) to f): digestion of ORN-2 with 0.25 M (b) and 0.5 M (c) NaOH solutions or 25% ammonia solution (d-f) in various conditions.

Screening of the gradient conditions with single 5'-monophosphates showed that when quite high concentrations of the monophosphates were used (µM-mM), the masses of the homo-dimers and homo-trimers were detected. This suggested that the cross-linking experiment would most likely need

to be performed on a much higher scale in order to detect the cross-linked dimer/trimer after digestion. Further tests are necessary to explore these findings. Alternatively, a more sensitive apparatus can be applied for mass detection. It also needs to be confirmed if the postulated digestion conditions do not degrade the trioxsalen moiety.

Contributions: AGp dimer was synthesized by Anna Knörlein. Digestion assay was designed by Anna L. Malinowska and performed by Jonas Römer and Anna L. Malinowska. Gradient for the IE-HPLC analysis was adopted from the other assay established by Anna Knörlein. LC-MS tests were performed by Anna L. Malinowska. Monophosphates were provided by Anna Knörlein.

V.2. Synthesis of the miR-CLIP probes

Table 23: Summary of the synthesized pre-miRNAs, miR-CLIP probes, mono-labelled probes and their counter-strands. Trioxsalen was incorporated in a form of the modified adenosine phosphoramidite (**X**; fully protected adenosine phosphoramidite with the trioxsalen attached pre-synthetically at the 2'-*O*-position *via* CuAAC reaction). Biotin was introduced post-synthetically (CuAAC reaction on the solid support) at the 2'-*O*-position of cytidine (**V**), uridine (**Y**), adenosine (**W**) or incorporated as a biotin dT phosphoramidite (**Z**) during the oligoribonucleotide synthesis.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found
miR-124-3p- and miR-132-3p-related probes – a collaboration with Dr Y. Wang					
1	ORN-57	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCAUA CAAUUAAGGCACGCGGUGAAUGCC	59	18903.4	18901.5
2	ORN-58	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCAUA CAAUUAAGG**V**ACGCGGUG**X**AUGCC	59	19707.4	19706.1
3	ORN-59	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCAUA CAAUUA**X**GGCACG**V**GGUGAAUGCC	59	19707.4	19707.4
4	ORN-60	ACCGUGGCUUUCGAUUGUUACUGUGGGAACUGGA GGUAAC**X**GUCUACAGC**V**AUGGUCG	58	19467.2	19464.5
miR-155-related probes – a collaboration with A. Łaski					
5	ORN-61	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUCCAA CUGACUCCUACAUAUUAGCAUUAACA	61	19383.6	19382.9
6	ORN-62	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUCCAA CUGACUCCUAC**X**UAUUAGCAU**Y**AACA	61	20187.6	20186.3
7	ORN-63	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUCCAA CUGACUCCUAC**X**UAUUAGCAU**Z**AACA	61	20083.5	20082.5
8	ORN-64	UUA**X**UGCUAAUCGUGAUAGGGGUU	24	8017.1	8016.1
9	ORN-65	UGCACGCAGCAUUAA	15	4759.0	4758.3
10	ORN-66	CUCCUAC**X**UAUUAGCAUUAACA	22	7212.6	7211.7
11	ORN-67	GCUAAUAUGUAGGAG	15	4840.0	4839.3
miR-106a-5p- and miR-106b-5p-related probes – a collaboration with Dr M. Lucic					
12	ORN-68	AAA**X**GUGCUUACAGUGC**W**GGUAGCUUUUUGAGAU CUACUGCAAUGUAAGCACUUCUUAC	59	19655.3	19654.0
13	ORN-69	AAA**X**GUGCUUACAGUGC**W**GGUAG	23	8238.6	8237.4
14	ORN-70	UAA**X**GUGCUGACAGUGC**W**GAUAGUGGUCCUCUCC GUGCUACCGCACUGUGGGUACUUGCUGC	62	20664.9	20663.7
15	ORN-71	UAA**X**GUGCUGACAGUGC**W**GAU	21	7564.2	7563.1
16	ORN-72	CUGCACUGUCAGCACUUUAUG	22	6610.0	6608.8
17	ORN-73	CCGCACUGUGGGUACUUGCUG	21	6681.0	6680.0
18	ORN-74	ACCUGCACUGUAAGCACUUUUAC	23	7228.4	7227.4
19	ORN-75	CUGCAAUGUAAGCACUUCUUAC	22	6923.2	6922.2

As part of a collaboration within the laboratory, the miR-CLIP probes (in the form of both, the pre-miRNA and miRNA duplex) were prepared (**Table 23**).

V.2.1. miR-124-3p- and miR-132-3p-related probes

To further explore the application of the miR-CLIP technique, the pre-miR-CLIP probes for the miR-124-3p and miR-132-3p were synthesized and used in the miR-CLIP experiment performed by Dr Yuluan Wang [287]. The novelty of these probes, prepared according to the original design [84], was related to the fact that the approach was extended for the 3p miRNAs for the first time.

To fulfil its function, the pre-miR-CLIP probe needs to be first recognized and cleaved by the Dicer-TRBP complex and then incorporated into the RISC (**Figure 49**). It was shown previously that cellular RISC machinery is sensitive and do not tolerate modifications at certain positions within the probes [84, 337, 338]. The proper processing is particularly important in the case of 3p miRNAs, as the Dicer-mediated cleavage step can result in the formation of the 5'-isomiRs targeting different genes. Therefore, the assessment of the functionality of the new probes is very important. For this purpose, a luciferase assy is often employed.
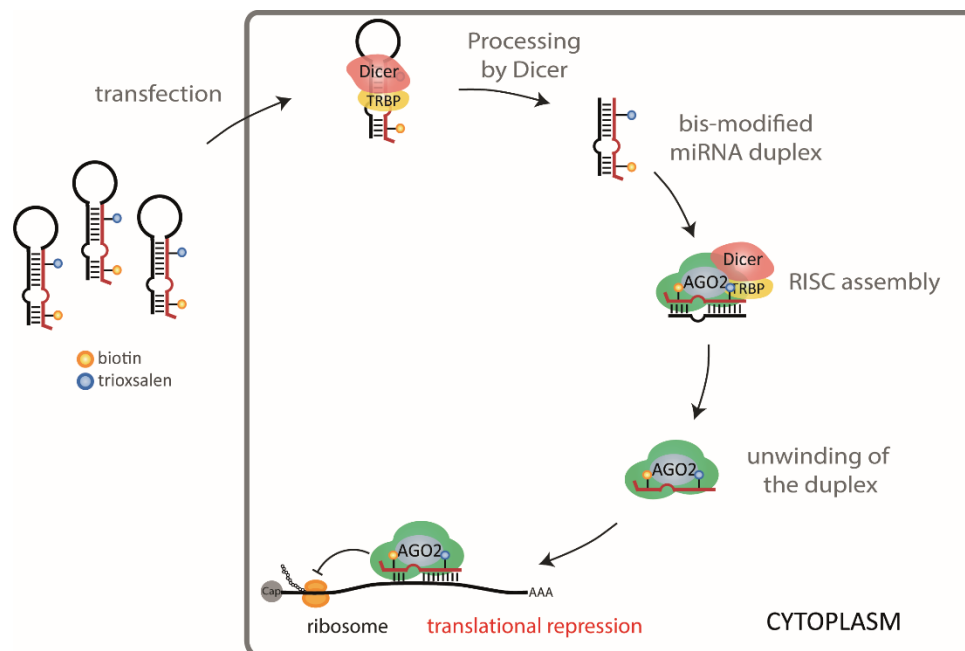
Figure 49: The processing and RISC loading of the pre-miR-CLIP probes in cells necessary for their biological activity.

In the luciferase assay, the putative binding site for the miRNA of interest is introduced into the regulatory elements (usually 3'-UTR) of the reporter gene such as *Renilla* luciferase gene [339, 340]. Upon transfection with the miRNA, the luminescent signal of the reporter protein is measured. When binding to its target site, miRNA interferes with the translation process which is reflected in a decrease in the generated signal. In a dual luciferase assay, the firefly luciferase reporter is used simultaneously as an intraplasmid standardization control for the reliable quantification of the silencing effect. Significantly reduced activity of the bis-labelled probe in comparison with the positive controls upon targeting of a fully complementary binding site would suggest that the modifications interfere with the formation of the functional RISC and/or proper processing of the probe.

Initially, four pre-miR-CLIP probes for pre-miR-124 and two pre-miR-CLIP probes for pre-miR-132 were synthesized (synthesis performed by Dr Ugo Pradère). As mentioned in section II.3.3, the probes for the miR-132-3p had the trioxsalen attached within the seed region while probes for the miR-124-3p bore the cross-linker either within the seed or in the 3'-region of the miRNA. Verification whether the positioning of the cross-linker has an influence on the captured targetome would contribute to the characterization of the probes. Based on the activity of the probes in biological assays (dual luciferase assay, reverse transcription quantitative polymerase chain reaction (RT-qPCR) and Western Blot) compared to the activity of the unmodified synthetic pre-miRNAs, two pre-miR-CLIP miR-124 (ORN-58 and ORN-59) and one pre-miR-CLIP miR-132 (ORN-60) probes were selected for further use in the miR-CLIP assay [287].

The miR-CLIP experiment performed by Dr Wang allowed to identify the predicted and not reported targets of miR-124-3p and miR-132-3p. Both pre-miR-CLIP miR-124 probes were functional and captured partially overlapping targets. Surprisingly, the probe with the 'non-canonical' design (ORN-58; trioxsalen introduced in the 3'-region of the miRNA) was able to capture 9% more targets with the predicted binding sites than the 'canonical' probe (ORN-59) [287]. This appeared to be in accordance with the results from the *in vitro* cross-linking assay with the trioxsalen mono-labelled probes. The cross-linking efficiencies with the ORN-29 (corresponding to ORN-58) seemed to be higher than with the ORN-30 (matching ORN-59), at least when the counter-strands mimicking the native targets were used (described in section II.3.3.1). Moreover, the results from the miR-CLIP experiment with the pre-miR-CLIP probes for miR-124 confirmed that the processing of the probes yielded two 5'-isomiRs of the mature miR-124, the 'canonical' miR-124-3p and *iso*-miR-124-3p (with additional uridine), having non-identical targetomes.

Contributions: the synthesis of the pre-miR-CLIP probes was initially performed by Dr Ugo Pradère and took over by Anna L. Malinowska as a contribution to the miR-CLIP project on miR-124-3p and miR-132-3p designed and led by Dr Yuluan Wang.

This collaboration resulted in a manuscript which is currently in the submission process:
Wang, Y.; Soneson, C.; Malinowska, A. L.; Laski, A.; Ghosh, S.; Kanitz, A.; Gebert, L. F.; Robinson, M. D.; Hall, J.; 'MiR-CLIP reveals *iso*-miR selective regulation in the miR-124 targetome'.

V.2.2. miR-155-related probes

Following the hypothesis generated based on the results from the *in vitro* cross-linking with the 2'-*O*-modified mono-labelled probes (described in section II.4) would require a substantial change in the design of the bis-labelled pre-miR-CLIP probes. So far, adenosine phosphoramidite bearing a pre-CLICKed trioxsalen at the 2'-*O*-position was introduced into the oligoribonucleotide sequence during the solid-phase synthesis. Simultaneous incorporation of the phosphoramidite (adenosine or cytidine) bearing a 2'-*O*-propargyl functionality enabled post-synthetic functionalization with a biotin at the selected position within the sequence, yielding the bis-modified probe (**Figure 50a**). Inserting of the trioxsalen at the position two base pairs downstream from the possible cross-linking site (outlined in **Figure 23** in section II.3.4) would mean a need to append the trioxsalen to different bases, not only adenosine. This, in turn, would demand a synthesis of all four trioxsalen-bearing fully protected phosphoramidites. Alternatively, the functionalization strategy could be reversed. The cross-linker could be CLICKed post-synthetically, and the biotin moiety could be introduced as a phosphoramidite. Such an approach was tested.
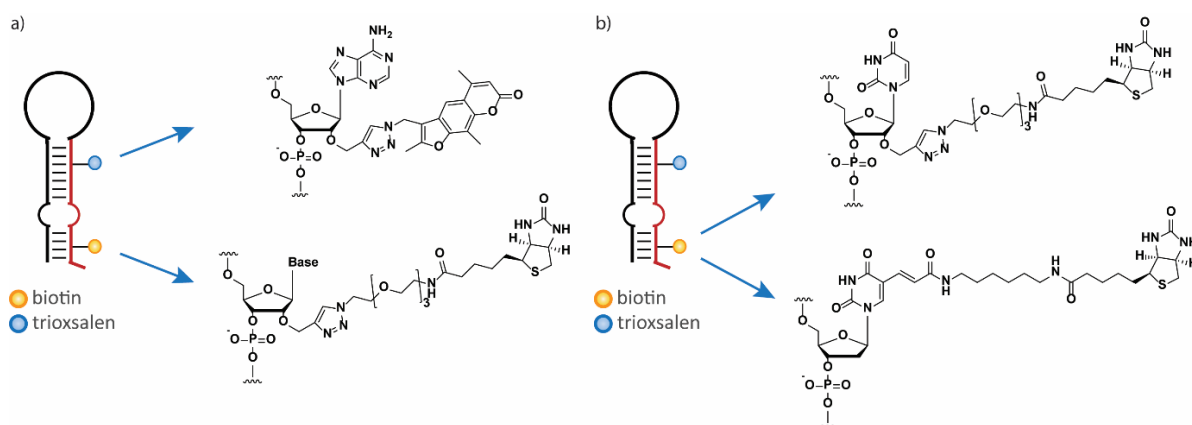
Figure 50: Schematic representation of the pre-miR-CLIP probes and their modifications; a) a 'classical' bis-labelled pre-miR-CLIP probe; b) the pre-miR-CLIP probe for the miR-155-3p with the biotin either CLICKed to the 2'-*O*-propargyl group of uridine (ORN-62) or introduced in a form of the biotin dT phosphoramidite (ORN-63) at the same position within the sequence.

Commercially available biotin deoxythymidine (dT) phosphoramidite (ChemGenes) was applied as a carrier of the biotin moiety, in which the modification was attached at the C^5 position of the nucleobase *via* relatively long and flexible linker. Although the use of this type of spacer should locate the modification away from the base, it might potentially interfere with the processing and loading of the probe. Moreover, the fact that this phosphoramidite was a deoxynucleotide could have an additional influence on the properties of the probe. To verify this, two analogous pre-miR-CLIP probes for miR-155-3p were synthesized. Both oligoribonucleotides had a trioxsalen at the 2'-*O*-position of adenosine (g_8 in the mature miRNA) and the biotin either CLICKed to the 2'-*O*-propargyl group of uridine (ORN-62) or introduced in a form of the biotin dT phosphoramidite (ORN-63) at the same position corresponding to g_{18} in the mature miRNA (**Figure 50b**).

To assess the functionality of the obtained probes, they were subsequently tested in a luciferase assay in HEK293T cells performed by Artur Łaski (**Figure 51**).
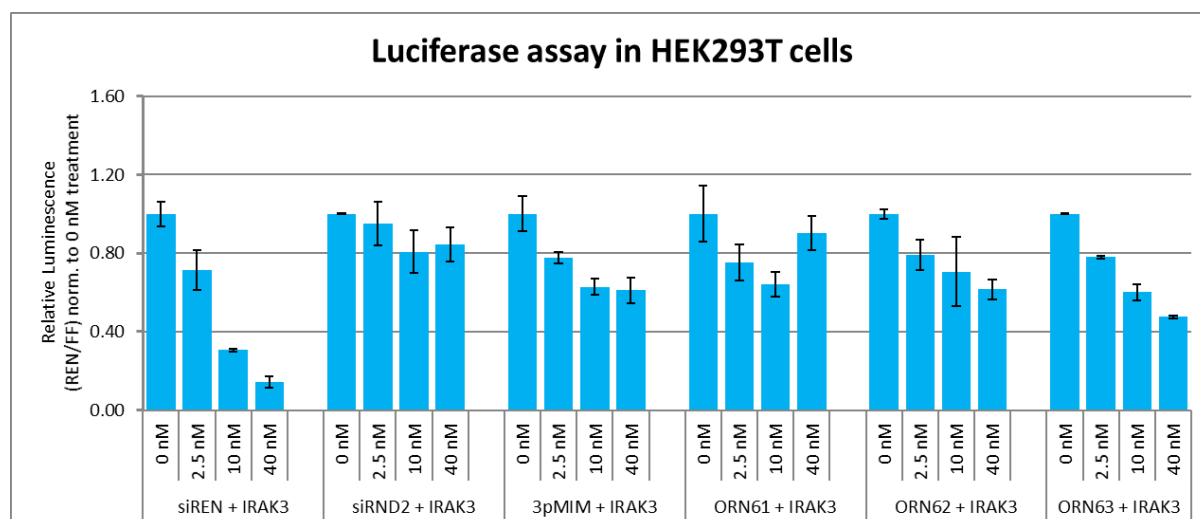
Figure 51: Regulation of the luciferase reporters by the pre-miR-CLIP probes for the miR-155-3p. HEK293T cells were co-transfected with the bis-modified pre-miR-CLIP probes for miR-155-3p or controls and the luciferase plasmid containing IRAK3 as a target site. Increasing concentrations of the indicated compounds were transfected (0 nM, 2.5 nM, 10 nM, 40 nM). N = 3. Error bars indicate standard deviation. siRND2 was used as a negative control. The assay was performed by Artur Łaski.

A binding site corresponding to the IRAK3 (Interleukin 1 Receptor Associated Kinase 3) gene (reported target site of the miR-155-3p [341]) was cloned into the 3'-UTR of the reporter (Promega's Dual-Glo assay vector psiCHECK-2™). siRenilla (siREN; directed against the coding region of the *Renilla* luciferase), miR-155-3p mimic (3pMIM; miRNA duplex ordered from QIAGEN) and unmodified pre-miR-155 (ORN-61) were used as positive controls, whereas siRND2 was used as a negative control.

Pleasingly, both of the probes caused downregulation of the target to a similar extent as the positive controls and in a dose-dependent manner (doses 0-40 nM). Surprisingly, the treatment with the new probe with dT biotin (ORN-63) seemed to result in higher repression of the target than in the case of the other probe (ORN-62). In contrary to the previous observations [287], unmodified pre-miR-155 was slightly less active than the miRNA mimic. This gave a first signal that the probes with the biotin introduced as the biotin dT phosphoramidite are tolerated by cellular processing and RISC machinery. However, both probes should be tested in other biological assays to exclude the possibility that the observed effect was false-positive.

Table 24: Summary of the *in vitro* photo-cross-linking experiments performed with the miR-155 mono-labelled probes. Samples were irradiated for 15 min with λ=365 nm UV light (reference samples were not irradiated). The concentration of each strand: 1 µM (D59) or 2 µM (D58). **X** designates adenosine with the trioxsalen attached at the 2'-*O*-position of the ribose. Underlined: mismatches and wobble base pairs. Fc ctstr: fully complementary counter-strand. Masses of the isolated peaks are presented in section VII.4.6.

Duplex	ORN	Sequence	Cross-linked product
		miR-155-5p and miR-155-3p	
D58	ORN-64 (probe)	5'-UUA**X**UGCUA<u>AAUC</u>GU<u>GA</u>UAGGGGUU-3'	not detected
	ORN-65 (SOCS1)	3'-AAUUACGA<u>C</u>G<u>C</u>ACGU-5'	
D59	ORN-66 (probe)	5'-CUCCUAC**X**UAUUAGCAUUAACA-3'	not detected
	ORN-67 (fc ctstr)	3'-GAGGAUGUAUAAUCG-5'	
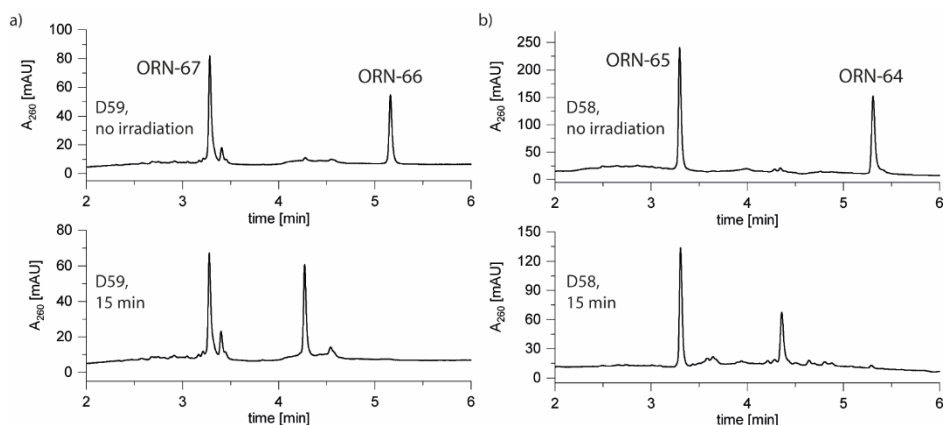
Figure 52: RP-HPLC chromatograms from the *in vitro* photo-cross-linking experiment with a) miR-155-3p (D59) and b) miR-155-5p (D58) probes. Top panels present non-irradiated samples (reference). Bottom panels represent samples irradiated for 15 min (λ=365 nm). No cross-linked product was detected.

To verify the cross-linking abilities of the new probe, its mono-labelled version (ORN-66) was prepared. A fully complementary counter-strand (ORN-67) was employed to create a duplex (**Table 24**). Despite the presence of uridine in the target strand two base pairs upstream from the modification site, no cross-linking after irradiation of the duplex for 15 min was observed (**Figure 52a** and **Figure 53**). The detected mass of the newly formed peak corresponded to the possible product of the intra-strand

cross-linking. The analysis of the sample containing an irradiated duplex by a 20% denaturing polyacrylamide gel electrophoresis showed the presence of a new band with mobility lower than the mobility of the single-strands but higher than the mobility of the RNA 40-mer used as a reference (data not shown). However, the identity of this band was not identified.

<div align="center">

fc ctstr (ORN-67):   3'-GAGGA$\overset{t_6}{U}$GUAUAAUCG-5'

miR-155-3p (ORN-66):   5'-CUCCUAC$\underset{g_8}{\textbf{X}}$UAUUAGCAUUAACA-3'

</div>
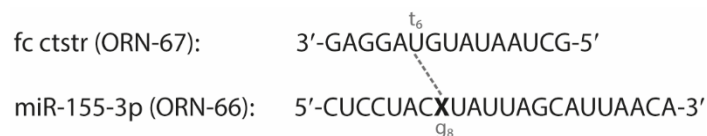
Figure 53: Sequence of the trioxsalen-modified miR-155-3p (ORN-66) and its fully complementary counter-strand (fc ctstr; ORN-67) used in the *in vitro* photo-cross-linking study. The putative cross-linking site is indicated by the dashed line. **X**: adenosine with trioxsalen at the 2'-*O*-position of the ribose.

The mono-labelled probe for miR-155-5p was also synthesized (ORN-64). In this case, the trioxsalen was placed at the g_4 position, following the design reported by Imig et al. [84]. Oligoribonucleotide with a sequence mimicking a reported target of miR-155-5p, SOCS1 (Suppressor of cytokine signalling 1) [342], was utilized as a counter-strand (ORN-65). Nevertheless, also for this sample no cross-linked product was detected (**Figure 52b**). This might be caused by a large number of mismatches present in this duplex structure which could have influenced its stability.

As discussed in the previous sections, the fact that in both cases no cross-linking was observed does not necessarily mean that the miR-CLIP probes will be non-functional in cells. The efficiency of the cross-linking may be too low to be detected *in vitro* but sufficient to be amplified and detected in the cellular assay.

Contributions: all of the oligoribonucleotides were synthesized by Anna L. Malinowska. The *in vitro* photo-cross-linking assay was performed by Anna L. Malinowska. The luciferase assay was performed by Artur Łaski.

V.2.3. miR-106a-5p- and miR-106b-5p-related probes

As a contribution to the project run by Dr Matije Lucic, miR-CLIP probes for miR-106a-5p (as a reference) and miR-106b-5p in the form of both, hairpins (ORN-68 and ORN-70 for miR-106a-5p and miR-106b-5p, respectively) and duplexes (ORN-69 and ORN-71 for miR-106a-5p and miR-106b-5p, respectively) were prepared (**Table 23**). The main advantage of the application of the hairpins as probes is that their cleavage by Dicer results in the formation of the isomiRs in the cell-dependent manner. This better reflects the natural conditions in cells. Nevertheless, duplex probes can be also applied and their preparation in easier as the synthesized sequence is shorter. Two counter-strands were prepared for each miRNA duplex probe: fully complementary ones (ORN-72 for the miR-106b-5p and ORN-74 for the miR-106a-5p) and the native ones (corresponding to the native passenger strands; ORN-73 for the miR-106b-5p and ORN-75 for the miR-106a-5p). The activity of the duplex probes seems to differ depending on the miRNA and should be assessed in biological assays before use [287]. In case of the miR-106a-5p and miR-106b-5p probes, hairpin probes were more active in the luciferase assay than their duplex counterparts (data not shown, assay performed by Artur Łaski).

Contributions: the synthesis of all of the oligoribonucleotides was performed by Anna L. Malinowska. The luciferase assay was performed by Artur Łaski.

V.3. Site-Specific Fluorophore Labeling of Guanosines in RNA G-Quadruplexes

G-quadruplexes are non-canonical, secondary structures which can be formed by guanine-rich DNA and RNA sequences [343]. G-quadruplexes have been reported to have important functions in various cellular processes [344, 345]. Different biophysical methods, such as UV melting or circular dichroism (CD) spectroscopy, have been developed to characterize G-quadruplex structures. However, the techniques which would allow to investigate the equilibrium of the formation of the G-quadruplexes are missing.

For this purpose, a FRET-based method was developed to enable the detection of the folding of the RNA into G-quadruplex and study the kinetics of this process. The reported technique employed probes which were site-specifically modified with the fluorophores: cyanine 3 (Cy3) and cyanine 5 (Cy5) at the 2'-*O*-positions of the selected guanosines participating in the formation of a G-quadruplex structure. Internal labelling was applied, as it provided more flexibility in the structural design of the probes. Besides, end-labelling with the fluorophores can have a stabilizing or destabilizing effect on the investigated structure [346-348]. To avoid the cross-labelling during the incorporation of two modifications and due to the instability of the Cy5 to the basic conditions of the oligonucleotide deprotection, 'stop-and-go' synthetic strategy with two consecutive CuAAC reactions was applied. The Cy3 (donor) was introduced by the CLICK reaction on the solid support with the interruption of the solid phase synthesis. Afterwards, the synthesis was resumed and the obtained oligonucleotide was cleaved from the solid support, deprotected and subjected to the CLICK reaction in solution to attach the Cy5 fluorophore [283]. Biophysical experiments showed that the applied labelling pattern did not interfere with the formation of the G-quadruplex structure. This approach allowed to obtain the library of mono- and/or bis-labelled G-quadruplex-forming oligoribonucleotides and their mutants. The FRET-based assay with the synthesized compounds enabled detection of the formation of the G-quadruplex structure in various conditions. In the future, the reported technique can be applied to study the formation of G-quadruplexes in long RNAs.

Contributions: Dr Helen L. Lightfoot, Timo Hagen and Anna L. Malinowska designed the project. Anna L. Malinowska performed the synthesis, purification and characterization of the oligoribonucleotides. Timo Hagen performed biophysical studies (UV melting, CD, FRET assay). The fully protected 2'-*O*-propargyl guanosine phosphoramidite was prepared previously by Martina Bigatti. Cy3 and Cy5 azides were synthesized previously by Dr Ugo Pradère.

This work resulted in the following publication [349]:
Hagen T.,* Malinowska A. L.,* Lightfoot H. L., Bigatti M., Hall J.; **2019**. 'Site-Specific Fluorophore Labeling of Guanosines in RNA G-Quadruplexes', *ACS Omega*, 4(5), 8472-8479.

* equal contribution

V.4. Additional publications

Other publications associated with this thesis:

Malinowska A. L., Hall J.; **2019**. 'CHAPTER 8: Research and Development of Oligonucleotides Targeting MicroRNAs (miRNAs)', *Advances in Nucleic Acid Therapeutics*, 151-180 [1];

Masliah G., Maris C., Konig S. L., Yulikov M., Aeschimann F., Malinowska A. L., Mabille J., Weiler J., Holla A., Hunziker J., Meisner-Kober N., Schuler B., Jeschke G., Allain F. H.; **2018**. 'Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains', *Embo j, 37*(6): e97089 [350];

Menzi M., Wild, B., Pradère U., Malinowska A. L., Brunschweiger A., Lightfoot H. L., Hall J.; **2017**. 'Towards Improved Oligonucleotide Therapeutics Through Faster Target Binding Kinetics', *Chemistry – A European Journal*, *23*(57), 14221-14230 [351];

Li M., Lightfoot H. L., Halloy F., Malinowska A. L., Berk C., Behera A., Schumperli D., Hall, J.; **2017**. 'Synthesis and cellular activity of stereochemically-pure 2'-*O*-(2-methoxyethyl)-phosphorothioate oligonucleotides', *Chemical Communications, 53*(3), 541-544 [352];

VI. General conclusions

A long term goal of this research investigation was to develop a strategy that would allow achieving a more versatile (sequence-independent) and efficient cross-linking between the miRNAs and their various RNA targets. It was envisioned that this could be accomplished through a better understanding of the cross-linking with the currently used miR-CLIP probes and by searching for alternative probe designs allowing more optimal positioning of the cross-linker.

The *in vitro* cross-linking study with miR-106a-5p analogues, in which systematic AU to CG mutations were introduced at selected positions within the duplex, suggested that the cross-linking process occurs between the photophore and uridines in the target strand positioned two or three base pairs upstream from the modification site. However, this effect seemed to be dependent on the local structure and region of the duplex. Extending the tests for other miRNAs helped to further verify previous observations. Therefore, it was proposed that it might be beneficial to shift the placement of the cross-linker with respect to the putative target uridine. These findings expanded the knowledge about the cross-linking process with the probes carrying the trioxsalen at the 2'-*O*-position of the ribose and underlined the complexity of the conscious probe design.

In search of the optimal cross-linking strategy, probes with a diazirine moiety localized within the minor or major groove of the duplex were tested. This approach occurred to be ineffective in the context of the RNA-RNA cross-linking, as the diazirine moiety most likely reacted with water. However, such probes may prove to be more practical in the detection of miRNA-protein interactions.

The triazole-based convertible nucleoside approach demonstrated to be useful for the preparation of the new probes, in which the trioxsalen was attached at the N^4 position of cytidine *via* linkers containing a different number (0-4) of ethylene glycol units. The main intention of the application of such probes was to increase the general probability of the cross-linking events with random uridines in the targets, as different linkers would place the cross-linker at different positions with respect to a duplex structure. While some of the linkers would be incapable of delivering the trioxsalen nearby uridines, the other ones might fulfil this task. The first experiments with new probes for miR-124-3p gave promising results however they need further evaluation. Probes for other miRNAs should be also prepared and tested to verify the general applicability of this approach.

miRNAs undoubtedly play a very important role in maintaining proper functioning of living organisms. A significant dysregulation in the miRNA levels has been associated with various diseases. The growing body of research shows the importance of non-canonical interactions of the miRNAs. This in turn highlights the need for the development of the new experimental methods that could be employed for the validation of such interactions. Hopefully, presented results will contribute to the design of the improved miR-CLIP probes which will allow to capture various miRNA targets, independently of the miRNA sequence.

VII. Experimental part

General experimental details

Thin-layer chromatography was done on silica gel 60 Å F254 aluminium sheets (Merck). Unless stated otherwise, the preparative flash column chromatography was carried out on silica gel, 60 Å (Fluka). For the compounds which synthesis has been described previously only ^{1}H NMR spectra were recorded. For the new compounds, ^{1}H and ^{13}C NMR spectra were measured. NMR spectra were recorded on a Bruker Av400 at a resonance frequency of 400 MHz (for ^{1}H NMR) or 101 MHz (for ^{13}C NMR). Solvent signals were used as internal standards. Chemical shifts (δ) are given in ppm. ^{1}H NMR spectra are described as follows: multiplicity, the coupling constant J [Hz], the number of protons (integral). For ^{13}C NMR spectra, only the chemical shifts of signals are given.

The following structures used in all of the studies were taken from the common laboratory stock (prepared by Martina Bigatti according to the published procedures [284, 349, 84]):
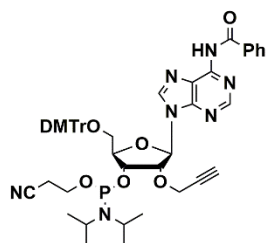
2'-O-propargyl-N4-benzoyladenosine phosphoramidite
[(2R,3R,4R,5R)-5-(6-benzamido-9H-purin-9-yl)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite]
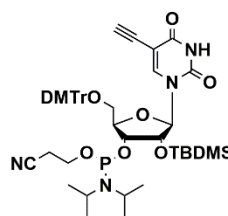
5-acetylene uridine phosphoramidite
[(2R,3R,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-5-(5-ethynyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite]
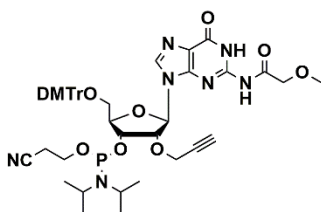
2'-O-propargyl-2-methoxyacetamideguanosine phosphoramidite [(2R,3R,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-5-(2-(2-methoxyacetamido)-6-oxo-1,6-dihydro-9H-purin-9-yl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite]
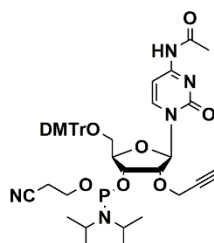
2'-O-propargyl-N4-acetylcytidine phosphoramidite
[(2R,3R,4R,5R)-5-(4-acetamido-2-oxopyrimidin-1(2H)-yl)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite]
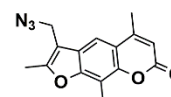
4'-azidomethyltrioxsalen
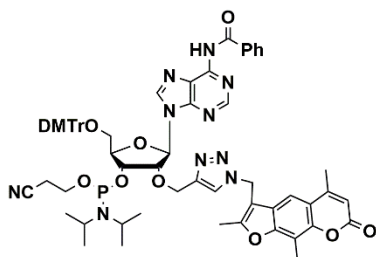
2'-O-trioxsalen-N4-benzoyladenosine phosphoramidite
[(2R,3R,4R,5R)-5-(6-benzamido-9H-purin-9-yl)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((1-((2,5,7,9-tetramethyl-7H-furo[3,2-g]chromen-3-yl)methyl)-1H-1,2,3-triazol-4-yl)methoxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite]
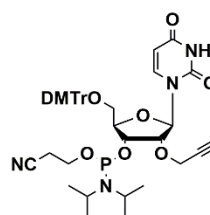
2'-O-propargyluridine phosphoramidite
(2R,3R,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite

The sequences of the synthetic miRNAs and pre-miRNAs were designed based on the miRNA and pre-miRNA sequences reported in the freely available miRBase database [353].

VII.1. Chapter II

VII.1.1. Oligoribonucleotide synthesis, deprotection and purification

Chemicals and solvents were purchased from Biosolve, Sigma-Aldrich, VWR, Acros Organics and TCI. RNA phosphoramidites were purchased from Thermo Fisher. Oligoribonucleotides were synthesized on a 50 nmol or 1 µmol scale with the MM12 synthesizer (Bio Automation Inc.) using 500 Å UnyLinker Controlled Pore Glass (CPG) (ChemGenes). Cleavage of the DMTr group was performed with 3% dichloroacetic acid in dichloromethane (DCM) (v/v). The RNA phosphoramidites were prepared as 0.08 M solutions in dry acetonitrile (ACN). 0.24 M solution of 5-benzylthio-1*H*-tetrazole (Carbosynth) in dry ACN was used as an activator. Failed sequences were capped with a 1:1 mixture of CapA (acetic acid anhydride in tetrahydrofuran) and CapB (*N*-methylimidazole and pyridine in tetrahydrofuran). The oxidizing step was performed with 0.02 M iodine in tetrahydrofuran/pyridine/water (7:2:1).

Synthesis parameters

50 nmol scale: deblock 3 x 20 s, coupling times: 2 x 120 s for the standard RNA phosphoramidites and 2 x 180 s for the modified phosphoramidites, capping: 1 x 30 s, oxidation: 1 x 40 s.
1 µmol scale: deblock 2 x 40 s, coupling times: 1 x 300 s for the standard RNA phosphoramidites and 1 x 360 s for the modified phosphoramidites, capping: 1 x 60 s, oxidation: 1 x 60 s.

After the solid-phase synthesis, the CPG was treated with gaseous methylamine for 1.5 h at 70°C (unmodified oligoribonucleotides) or with a mixture of 200 µL of ammonia solution (25% in water) and 200 µL of methylamine solution (40% in water) for 6 h at 35°C (modified oligoribonucleotides). After basic deprotection and cleavage from the solid support, the CPG was washed with 3 x 200 µL of water/EtOH (1:1) mixture. To the solution, 20 µL of 1 N Tris-base was added and it was evaporated to dryness in a SpeedVac. Desilylation was carried out by the treatment with 130 µL of a freshly prepared mixture of *N*-methyl-2-pyrrolidone (NMP) (60 µL), triethylamine (30 µL) and triethylamine trihydrofluoride (Et$_3$N x 3HF) (40 µL) at 70°C for 2 h. The reaction was quenched with trimethylethoxysilane (200 µL, 3 min, room temperature). Diethyl ether (200 µL, 5 min, room temperature) was then added and the mixture was vortexed and centrifuged.

In case of the 1 µmol scale synthesis, the CPG was divided into smaller portions (around 20 mg of the CPG each) and treated with a doubled amount of the reagents.

The precipitate was dissolved in 200 µL of ultrapure water and purified on an Agilent 1200 series preparative RP-HPLC on an XBridge OST C-18 column (10 x 50 mm, 2.5 µm; Waters) at 65°C with a flow rate 5 mL/min. Eluent A: 0.1 M aqueous triethylamine/acetic acid, pH 8.0; eluent B: 100% MeOH; gradient 20-60% B in 5 min. Fractions containing the DMTr-on product were collected, dried in a SpeedVac and treated with 200 µL of 40% aqueous acetic acid for 15 min at room temperature. Samples were concentrated in a SpeedVac, dissolved in 200 µL of ultrapure water and purified in DMTr-off mode by RP-HPLC on an XBridge OST C-18 column (10 x 50 mm, 2.5 µm; Waters) at 65°C with a flow rate 5 mL/min. Eluent A: 0.1 M aqueous triethylamine/acetic acid, pH 8.0; eluent B: 100% MeOH; gradient 5-30% B in 5 min. Fractions containing the product were collected, dried in a SpeedVac, re-dissolved in 200 µL of ultrapure water and analysed by LC-MS (Agilent 1200/6130 system) on an Acquity OST C-18 column (2.1 x 50 mm; Waters). The column oven was set to 65°C,

flow-rate: 0.3 mL/min. Eluent C: ultrapure water containing 0.4 M hexafluoroisopropanol (HFIP) and 15 mM triethylamine; eluent B: 100% MeOH; gradient 5-35% B in 14 min.

UV-absorption (λ=260 nm) of the final products was measured in duplicates (modified oligoribonucleotides) or triplicates (unmodified oligoribonucleotides) on a NanoDrop 2000 spectrophotometer (Fisher Scientific). Concentrations were calculated with in-house programmed software. No next-neighbour correction for the extinction coefficients of the modified bases was made (influence <10%).

VII.1.2. Post-synthetic modification of the 2′-O-propargyl–substituted oligoribonucleotides by CuAAC reaction

A copper(I)-catalysed azide-alkyne cycloaddition reaction between the alkyne-modified oligoribonucleotides and 4'-azidomethyltrioxsalen was performed as described previously with minor modifications [284, 84]. After the solid-phase synthesis, the CPG containing the alkynyl-modified RNA was suspended in 300 μL of water/PBS (1:1) mixture. Subsequently, 4'-azidomethyltrioxsalen (20 eq, 1 μmol in 60 μL of DMF), TBTA (10 eq, 500 nmol in 20 μL of DMF), Na-ascorbate (10 eq, 500 nmol, 10 μL of a solution containing 10 mg of Na-ascorbate in 1 mL of water) and $CuSO_4$ x $5H_2O$ (1 eq, 50 nmol, 10 μL of a solution containing 12.5 mg of $CuSO_4$ x $5H_2O$ in 10 mL of water) were added to the suspension in this order. All solutions were freshly prepared before use. The reaction mixture was shaken (1400 rpm) overnight at 45°C in an Eppendorf shaker. The CPG was filtered off and washed three times with 0.5 mL of each: DMF, 0.1 N aqueous EDTA, DMF, ACN, chloroform and dried under vacuum. Afterwards, the post-synthetically modified oligoribonucleotides were deprotected according to the standard procedure (described in section VII.1.1).

VII.1.3. In vitro photo-cross-linking experiments

The in vitro photo-cross-linking experiments were performed adapting the procedure published by Imig et al. [84].

VII.1.3.1. Protocol for the in vitro photo-cross-linking with D1-D10*

0.2 nmol of the trioxsalen-modified RNA and its unmodified counter-strand were mixed, dried, and re-dissolved in 200 μL of the phosphate (annealing) buffer (2.5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 100 mM NaCl and 0.1 mM Na_2EDTA) so that the final concentration of each strand was 1 μM. For annealing, the solution was heated to 95°C, held at that temperature for 5 min and cooled down to room temperature over 2 h. Samples were irradiated for 15 min (Bio-Link BLX, Vilber. UV source: 5 x 8-watt lamps, λ=365 nm, a distance of the plate from the lamp: 5 cm) in an open 24-well plate placed on ice. Then, samples were directly injected on the LC-MS (settings as described in section VII.1.1) with a gradient 5-40% B in 25 min. *Initial cross-linking experiments were run and analysed with slightly different conditions, as the optimization of the protocol was on-going.

VII.1.3.2. The general protocol established for the in vitro photo-cross-linking

0.2 or 1.0 nmol of the trioxsalen-modified RNA and its unmodified counter-strand were mixed, dried, and re-dissolved in 200 μL of the phosphate (annealing) buffer (2.5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 100 mM NaCl and 0.1 mM Na_2EDTA) so that the final concentration of each strand was 1 μM or 5 μM. For annealing, the solution was heated to 95°C, held at that temperature for 5 min and cooled down to room temperature over 2 h. Samples were irradiated for 5, 10, 15 or 30 min (Bio-Link BLX, Vilber. UV source: 5 x 8-watt lamps, λ=365 nm, a distance of the plate from the lamp: 5 cm) in an open 24-

well plate placed on ice. Then, the samples were purified by RP-HPLC (settings as described in section VII.1.1) using a gradient 1-60% B in 12 min. Collected fractions were dried, re-dissolved in ultrapure water and analysed by LC-MS (settings as described in section VII.1.1) with a gradient 5-60% B in 14 min.

In some cases, it was not possible to deconvolute/assign the mass of the analysed peaks. For some of the compounds, sodium and potassium adducts were observed.

VII.1.4. Chromatograms from the *in vitro* photo-cross-linking experiments
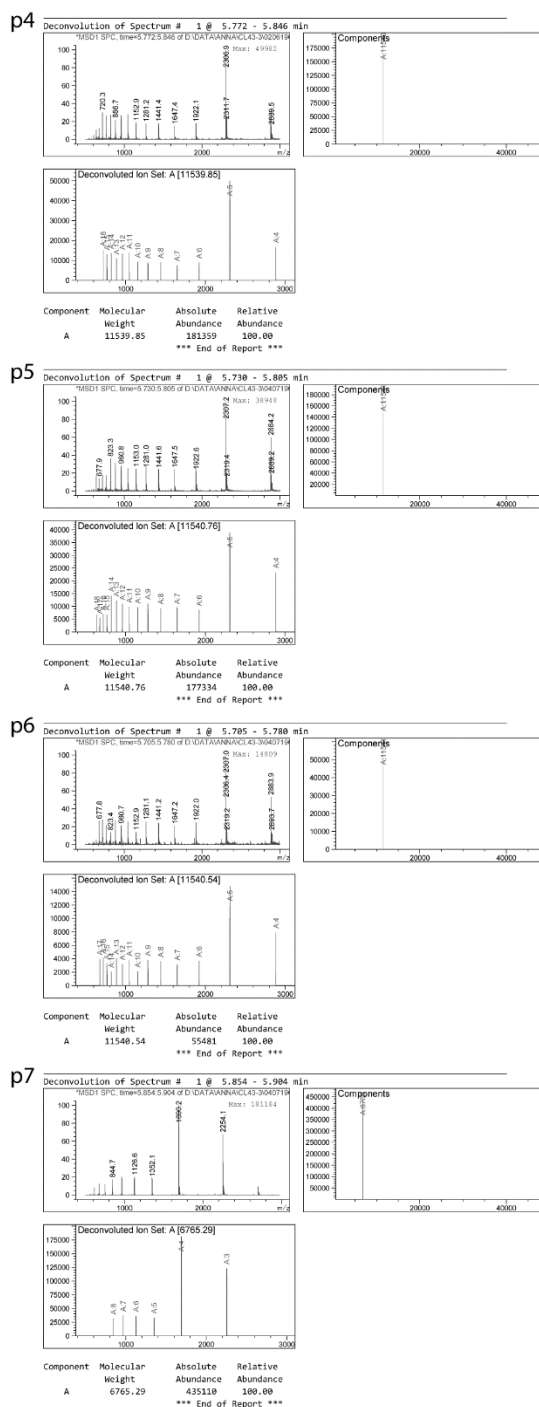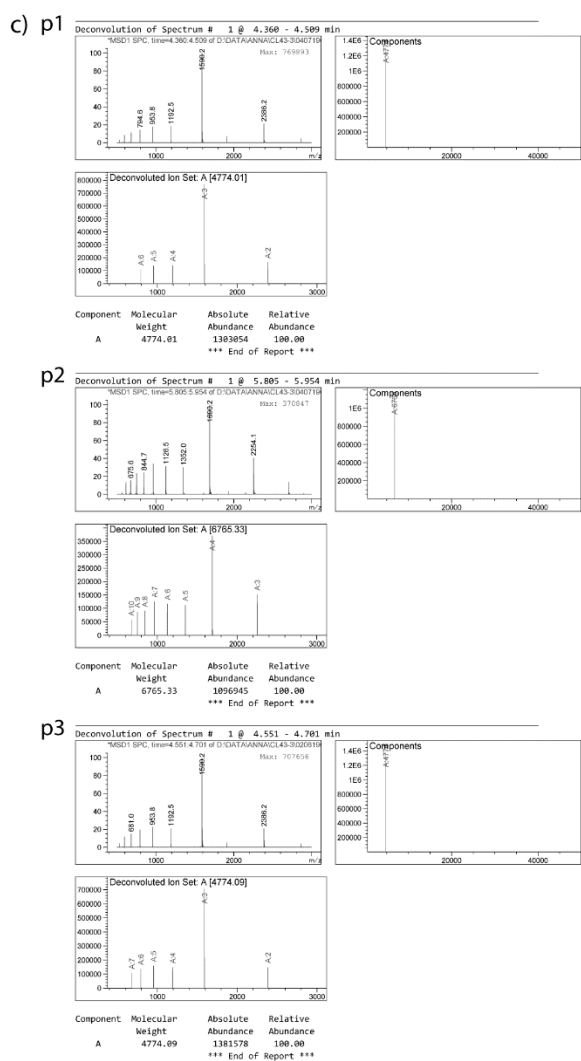
VII.1.4.1. miR-106-5p analogues

D1
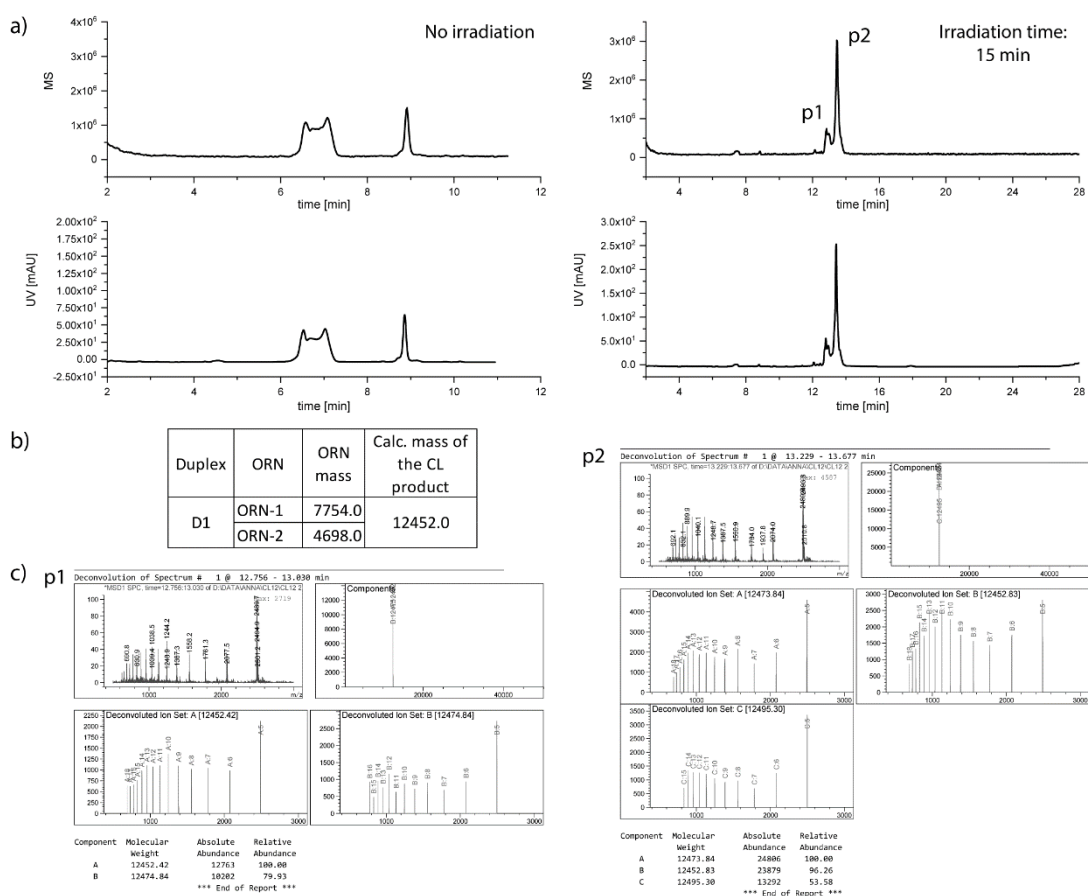
Figure 54: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D1; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
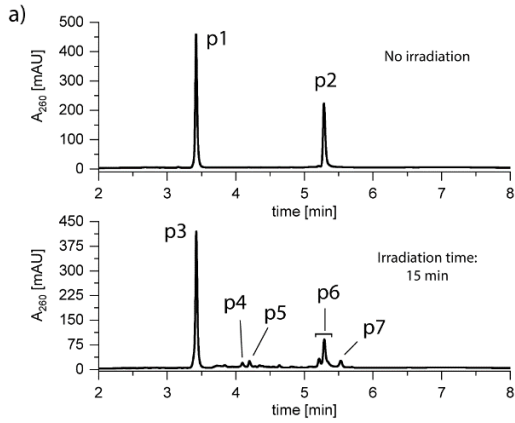
D2

a)

b)
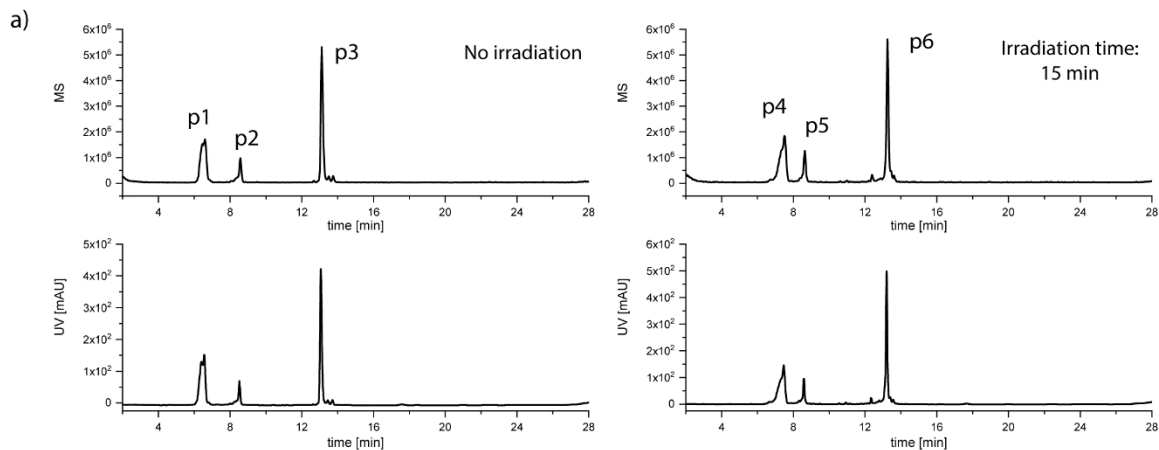
Duplex	ORN	ORN mass	Calc. mass of the CL product
D2	ORN-3	7732.0	12468.9
	ORN-4	4736.9	

c) p1

p2

p3

p4

p5

p6
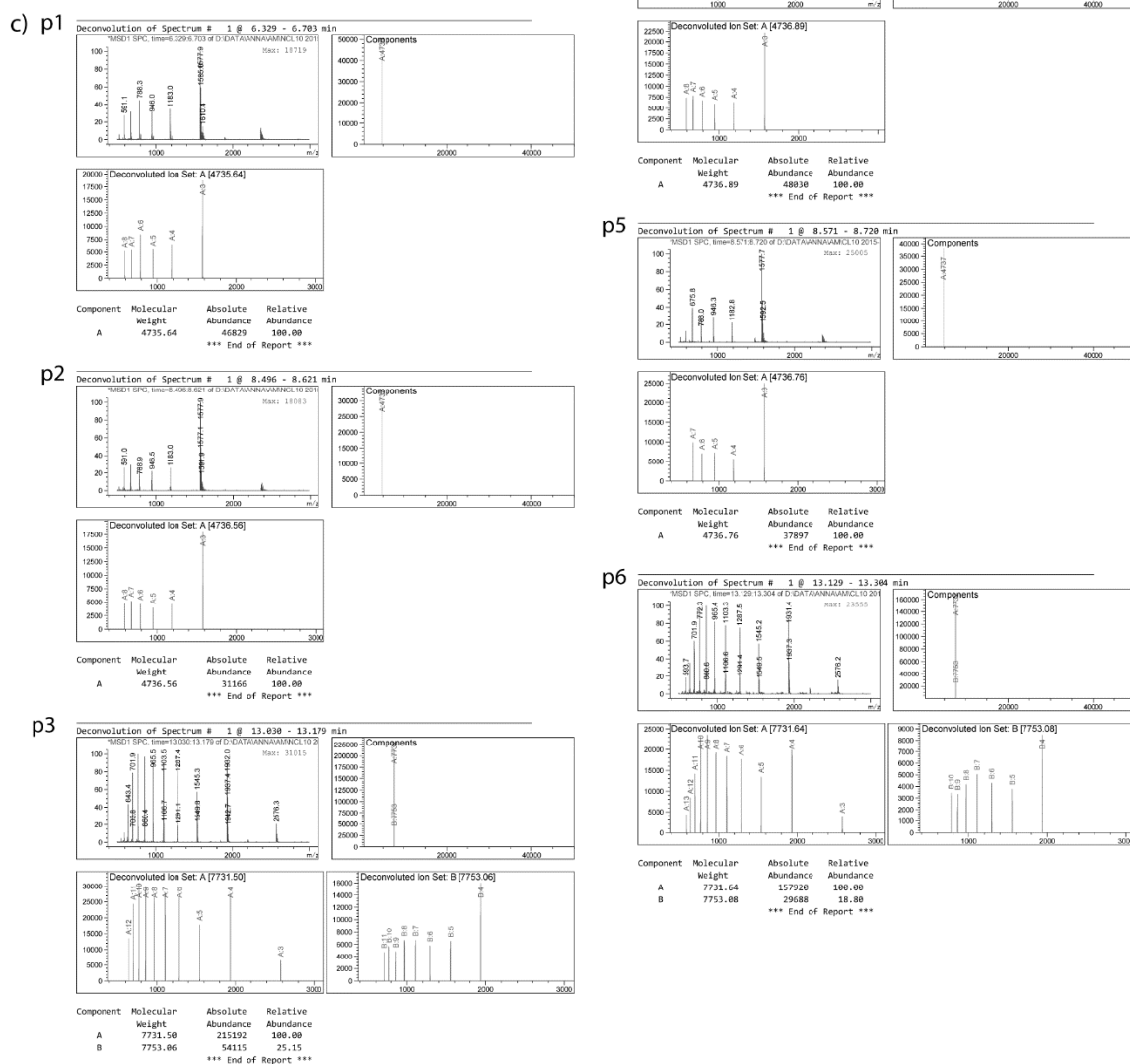
Figure 55: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D2; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D3

a)

b)
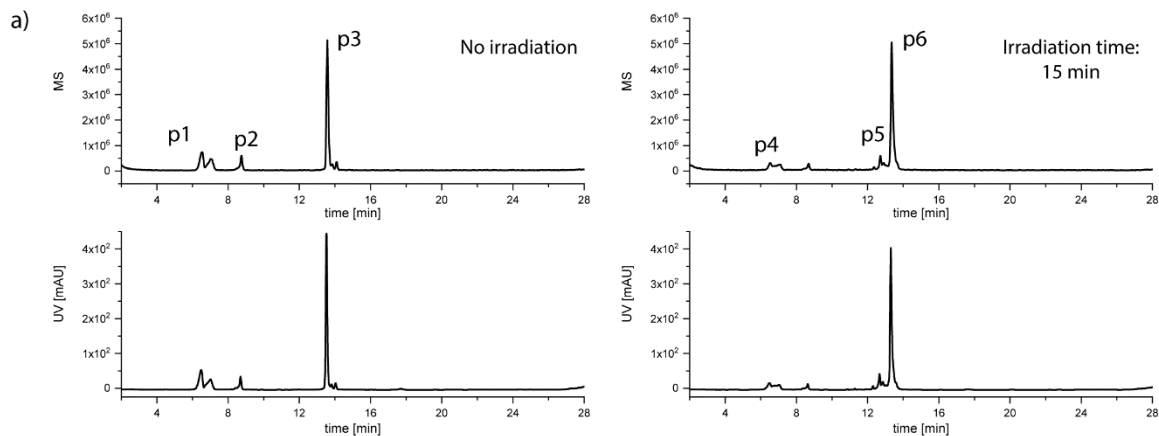
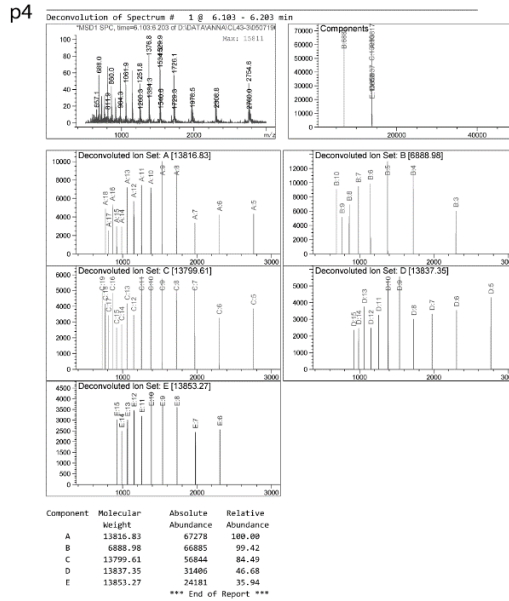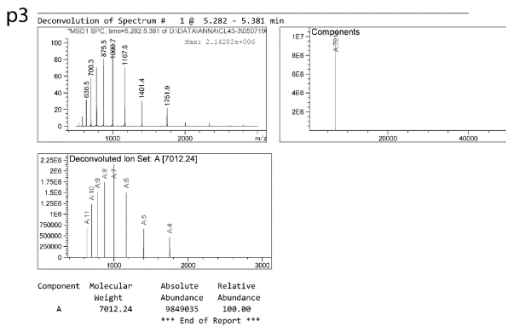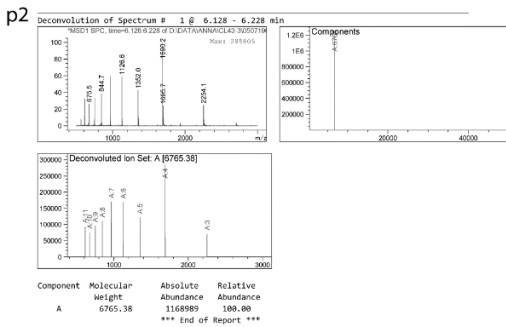| Duplex	ORN	ORN mass	Calc. mass of the CL product
D3	ORN-5	7732.0	12468.9
	ORN-6	4736.9	

c)
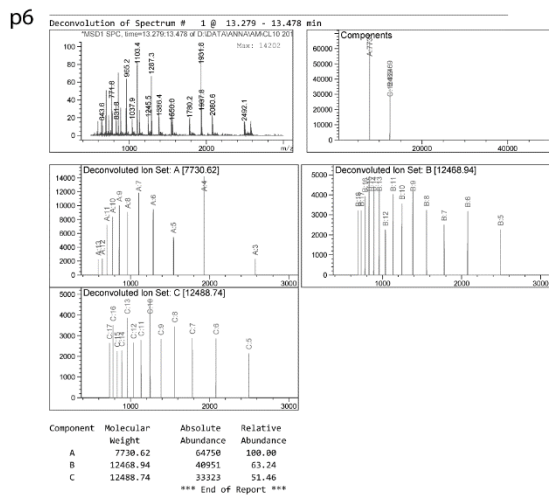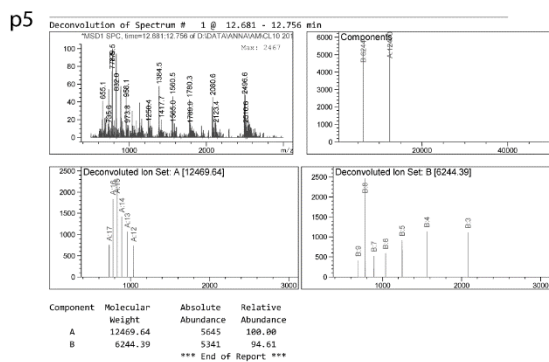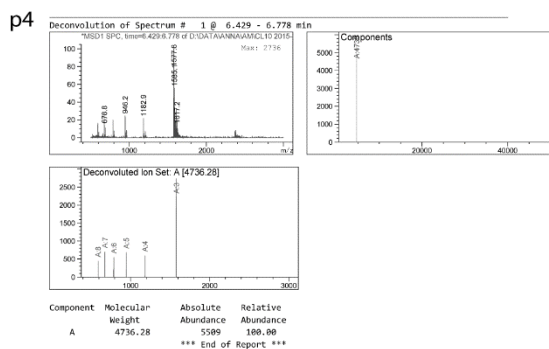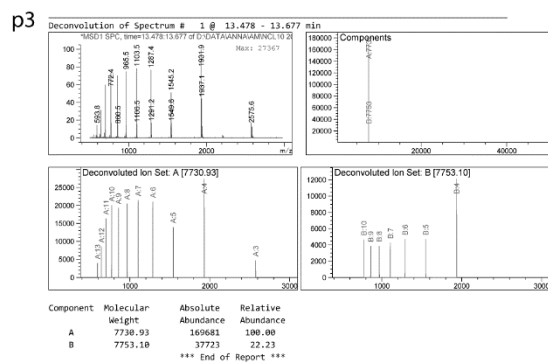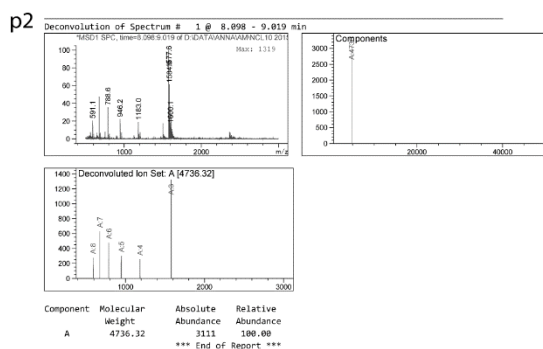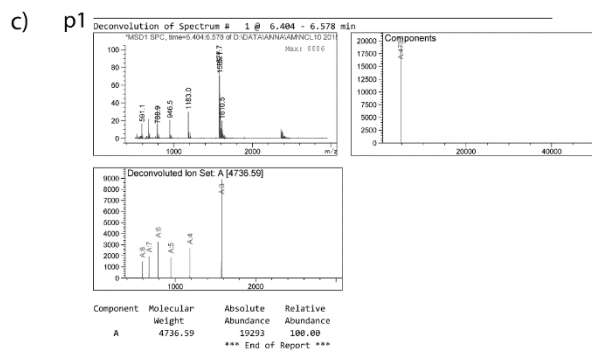
Figure 56: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D3; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D4

a)

b)
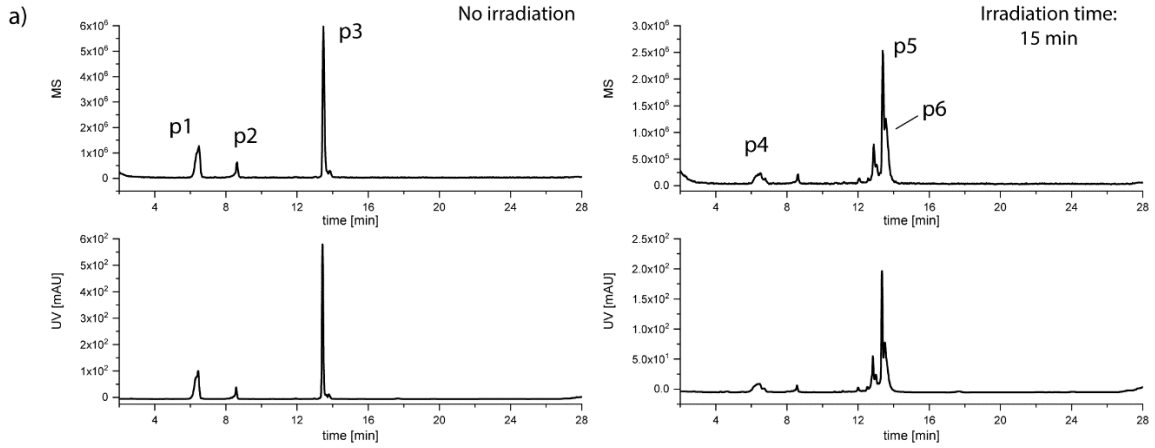
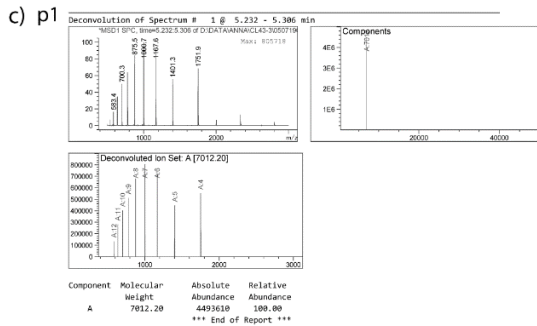| Duplex	ORN	ORN mass	Calc. mass of the CL product
D4	ORN-7	7732.0	12468.9
	ORN-8	4736.9	

c)
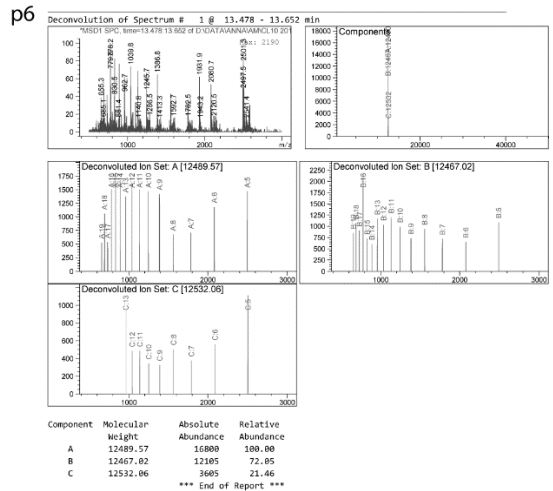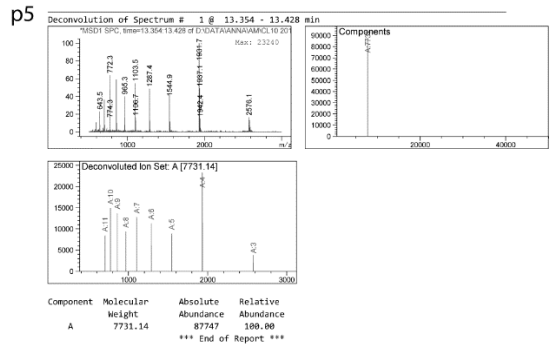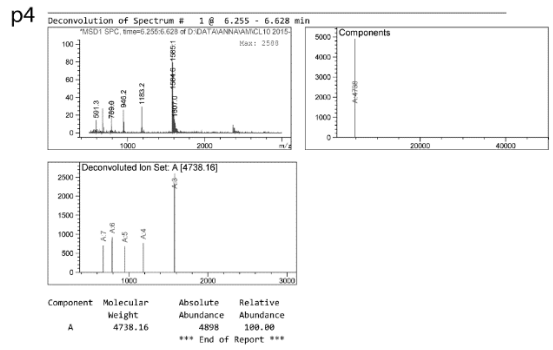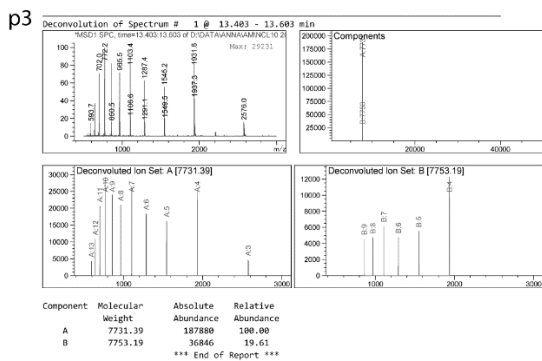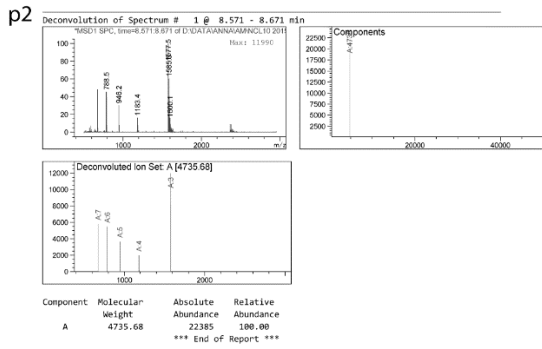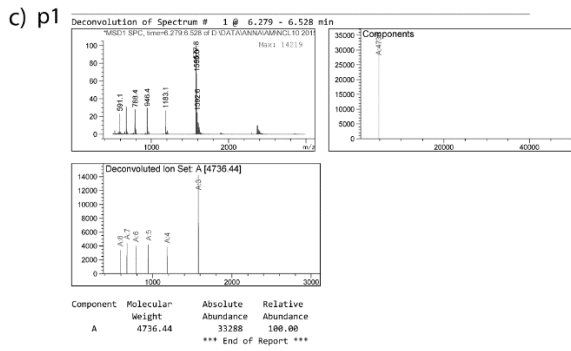
Figure 57: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D4; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
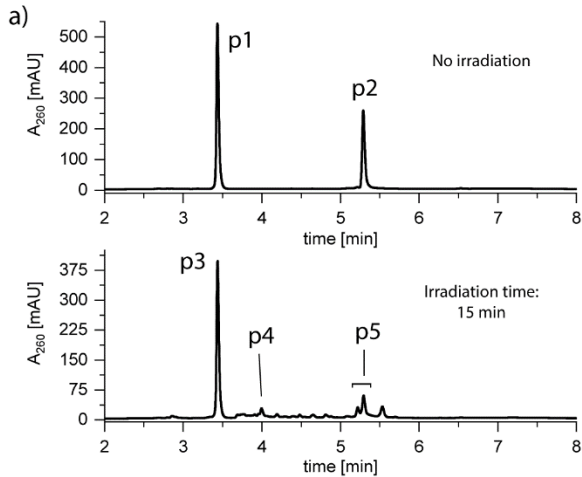
D5
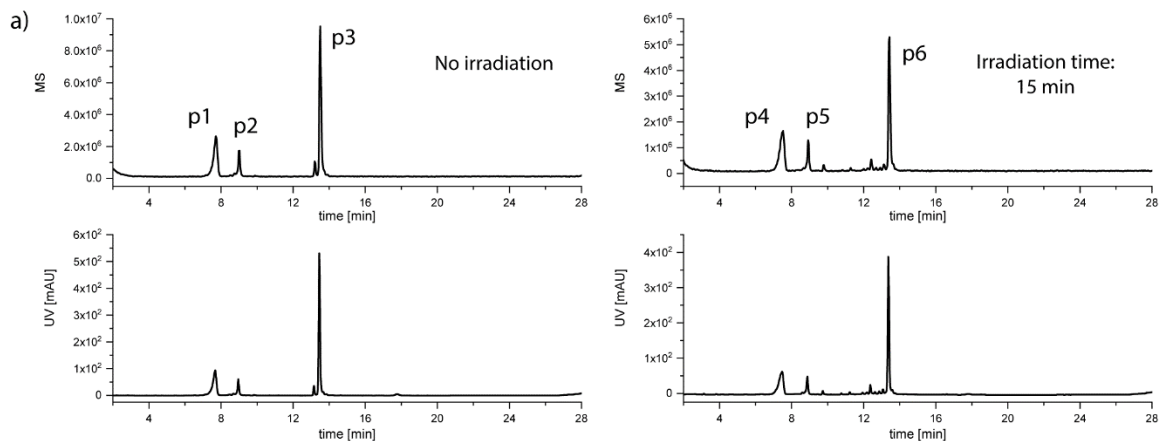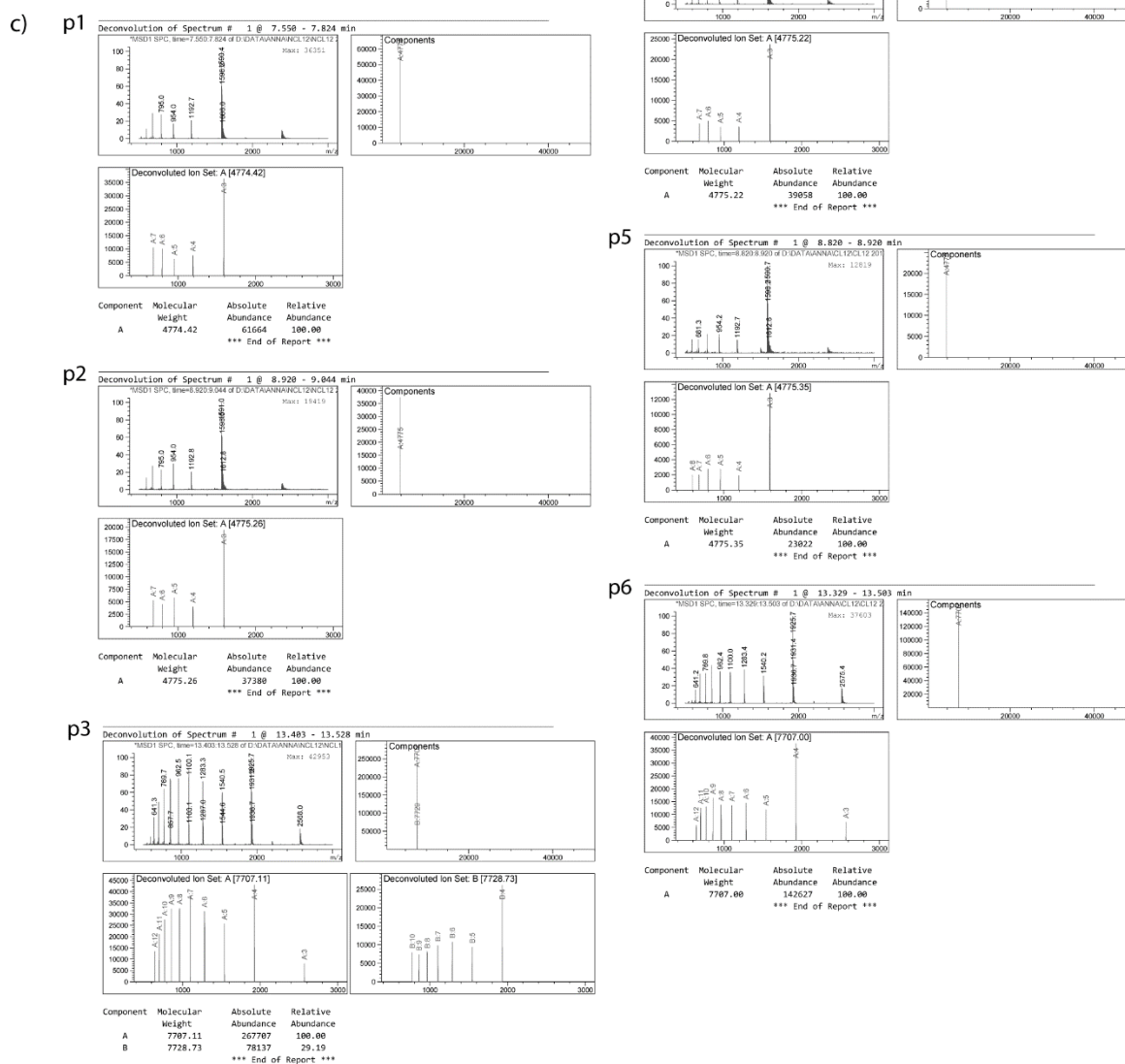
Figure 58: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D5; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D6

a)

b)
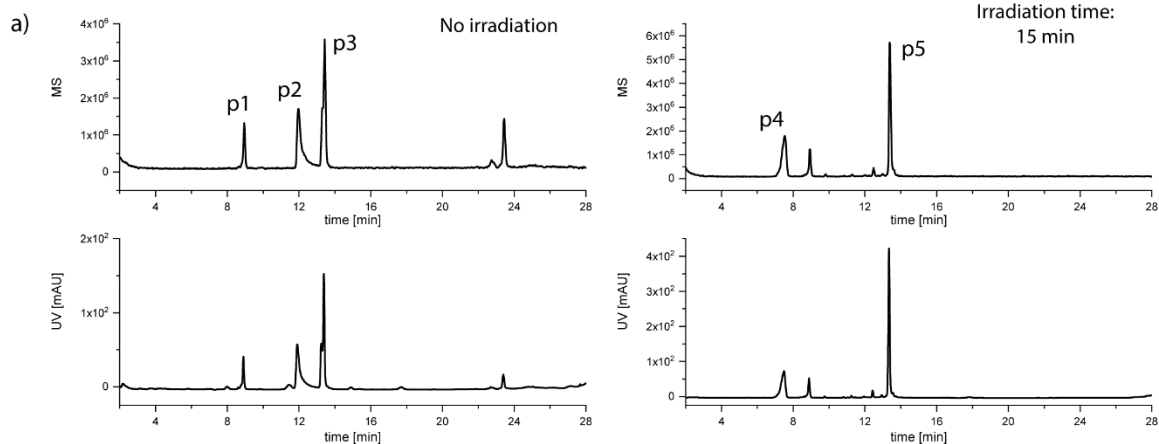
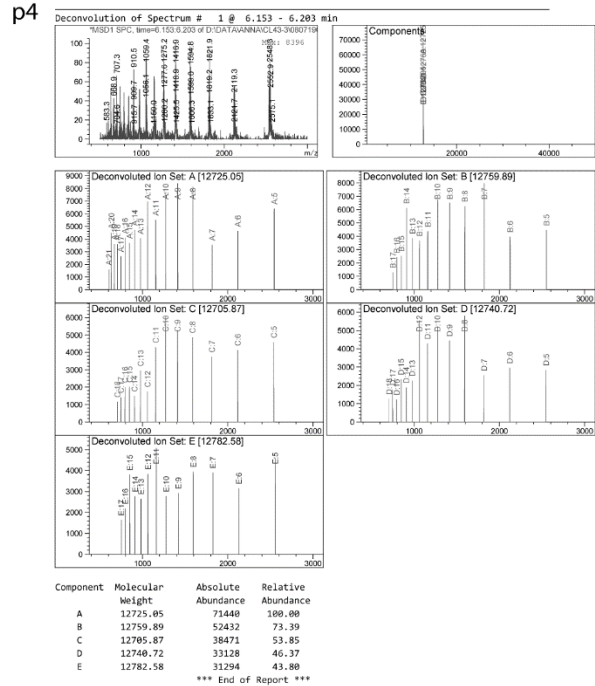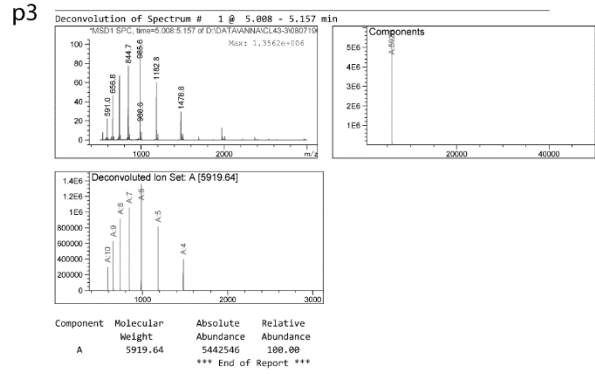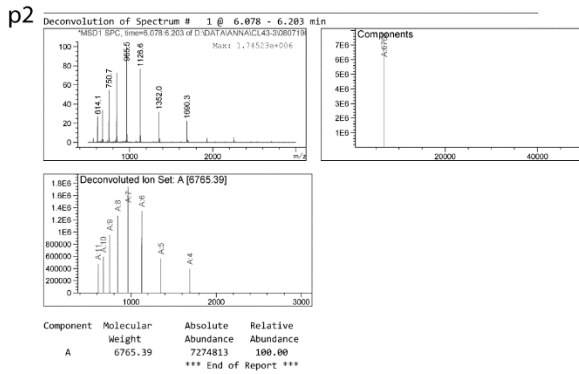| Duplex	ORN	ORN mass	Calc. mass of the CL product
D6	ORN-11	7707.8	12483.8
	ORN-12	4776.0	

c)
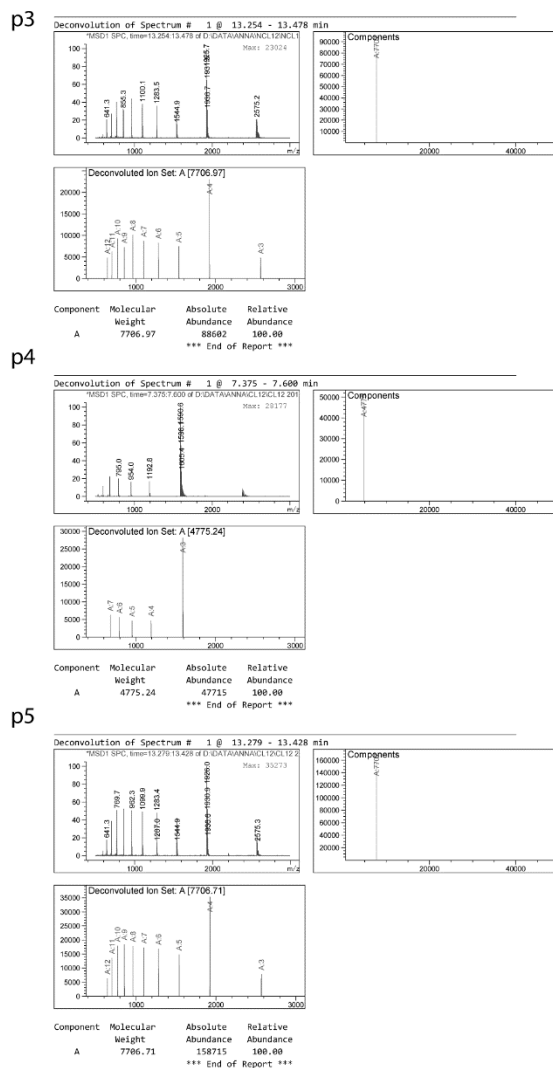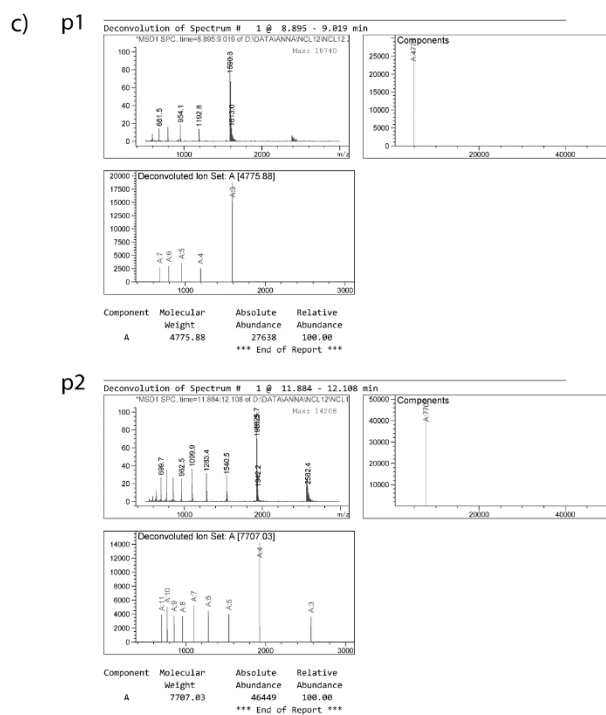
Figure 59: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D6; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D7

a)

b)

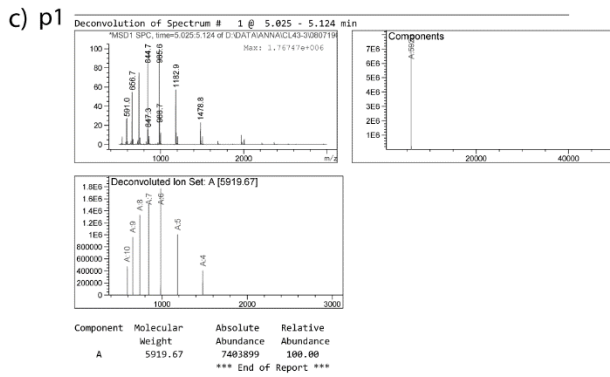| Duplex	ORN	ORN mass	Calc. mass of the CL product
D7	ORN-13	7707.8	12483.8
	ORN-14	4776.0	

c)
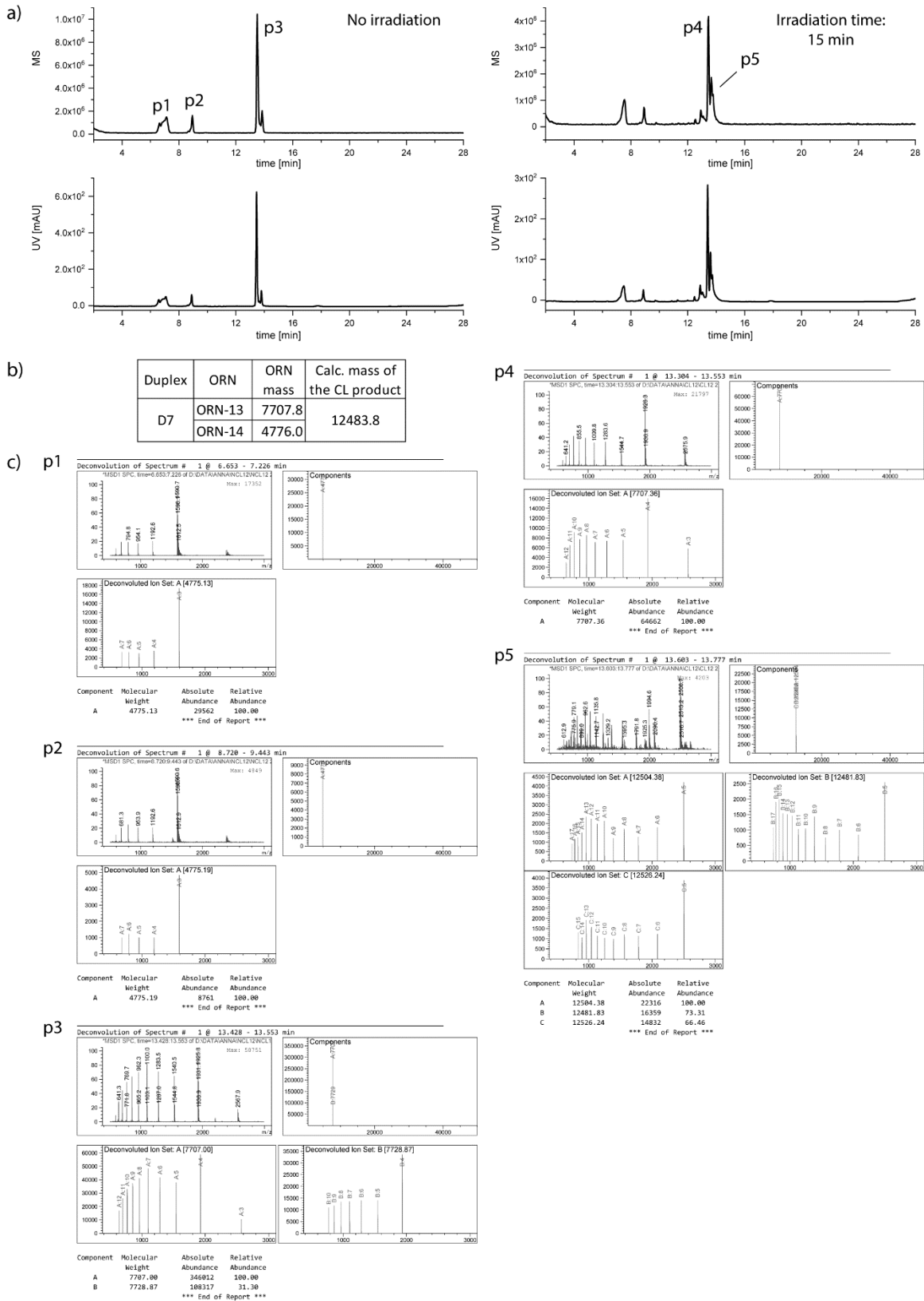
Figure 60: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D7; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
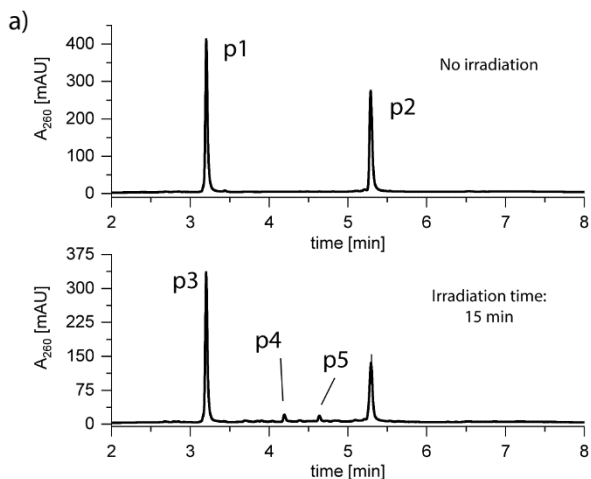
D8

a)

b)

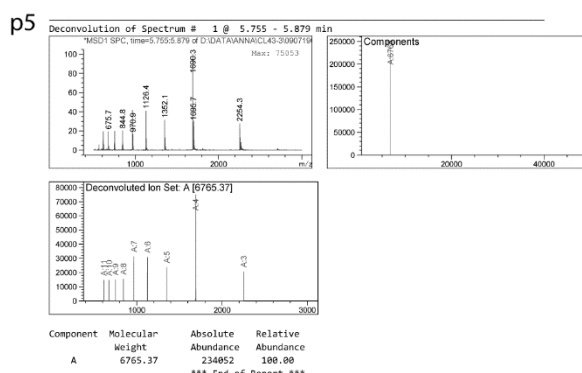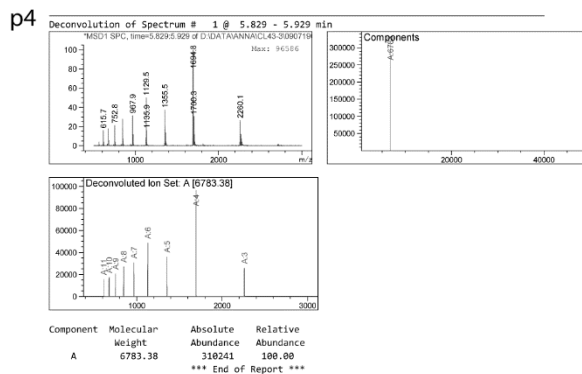| Duplex	ORN	ORN mass	Calc. mass of the CL product
D8	ORN-15	7683.8	12498.8
	ORN-16	4815.0	

c)
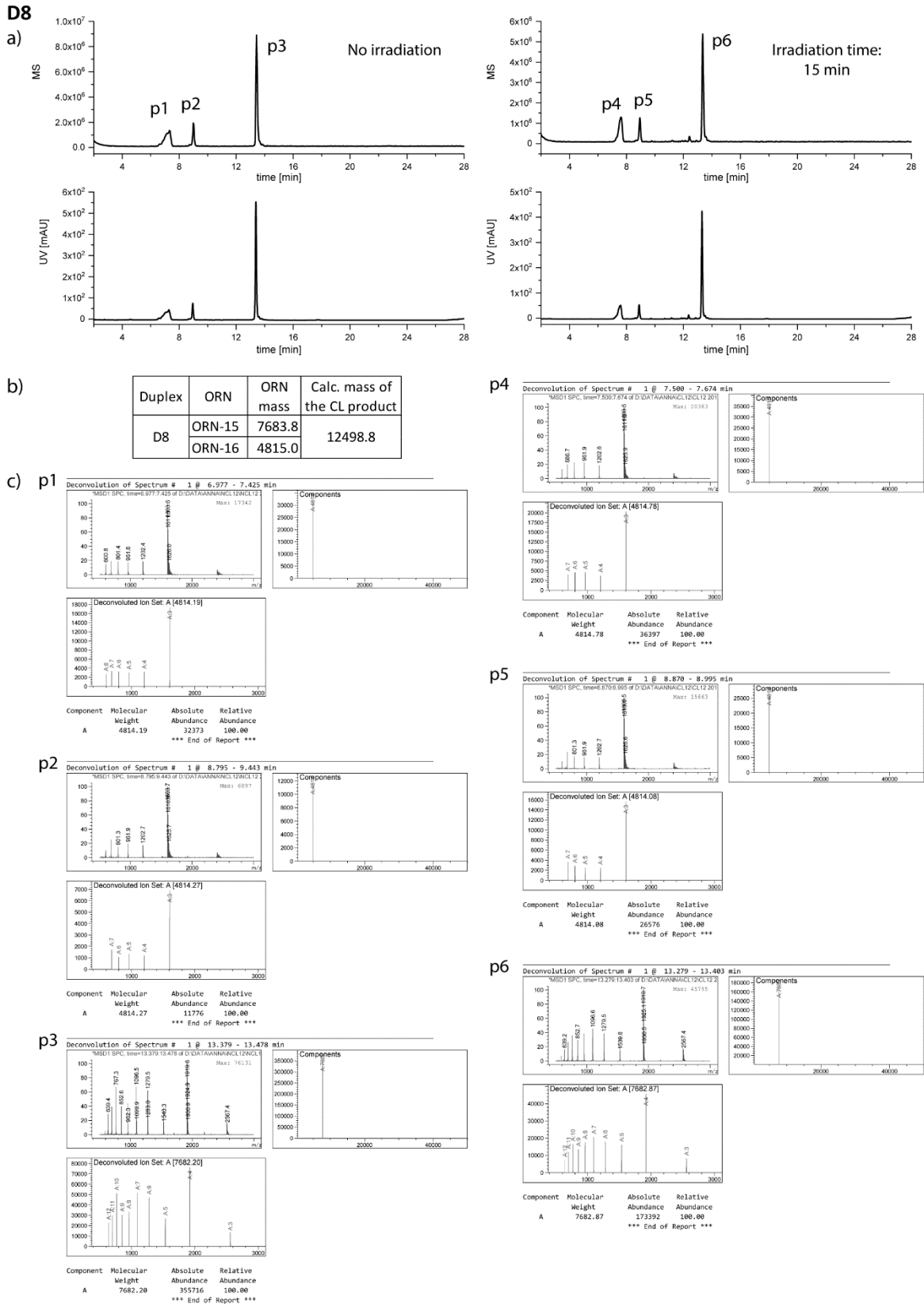
Figure 61: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D8; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D9

a)

b)
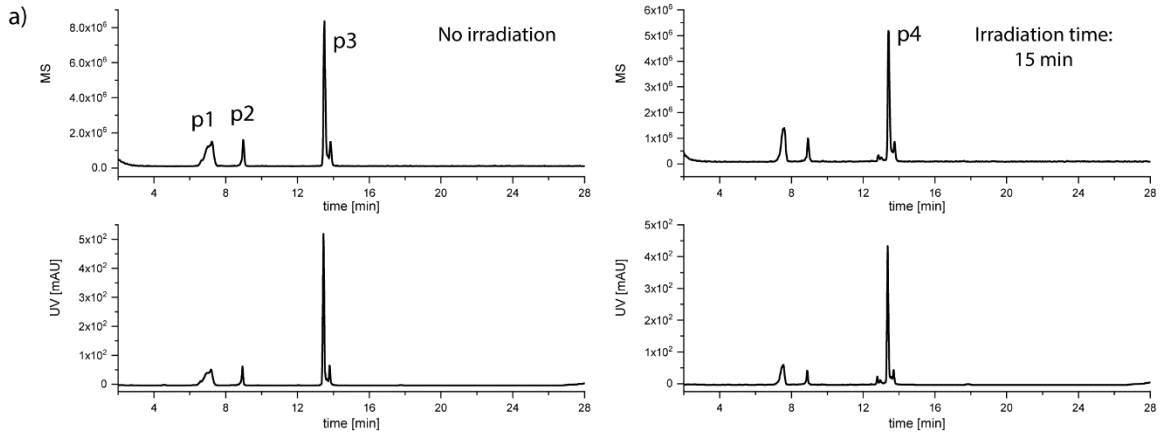
Duplex	ORN	ORN mass	Calc. mass of the CL product
D9	ORN-13	7707.8	12482.8
	ORN-17	4775.0	

c)
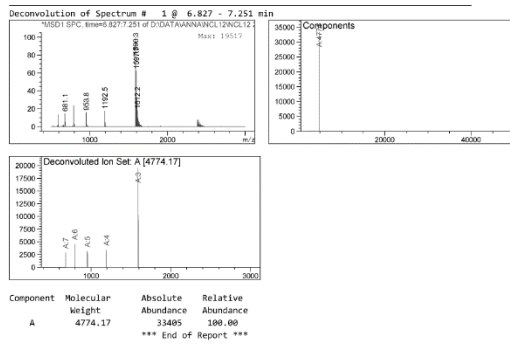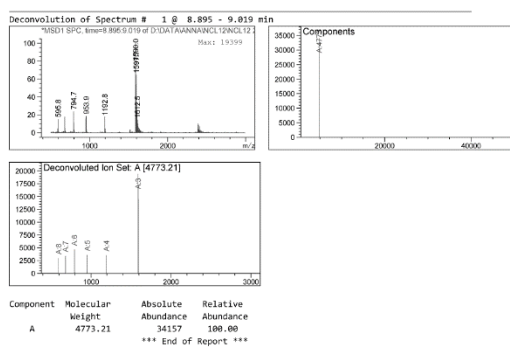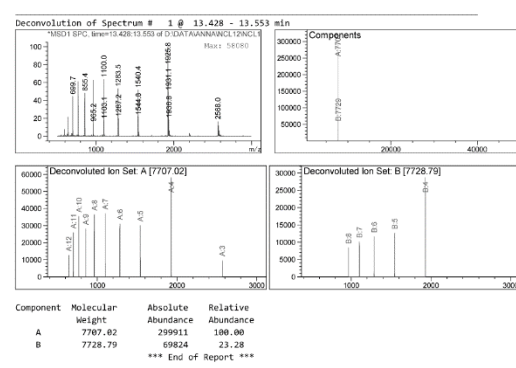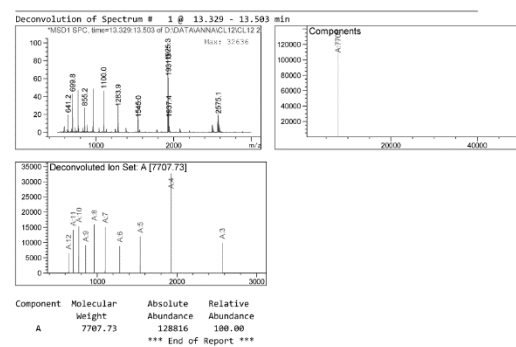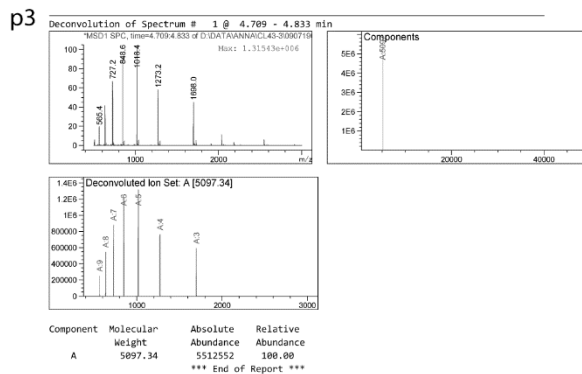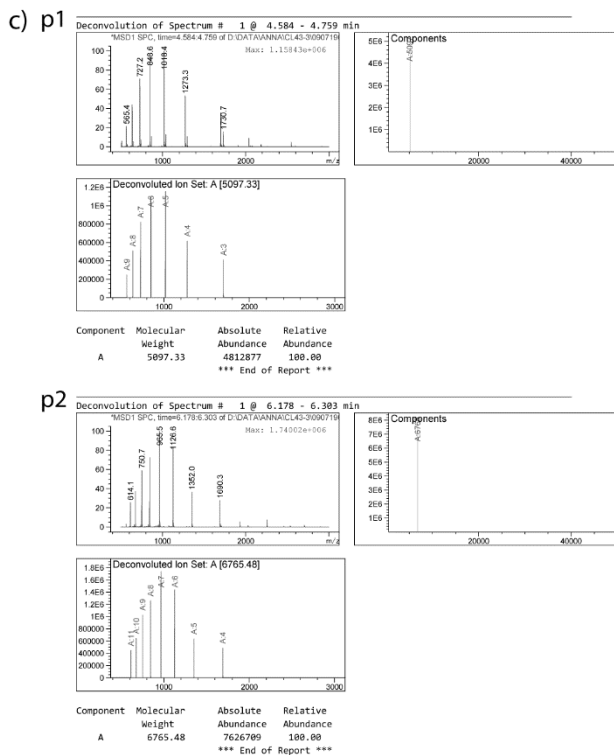

Figure 62: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D9; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D10

a)

b)

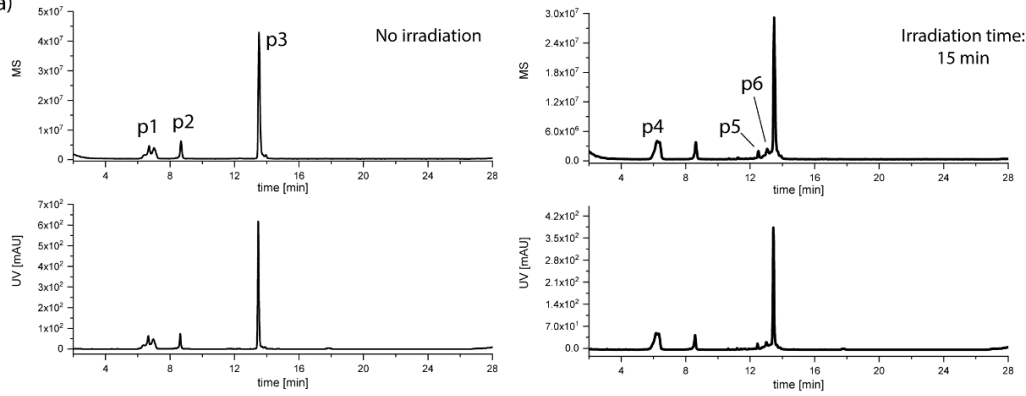| Duplex	ORN	ORN mass	Calc. mass of the CL product
D10	ORN-11	7707.8	12482.8
	ORN-18	4775.0	

c)
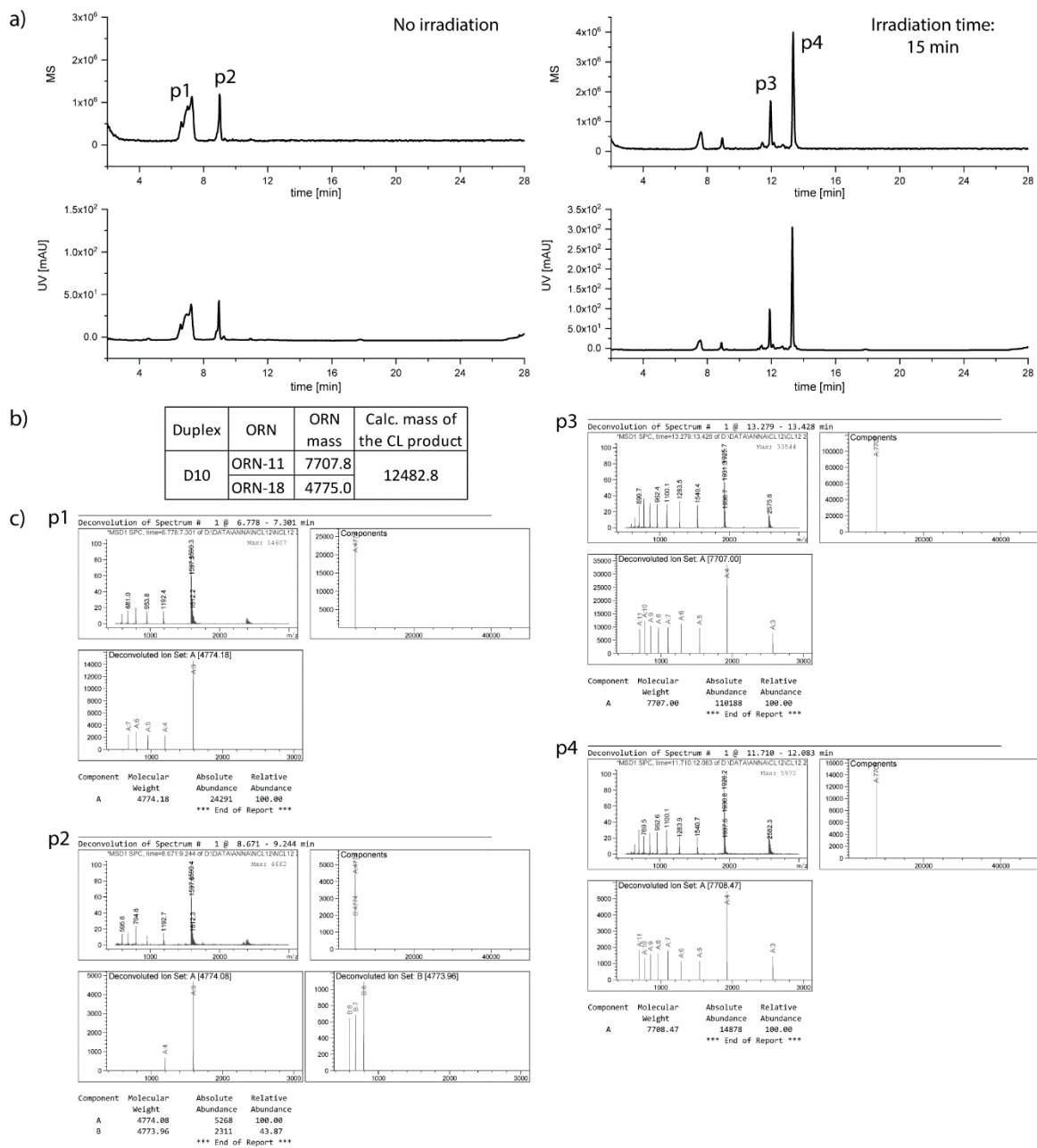
Figure 63: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D10; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D11

a)

b)
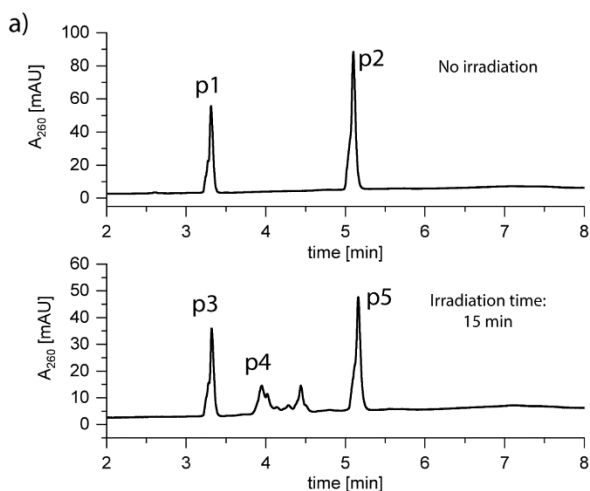
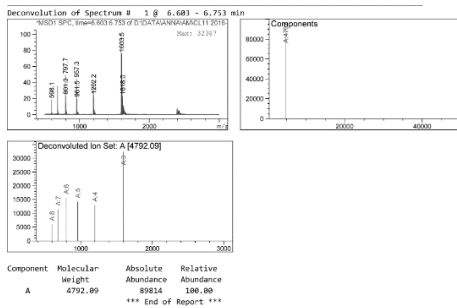| Duplex	ORN	ORN mass	Calc. mass of the CL product
D11	ORN-19	7756.0	12454.0
	ORN-2	4698.0	

c)
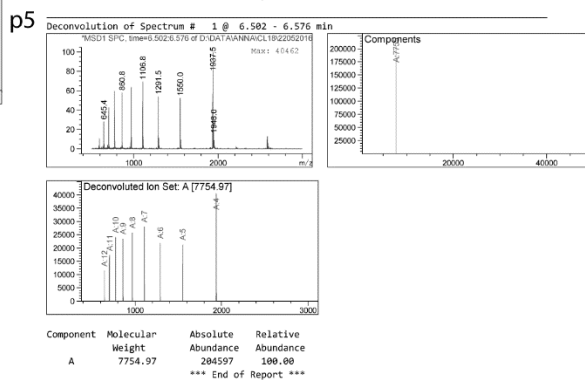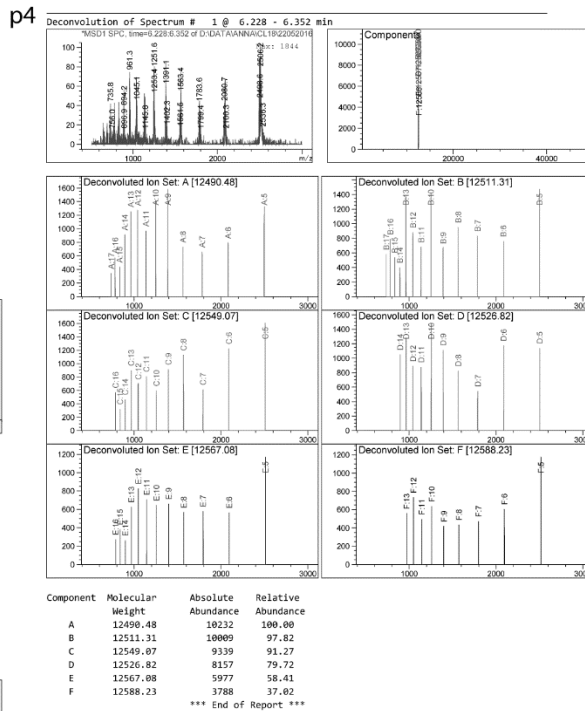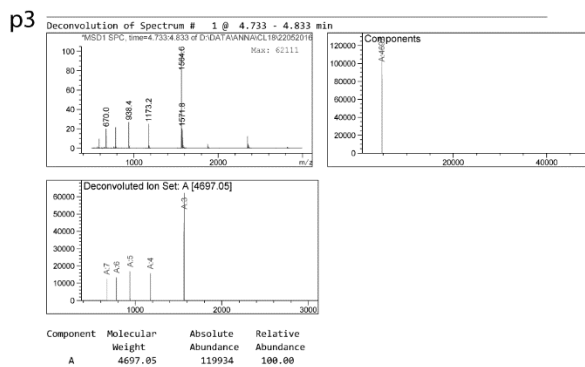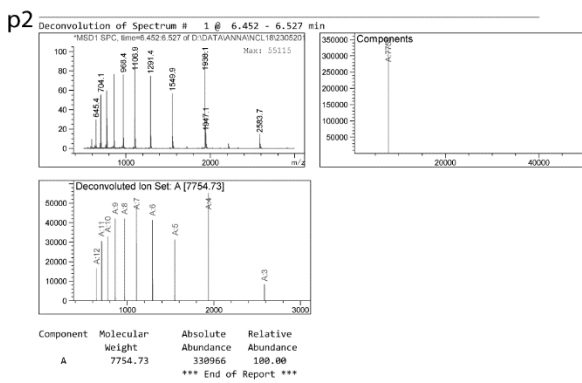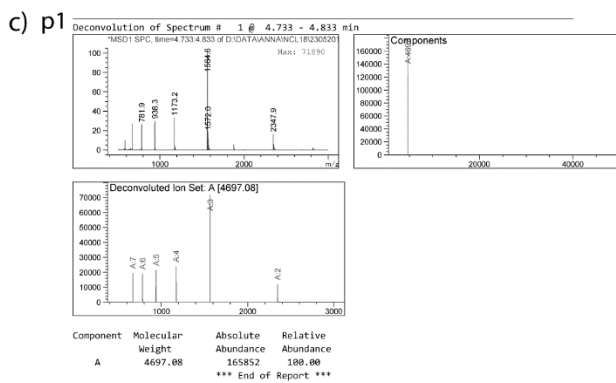
Figure 64: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D11; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D12

a)

b)
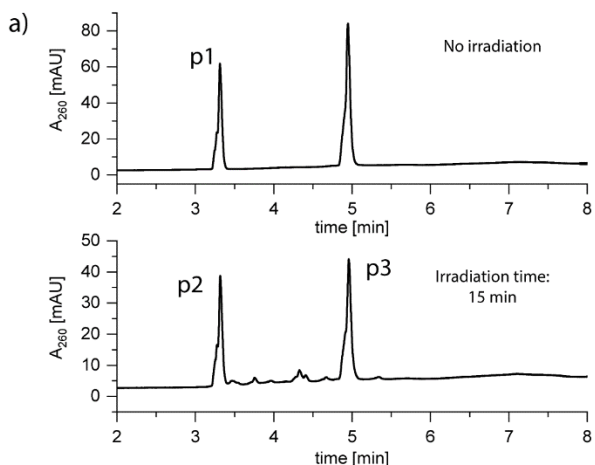
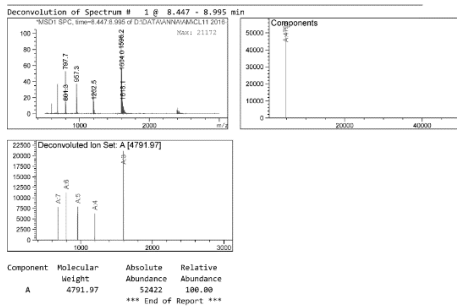| Duplex	ORN	ORN mass	Calc. mass of the CL product
D12	ORN-20	7683.9	12498.9
	ORN-21	4815.0	

c)
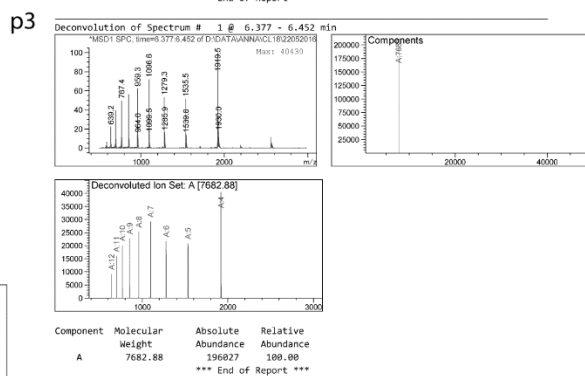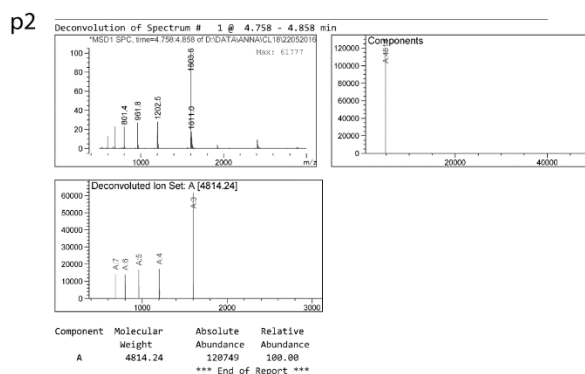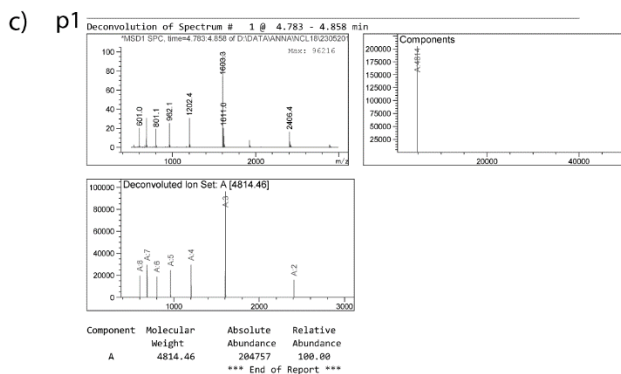
Figure 65: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D12; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D13

a)

b)
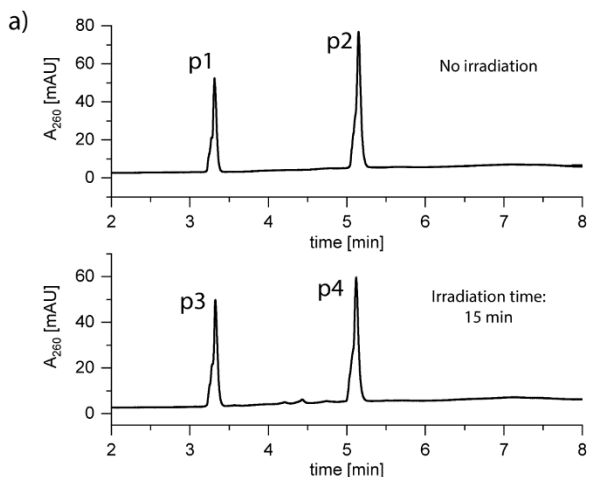
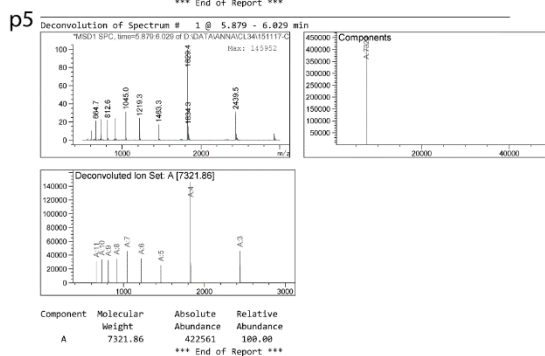| Duplex	ORN	ORN mass	Calc. mass of the CL product
D13	ORN-22	7756.0	12454.0
	ORN-2	4698.0	

c)
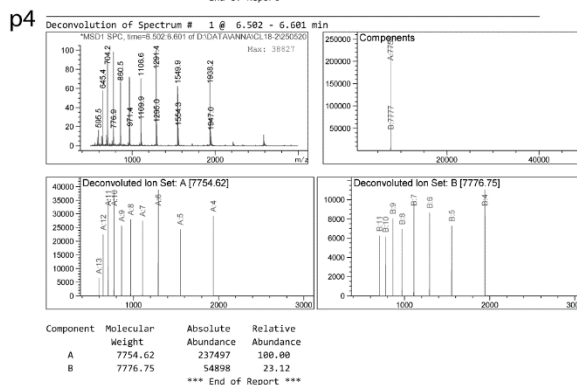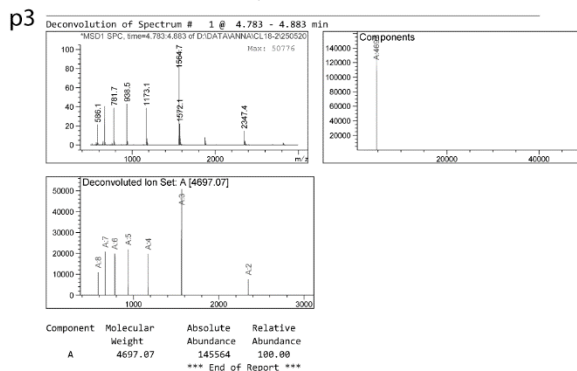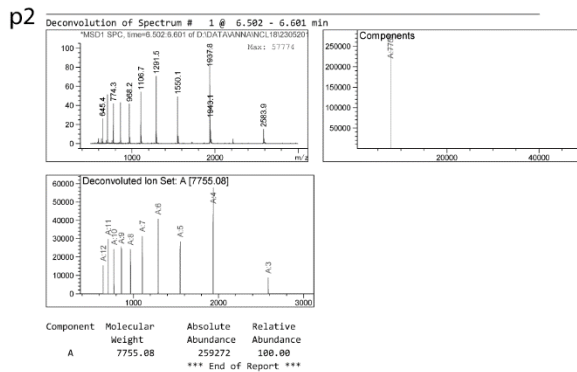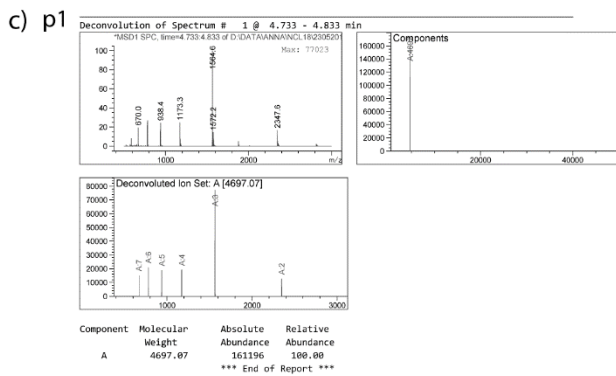
Figure 66: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D13; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D14

a)

b)
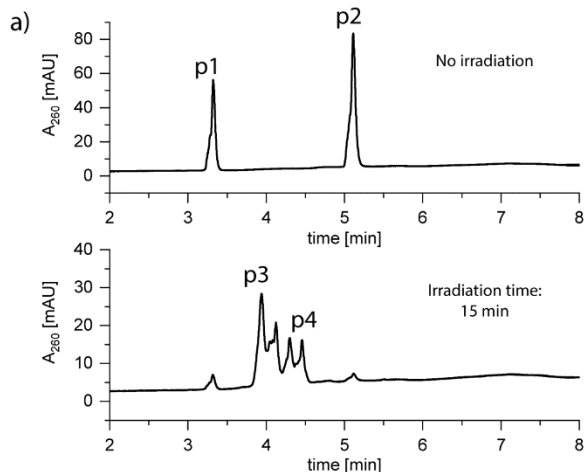
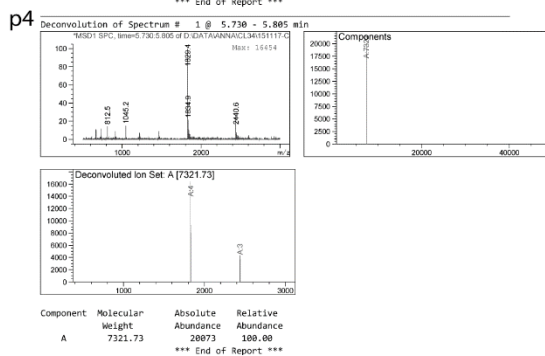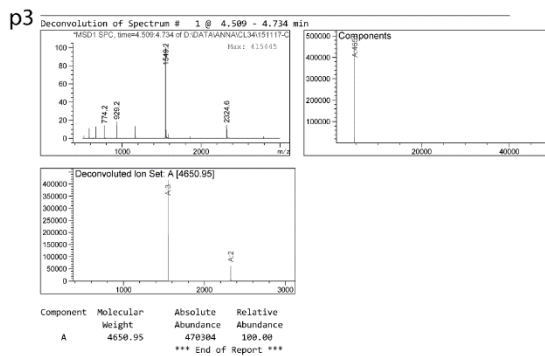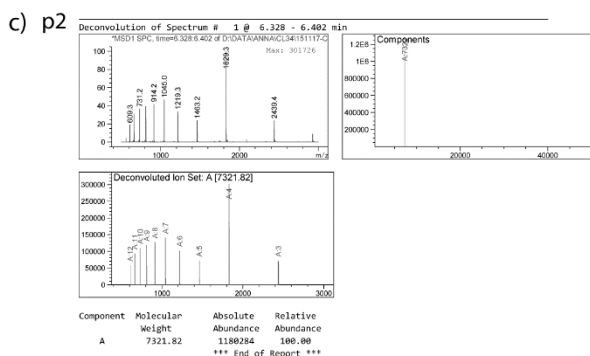| Duplex	ORN	ORN mass	Calc. mass of the CL product
D14	ORN-23	7707.9	12483.8
	ORN-24	4775.9	

c) p1

p2

p3

p4
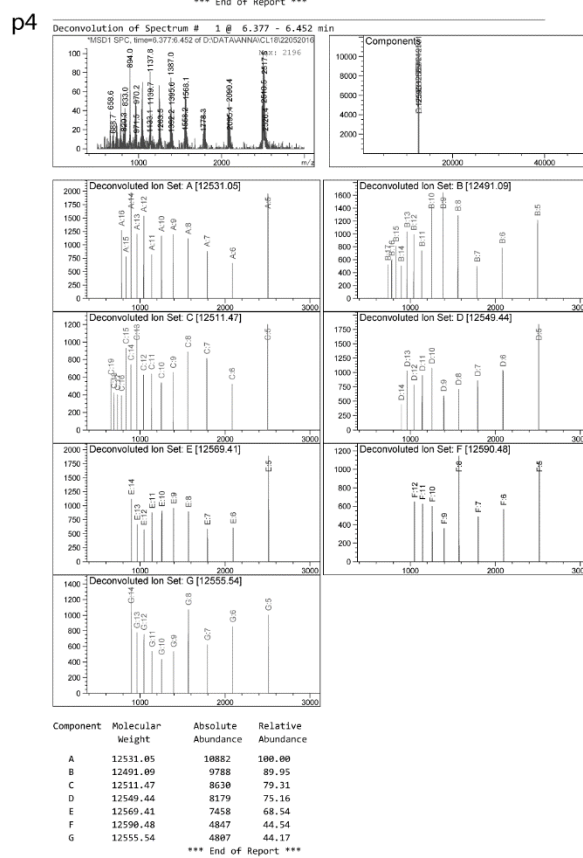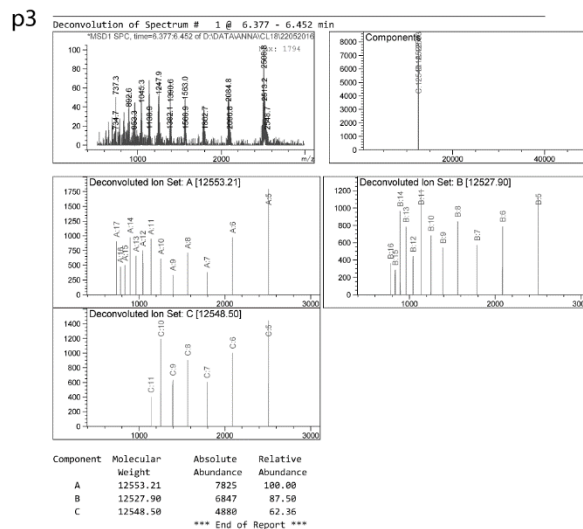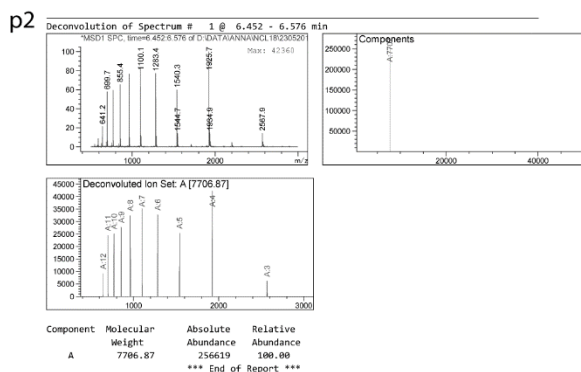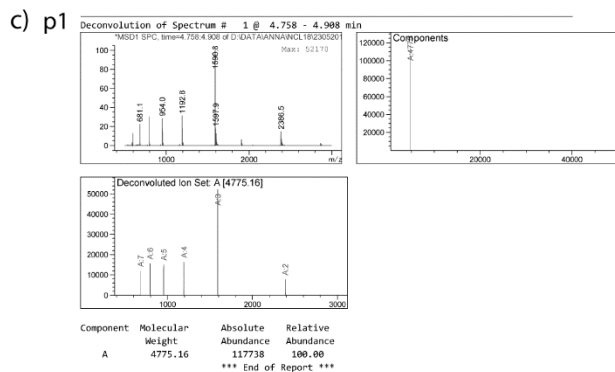
Figure 67: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D14; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D15

a)

b)
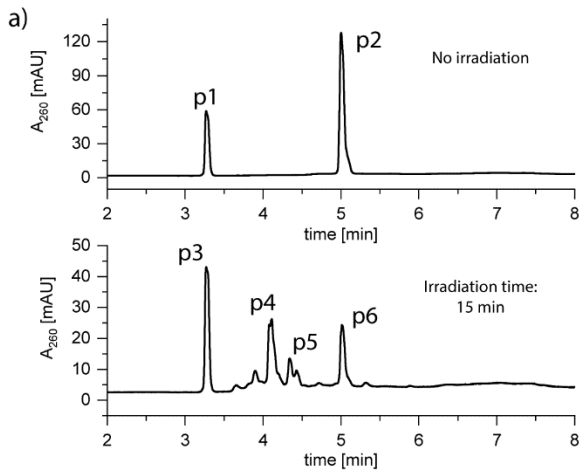
Duplex	ORN	ORN mass	Calc. mass of the CL product
D15	ORN-25	7754.0	12452.0
	ORN-2	4698.0	

c)
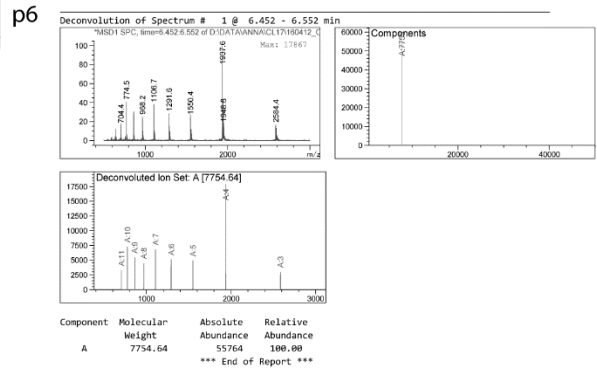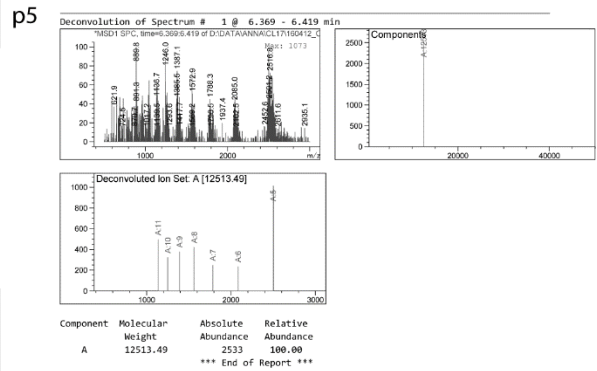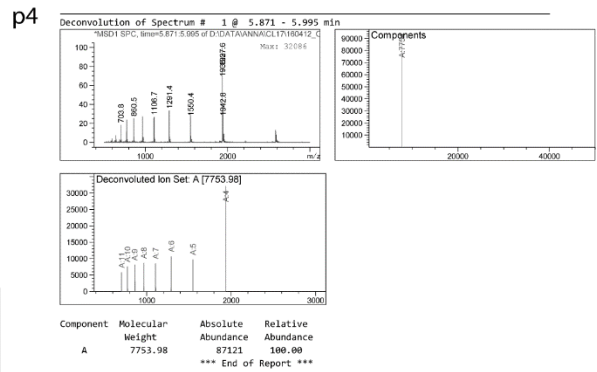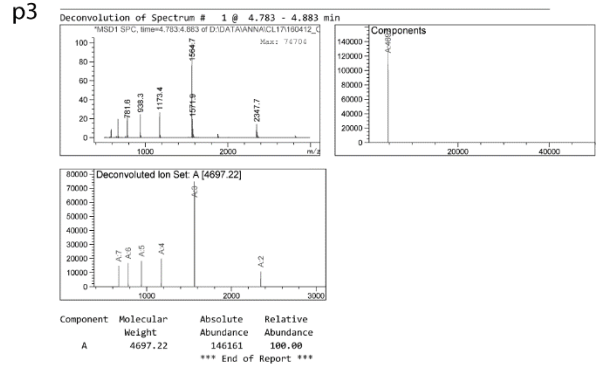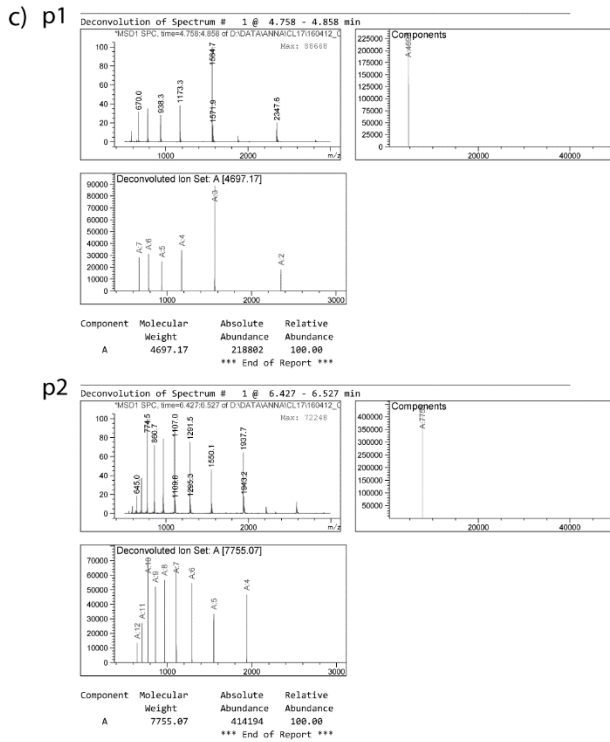
Figure 68: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D15; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
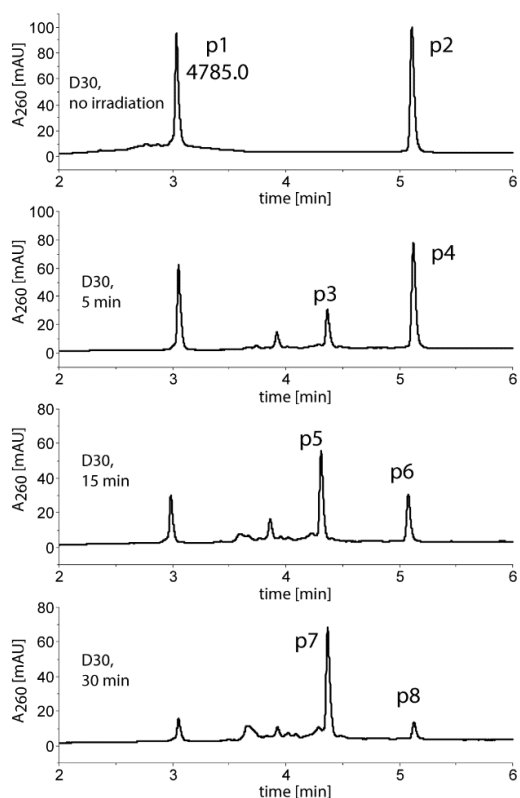
D16

a)

b)
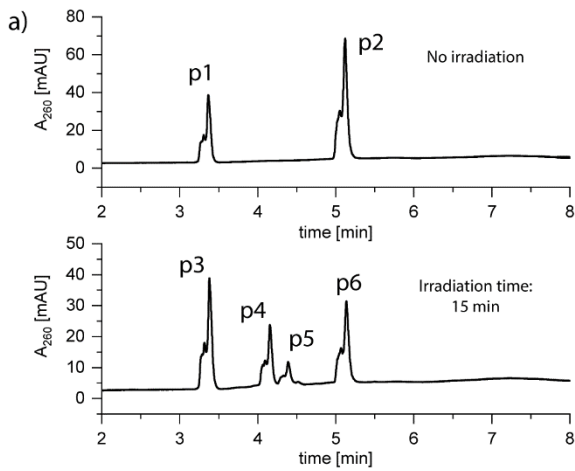
Duplex	ORN	ORN mass	Calc. mass of the CL product
D16	ORN-26	7756.0	12454.0
	ORN-2	4698.0	

c) p1
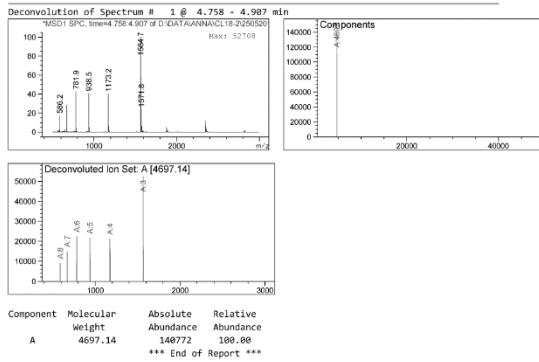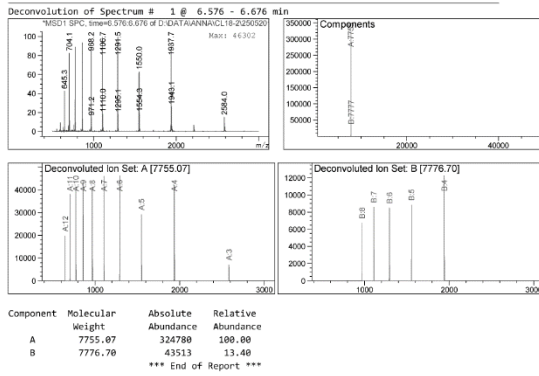
p3
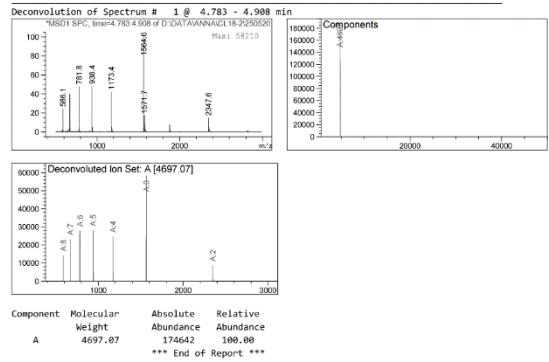
p4
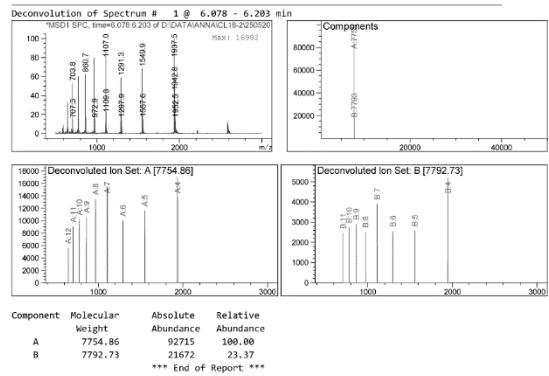
p5
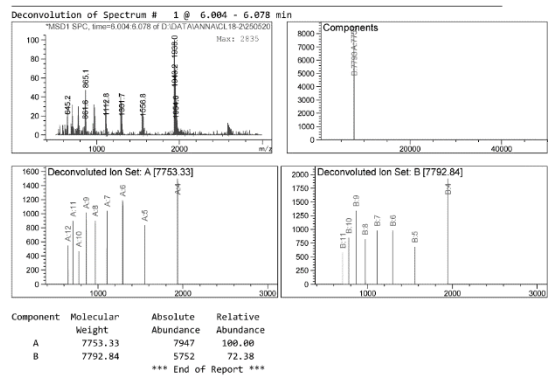
p6
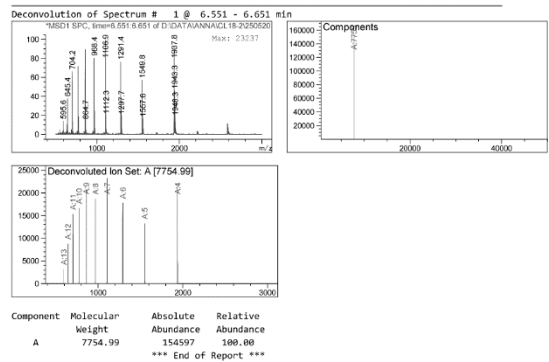
Figure 69: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D16; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
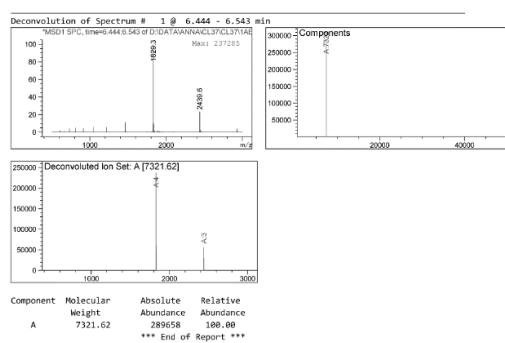
D17
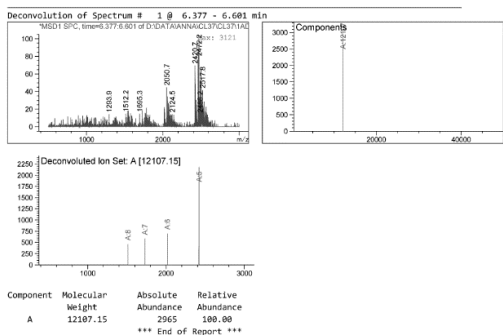
a)

b)
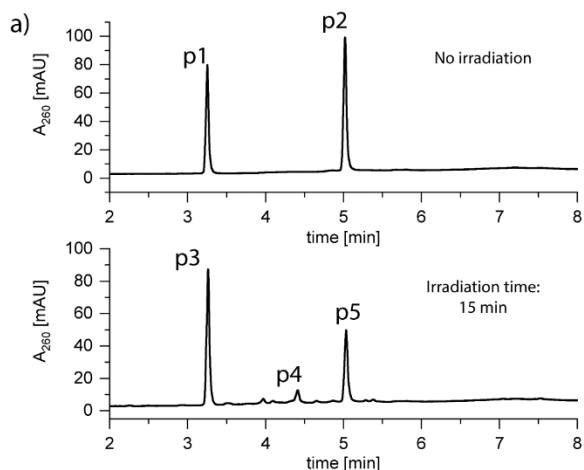
Duplex	ORN	ORN mass	Calc. mass of the CL product
D17	ORN-27	7722.9	12513.9
	ORN-28	4791.0	

c)
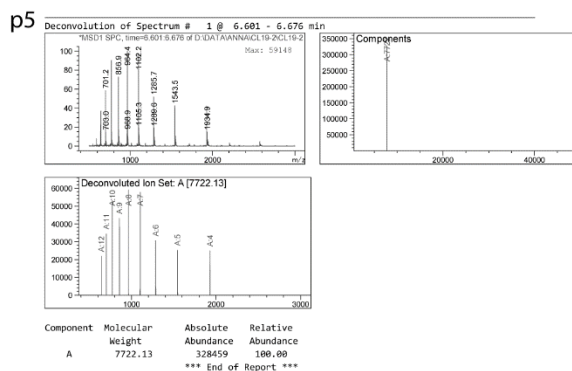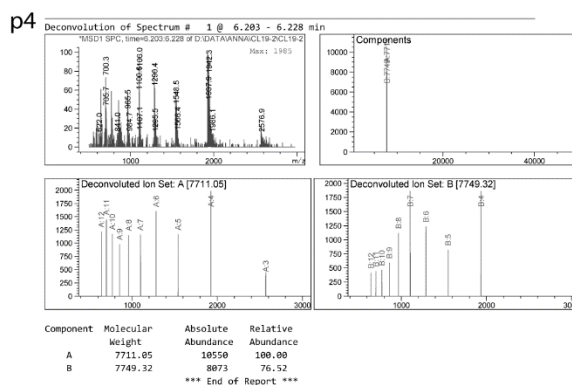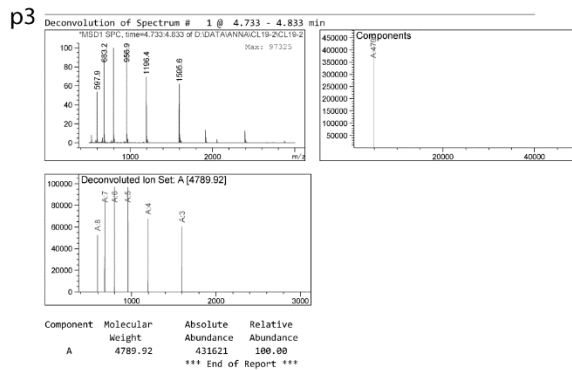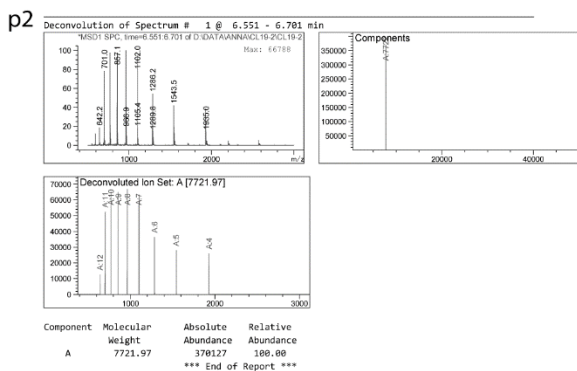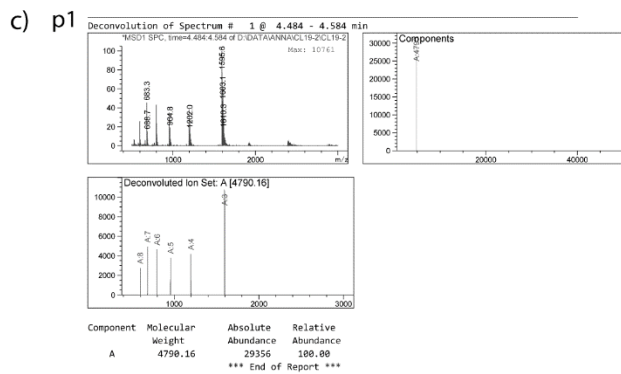
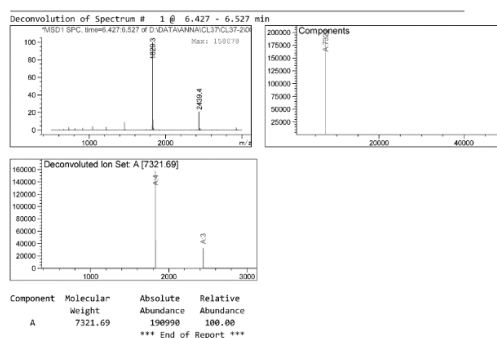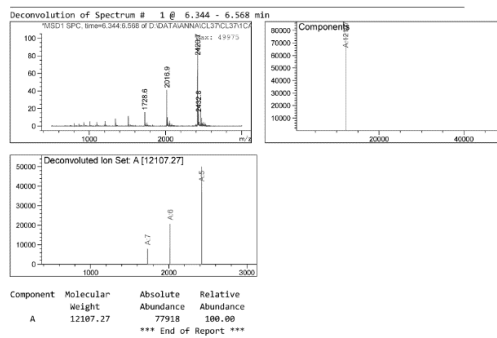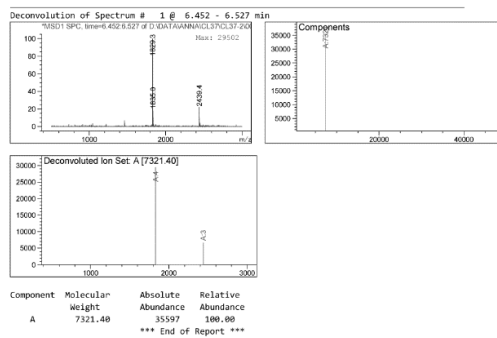**Figure 70:** a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D17; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.1.4.2. miR-124-3p

ORN-29 and ORN-30 irradiation

a)

b)

ORN	ORN mass
ORN-29	6766.4
ORN-30	6766.4

c)
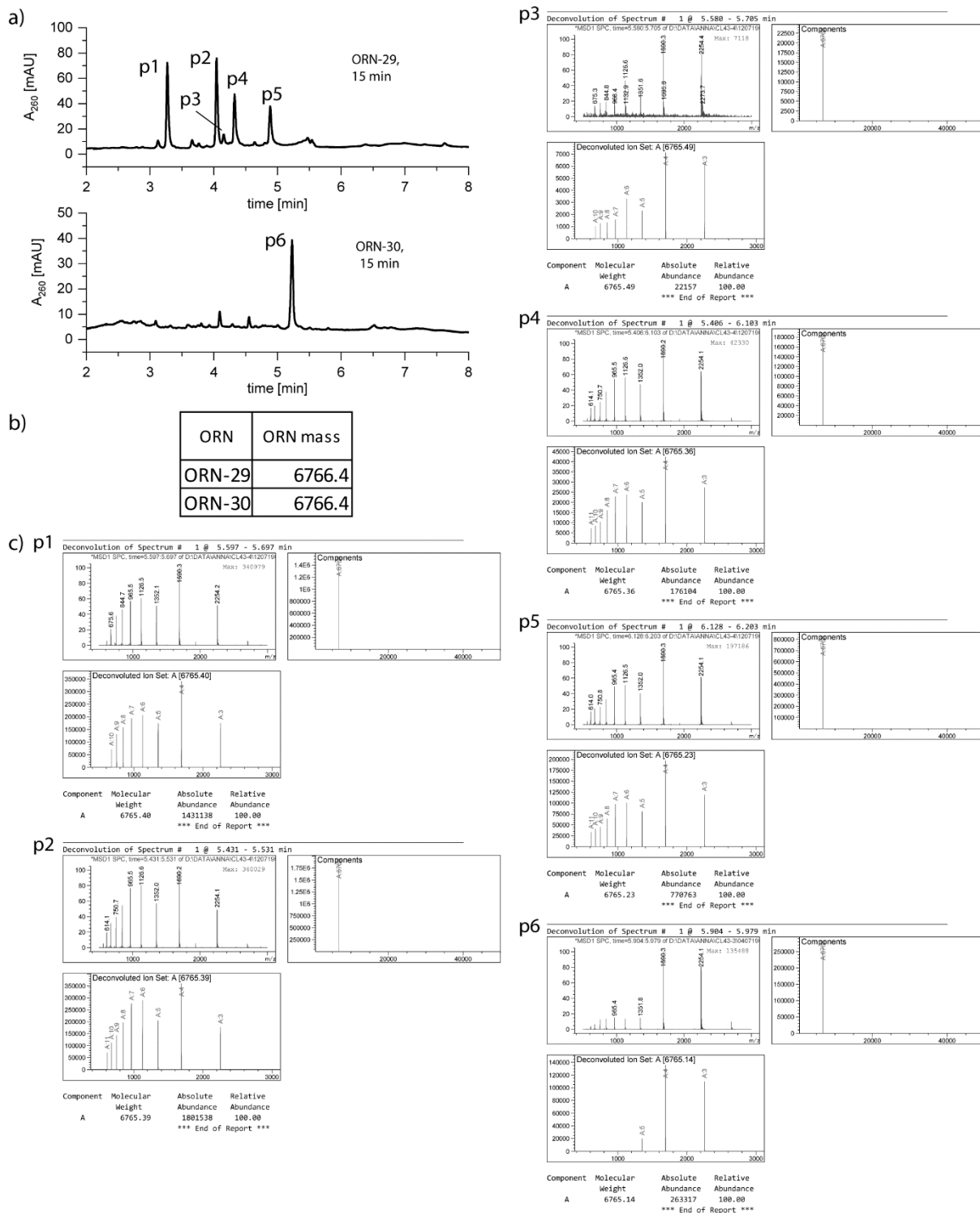
Figure 71: a) RP-HPLC chromatograms from the irradiation of ORN-29 and ORN-30 (single-strands); b) calculated masses of the ORNs; c) deconvoluted masses of the peaks indicated on the chromatograms.

ORN-31 to ORN-35 irradiation

a)

p2

p3

p4

p5
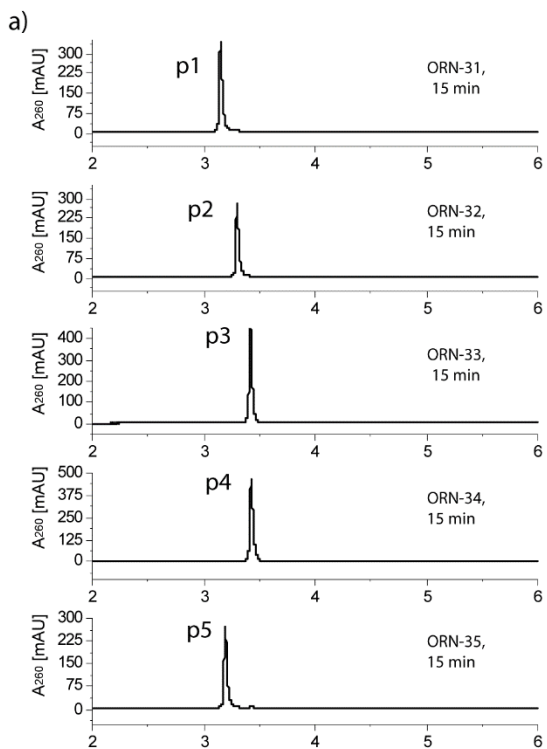
b)

ORN	ORN mass
ORN-31	4687.9
ORN-32	4775.0
ORN-33	7013.2
ORN-34	5920.6
ORN-35	5098.1

c) p1
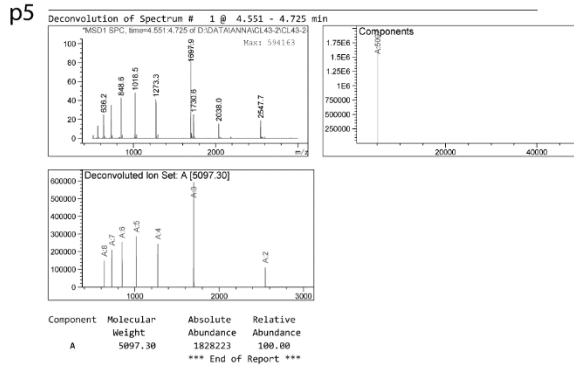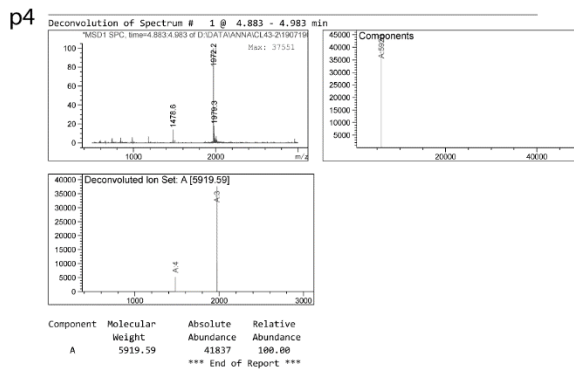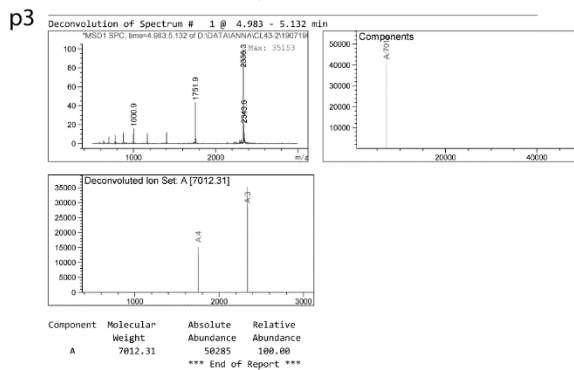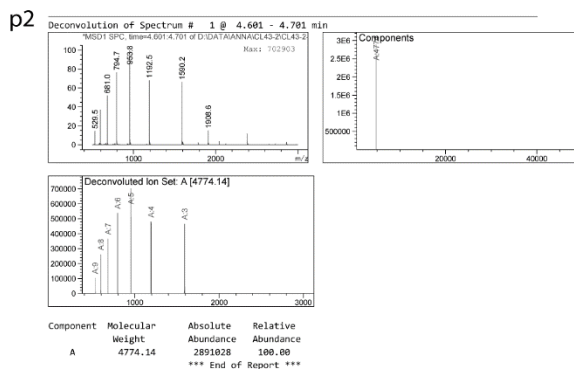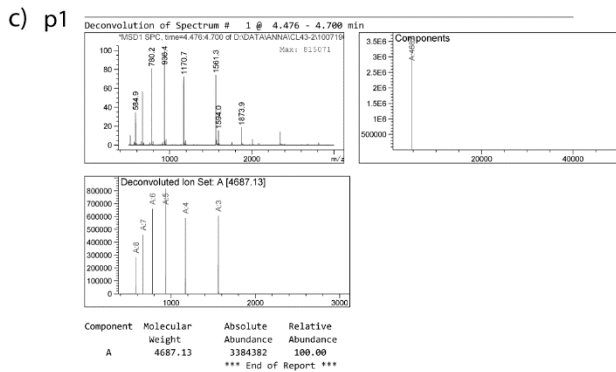
Figure 72: a) RP-HPLC chromatograms from the irradiation of ORN-31 to ORN-35 (single-strands); b) calculated masses of the ORNs; c) deconvoluted masses of the peaks indicated on the chromatograms.

D18

a)

b)
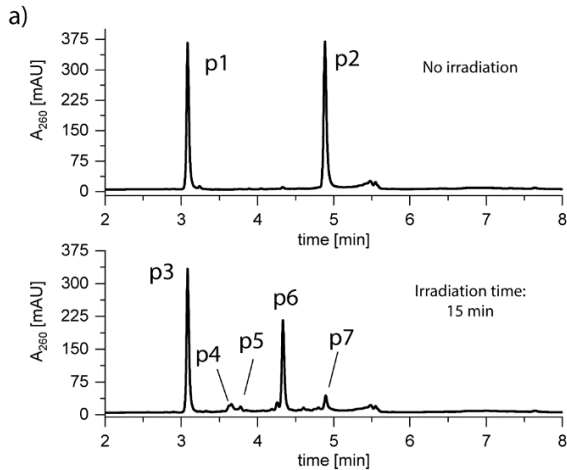
Duplex	ORN	ORN mass	Calc. mass of the CL product
D18	ORN-29	6766.4	11454.3
	ORN-31	4687.9	

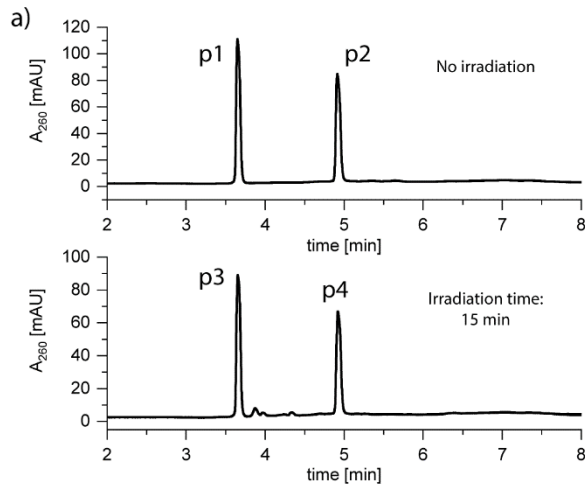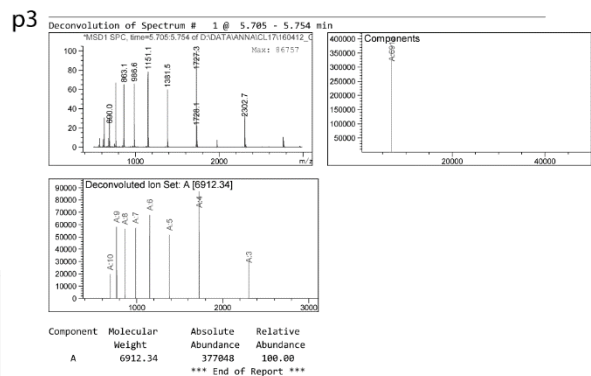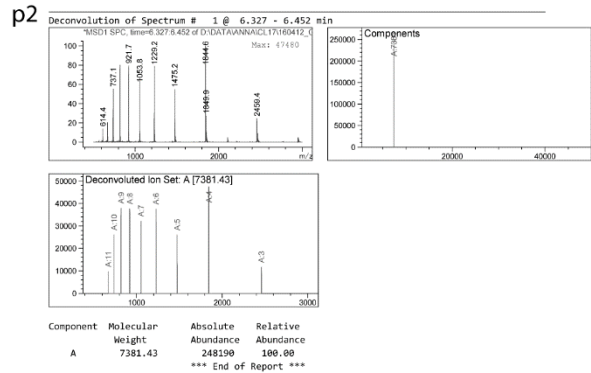c)

p1

p2

p3

p4

p5

p6

p7
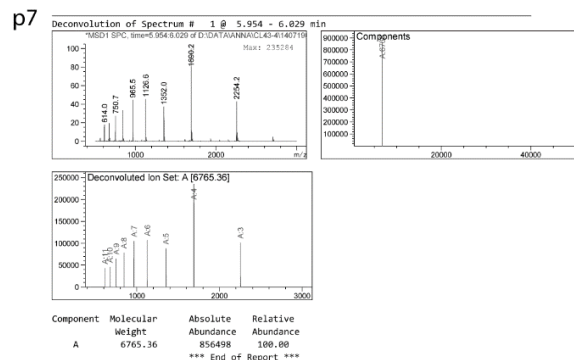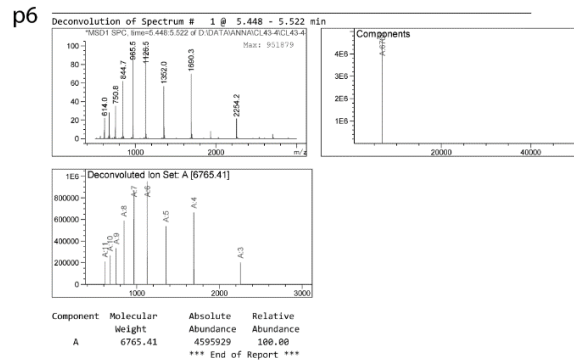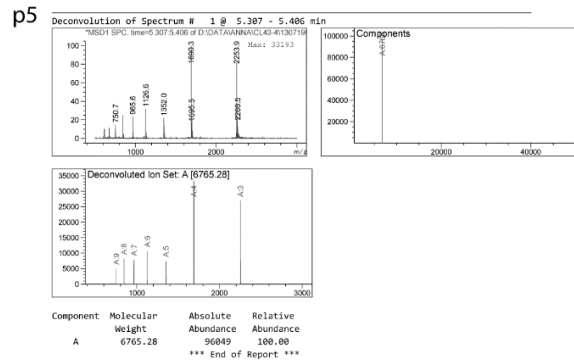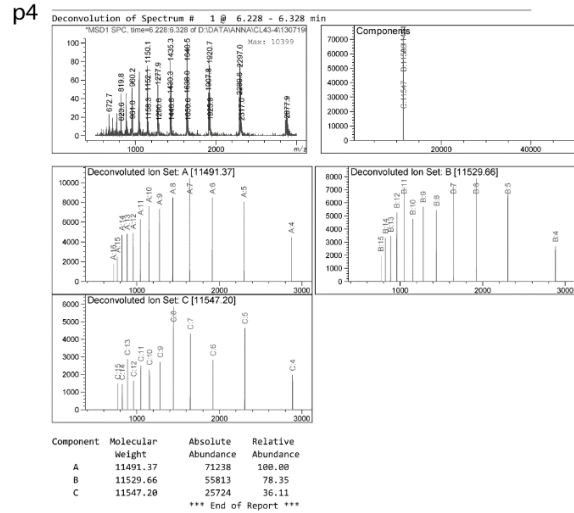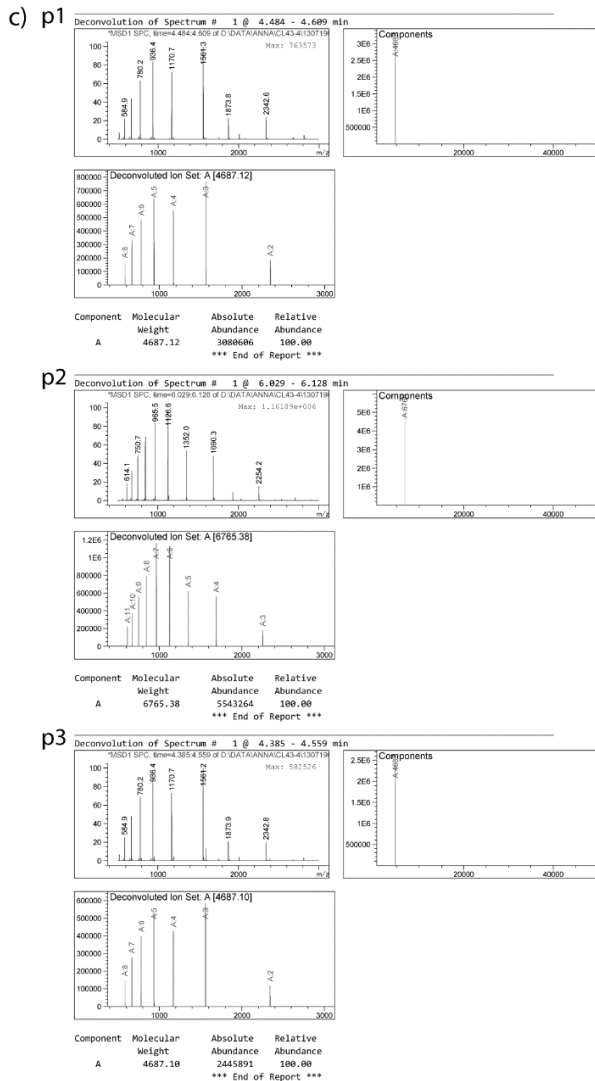
Figure 73: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D18; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D19

a)

b)

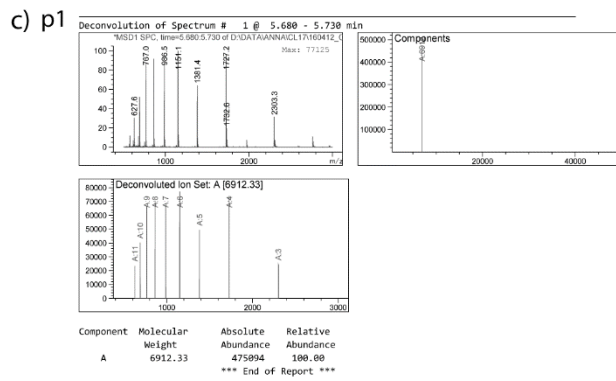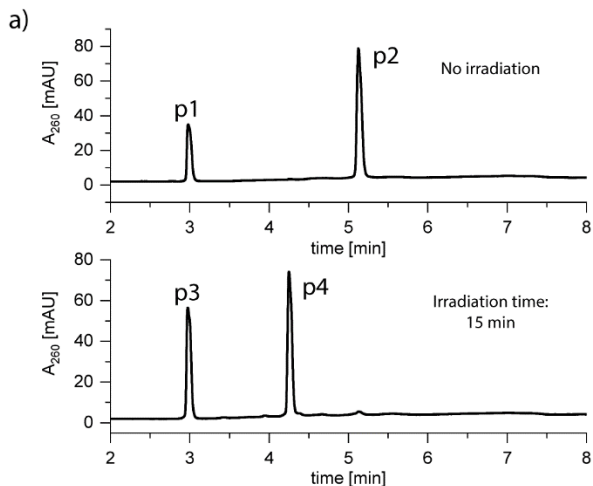| Duplex	ORN	ORN mass	Calc. mass of the CL product
D19	ORN-29	6766.4	11541.4
	ORN-32	4775.0	

c)
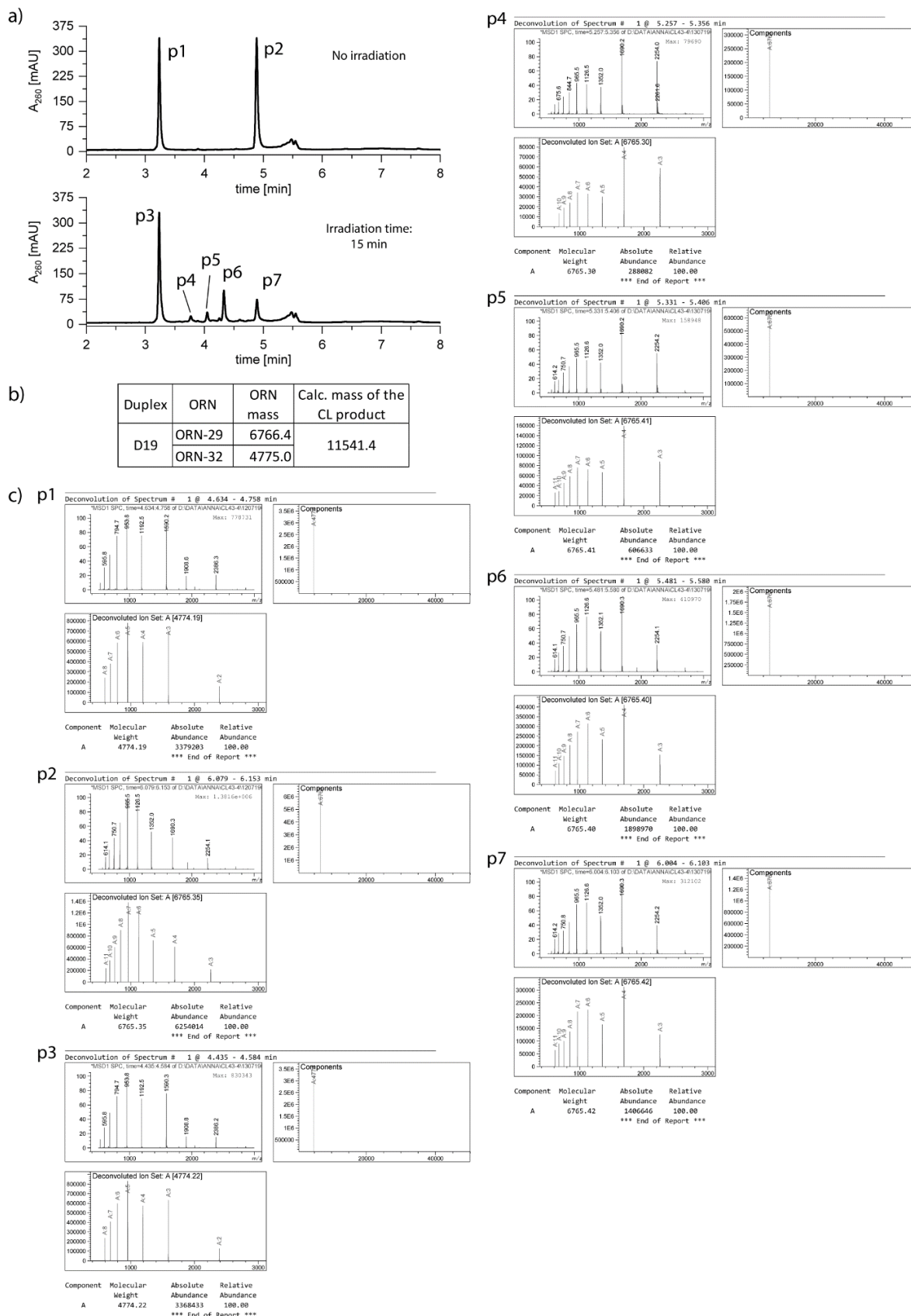
Figure 74: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D19; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D20

Figure 75: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D20; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D21

a)

b)
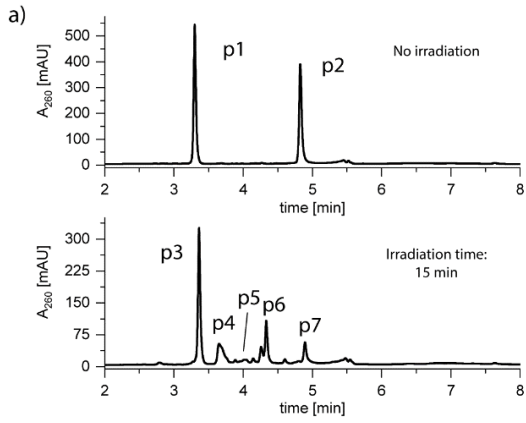
Duplex	ORN	ORN mass	Calc. mass of the CL product
D21	ORN-29	6766.4	12687.0
	ORN-34	5920.6	

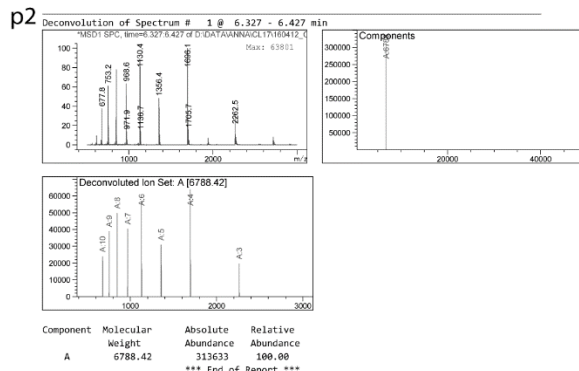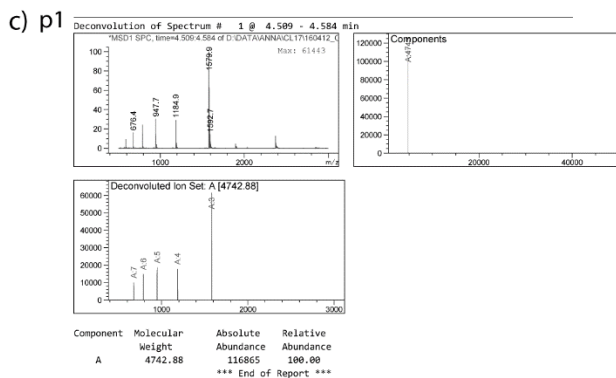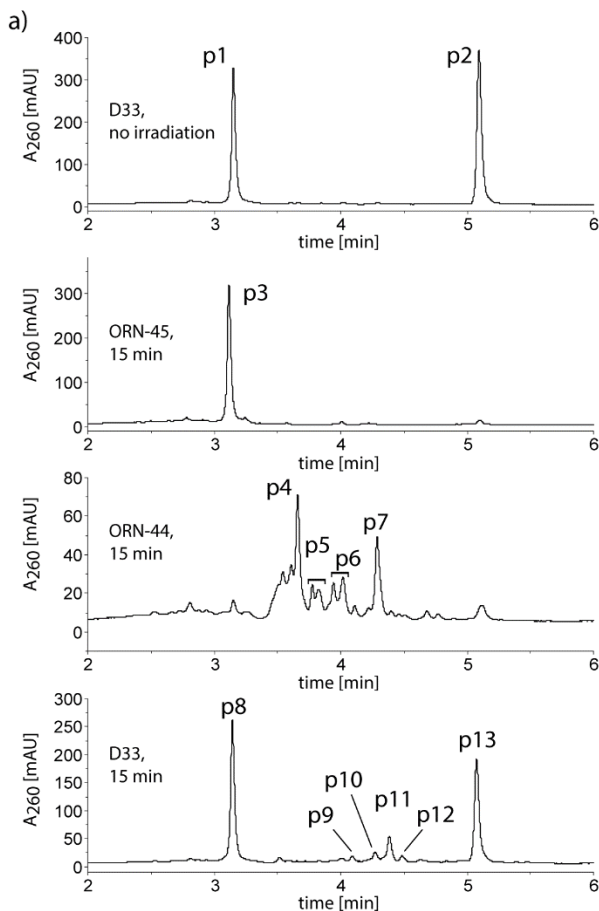c) p1

p2

p3

p4

p5

p6

p7
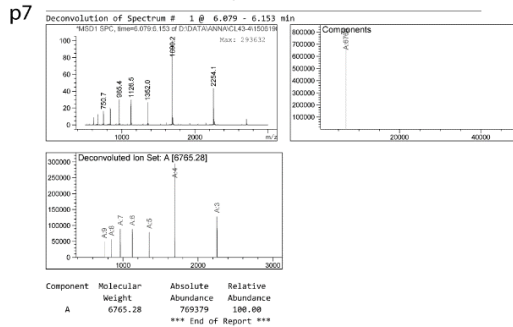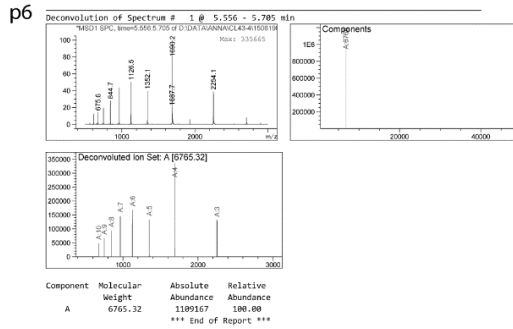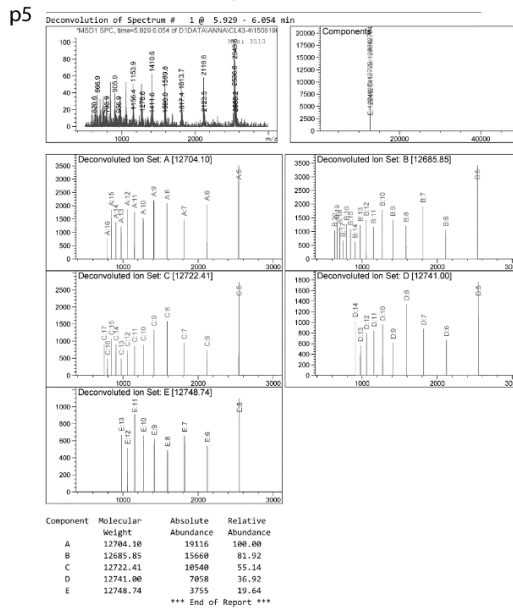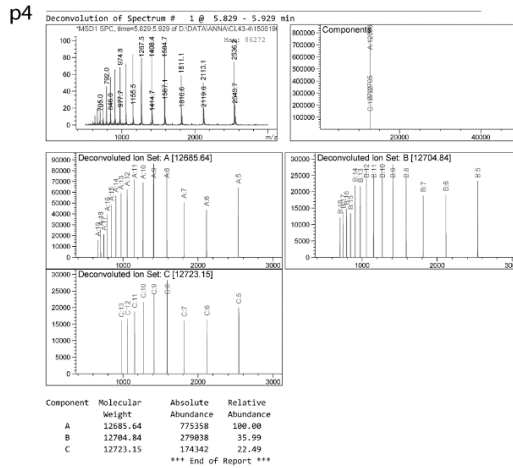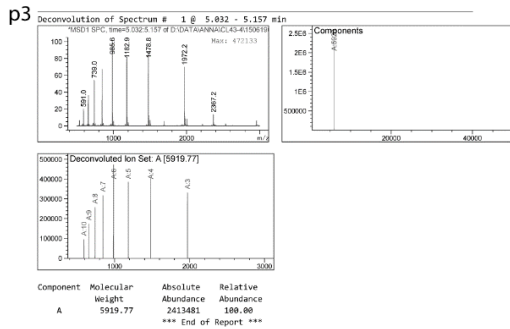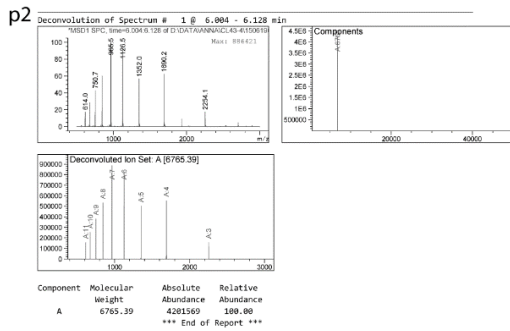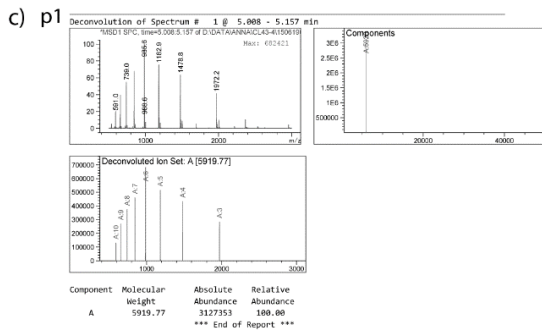
Figure 76: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D21; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D22

a)

b)
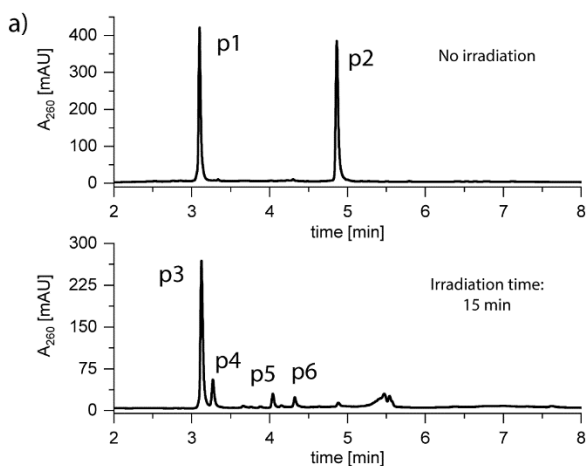
Duplex	ORN	ORN mass	Calc. mass of the CL product
D22	ORN-29	6766.4	11864.5
	ORN-35	5098.1	

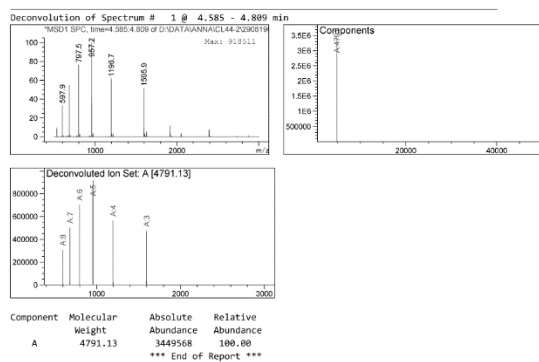c) p1

p2

p3

p4

p5

p6
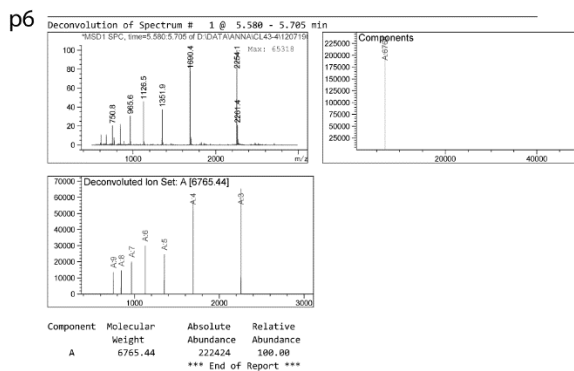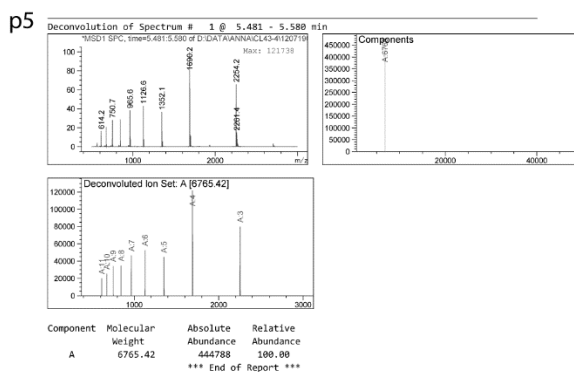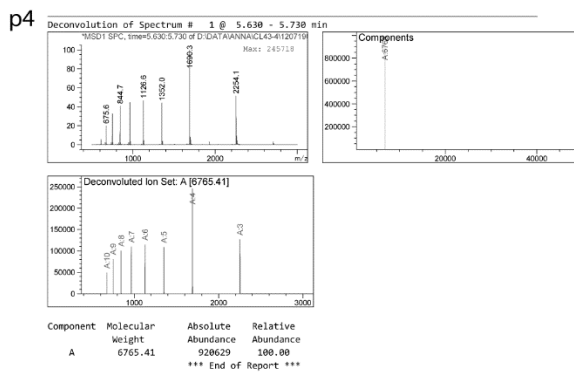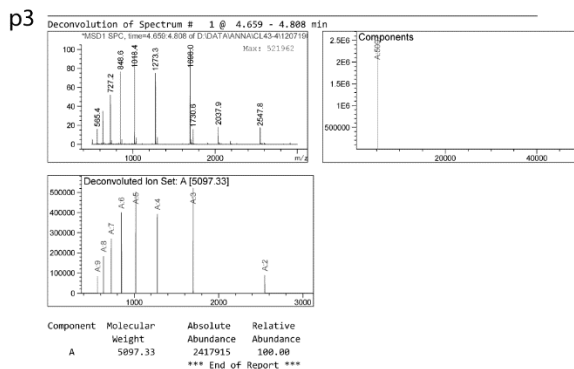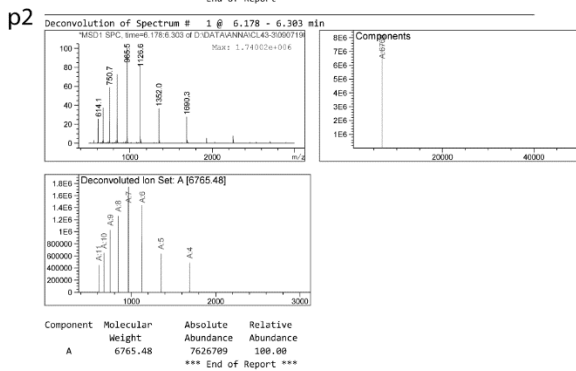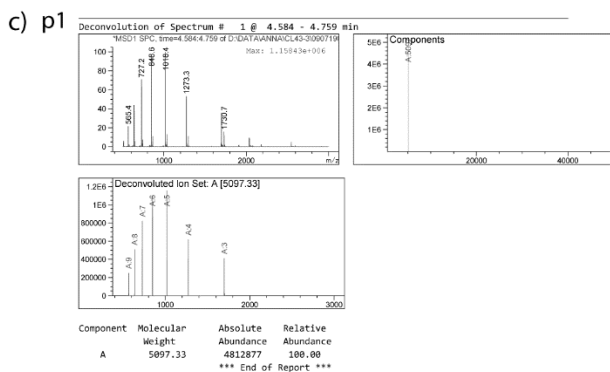
Figure 77: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D22; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
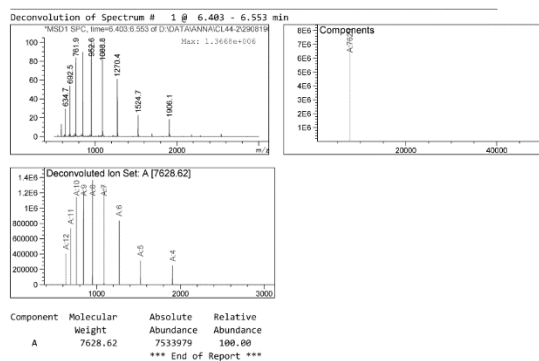
D23
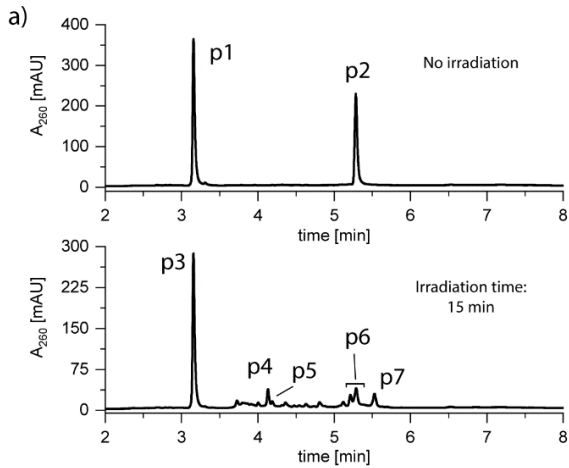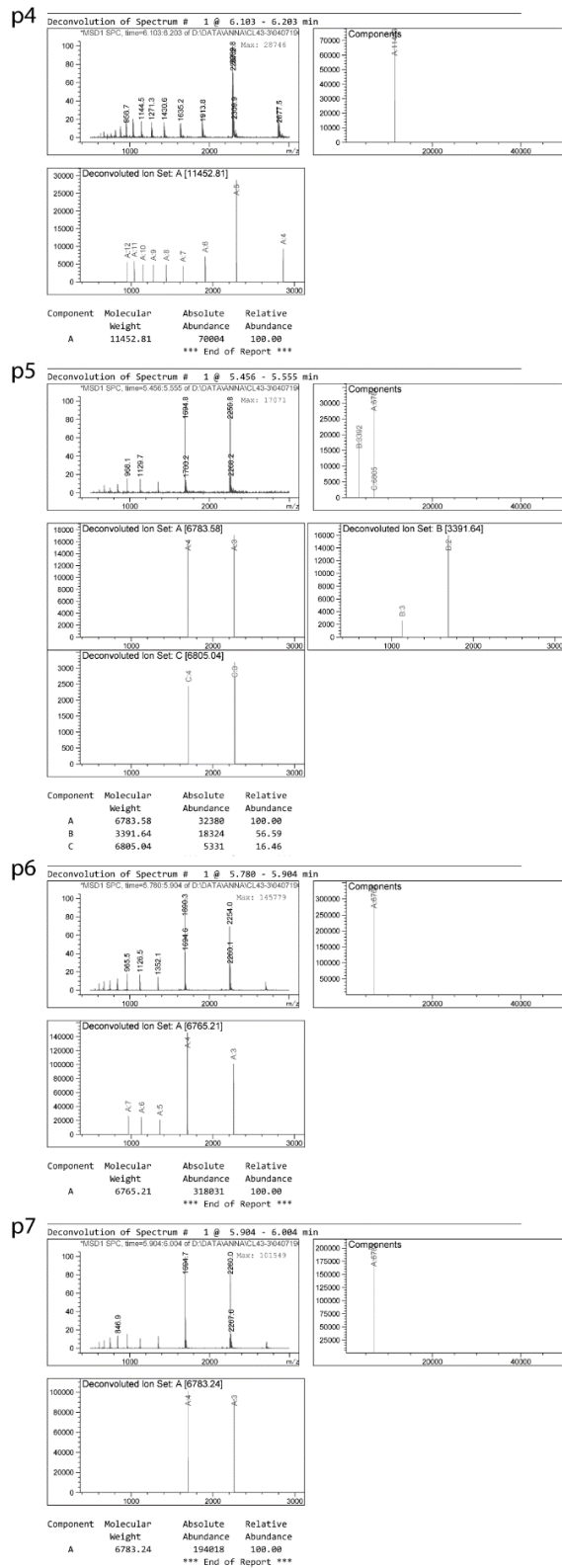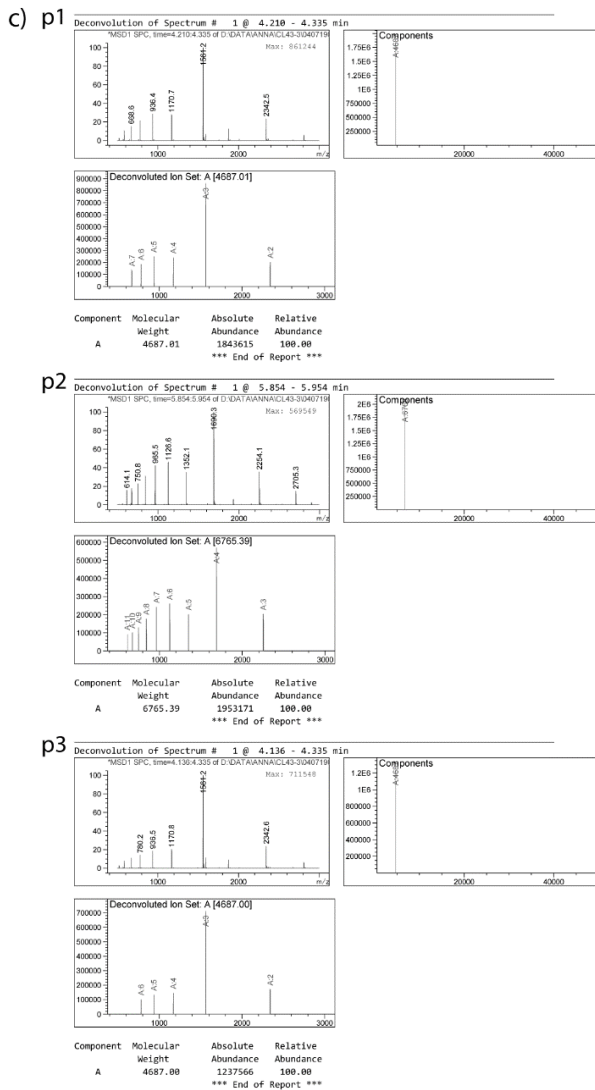
Figure 78: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D23; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D24

a)

b)
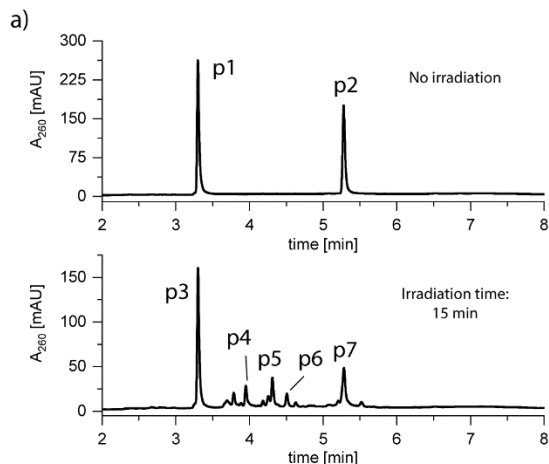
Duplex	ORN	ORN mass	Calc. mass of the CL product
D24	ORN-30	6766.4	11541.4
	ORN-32	4775.0	

c)

Figure 79: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D24; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D25

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D25	ORN-30	6766.4	13779.6
	ORN-33	7013.2	

c)
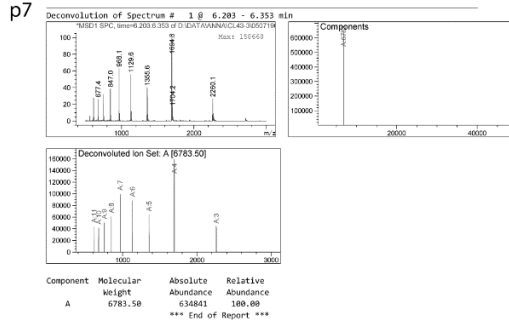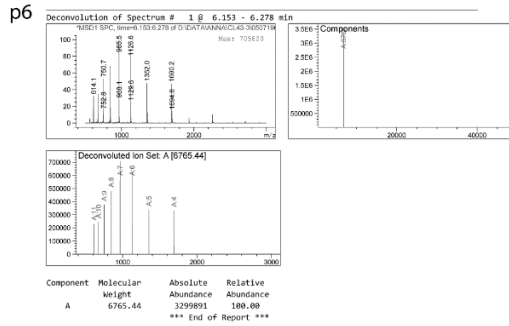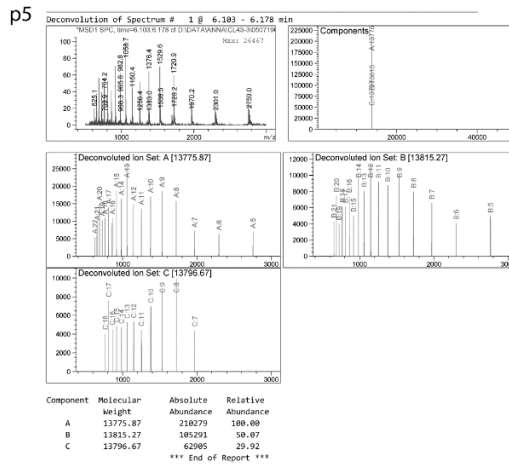
Figure 80: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D25; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D26

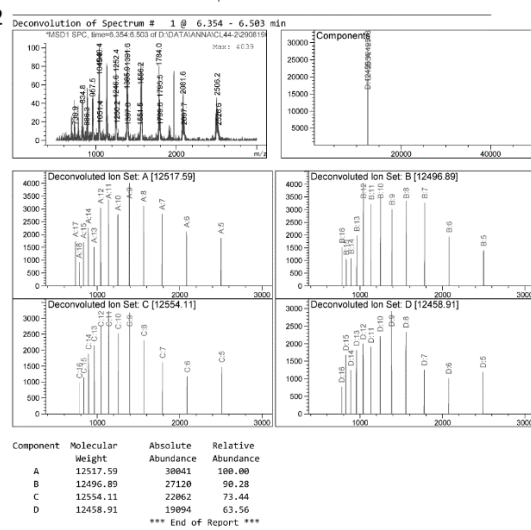**Figure 81:** a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D26; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D27

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D27	ORN-30	6766.4	11864.5
	ORN-35	5098.1	

c)
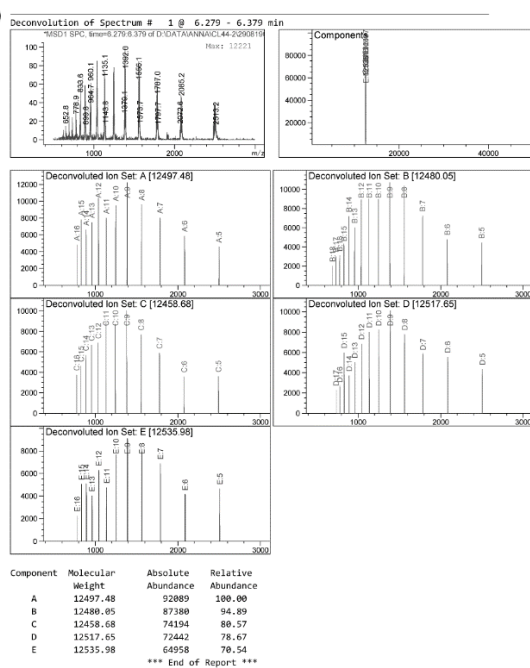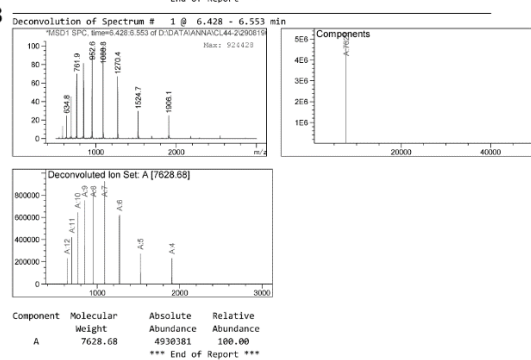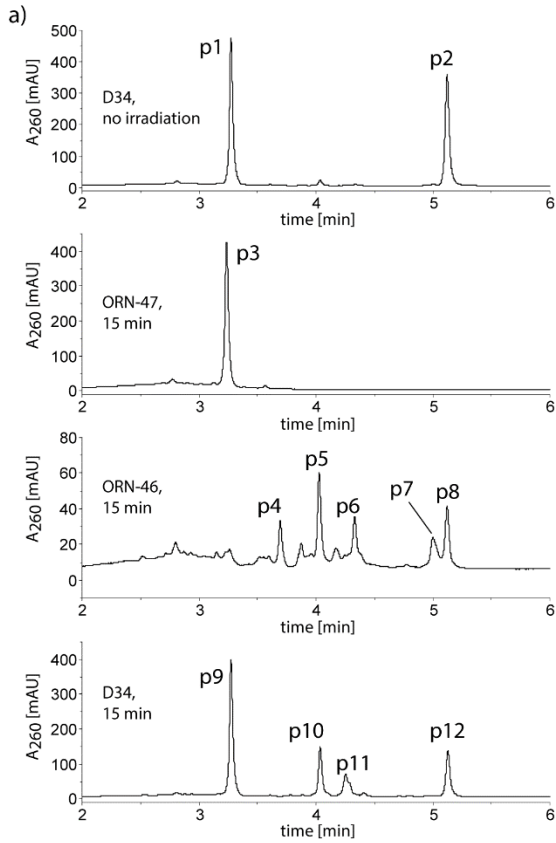

Figure 82: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D27; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.1.4.3. miR-132-3p

D28

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D28	ORN-36	7322.7	12115.6
	ORN-37	4792.9	

c) p1

p2

p3
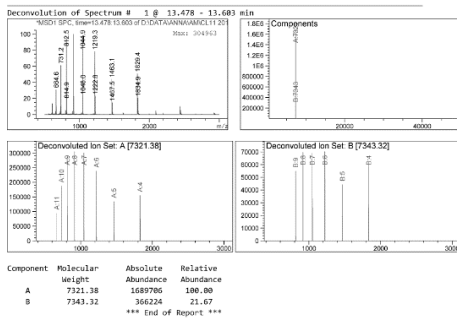
p4
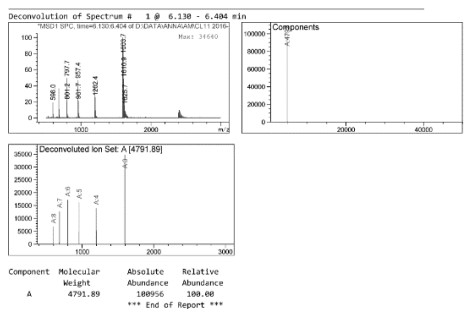
p5
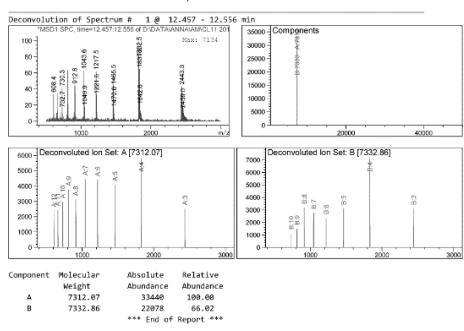
p6
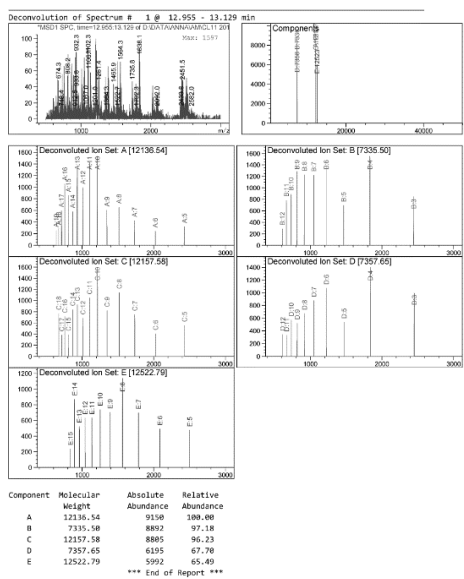
Figure 83: a) LC-MS chromatograms from the *in vitro* cross-linking experiment with duplex D28; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D29

a)

b)
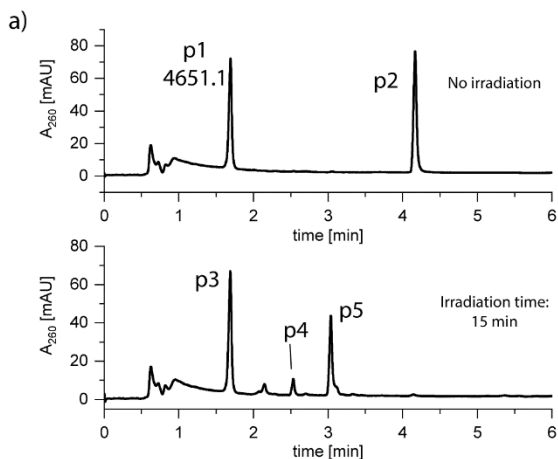
Duplex	ORN	ORN mass	Calc. mass of the CL product
D29	ORN-36	7322.7	11974.5
	ORN-38	4651.8	

c)

Figure 84: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D29; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D30

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D30	ORN-36	7322.7	12108.6
	ORN-39	4785.9	

c) p2

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7321.62	289658	100.00

*** End of Report ***

p3

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	12107.15	2965	100.00

*** End of Report ***

p4

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7321.50	386294	100.00

*** End of Report ***

p5

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	12107.16	27002	100.00

*** End of Report ***

p6

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7321.69	190990	100.00

*** End of Report ***

p7

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	12107.27	77918	100.00

*** End of Report ***

p8

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7321.40	35597	100.00

*** End of Report ***

Figure 85: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D30; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.1.4.4. Miscellaneous miRNAs

D31
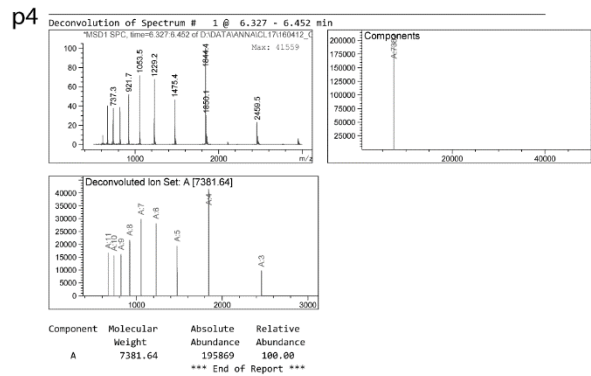
Figure 86: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D31; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D32

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D32	ORN-42	6789.4	11533.3
	ORN-43	4743.9	

c) p1
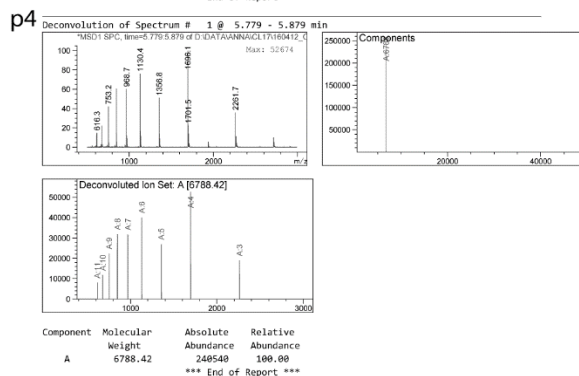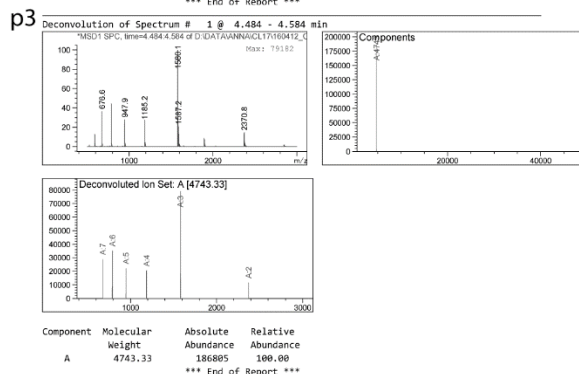
Figure 87: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D32; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

122

D33

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D33	ORN-44	7629.8	12421.7
	ORN-45	4791.9	

c) p1

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4791.13	3449568	100.00

*** End of Report ***

p2

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7628.62	7533979	100.00

*** End of Report ***

p3

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4791.10	3047500	100.00

*** End of Report ***

p4

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7629.30	30830	100.00

*** End of Report ***

p5

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7628.75	701750	100.00

*** End of Report ***

p6

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7628.85	505037	100.00

*** End of Report ***

p7

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7628.72	1586385	100.00

*** End of Report ***

D33 cd

p8

p11

p9

p12

p13

p10

Figure 88: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D33; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D34

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D34	ORN-46	7325.7	12085.6
	ORN-47	4759.9	

c) p1

D34 cd

p8

p9

p10

p11

p12
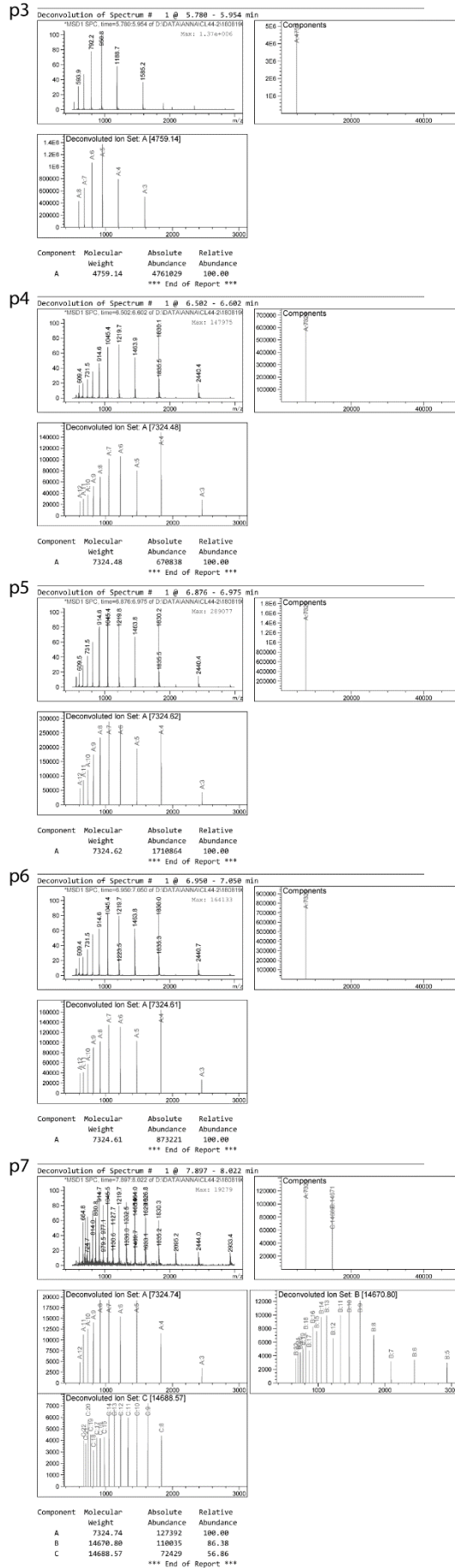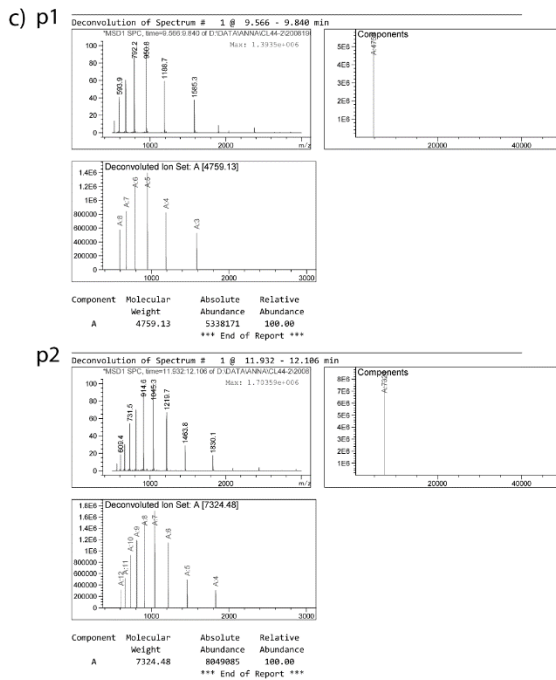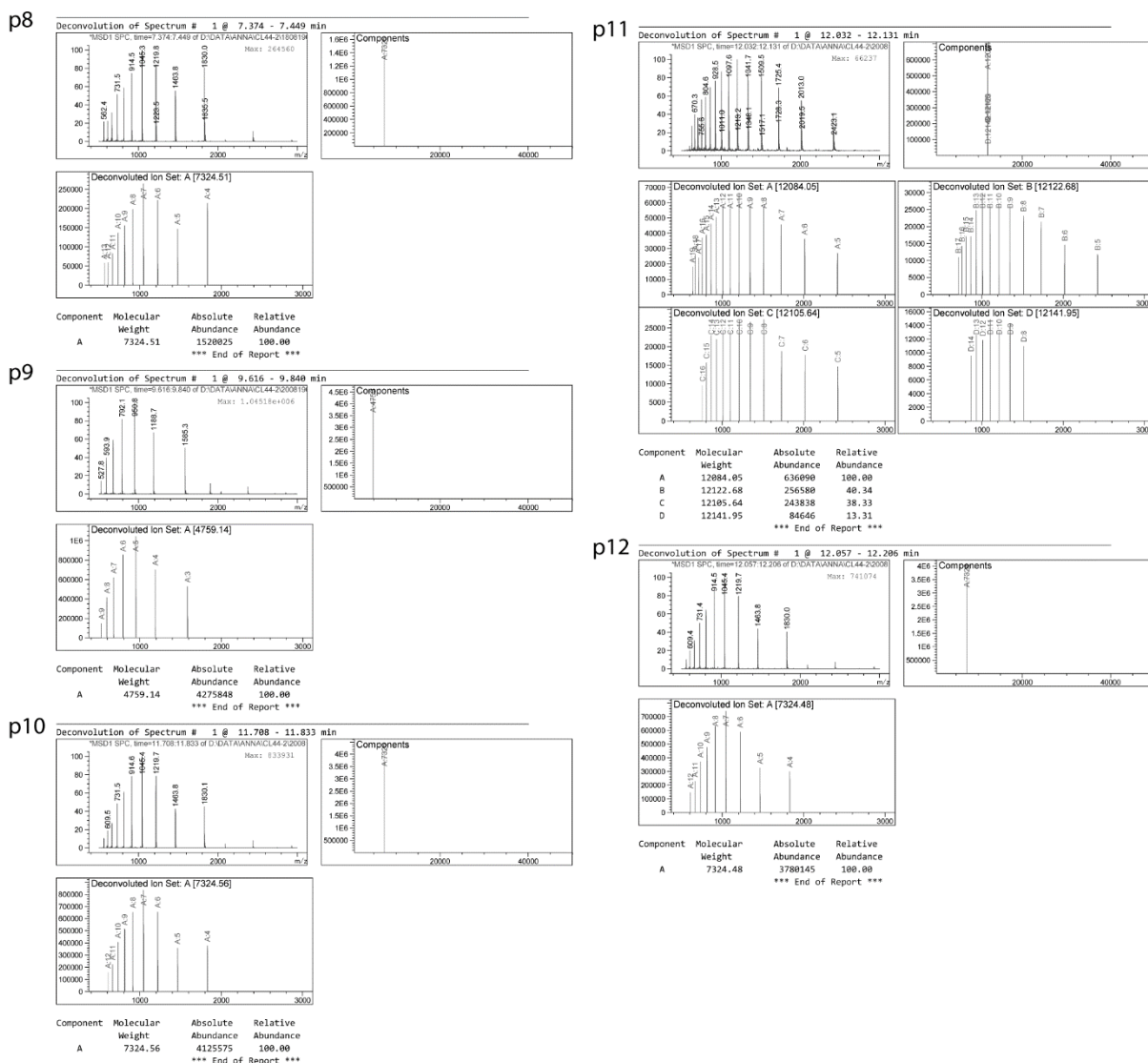
Figure 89: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D34; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.1.5. 20% Denaturing polyacrylamide gel electrophoresis

The general protocol for the denaturing polyacrylamide gel electrophoresis was provided by Artur Łaski.

Glass plates (10 cm cell) and combs (0.75 mm) were washed with RNase Zip, ultrapure water, EtOH and dried before use. 20% polyacrylamide gel was prepared by mixing 10 mL of the Gel Mix (pre-mixed solution containing 50 mL of 40% acrylamide/ bis-acrylamide 29/1 and 25 mL of 2 x Tris-borate-EDTA buffer), 60 µL of the freshly prepared 10% ammonium peroxodisulphate and 10 µL of the tetramethylethylenediamine. The mixture was then poured and left to polymerize.

100 µL of the sample from the cross-linking experiment (either irradiated or non-irradiated) was taken, concentrated, dissolved in 5 µL of ultrapure water and mixed with 5 µL of the loading dye (formamide, 0.5% EDTA, 0.1% xylene cyanol and 0.1% bromphenol blue). Before loading on the gel, samples were denatured by heating at 90°C for 3 min. The gel was pre-run at 200 V for 30 min. After loading, the gel

was run in 0.5 x Tris-borate-EDTA buffer at 170 V for the first 15 min and then at 200 V for 1.5 h. The gel was treated with the staining solution (5 μL of the SYBR® Gold Nucleic Acid Gel Stain, cat. no. S11494, Thermo Fisher Scientific, in 50 mL of 0.5 x Tris-borate-EDTA buffer) with gentle mixing for 40 min. Then, the gel was washed with a running buffer and visualized under the UV light (BioRad, Molecular Imager ChemiDoc XRS System, cat. no. 1708195EDU).

VII.2. Chapter III

VII.2.1. Chemical synthesis

tert-butyl((4-iodobenzyl)oxy)dimethylsilane (1)
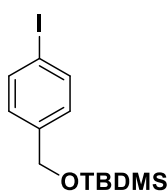
Chemical Formula: $C_{13}H_{21}IOSi$
Molecular Weight: 348.30

The compound was synthesized following the published procedure [278]. A mixture of 4-iodobenzyl alcohol (3.00 g, 12.8 mmol, 1.00 eq), TBDMSCl (2.13 g, 14.1 mmol, 1.10 eq) and imidazole (1.92 g, 28.2 mmol, 2.20 eq) was dissolved in DMF (25 mL) and stirred for 2 h at room temperature. Then, the reaction mixture was concentrated on the rotary evaporator. The liquid residue was taken into EtOAc and washed with water (x3), saturated aqueous $NaHCO_3$ (x3) and brine (x3). The organic phase was dried over Na_2SO_4 and concentrated under reduced pressure. The crude compound was purified by flash column chromatography (gradient: hexane/EtOAc 96/4) to give a clear oil which solidified as a white, oily solid upon placement in the freezer (**1**, 4.36 g, 98%)

¹H NMR (400 MHz, Chloroform-_d_): δ = 7.65 (d, _J_ = 8.3 Hz, 2H), 7.07 (d, _J_ = 8.4 Hz, 2H), 4.67 (s, 2H), 0.93 (s, 9H), 0.09 (s, 6H).

1-(4-(((_tert_-butyldimethylsilyl)oxy)methyl)phenyl)-2,2,2-trifluoroethan-1-one (2)
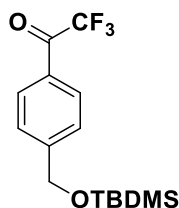
Chemical Formula: $C_{15}H_{21}F_3O_2Si$
Molecular Weight: 318.41

The compound was synthesized modifying the published procedure [278]. To a clear solution of **1** (4.00 g, 11.5 mmol, 1.00 eq) in dry THF (25 mL) at -95°C, _n_-BuLi (1.6 M solution in _n_-hexane; 16 mL, 25.3 mmol, 2.20 eq) was added dropwise over a period of 40 min. During the addition of _n_-BuLi, the temperature was kept below -90°C. The solution was then stirred for 30 min at -90°C. Subsequently, ethyl trifluoroacetate (3.2 mL, 26.4 mmol, 2.30 eq) was added slowly over a period of 15-20 min. After stirring for 1 h below -90°C, the reaction was slowly quenched with saturated NH_4Cl (10 mL). The reaction mixture was stirred for 1 h at around -80°C and then wormed up to room temperature throughout 1.5 h. The reaction mixture was diluted with water and extracted with hexane (x4). The combined organic layers were washed with saturated aqueous $NaHCO_3$ (x3), brine (x3), dried over Na_2SO_4 and concentrated under reduced pressure. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 100/0 to 95/5) to obtain **2** as a transparent oil (2.81 g, 77 %).

¹H NMR (400 MHz, Chloroform-_d_): δ = 8.05 (d, _J_ = 7.8 Hz, 2H), 7.50 (d, _J_ = 8.6 Hz, 2H), 4.83 (s, 2H), 0.96 (s, 9H), 0.12 (s, 6H).

1-(4-(((*tert*-butyldimethylsilyl)oxy)methyl)phenyl)-2,2,2-trifluoroethan-1-one *O*-tosyl oxime (3)
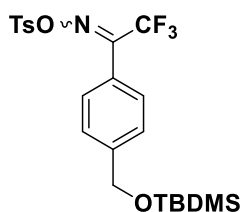
Chemical Formula: $C_{22}H_{28}F_3NO_4SSi$
Molecular Weight: 487.61

The compound was synthesized according to the published procedure with minor modifications [278]. The trifluoroacetophenone (**2**) (2.80 g, 8.79 mmol, 1.00 eq) was dissolved in pyridine/EtOH 1/1 mixture (22 mL). $HONH_2$ x HCl (0.90 g, 13.1 mmol, 1.49 eq) was added and the resulting mixture was stirred at 60°C overnight. Afterwards, the reaction mixture was cooled down to room temperature and concentrated on the rotary evaporator. The crude oily residue was dissolved in dry DCM (18 mL) and then distilled triethylamine (4.5 mL, 32.4 mmol, 3.68 eq), a catalytic amount of DMAP, and *p*-toluenesulphonyl chloride (3.30 g, 19.4 mmol, 2.21 eq) were added. The resulting thick reaction mixture was stirred overnight at room temperature. After the addition of *p*-toluenesulphonyl chloride, the mixture changed colour to orange, and then to brown. The reaction mixture was concentrated under reduced pressure and the residue was dissolved in DCM. The organic layer was washed with saturated aqueous $NaHCO_3$ (x3), brine (x3), dried over Na_2SO_4 and concentrated on the rotary evaporator to give a maroon oil. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 100/0 to 90/10) to give compound **3** in the form of a transparent, thick oil (3.18 g, 74%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.94 − 7.86 (m, 2H), 7.45 − 7.35 (m, 6H), 4.77 (d, *J* = 6.5 Hz, 2H), 2.47 (d, *J* = 7.0 Hz, 3H), 0.95 (d, *J* = 4.8 Hz, 9H), 0.12 (d, *J* = 6.3 Hz, 6H).

3-(4-(((*tert*-butyldimethylsilyl)oxy)methyl)phenyl)-3-(trifluoromethyl)diaziridine (4)
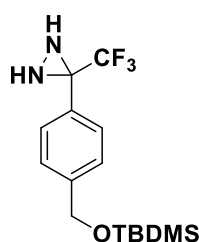
Chemical Formula: $C_{15}H_{23}F_3N_2OSi$
Molecular Weight: 332.44

The compound was synthesized following the published procedure [278]. The tosylated oxime (**3**) (2.80 g, 5.74 mmol, 1.0 eq) was dissolved in THF (5.4 mL) and transferred to the microwave reactor (maximal volume: 20 mL). The reactor was sealed, cooled down to -78°C and 32% aqueous ammonia (10 mL) was added to the reaction mixture under argon atmosphere. The resulting solution was slowly warmed up to room temperature and stirred for around 60 h. Then, the mixture was cooled down to -78°C again and a balloon with argon was connected to the reactor to even out the pressure. After a few minutes, the reaction mixture was warmed up to room temperature and the reactor was opened to allow the excess NH_3 to escape slowly. The mixture was partitioned between EtOAc and water. The organic layer was washed with saturated aqueous $NaHCO_3$ (x3), brine (x3) dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 100/0 to 95/5) to give **4** as a transparent oil (0.49 g, 89%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.58 (d, *J* = 8.1 Hz, 2H), 7.38 (d, *J* = 8.5 Hz, 2H), 4.76 (s, 2H), 2.76 (s, 1H), 2.17 (s, 1H), 0.95 (s, 9H), 0.11 (s, 6H).

3-(4-(((*tert*-butyldimethylsilyl)oxy)methyl)phenyl)-3-(trifluoromethyl)-3*H*-diazirine (5)
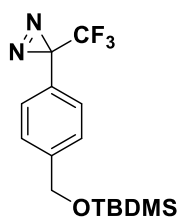
Chemical Formula: $C_{15}H_{21}F_3N_2OSi$
Molecular Weight: 330.43

The compound was synthesized according to the reported protocol [278]. The diaziridine (**4**) (1.70 g, 5.11 mmol, 1.00 eq) was dissolved in MeOH (17 mL). Triethylamine (1.7 mL, 12.2 mmol, 2.39 eq) and I_2 (1.43 g, 5.63 mmol, 1.10 eq) were subsequently added (upon addition of I_2 the solution changed the colour to brown/violet-brown). The mixture was stirred overnight at room temperature. Then, the reaction mixture was partitioned between EtOAc and aqueous $Na_2S_2O_3$. The organic layer was washed with brine (x3), dried over Na_2SO_4 and concentrated under reduced pressure. The crude compound was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 100/0 to 97/3) to give the protected diazirine (**5**) as a transparent oil (1.60 g, 95%).

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.35 (d, *J* = 8.5 Hz, 2H), 7.16 (d, *J* = 8.1 Hz, 2H), 4.74 (s, 2H), 0.94 (s, 9H), 0.09 (s, 6H).

(4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)phenyl)methanol (6)
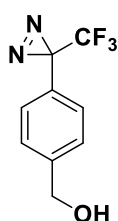
Chemical Formula: $C_9H_7F_3N_2O$
Molecular Weight: 216.16

The compound was synthesized following the published procedure [278]. To a solution of the protected diazirine (**5**) (0.73 g, 2.22 mmol, 1.00 eq) in THF (9.4 mL), TBAF (2.4 mL, 2.44 mmol, 1.10 eq) was added and the reaction mixture was stirred at room temperature for 1.5 h. Then, the reaction mixture was concentrated on the rotary evaporator. As the ^{1}H NMR of the crude sample showed that it was a mixture of the protected and deprotected compound, the deprotection was repeated. The crude was dissolved in THF (9.4 mL) and TBAF (2.4 mL, 2.44 mmol, 1.10 eq) was added. The reaction mixture was stirred at room temperature for 2 h and then the solvent was evaporated under reduced pressure. The crude product was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 80/20 to 70/30) to give the deprotected diazirine (**6**) in the form of a bright yellow oil, which crystallized as a beige solid in the freezer (0.38 g, 80%).

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.40 (d, *J* = 8.6 Hz, 2H), 7.20 (d, *J* = 8.0 Hz, 2H), 4.72 (s, 2H), 1.75 (s, 1H).

4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl methanesulphonate (7)
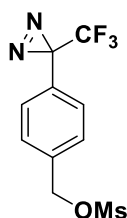
Chemical Formula: $C_{10}H_9F_3N_2O_3S$
Molecular Weight: 294.25

The compound was synthesized according to the published procedure [313]. To a solution of a diazirine (**6**) (0.10 g, 0.45 mmol, 1.00 eq) in dry DCM (1.7 mL) triethylamine (0.17 mL, 1.22 mmol, 2.70 eq) and methane-sulphonyl chloride (0.07 mL, 0.95 mmol, 2.10 eq) were added at 0°C and the reaction mixture was stirred at 0°C for 15 min. Then, the reaction mixture was concentrated on the rotary evaporator. The crude product was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 80/20 to 75/25) to obtain the mesylated product (**7**) as a dark-orange, transparent oil (0.07 g; 52%).

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.46 (d, *J* = 8.5 Hz, 2H), 7.24 (d, *J* = 8.1 Hz, 2H), 5.24 (s, 2H), 2.97 (s, 3H).

3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3*H*-diazirine (8)
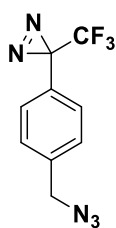
Chemical Formula: $C_9H_6F_3N_5$
Molecular Weight: 241.18

The compound was synthesized according to the reported protocol [313]. To a solution of **7** (0.05 g, 0.17 mmol, 1.00 eq) in DMF (1.7 mL) sodium azide (0.03 g, 0.50 mmol, 3.00 eq) and TBAI (0.01 g, 0.02 mmol, 0.10 eq) were added at room temperature. Then the reaction mixture was heated to 60°C and stirred for 5 h. The reaction mixture was concentrated on the rotary evaporator and partitioned between diethyl ether and water. The organic phase was washed with water (x2), brine (x3), dried over Na_2SO_4 and concentrated under reduced pressure to give the slightly impure product, 3-(4-(azidomethyl)phenyl)-3-(trifluoromethyl)-3*H*-diazirine, **8** in the form of a yellow oil (0.03 g, 61%).

1H NMR (400 MHz, Chloroform-*d*): δ = 7.36 (d, *J* = 8.4 Hz, 2H), 7.22 (d, *J* = 8.1 Hz, 2H), 4.37 (s, 2H).

(2*R*,3*R*,4*R*,5*R*)-2-((benzoyloxy)methyl)-5-((4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl)oxy)tetra-hydrofuran-3,4-diyl dibenzoate (9)
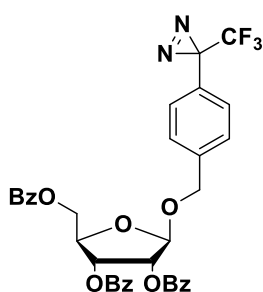
Chemical Formula: $C_{35}H_{27}F_3N_2O_8$
Molecular Weight: 660.60

The compound was synthesized following the published procedure [278]. The glassware was pri-dried in the oven before use. 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-*β*-D-ribofuranose (0.95 g, 1.88 mmol, 1.10 eq) was dissolved in dry DCM (10 mL) and cooled down to -30°C. Trimethylsilyl trifluoromethanesulphonate (TMSOTf) (0.29 mL, 1.61 mmol, 0.94 eq) was added dropwise, followed by the addition of **6** (0.37 g, 1.71 mmol, 1.00 eq) dissolved in dry DCM (6.0 mL) upon the previous co-evaporation with dry THF. The reaction mixture was stirred at -30°C for 2.5 h. The reaction was quenched at -30°C with saturated aqueous $NaHCO_3$ (10 mL). The reaction mixture was partitioned between EtOAc and water. The organic layer was washed with saturated aqueous $NaHCO_3$ (x3), brine (x3), dried over Na_2SO_4 and concentrated under reduced pressure. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 95/5 to 80/20) to give a beige, crystalline-like glycosylation product **9** (1.07 g, 95%; traces of the solvent still present).

1H NMR (400 MHz, Chloroform-*d*): δ = 8.05 − 7.97 (m, 4H), 7.89 (dd, *J* = 8.4, 1.3 Hz, 2H), 7.58 (t, *J* = 7.5 Hz, 1H), 7.55 − 7.47 (m, 2H), 7.43 (t, *J* = 7.8 Hz, 2H), 7.36 − 7.26 (m, 6H), 7.12 (d, *J* = 8.0 Hz, 2H), 5.96 − 5.89 (m, 1H), 5.77 (d, *J* = 4.9 Hz, 1H), 5.31 (s, 1H), 4.83 − 4.73 (m, 3H), 4.57 (d, *J* = 12.3 Hz, 1H), 4.55 − 4.48 (m, 1H).

(2*R*,3*S*,4*R*,5*R*)-2-(hydroxymethyl)-5-((4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl)oxy)tetra-hydrofuran-3,4-diol (10)

The compound was synthesized following the published procedure [278]. The glassware was pri-dried in the oven before use. Compound **9** (0.96 g, 1.45 mmol, 1.00 eq) was dissolved in dry MeOH (5.0 mL). In a separate vial, a solution of sodium methoxide in dry MeOH was prepared (1.2 g/ 3 mL) and added to the reaction mixture. Additional portions of sodium methoxide were added after 2 h and 5 h of stirring. After another 2 h of stirring, the reaction was quenched with aqueous saturated NH_4Cl (5.0 mL). The reaction mixture was concentrated on the rotary evaporator. The crude was dissolved in DCM, filtered and concentrated under reduced pressure. The product was suspended on the silica gel and purified by flash column chromatography (gradient: chloroform/MeOH: 100/0 to 85/15) to give the deprotected product **10** as a white solid (0.48 g, 94%).

Chemical Formula: $C_{14}H_{15}F_3N_2O_5$
Molecular Weight: 348.28

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.35 (d, *J* = 8.5 Hz, 2H), 7.19 (d, *J* = 8.1 Hz, 2H), 5.04 (s, 1H), 4.77 (d, *J* = 12.2 Hz, 1H), 4.55 (d, *J* = 12.2 Hz, 1H), 4.41 (t, *J* = 5.5 Hz, 1H), 4.16 – 4.08 (m, 2H), 3.82 (dd, *J* = 11.9, 3.4 Hz, 1H), 3.68 (dd, *J* = 11.9, 4.1 Hz, 1H).

(2*R*,3*S*,4*R*,5*R*)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-5-((4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl)oxy)tetrahydrofuran-3,4-diol (11)

The compound was synthesized according to the published procedure [278]. The glassware was pri-dried in the oven before use. Compound **10** (0.28 g, 0.80 mmol, 1.00 eq) was co-evaporated with a dry pyridine twice and then dissolved in dry pyridine (4.0 mL). DMTrCl (0.33 g, 0.96 mmol, 1.20 eq) was added in three portions within one hour and the reaction mixture was stirred for 3 h at room temperature. The reaction mixture was quenched with 0.5 mL of EtOH, concentrated and partitioned between EtOAc and aqueous saturated $NaHCO_3$. The organic layer was washed with brine (x1), dried over Na_2SO_4 and concentrated under reduced pressure. The crude was combined with the crude from other reaction, suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 75/25 to 60/40 + 1% of triethylamine) to yield the 5'-*O*-DMTr-protected product **11** as a beige foam (total yield from the combined reactions: 0.77 g, 75%).

Chemical Formula: $C_{35}H_{33}F_3N_2O_7$
Molecular Weight: 650.65

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.45 (d, *J* = 7.0 Hz, 2H), 7.33 (d, *J* = 8.9 Hz, 4H), 7.28 – 7.19 (m, 5H), 7.11 (d, *J* = 8.1 Hz, 2H), 6.79 (dd, *J* = 8.9, 3.2 Hz, 4H), 5.03 (s, 1H), 4.70 (d, *J* = 12.1 Hz, 1H), 4.42 (d, *J* = 12.0 Hz, 1H), 4.33 (dd, *J* = 6.5, 4.8 Hz, 1H), 4.16 – 4.09 (m, 2H), 3.77 (d, *J* = 2.0 Hz, 6H), 3.31 (dd, *J* = 5.2, 2.1 Hz, 2H).

(2*R*,3*R*,4*R*,5*R*)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((*tert*-butyldimethyl-silyl)oxy)-5-((4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl)oxy)tetrahydrofuran-3-ol (12)
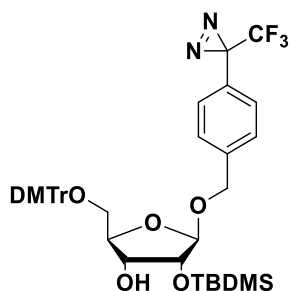
Chemical Formula: $C_{41}H_{47}F_3N_2O_7Si$
Molecular Weight: 764.91

The compound was synthesized following the reported protocol [278]. The glassware was pri-dried in the oven before use. Compound **11** (around 0.05 g, 0.08 mmol, 1.00 eq) was dissolved in dry pyridine (2.0 mL). Subsequently, imidazole (0.02 g, 0.24 mmol, 3.10 eq) and TBDMSCl (0.02 g, 0.11 mmol, 1.40 eq) were added and the reaction mixture was stirred for 4 h at 0°C. The reaction mixture was concentrated on the rotary evaporator, dissolved in EtOAc and partitioned between EtOAc and saturated aqueous NaHCO$_3$. The organic layer was washed with brine (x1), dried over Na$_2$SO$_4$ and concentrated under reduced pressure. Due to the small amount of the product, the crude was purified by preparative TLC (TLC Silica gel 60 F$_{254}$ glass plates, Merck. Eluent composition: hexane/EtOAc 80/20 + 1% of triethylamine). Silica gel containing the separated product was collected and suspended in EtOAc. The suspension was stirred for 2 h and then filtered. The filtrate was concentrated on the rotary evaporator to give the 2'-*O*-TBDMS-protected product **12** (16 mg, 27%) with traces of the solvent still present.

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.48 (d, *J* = 7.1 Hz, 2H), 7.35 (d, *J* = 8.9 Hz, 4H), 7.28 – 7.21 (m, 5H), 7.12 (d, *J* = 8.1 Hz, 2H), 6.77 (dd, *J* = 9.0, 2.3 Hz, 4H), 4.95 (d, *J* = 1.3 Hz, 1H), 4.78 (d, *J* = 12.2 Hz, 1H), 4.46 (d, *J* = 12.2 Hz, 1H), 4.19 (dd, *J* = 4.6, 1.5 Hz, 1H), 4.17 – 4.08 (m, 2H), 3.76 (d, *J* = 1.5 Hz, 6H), 3.35 (dd, *J* = 10.0, 3.4 Hz, 1H), 3.17 – 3.12 (m, 1H), 2.45 (d, *J* = 7.6 Hz, 1H), 0.90 (s, 9H), 0.10 (d, *J* = 7.1 Hz, 6H).

2,2,2-trifluoro-1-(4-iodophenyl)ethan-1-one (13)
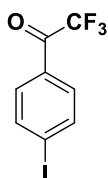
Chemical Formula: $C_8H_4F_3IO$
Molecular Weight: 300.02

The compound was synthesized according to the published procedure [314]. *n*-BuLi (1.6 M in *n*-hexane; 30 mL, 0.05 mol, 1.60 eq) was added slowly into dry THF (180 mL) with stirring under argon atmosphere. The resulting solution was cooled down to −90°C. Then, a solution of 1,4-diiodobenzene (10.0 g, 0.03 mol, 1.00 eq) in dry THF (25 mL) was added dropwise over 20 min. The resulting solution was stirred at around −90°C for 1 h. Subsequently, ethyl trifluoroacetate (11.2 mL, 0.09 mol, 3.10 eq) was added dropwise. The reaction mixture was stirred at around −90°C for 10 min and then warmed up to room temperature over a period of 3 h. After 1 h, saturated aqueous NH$_4$Cl (5.0 mL) was added to quench the reaction. The reaction mixture was divided between water (100 mL) and hexane (100 mL). The organic layer was washed with brine (x2), dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 99/1 to 95/5) to yield a slightly yellow, transparent liquid as a product (**13**) (4.30 g, 47%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.98 – 7.90 (m, 2H), 7.80 – 7.72 (m, 2H).

3-(4-iodophenyl)-3-(trifluoromethyl)-3*H*-diazirine (14)
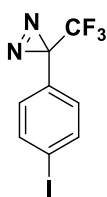
Chemical Formula: $C_8H_4F_3IN_2$
Molecular Weight: 312.03

The compound was synthesized following the published procedure [314].

Step 1. A mixture of **13** (2.00 g, 6.67 mmol, 1.00 eq), $HONH_2$ x HCl (0.79 g, 11.3 mmol, 1.70 eq) and dry pyridine (0.8 mL) in MeOH (3.8 mL) was heated under reflux for 2.5 h. After this time, the reaction mixture was cooled down to room temperature and divided between diethyl ether and water. The organic layer was washed with 1 M HCl (x2), water and brine. The organic layer was dried over Na_2SO_4 and concentrated under reduced pressure.

Step 2. The crude oxime obtained in the previous step was dissolved in DCM (1.0 mL), cooled in the ice bath and the solution of *p*-toluenesulphonyl chloride (1.91 g, 10.0 mmol, 1.50 eq) in pyridine (1.6 mL) was added dropwise. The reaction mixture was stirred overnight at room temperature and then concentrated on the rotary evaporator.

Step 3. The crude tosylated oxime was dissolved in THF (35 mL) and equal volumes of the solution were transferred to three microwave reactors. Reactors were sealed, reaction mixtures cooled down to -78°C and 32% aqueous ammonia (32 mL in total) was added to each reactor under argon atmosphere. The resulting solutions were slowly warmed up to room temperature and stirred for 48 h. Then, the reaction mixtures were cooled down to -78°C again and a balloon with argon was connected to each reactor to even out the pressure. After a few minutes, reactors were opened to allow the excess NH_3 to escape slowly and the reactions were warmed up. The crude reaction mixtures were combined and partitioned between EtOAc and water. The organic layer was washed with $NaHCO_3$ (x3), brine (x3), dried over Na_2SO_4 and concentrated under reduced pressure.

Step 4. The crude was dissolved in dry MeOH (15 mL). Triethylamine (4.6 mL, 33.0 mmol, 3.30 eq) and I_2 (1.86 g, 7.33 mmol, 1.10 eq) were added and the reaction mixture was stirred overnight at room temperature. Then, the mixture was partitioned between EtOAc and aqueous $Na_2S_2O_3$. The organic layer was washed with $NaHCO_3$ and brine (x3), dried over Na_2SO_4 and concentrated on the rotary evaporator. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: hexane/EtOAc 100/0 to 98/2) to give the diazirine **14** as a product (1.56 g, 75%).

¹H NMR (400 MHz, Chloroform-*d*): *δ* = 7.76 – 7.72 (m, 2H), 6.95 – 6.91 (m, 2H).

1-((2*R*,4*S*,5*R*)-4-((*tert*-butyldimethylsilyl)oxy)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetra-hydrofuran-2-yl)-5-methylpyrimidine-2,4(1*H*,3*H*)-dione (15)
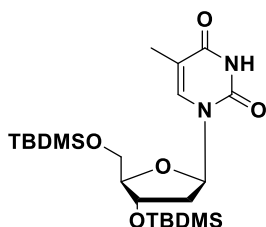
Chemical Formula: $C_{22}H_{42}N_2O_5Si_2$
Molecular Weight: 470.76

The compound was synthesized according to the published procedure [315]. Thymidine (4.00 g, 16.5 mmol, 1.00 eq) was dissolved in dry DMF (16 mL) and stirred for 20 min. Imidazole (6.75 g, 99.1 mmol, 6.00 eq) and TBDMSCl (7.47 g, 49.5 mmol, 3.00 eq) were subsequently added and the reaction mixture was stirred for 3.5 h at room temperature. After this time, the reaction mixture was concentrated on the rotary evaporator. The liquid residue was taken into EtOAc and washed with saturated aqueous $NaHCO_3$ (x3), brine (x2), dried over Na_2SO_4 and concentrated under reduced pressure.

The crude was suspended on the silica gel and purified by flash column chromatography (gradient: DCM/MeOH 100/0 to 97/3) which gave an oil as a product. The oil was co-evaporated few times with

dry ACN and dried overnight on the high-vacuum pump to yield white solid as a product (**15**) (7.60 g, 98%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 8.34 (s, 1H), 7.47 (s, 1H), 6.33 (dd, *J* = 8.0, 5.8 Hz, 1H), 4.40 (dt, *J* = 5.5, 2.5 Hz, 1H), 3.93 (q, *J* = 2.5 Hz, 1H), 3.81 (ddd, *J* = 43.2, 11.4, 2.5 Hz, 2H), 2.24 (ddd, *J* = 13.1, 5.8, 2.6 Hz, 1H), 2.00 (ddd, *J* = 13.3, 8.0, 6.0 Hz, 1H), 1.91 (d, *J* = 1.1 Hz, 3H), 0.93 (s, 9H), 0.89 (s, 9H), 0.11 (d, *J* = 0.9 Hz, 6H), 0.08 (d, *J* = 2.8 Hz, 6H).

tert-butyl(((2*R*,3*S*)-3-((*tert*-butyldimethylsilyl)oxy)-2,3-dihydrofuran-2-yl)methoxy)-dimethylsilane (16)

TBDMSO

OTBDMS
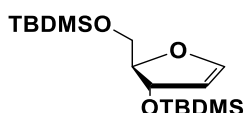
Chemical Formula: $C_{17}H_{36}O_3Si_2$
Molecular Weight: 344.64

The compound was synthesized according to the published procedure [315]. To the compound **15** (2.00 g, 4.25 mmol, 1.00 eq) hexamethyldisilazane (HMDS) (15 mL, 90.0 mmol, 21.1 eq) was added and the suspension was stirred for around 10 min. Then, $(NH_4)_2SO_4$ (1.16 g, 8.78 mmol, 2.07 eq) was added and the reaction mixture was heated to 125°C. Upon heating, **15** slowly dissolved in the HMDS. The reaction mixture was refluxed for 4 h. Then, the reaction mixture was cooled down to room temperature and concentrated on the rotary evaporator. The liquid residue was taken into hexane and washed with saturated aqueous $NaHCO_3$ (x3), brine (x2), dried over Na_2SO_4 and concentrated under reduced pressure. The crude was suspended on the silica gel and purified by flash column chromatography (gradient: diethyl ether/hexane 1/2 to 1/1) to give a brown oil as a product (**16**) (1.03 g, 71%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 6.49 – 6.46 (m, 1H), 5.01 (t, *J* = 2.6 Hz, 1H), 4.86 (t, *J* = 2.6 Hz, 1H), 4.29 (td, *J* = 6.0, 2.8 Hz, 1H), 3.69 (dd, *J* = 10.7, 5.7 Hz, 1H), 3.51 (dd, *J* = 10.7, 6.3 Hz, 1H), 0.90 (s, 9H), 0.89 (s, 9H), 0.09 – 0.08 (m, 6H), 0.07 (d, *J* = 3.7 Hz, 6H).

VII.2.2. Oligoribonucleotide synthesis, functionalization, deprotection and purification

Oligoribonucleotide synthesis, functionalization (CuAAC reaction), deprotection and purification were performed on a 50 nmol scale as described in sections VII.1.1 and VII.1.2.

VII.2.3. *In vitro* photo-cross-linking

The *in vitro* photo-cross-linking assay was performed as described in section VII.1.3.2. 1 μM concentration of each strand was used. Samples were irradiated for 5, 15 or 30 min.

In some cases, it was not possible to deconvolute/assign the masses of the analysed peaks. For some of the compounds, sodium and potassium adducts were observed.

VII.2.4. Thermal stability studies

UV melting profiles were recorded on a Cary 300 instrument (Varian, Palo Alto, CA) equipped with a thermocontroller. Calculated volumes of the ORNs dissolved in water were transferred to screw cap tubes, dried and re-dissolved in a final volume of 100 μL of the melting buffer (100 mM NaCl, 10 mM Na-phosphate buffer, 0.1 mM Na_2EDTA, pH 7.0) at a final concentration of 2 μM. Samples were transferred to quartz cuvettes and annealed by heating to 90°C (5°C/min) and cooling down to 20°C. Melting curves were obtained by the measurement of the A_{260} absorption during the heating step (heating to 90°C, 0.5°C/min), which was repeated three times. Melting temperature was calculated as the average value from the maxima of the first derivatives of 3 consecutive melting curves.

VII.2.5. Chromatograms from the *in vitro* photo-cross-linking experiments

D35

a)

b)
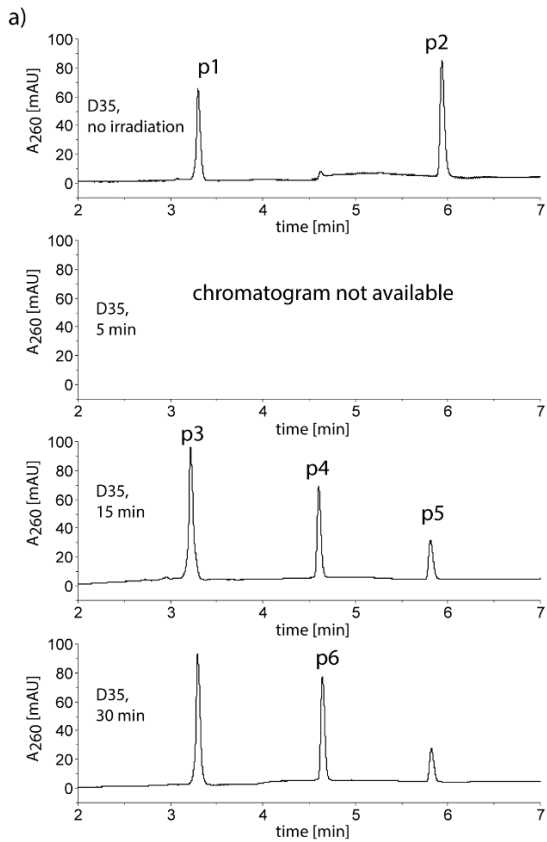
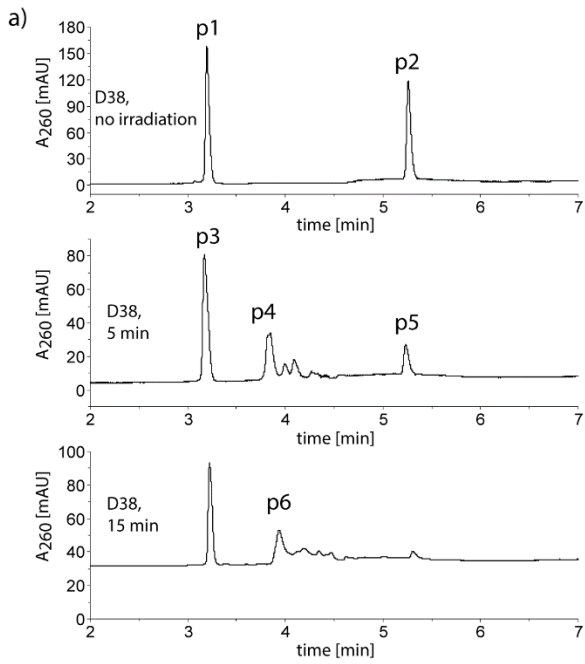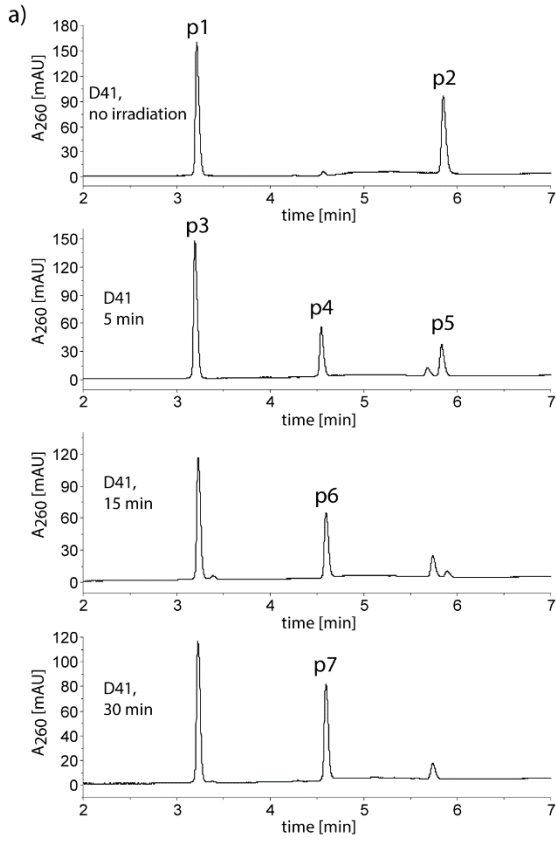| Duplex	ORN	ORN mass	Calc. mass of the CL product
D35	ORN-48	7713.8	12383.7
	ORN-2	4697.9	

c) p1

p3

p4

p5

p6
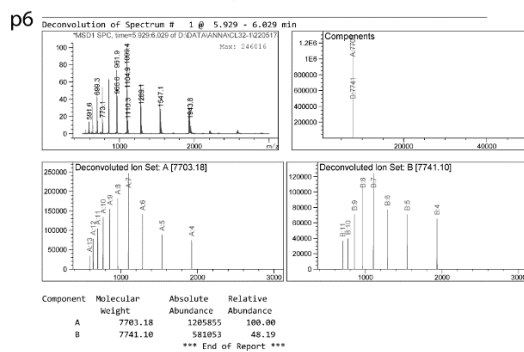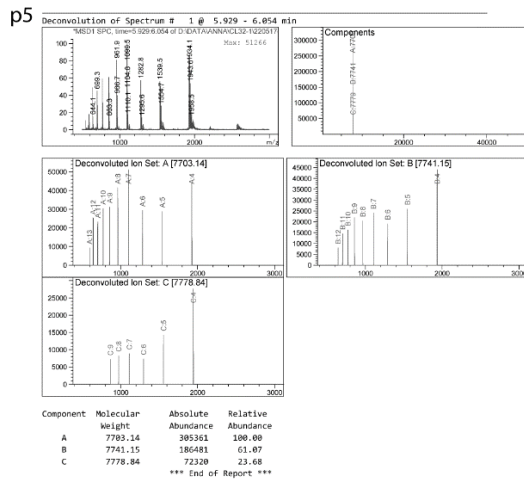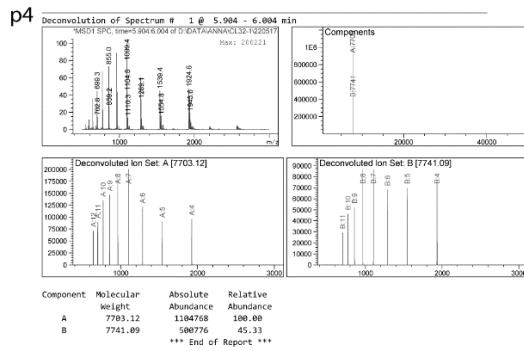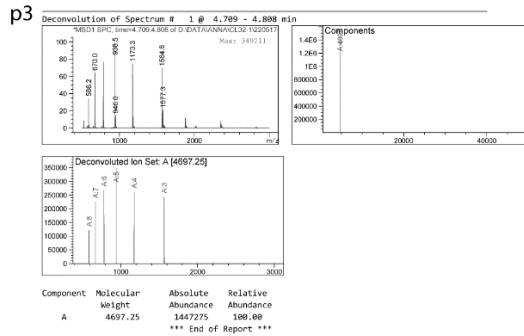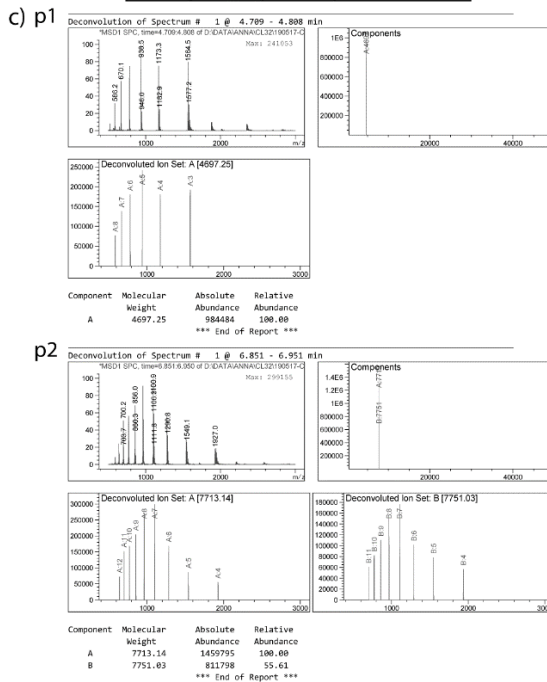
Figure 90: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D35; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

135

D36

a)

b)
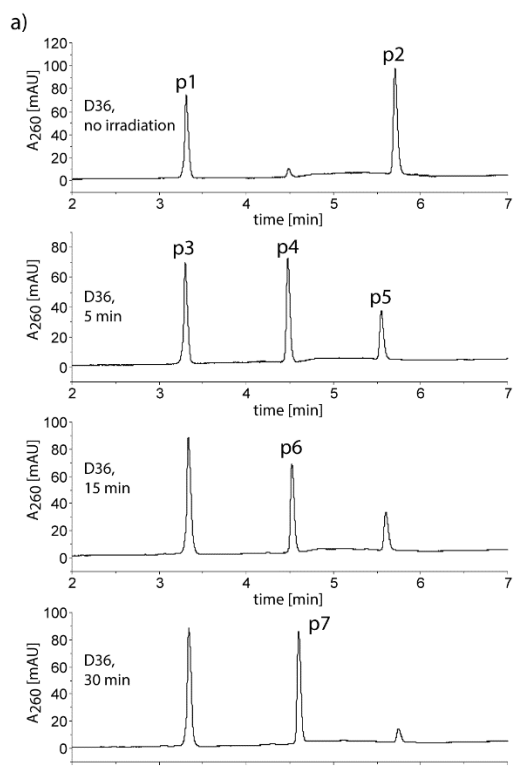
Duplex	ORN	ORN mass	Calc. mass of the CL product
D36	ORN-49	7713.8	12383.7
	ORN-2	4697.9	

c) p1

p2

p3

p4

p5

p6

p7
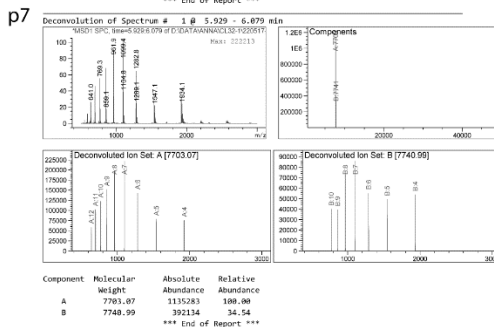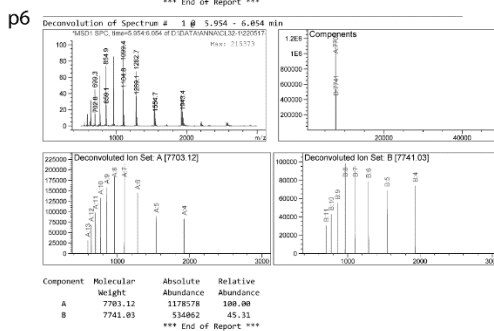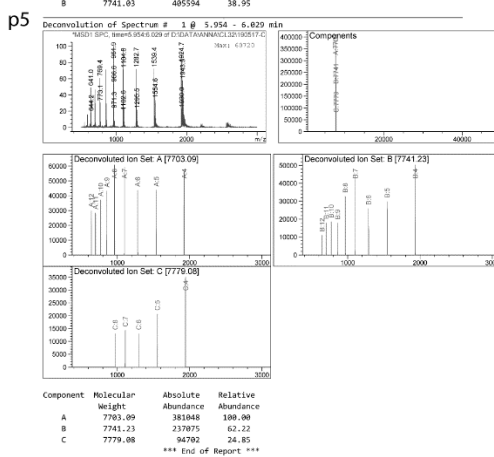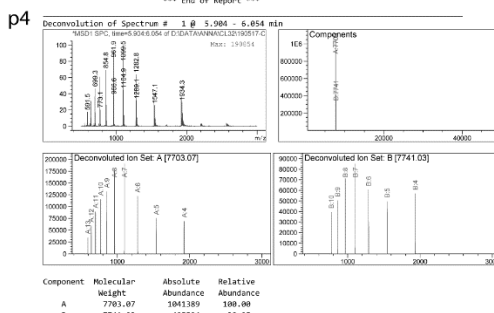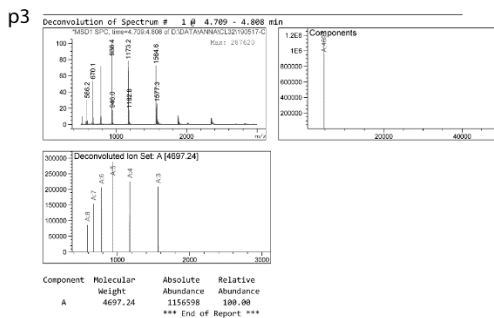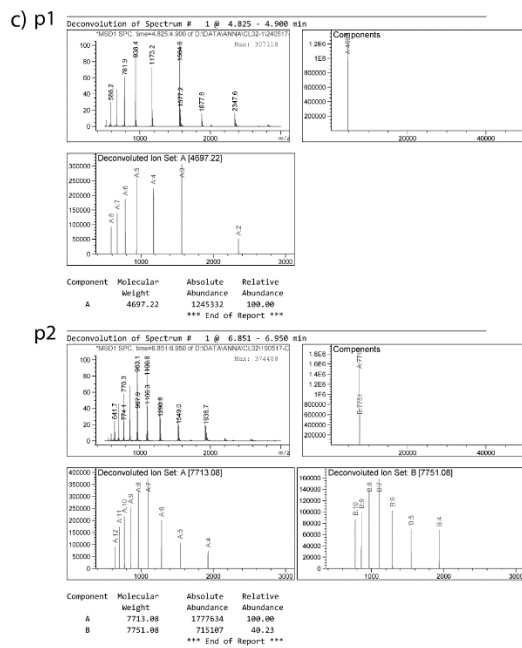
Figure 91: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D36; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

136

D37

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D37	ORN-50	7699.8	12369.7
	ORN-2	4697.9	

c)
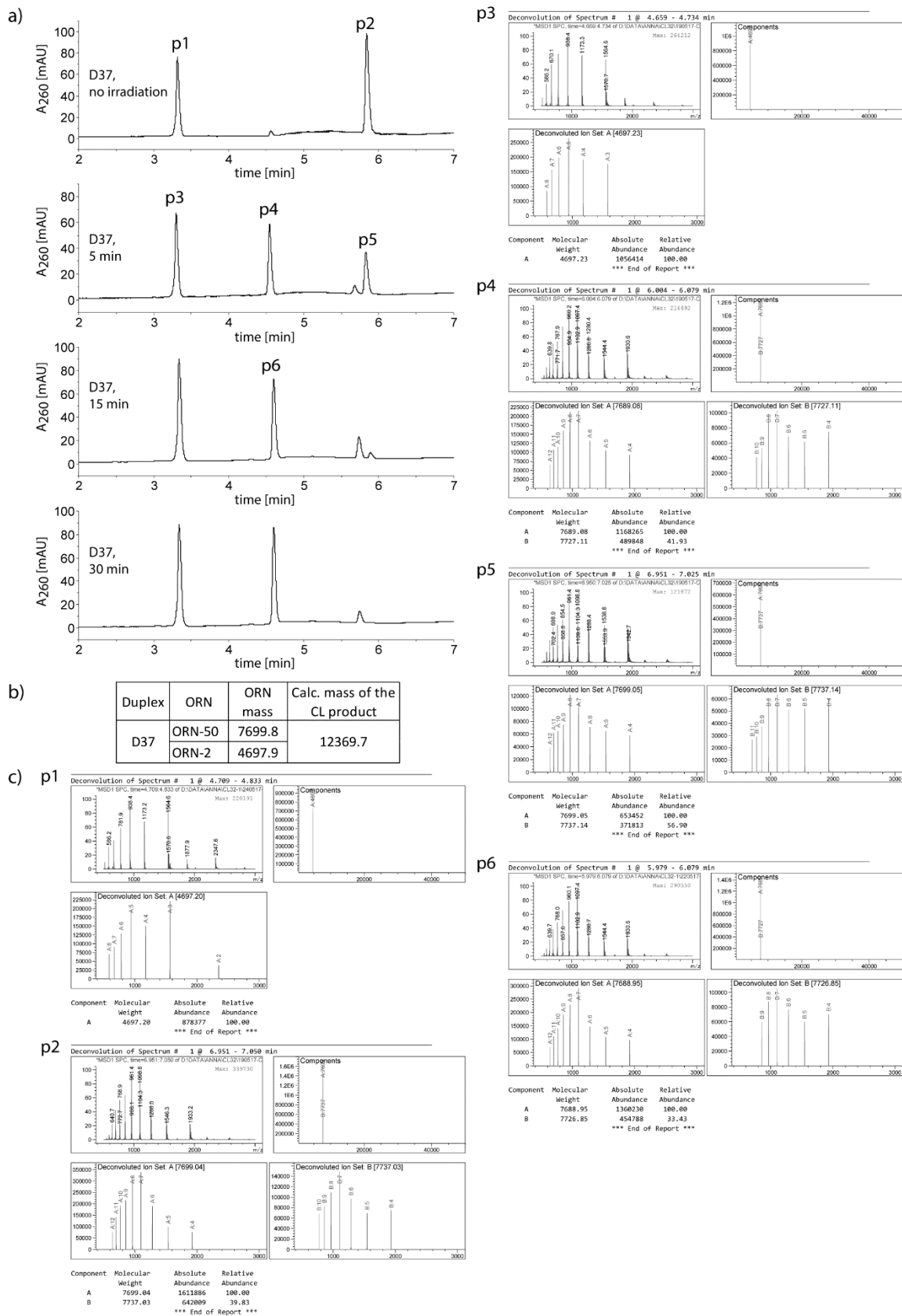
Figure 92: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D37; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D38
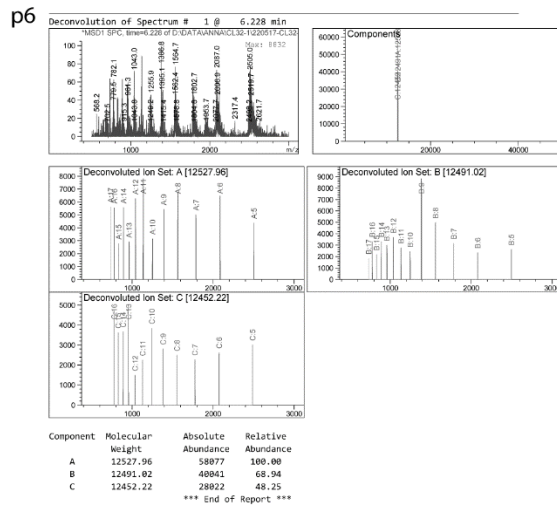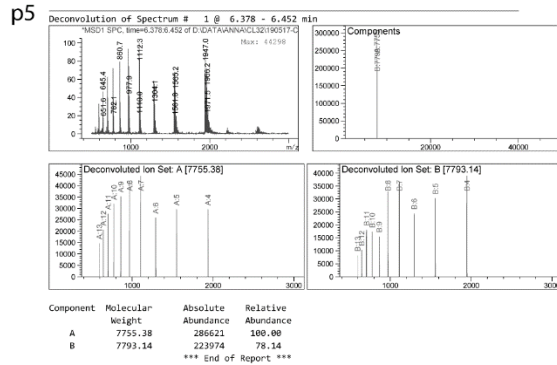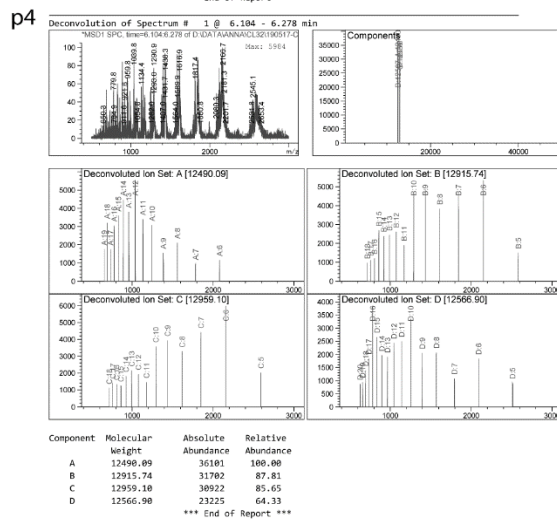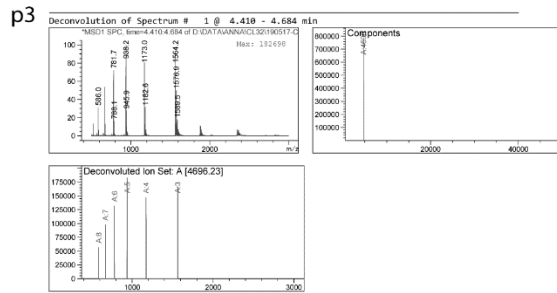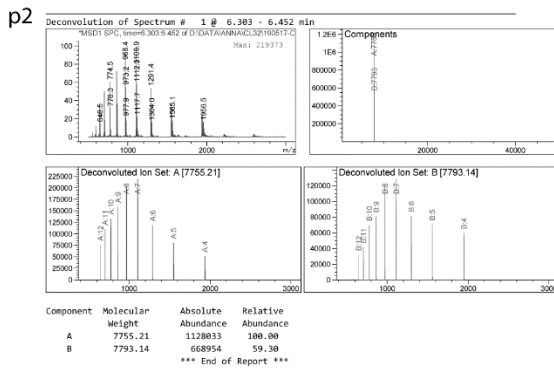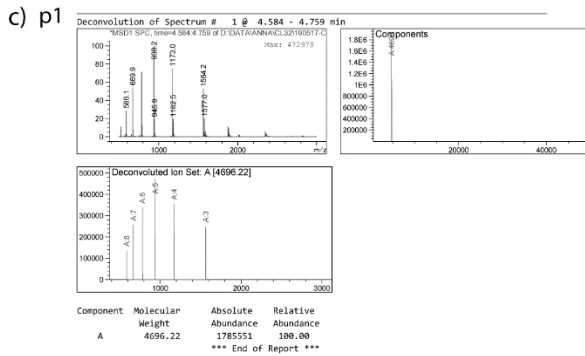
Figure 93: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D38; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D39

a)

b)
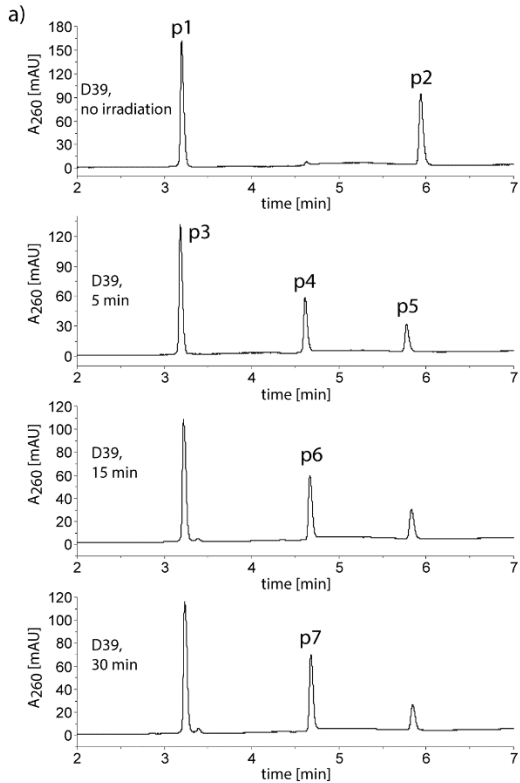
Duplex	ORN	ORN mass	Calc. mass of the CL product
D39	ORN-48	7713.8	12382.7
	ORN-51	4696.9	

c) p1

p2

p3

p4

p5

p6

p7
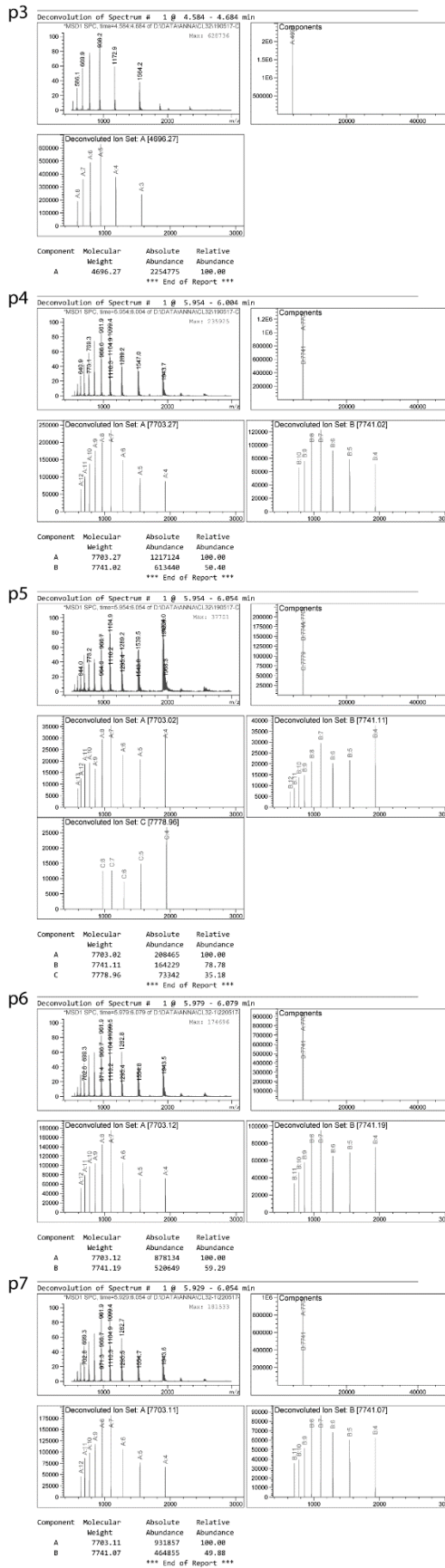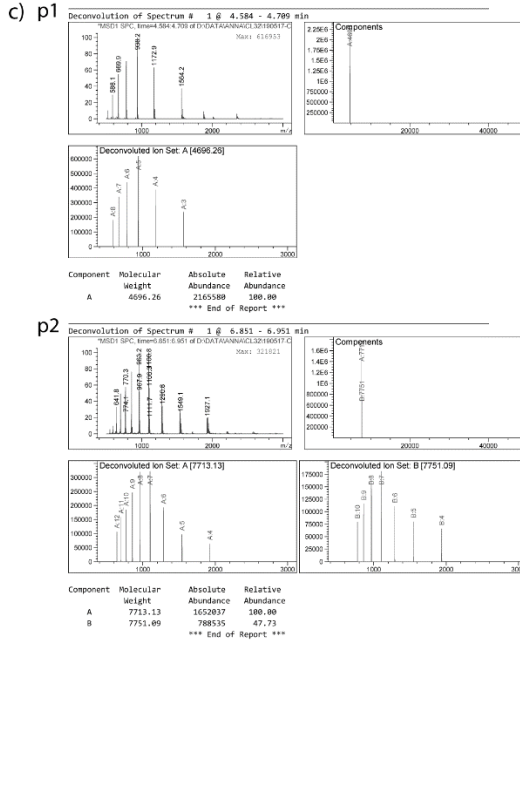
Figure 94: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D39; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
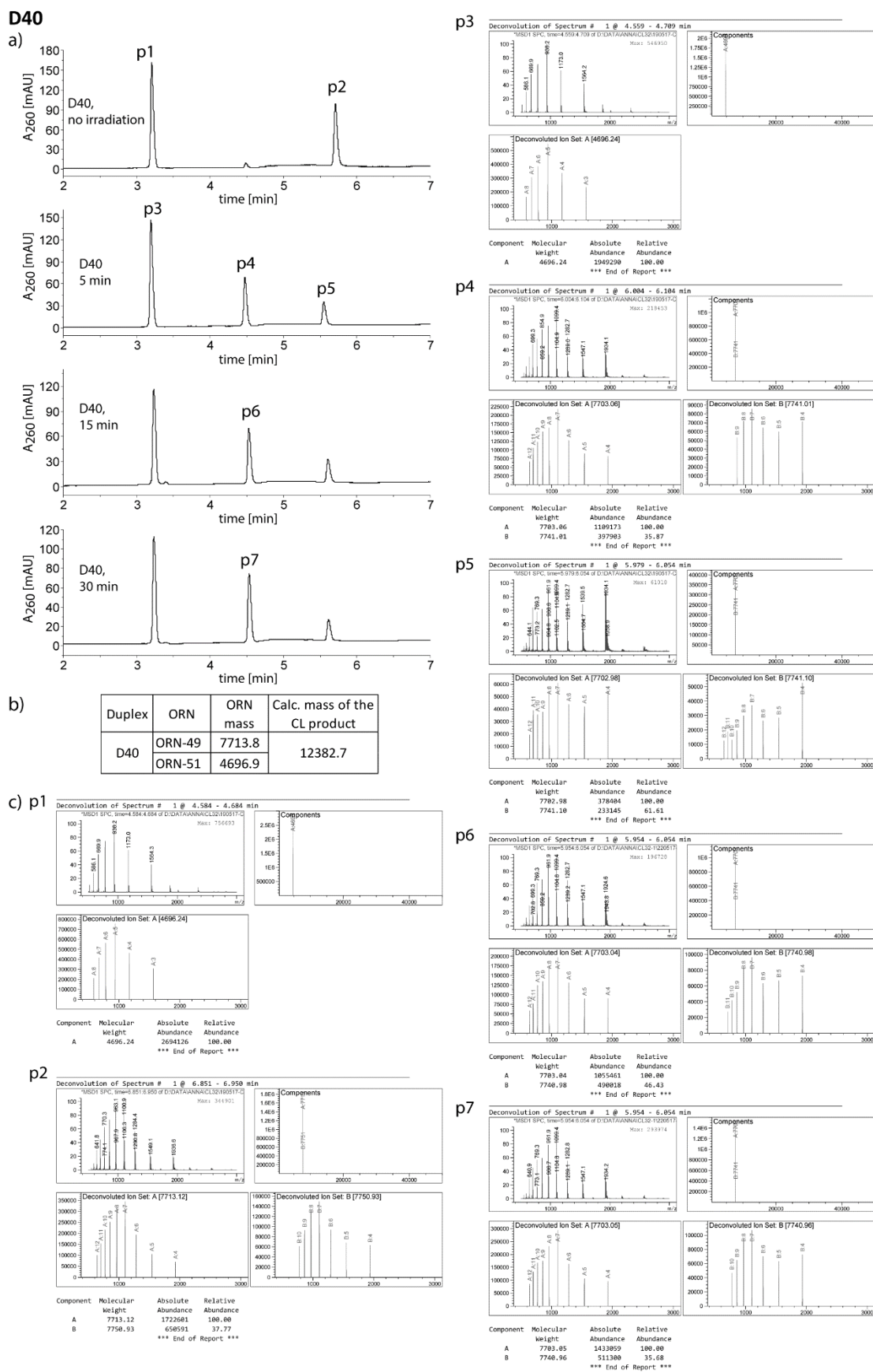
Figure 95: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D40; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D41

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D41	ORN-50	7699.8	12368.7
	ORN-51	4696.9	

c) p1
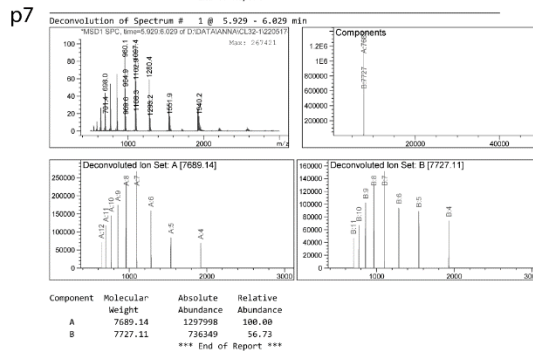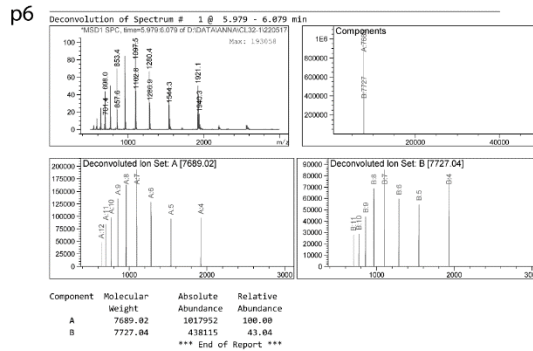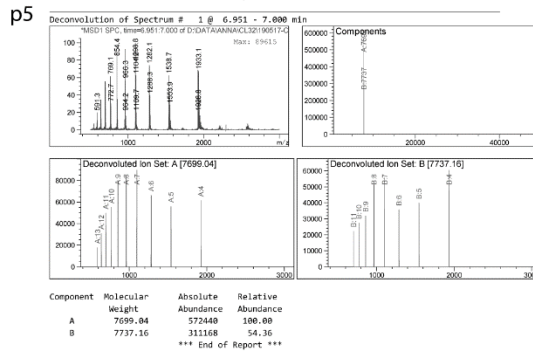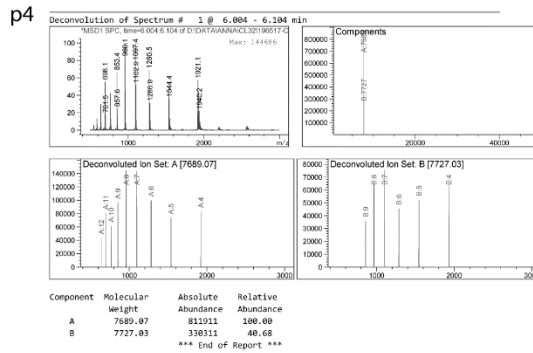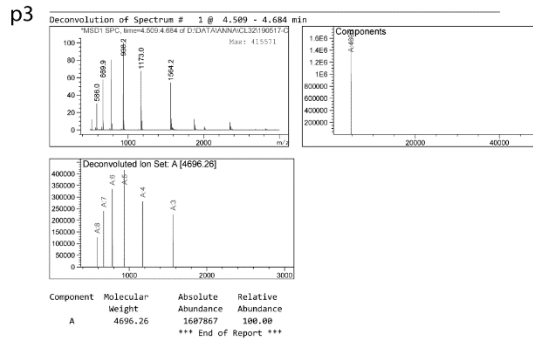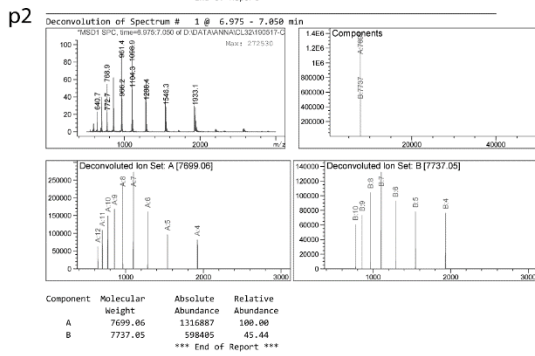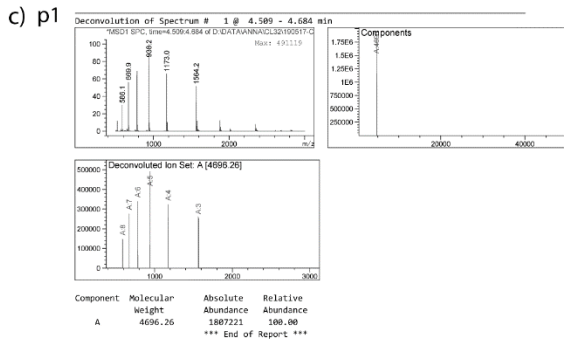
Figure 96: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D41; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.3. Chapter IV

VII.3.1. Chemical synthesis

4'-Chloromethyl-4,5',8-trimethylpsoralen (17)
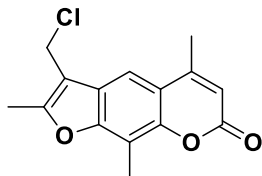
Chemical Formula: $C_{15}H_{13}ClO_3$
Molecular Weight: 276.72

The compound was prepared following the published procedure [284]. To a solution of the 1,3,5-trioxane (1.20 g, 13.3 mmol, 1.01 eq) in 12 N HCl (75 mL) trioxsalen (3.00 g, 13.1 mmol, 1.00 eq) was added. The reaction mixture (a milky suspension) was stirred vigorously at room temperature overnight. Then, the reaction mixture was filtered and the precipitate was washed with water. The solid was dissolved in DCM (120 mL) and washed with water (x3) until the neutral pH was reached. The organic phase was dried over Na_2SO_4, concentrated under reduced pressure and dried on the high-vacuum pump overnight. 3.09 g (86%) of the desired product **17** (light white solid) slightly contaminated with the starting material (10%) was obtained.

1H NMR (400 MHz, Chloroform-*d*): δ = 7.60 (s, 1H), 6.27 (s, 1H), 4.74 (s, 2H), 2.58 (s, 3H), 2.55 – 2.50 (m, 6H).

4'-[(4-Hydroxy-2-oxa)butyl]-4,5',8-trimethylpsoralen (18a)
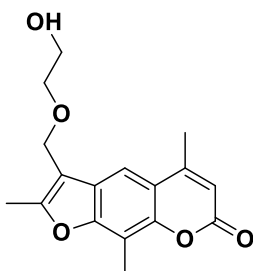
Chemical Formula: $C_{17}H_{18}O_5$
Molecular Weight: 302.33

The synthesis was performed following the previously reported protocol [327]. 4'-Chloromethyl-4,5',8-trimethylpsoralen (**17**) (1.00 g, 3.61 mmol, 1.00 eq) and ethylene glycol (12.2 mL, 0.22 mol, 60.0 eq) were heated in acetone (13 mL) to 50-60°C and stirred at this temperature for 4 h. After 2 h, the white suspension turned into a clear, light yellow solution. The acetone and ethylene glycol were removed on the rotary evaporator and water (50 mL) was added to the residue. The resultant suspension was filtered, washed with cold water and dried. The crude product was dissolved in a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 95/5) to give the product **18a** (0.86 g, 78%).

1H NMR (400 MHz, Chloroform-*d*): δ = 7.54 (s, 1H), 6.15 (s, 1H), 4.65 (s, 2H), 3.79 – 3.71 (m, 2H), 3.63 – 3.55 (m, 2H), 2.52 – 2.45 (m, 6H), 2.44 (s, 3H), 2.30 (s, 1H).

4'-[(4-Methanesulphonyloxy-2-oxa)butyl]-4,5',8-trimethylpsoralen (19a)
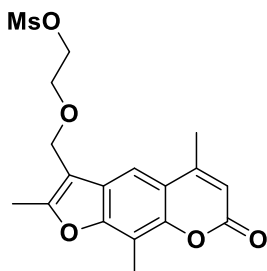
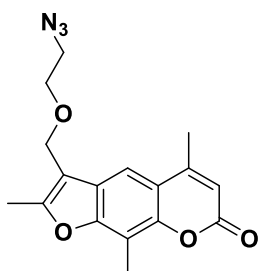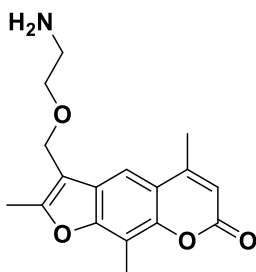Chemical Formula: $C_{18}H_{20}O_7S$
Molecular Weight: 380.41

The synthesis was performed following the previously reported protocol [327]. 4'-[(4-Hydroxy-2-oxa)butyl]-4,5',8-trimethylpsoralen (**18a**) (0.91 g, 2.99 mmol, 1.00 eq, purity 80%) was dissolved in dry DCM (9.5 mL) under argon atmosphere. Triethylamine (0.78 mL, 5.60 mmol, 1.87 eq) was then added. The reaction mixture was flushed with argon and cooled down to 0°C. After a few minutes of stirring, methanesulphonyl chloride (0.32 mL, 4.19 mmol, 1.40 eq) was added dropwise, keeping the temperature below 10°C. When the addition was completed, the reaction mixture was stirred for 25 minutes at around 0°C and then warmed up to room temperature and stirred overnight. The reaction mixture was diluted with DCM, washed with water, dried over Na_2SO_4

and concentrated under reduced pressure. After additional drying on the high-vacuum pump, 4'-[(4-methanesulphonyloxy-2-oxa)butyl]-4,5',8-trimethylpsoralen (**19a**) (1.04 g, 91%; the impurity from the starting material still present) was obtained as a beige solid. The crude product was used in the next step.

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.61 (s, 1H), 6.25 (s, 1H), 4.71 (s, 2H), 4.42 – 4.36 (m, 2H), 3.79 – 3.73 (m, 2H), 2.99 (s, 3H), 2.57 (s, 3H), 2.52 – 2.50 (m, 6H).
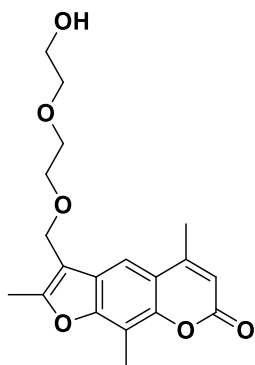
4'-[(4-Azido-2-oxa)butyl]-4,5',8-trimethylpsoralen (20a)

Chemical Formula: $C_{17}H_{17}N_3O_4$
Molecular Weight: 327.34

The synthesis was performed following the previously reported protocol with small changes [327]. 4'-[(4-Methanesulphonyloxy-2-oxa)butyl]-4,5',8-trimethylpsoralen (**19a**) (1.00 g, 2.63 mmol, 1.00 eq, purity 78%) and sodium azide (0.34 g, 5.26 mmol, 2.00 eq) were refluxed in ethanol (7.0 mL) for few hours and then cooled down to room temperature and stirred overnight. 1.00 eq (0.17 g, 2.63 mmol) of sodium azide was subsequently added and the reaction mixture was refluxed for 2 h and then stirred overnight at 55°C. The reaction mixture was cooled down to room temperature and cold water (10 mL) was added. The suspension was filtered. The crude solid was washed with cold water and dried in a desiccator overnight to give the product (**20a**) (0.73 g, 85%; the impurity from the starting material still present), which was used in the next step without further purification.

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.65 (s, 1H), 6.24 (s, 1H), 4.69 (s, 2H), 3.69 – 3.64 (m, 2H), 3.43 – 3.38 (m, 2H), 2.57 (s, 3H), 2.50 (s, 6H).

4'-[(4-Amino-2-oxa)butyl]-4,5',8-trimethylpsoralen (21a)

Chemical Formula: $C_{17}H_{19}NO_4$
Molecular Weight: 301.34

The synthesis was performed following the previously reported protocol [327]. 4'-[(4-Azido-2-oxa)butyl]-4,5',8-trimethylpsoralen (**20a**) (0.65 g; 1.99 mmol, 1.00 eq) was dissolved in dry THF (4.0 mL). Triphenylphosphine (0.63 g; 2.38 mmol, 1.20 eq) and two drops of water were added to the stirred solution. After stirring at room temperature overnight, the reaction mixture was concentrated under reduced pressure to give a yellow solid. The residue was dissolved in DCM and extracted with 1% aqueous HCl solution (x4). Combined acidic layers were carefully treated with K_2CO_3 until saturation (pH~11). The water phase was extracted with DCM (x4). Combined organic layers were washed with brine, dried over Na_2SO_4 and concentrated on the rotary evaporator to give the slightly impure final product (**21a**) as a pale yellow solid (0.38 g, 68%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.60 (s, 1H), 6.26 – 6.20 (m, 1H), 4.66 (s, 2H), 3.53 (t, *J* = 5.2 Hz, 2H), 2.90 (s, 2H), 2.57 (s, 3H), 2.49 (s, 6H).
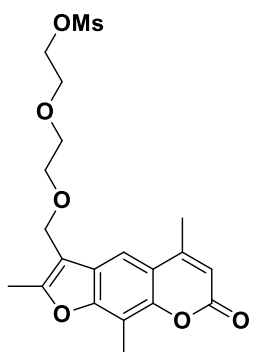
4'-[(7-Hydroxy-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (18b)

Chemical Formula: $C_{19}H_{22}O_6$
Molecular Weight: 346.38

The synthesis was performed following the previously reported protocol with minor changes [327]. 4'-Chloromethyl-4,5',8-trimethylpsoralen (**17**) (1.20 g, 4.34 mmol, 1.00 eq) and diethylene glycol (29 mL, 0.30 mol, 70.0 eq) were heated in acetone (26 mL) to 50-60°C and stirred at this temperature overnight. In the meantime, the white suspension turned into a clear brown solution. The acetone and diethylene glycol were partially removed on the rotary evaporator. The residue was diluted with chloroform. The organic phase was washed several times with water, brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was dissolved in a minimal amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 95/5) to give the product **18b** (1.36 g, 91%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.61 (s, 1H), 6.23 (s, 1H), 4.69 (s, 2H), 3.74 – 3.66 (m, 4H), 3.65 – 3.57 (m, 4H), 2.56 (s, 3H), 2.49 (d, *J* = 1.3 Hz, 6H).

4'-[(7-Methanesulphonyloxy -2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (19b)

Chemical Formula: $C_{20}H_{24}O_8S$
Molecular Weight: 424.46

The synthesis was performed following the previously reported protocol [327]. 4'-[(7-Hydroxy-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (**18b**) (1.35 g, 4.47 mmol, 1.00 eq) was dissolved in DCM (5.0 mL) under argon atmosphere. Triethylamine (1.0 mL, 7.14 mmol, 1.60 eq) was then added and the reaction mixture was flushed with argon and cooled down to 0°C. Subsequently, methanesulphonyl chloride (0.48 mL, 6.25 mmol, 1.40 eq) was added dropwise. When the addition was completed, the reaction mixture was stirred for 15 more minutes at around 0°C and then it was left to warm up to room temperature overnight. The reaction mixture was diluted with DCM, washed with water (x3), brine, dried over Na_2SO_4 and concentrated on the rotary evaporator to giv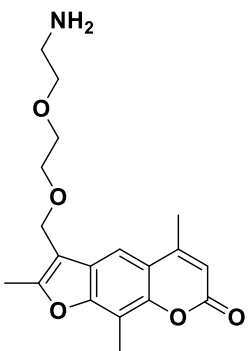e the product (**19b**) as a dark brown oil (1.66 g, 87%). The crude product was used in the next step without further purification.

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.59 (s, 1H), 6.23 (s, 1H), 4.68 (s, 2H), 4.38 – 4.33 (m, 2H), 3.80 – 3.74 (m, 2H), 3.68 (d, *J* = 4.3 Hz, 2H), 3.63 (dd, *J* = 6.0, 3.0 Hz, 2H), 3.00 (s, 3H), 2.56 (s, 3H), 2.49 (s, 6H).

4'-[(7-Azido-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (20b)

The synthesis was performed following the previously reported protocol [327]. 4'-[(7-Methanesulphonyloxy-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (**19b**) (1.80 g, 4.24 mmol, 1.00 eq) and sodium azide (0.55 g, 8.48 mmol, 2.00 eq) were refluxed in ethanol (19 mL) overnight. The reaction mixture was cooled down to room temperature and cold water (20 mL) was added. The suspension was filtered and the crude solid was washed with cold water and dried. The crude product was dissolved in the minimal amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 95/5) to give the slightly impure product (**20b**) as an orange oil (1.25 g, 80%).

Chemical Formula: $C_{19}H_{21}N_3O_5$
Molecular Weight: 371.39

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.61 (s, 1H), 6.25 – 6.19 (m, 1H), 4.69 (s, 2H), 3.66 (td, *J* = 4.8, 3.3 Hz, 6H), 3.42 – 3.33 (m, 2H), 2.55 (s, 3H), 2.49 (s, 6H).

4'-[(7-Amino-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (21b)

The compound was prepared according to the published procedure [327]. 4'-[(7-Azido-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (**20b**) (0.75 g; 2.02 mmol, 1.00 eq) was dissolved in dry THF (4.0 mL). Triphenylphosphine (0.64 g; 2.42 mmol, 1.20 eq) and two drops of water were added to the stirred solution. After stirring at room temperature overnight, the reaction mixture was concentrated to give a yellow residue. The residue was dissolved in DCM and extracted with 1% aqueous HCl solution (x3). Combined acidic layers were carefully treated with 5 M aqueous NaOH solution until the basic pH (pH~10) was reached. The water layer was extracted with DCM (x4). Combined organic layers were washed with water, brine, dried over Na_2SO_4 and concentrated under reduced pressure to give the slightly impure 4'-[(7-amino-2,5-oxa)heptyl]-4,5',8-trimethylpsoralen (**21b**) as a pale yellow oil, which after long drying on the high-vacuum pump partially solidified (0.41 g; 58%).
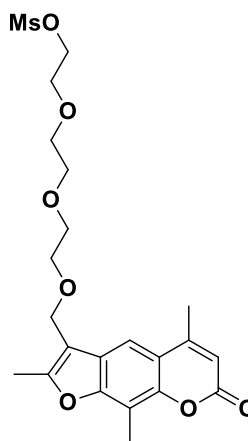
Chemical Formula: $C_{19}H_{23}NO_5$
Molecular Weight: 345.40

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.62 (s, 1H), 6.25 – 6.22 (m, 1H), 4.69 (s, 2H), 3.63 (s, 4H), 3.52 (t, *J* = 5.2 Hz, 2H), 2.88 (t, *J* = 5.2 Hz, 2H), 2.56 (s, 3H), 2.49 (s, 6H).

4'-[(10-Hydroxy-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (18c)
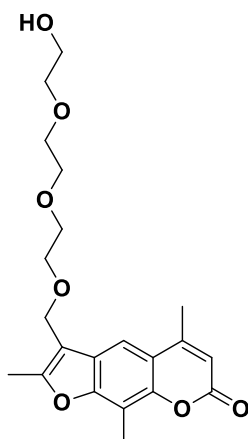
Chemical Formula: $C_{21}H_{26}O_7$
Molecular Weight: 390.43

The compound was synthesized adjusting the published procedure [327]. 4'-Chloromethyl-4,5',8-trimethylpsoralen (**17**) (1.00 g, 3.61 mmol, 1.00 eq) and triethylene glycol (38.0 mL, 0.29 mol, 80.00 eq) were heated in acetone (22 mL) to 50-60°C and stirred at this temperature for around 48 h. In the meantime, the white suspension turned into a clear brown solution. The acetone and triethylene glycol were partially removed on the rotary evaporator. The residue was diluted with DCM. The organic phase was washed several times with water, brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 93/7) to give the slightly impure product (an excess of the triethylene glycol present) (**18c**) in the form of a dark-brown oil (1.08 g, 77%).

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.64 (s, 1H), 6.23 (d, *J* = 1.1 Hz, 1H), 4.68 (s, 2H), 3.73 – 3.66 (m, 6H), 3.64 – 3.57 (m, 6H), 2.56 (s, 3H), 2.51 – 2.47 (m, 6H). **¹³C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.2, 154.8, 153.6, 149.4, 125.2, 116.3, 112.9, 2 x 111.9, 109.2, 72.6, 70.8, 70.7, 70.5, 69.0, 63.5, 61.9, 19.5, 12.5, 8.6.

4'-[(10-Methanesulphonyloxy-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (19c)

Chemical Formula: $C_{22}H_{28}O_9S$
Molecular Weight: 468.52

The compound was obtained adapting the published procedure [327]. 4'-[(10-Hydroxy-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (**18c**) (1.00 g, 2.56 mmol, 1.00 eq) was dissolved in DCM (8.0 mL) under argon atmosphere. Triethylamine (0.57 mL, 4.10 mmol, 1.60 eq) was then added and the reaction mixture was flushed with argon and cooled down to 0°C. Subsequently, methanesulphonyl chloride (0.28 mL, 3.59 mmol, 1.40 eq) was added dropwise. When the addition was completed, the reaction mixture was stirred for 15 minutes at around 0°C and then was left to warm up to room temperature overnight. The reaction mixture was diluted with DCM, washed with water (x3), brine, dried over anhydrous Na_2SO_4 and concentrated under reduced pressure. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 90/10) to give the slightly impure mesylated product (**19c**) as an amber oil (1.16 g; 97%).
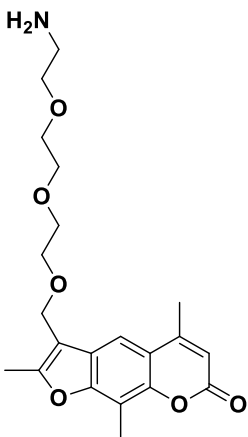
¹H NMR (400 MHz, Chloroform-*d*): δ = 7.61 (s, 1H), 6.24 (d, *J* = 1.2 Hz, 1H), 4.68 (s, 2H), 4.36 – 4.32 (m, 2H), 3.76 – 3.72 (m, 2H), 3.66 – 3.61 (m, 8H), 3.03 (s, 3H), 2.57 (s, 3H), 2.51 – 2.48 (m, 6H). **¹³C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.2, 154.8, 153.5, 149.4, 125.2, 116.3, 113.0, 111.9, 111.8, 109.3, 70.9, 70.8, 70.7, 69.2, 69.2, 69.0, 63.5, 37.8, 19.5, 12.5, 8.6.

4'-[(10-Azido-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (20c)

The compound was synthesized adapting the published procedure [327]. 4'-[(10-Methanesulphonyloxy-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (**19c**) (1.10 g, 2.35 mmol, 1.00 eq) and sodium azide (0.31 g, 4.70 mmol, 2.00 eq) were refluxed in ethanol (12 mL) for 20 h. Then, another portion of sodium azide (0.15 g, 2.31 mmol, 0.98 eq) was added and the reaction mixture was refluxed for 24 h. The reaction mixture was then cooled down to room temperature and washed with cold water (x3), brine, dried over Na_2SO_4 and concentrated on the rotary evaporator. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 93/7) to give an amber oil as the slightly impure product (**20c**) (0.84 g, 86%).

Chemical Formula: $C_{21}H_{25}N_3O_6$
Molecular Weight: 415.45

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.63 (s, 1H), 6.24 (d, *J* = 1.1 Hz, 1H), 4.69 (s, 2H), 3.70 – 3.66 (m, 5H), 3.64 (ddd, *J* = 6.6, 4.8, 2.7 Hz, 5H), 3.38 – 3.32 (m, 2H), 2.57 (s, 3H), 2.52 – 2.46 (m, 6H). **^{13}C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.1, 154.9, 153.5, 149.4, 125.3, 116.3, 112.9, 112.0, 111.9, 109.3, 71.0, 70.9, 70.8, 70.2, 69.1, 63.5, 50.8, 19.5, 12.5, 8.6.

4'-[(10-Amino-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (21c)

The compound was prepared adapting the published procedure [327]. 4'-[(10-azido-2,5,8-oxa)decyl]-4,5',8-trimethylpsoralen (**20c**) (0.70 g; 1.68 mmol, 1.00 eq) was dissolved in dry THF (4.0 mL). Triphenylphosphine (0.53 g; 2.02 mmol, 1.20 eq) and two drops of water were added to the stirred solution. After stirring at room temperature overnight, the reaction mixture was concentrated under reduced pressure. The residue was dissolved in DCM and extracted with 1% aqueous HCl solution (x4). Combined acidic layers were carefully treated with 5 M aqueous NaOH solution until the basic pH (pH~10) was reached. The water layer was extracted with DCM (x4). Combined organic layers were washed with water, brine, dried over Na_2SO_4, and concentrated on the rotary evaporator to give the free amine. The crude product was diluted with DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 70/30). After drying on the high-vacuum

Chemical Formula: $C_{21}H_{27}NO_6$
Molecular Weight: 389.45

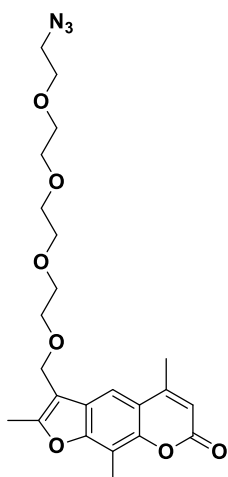pump, the desired final product **21c** was obtained in the form of a dark-yellow oil (0.51 g, 78%).

^{1}H NMR (400 MHz, Chloroform-*d*): δ = 7.62 (s, 1H), 6.25 – 6.20 (m, 1H), 4.70 (s, 2H), 3.66 – 3.60 (m, 8H), 3.51 (t, J = 5.2 Hz, 2H), 2.87 (t, J = 5.2 Hz, 2H), 2.62 (s, 2H), 2.55 (s, 3H), 2.51 – 2.46 (m, 6H). **^{13}C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.2, 154.8, 153.5, 149.3, 125.2, 116.3, 112.9, 111.9, 111.9, 109.2, 72.7, 70.9, 70.7, 70.4, 69.0, 63.4, 41.6, 19.6, 12.5, 8.6.

4'-[(13-Hydroxy-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (18d)
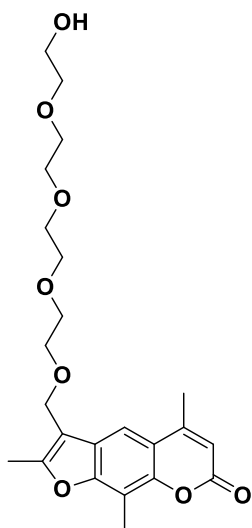
Chemical Formula: $C_{23}H_{30}O_8$
Molecular Weight: 434.49

The compound was prepared adjusting the published procedure [327]. 4'-Chloromethyl-4,5',8-trimethylpsoralen (17) (1.00 g, 3.61 mmol, 1.00 eq) and tetraethylene glycol (56 mL, 0.33 mol, 90.00 eq) were heated in acetone (30 mL) to 50-60°C and stirred at this temperature overnight. In the meantime, the white suspension turned into a clear, light brown solution. The acetone and tetraethylene glycol were partially removed on the rotary evaporator. The residue was diluted with chloroform. The organic phase was washed several times with water, brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was dried on the high-vacuum pump. The ^{1}H NMR spectrum showed that the starting material was still present. The crude and tetraethylene glycol (25 mL 0.15 mol, 40.00 eq) were heated in acetone (15 mL) to 50-60°C and stirred at this temperature overnight. The reaction mixture was concentrated on the rotary evaporator and the residue was diluted with DCM. The organic phase was washed with water (x4), brine (x2), dried over Na_2SO_4 and concentrated. The crude product was dried overnight on the high-vacuum pump. The crude product was diluted with DCM and purified (twice) using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 90/10) to give the product (18d) with a small excess of the tetraethylene glycol (dark-brown oil; 1.34 g, 87%).

^{1}H NMR (400 MHz, Chloroform-d): δ = 7.63 (s, 1H), 6.23 (d, J = 1.1 Hz, 1H), 4.69 (s, 2H), 3.71 – 3.68 (m, 2H), 3.67 – 3.62 (m, 12H), 3.60 – 3.57 (m, 2H), 2.56 (s, 3H), 2.51 – 2.46 (m, 6H). **^{13}C NMR** (101 MHz, Chloroform-d): δ = 161.7, 155.2, 154.8, 153.6, 149.4, 125.3, 116.3, 112.9, 112.0, 111.9, 109.2, 72.6, 70.9, 70.8, 2 x 70.7, 70.4, 69.1, 63.5, 61.9, 19.6, 12.5, 8.6.

4'-[(13-Methanesulphonyloxy-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (19d)
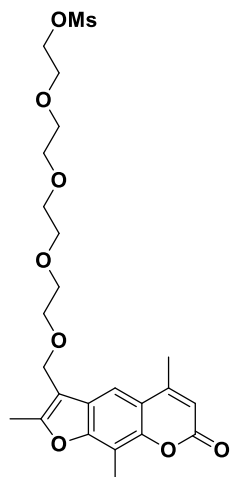
Chemical Formula: $C_{24}H_{32}O_{10}S$
Molecular Weight: 512.57

The compound was prepared adapting the published procedure [327]. 4'-[(13-Hydroxy-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (18d) (1.30 g, 2.99 mmol, 1.00 eq) was dissolved in DCM (10 mL) under argon atmosphere. Triethylamine (0.67 mL, 4.79 mmol, 1.60 eq) was then added and the reaction mixture was flushed with argon and cooled down to 0°C. Then, methanesulphonyl chloride (0.32 mL, 4.19 mmol, 1.40 eq) was added dropwise. When the addition was completed, the reaction mixture was stirred for 15 minutes at around 0°C and then was left to warm up to room temperature overnight. The reaction mixture was diluted with DCM, washed with water (x3), brine (x2), dried over Na_2SO_4 and concentrated on the rotary evaporator to give the mesylated product (19d) as a dark-brown oil (1.50 g, quant.). The crude product was used in the next step without further purification.
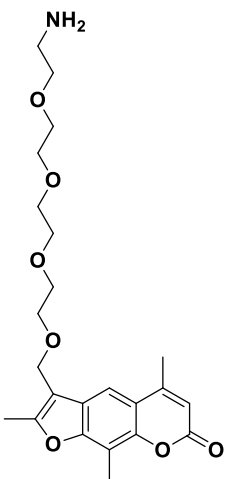
^{1}H NMR (400 MHz, Chloroform-d): δ = 7.62 (s, 1H), 6.26 – 6.21 (m, 1H), 4.69 (s, 2H), 4.37 – 4.32 (m, 2H), 3.76 – 3.70 (m, 2H), 3.66 – 3.61 (m, 12H), 3.06 (s, 3H), 2.56 (s, 3H), 2.51 – 2.48 (m, 6H). **^{13}C NMR** (101 MHz, Chloroform-d): δ = 161.7, 155.2, 154.8, 153.5, 149.4, 125.3, 116.3, 112.9, 112.0, 111.9, 109.2, 70.9, 3 x 70.7, 70.6, 69.4, 69.1, 69.0, 63.4, 37.9, 19.5, 12.5, 8.6.

4'-[(13-Azido-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (20d)

The compound was prepared adjusting the published procedure [327]. 4'-[(13-Methanesulphonyloxy-2,5,8,11-oxa)tridecyl-4,5',8-trimethylpsoralen (**19d**) (1.40 g, 2.73 mmol, 1.00 eq) and sodium azide (0.36 g, 5.46 mmol, 2.00 eq) were refluxed in ethanol (14 mL) for 20 h. Then, another portion of sodium azide (0.18 g, 2.73 mmol, 1.00 eq) was added and the reaction mixture was refluxed for 18 h. The reaction mixture was cooled down to room temperature and washed with cold water (x3), brine, dried over Na_2SO_4 and concentrated on the rotary evaporator. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 93/7) to give a dark-brown oil as the desired product (**20d**) (1.09 g, 87%).

Chemical Formula: $C_{23}H_{29}N_3O_7$
Molecular Weight: 459.50

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.63 (s, 1H), 6.23 (s, 1H), 4.69 (s, 2H), 3.67 − 3.62 (m, 14H), 3.39 − 3.33 (m, 2H), 2.57 (s, 3H), 2.51 − 2.47 (m, 6H). **¹³C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.1, 154.9, 153.5, 149.4, 125.3, 116.3, 112.9, 112.0, 111.9, 109.2, 70.9, 4 x 70.8, 70.2, 69.1, 63.5, 50.8, 19.6, 12.5, 8.6.

4'-[(13-Amino-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (21d)

The compound was prepared adapting the published procedure [327]. 4'-(13-Azido-2,5,8,11-oxa)tridecyl]-4,5',8-trimethylpsoralen (**20d**) (0.90 g, 1.96 mmol, 1.00 eq) was dissolved in dry THF (4.0 mL). Triphenylphosphine (0.62 g; 2.35 mmol, 1.20 eq) and two drops of water were added to the stirred solution. After stirring at room temperature overnight, the reaction mixture was concentrated on the rotary evaporator to give a yellow residue. The residue was dissolved in DCM and extracted with 1% aqueous HCl solution (x4). Combined acidic layers were carefully treated with 5 M aqueous NaOH solution until basic pH (pH~10) was reached. The water layer was extracted with DCM (x5). Combined organic layers were washed with water, brine, dried over Na_2SO_4 and concentrated under reduced pressure to give the free amine as the final product **21d** (brown oil; 0.53 g; 63%).
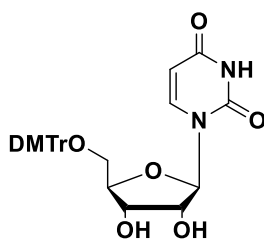
Chemical Formula: $C_{23}H_{31}NO_7$
Molecular Weight: 433.50

¹H NMR (400 MHz, Chloroform-*d*): δ = 7.63 (s, 1H), 6.24 (s, 1H), 4.69 (s, 2H), 3.68 − 3.60 (m, 12H), 3.49 (t, *J* = 5.2 Hz, 2H), 2.81 (t, *J* = 5.1 Hz, 2H), 2.57 (s, 3H), 2.50 (d, *J* = 4.8 Hz, 6H). **¹³C NMR** (101 MHz, Chloroform-*d*): δ = 161.7, 155.2, 154.9, 153.6, 149.4, 125.2, 116.4, 113.0, 112.0, 111.9, 109.3, 73.2, 70.9, 3 x 70.7, 70.4, 69.1, 63.5, 41.8, 19.6, 12.5, 8.6.

4'-Aminomethyl-4,5',8-trimethylpsoralen (22)
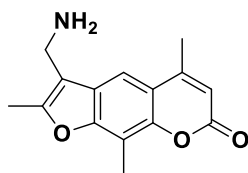
Chemical Formula: $C_{15}H_{15}NO_3$
Molecular Weight: 257.29

The compound was synthesized following the published procedure [252]. 4'-Chloromethyl-4,5',8-trimethylpsoralen (**17**) (0.30 g, 1.08 mmol, 1.00 eq), potassium phthalimide (0.24 g, 1.30 mmol, 1.20 eq) and dry DMF (15 mL) were heated to 100°C and stirred for 6 h. After this time, the reaction mixture was cooled down to room temperature and concentrated on the high-vacuum rotary evaporator. The resulting yellow residue was dissolved in chloroform and washed with water (x3). The organic phase was dried over Na_2SO_4 and concentrated under reduced pressure to give the crude product, a pale yellow solid. The crude was suspended in EtOH (50 mL) and 50-60% hydrazine hydrate (2.2 mL) was added. The reaction mixture was refluxed for 6 h. Then, the reaction mixture was cooled down to room temperature and concentrated on the rotary evaporator to a crude solid mass. The solid was suspended in 0.1 N aqueous NaOH and extracted with DCM (x4). All organic phases were combined, washed with water, dried over Na_2SO_4 and concentrated under reduced pressure. The crude was suspended on the silica gel and purified by column chromatography (gradient: DCM/MeOH 100/0 to 90/10 + 1% of triethylamine) to give **22** in the form of a pale yellow powder (0.21 g, 76%).

^{1}H NMR (400 MHz, Methanol-d_4): δ = 7.86 (s, 1H), 6.26 (s, 1H), 4.06 (s, 2H), 2.56 (s, 3H), 2.52 (s, 3H), 2.50 (s, 3H).

1-((2R,3R,4S,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3,4-dihydroxytetrahydro-furan-2-yl)pyrimidine-2,4(1H,3H)-dione (23)

Chemical Formula: $C_{30}H_{30}N_2O_8$
Molecular Weight: 546.58

The synthesis was performed following the previously reported protocol [354]. To a solution of uridine (8.00 g, 32.8 mmol, 1.00 eq) in dry pyridine (65 mL) DMTrCl (13.88 g, 41.0 mmol, 1.25 eq) was added in multiple portions over a period of 1.5 h. The reaction mixture was then stirred for 1 h at room temperature. The mixture was diluted with DCM and washed with water (x3). The organic layer was dried over Na_2SO_4 and concentrated on the rotary evaporator. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: DCM/MeOH 100/0 to 90/10 + 1% of triethylamine). The DMTr-protected product (**23**) was obtained in the form of slightly pink foam (12.81 g, 72%).

^{1}H NMR (400 MHz, Chloroform-d): δ = 8.01 (d, J = 8.1 Hz, 1H), 7.42 − 7.36 (m, 2H), 7.34 − 7.19 (m, 8H), 6.89 − 6.80 (m, 4H), 5.89 (d, J = 2.6 Hz, 1H), 5.37 (d, J = 8.1 Hz, 1H), 4.45 − 4.39 (m, 1H), 4.33 (dd, J = 5.1, 2.6 Hz, 1H), 4.18 (dt, J = 5.9, 2.4 Hz, 1H), 3.78 (s, 6H), 3.51 (qd, J = 11.1, 2.6 Hz, 2H).

1-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethyl-silyl)oxy)-4-hydroxytetrahydrofuran-2-yl)pyrimidine-2,4(1*H*,3*H*)-dione (24)
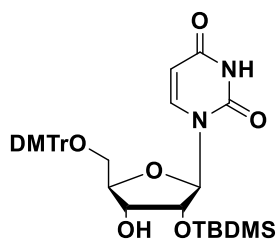
Chemical Formula: $C_{36}H_{44}N_2O_8Si$
Molecular Weight: 660.84

The synthesis was performed following the previously reported procedure [355]. 5'-*O*-DMTr-uridine (**23**) (8.00 g, 14.6 mmol, 1.00 eq) was dissolved in dry THF (98 mL). Dry pyridine (4.6 mL, 57.1 mmol, 3.90 eq) and $AgNO_3$ (2.73 g, 16.1 mmol, 1.10 eq) were added and stirring was continued for 20 min. Once the majority of the $AgNO_3$ dissolved (cloudy solution), TBDMSCl (2.43 g, 16.1 mmol, 1.10 eq) was added and the resulting milky solution was stirred at room temperature overnight. The reaction mixture was filtered into a 5% aqueous $NaHCO_3$ solution to avoid removal of the DMTr group. This mixture was concentrated on the rotary evaporator to remove the majority of the solvent and was then diluted with EtOAc and extracted with saturated aqueous $NaHCO_3$ (x3). The organic layer was dried over Na_2SO_4 and condensed to dryness. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: hexane/*tert*-butylmethyl ether 100/0 to 50/50). The 2'-*O*-TBDMS-protected product (**24**) was obtained as a white foam (4.87 g, 49%).

1H NMR (400 MHz, DMSO-d_6): δ = 11.37 (s, 1H), 7.74 (d, *J* = 8.1 Hz, 1H), 7.40 − 7.28 (m, 4H), 7.24 (dd, *J* = 8.9, 2.1 Hz, 5H), 6.90 (d, *J* = 8.9 Hz, 4H), 5.77 − 5.73 (m, 1H), 5.32 − 5.27 (m, 1H), 5.12 (d, *J* = 6.2 Hz, 1H), 4.23 − 4.18 (m, 1H), 4.10 − 4.03 (m, 1H), 4.01 − 3.96 (m, 1H), 3.74 (s, 6H), 3.30 (d, *J* = 4.2 Hz, 1H), 3.24 (dd, *J* = 10.7, 2.5 Hz, 1H), 0.85 (s, 9H), 0.06 (d, *J* = 3.6 Hz, 6H).

(2*R*,3*R*,4*R*,5*R*)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((*tert*-butyldimethylsilyl)oxy)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2*H*)-yl)tetrahydrofuran-3-yl acetate (25)
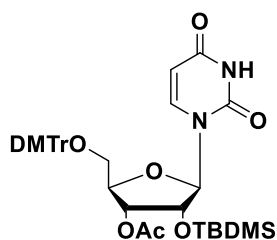
Chemical Formula: $C_{38}H_{46}N_2O_9Si$
Molecular Weight: 702.88

Compound **24** (1.00 g, 1.51 mmol, 1.00 eq) and DMAP (0.07 g, 0.57 mmol, 0.38 eq) were dissolved in dry pyridine (8.0 mL). Acetic anhydride (1.20 mL, 12.69 mmol, 8.39 eq) was added dropwise and the reaction mixture was stirred at room temperature overnight. Then, the reaction mixture was concentrated on the rotary evaporator. The crude was dissolved in DCM and washed with saturated $NaHCO_3$ (x2), brine (x2), dried with Na_2SO_4 and concentrated under reduced pressure. The crude product was diluted with DCM and purified using Isolera One flash chromatography system (gradient: hexane/EtOAc 9/1 to 1/1) to give the acetylated product (**25**) in the form of a white solid (0.98 g, 84%).

1H NMR (400 MHz, DMSO-d_6): δ = 11.49 (s, 1H), 7.71 (d, *J* = 8.1 Hz, 1H), 7.41 − 7.30 (m, 4H), 7.29 − 7.20 (m, 5H), 6.92 (d, *J* = 8.9 Hz, 4H), 5.81 (d, *J* = 6.2 Hz, 1H), 5.51 (d, *J* = 8.1 Hz, 1H), 5.19 (dd, *J* = 5.4, 3.6 Hz, 1H), 4.53 (t, *J* = 5.9 Hz, 1H), 4.16 (q, *J* = 3.7 Hz, 1H), 3.75 (s, 6H), 3.38 (dd, *J* = 10.7, 4.5 Hz, 1H), 3.28 (dd, *J* = 10.7, 3.3 Hz, 1H), 2.07 (s, 3H), 0.81 (s, 9H), -0.00 (d, *J* = 7.2 Hz, 6H). **13C NMR** (101 MHz, DMSO-d_6): δ = 169.3, 162.7, 2 x 158.2, 150.5, 140.0, 134.9, 4 x 129.7, 4 x 128.0, 2 x 127.6, 126.9, 4 x 113.3, 102.3, 87.5, 86.2, 80.6, 72.7, 71.5, 63.0, 2 x 55.1, 3 x 25.3, 2x 20.6, -5.4, -5.5.

(2R,3R,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((*tert*-butyldimethylsilyl)oxy)-5-(2-oxo-4-(1*H*-1,2,4-triazol-1-yl)-3,4-dihydropyrimidin-1(2*H*)-yl)tetrahydrofuran-3-yl acetate (26)
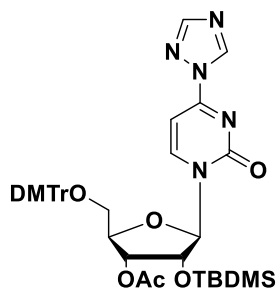
Chemical Formula: $C_{40}H_{49}N_5O_8Si$
Molecular Weight: 755.94

The product was obtained adapting the procedure published by Wiesmayr et al. [356]. Compound **25** (0.60 g, 0.85 mmol, 1.00 eq) and 1,2,4-triazole (0.94 g, 13.7 mmol, 16.00 eq) were suspended in anhydrous acetonitrile (7.5 mL) and triethylamine (2.74 mL, 19.6 mmol, 23.00 eq) under argon atmosphere and stirred for 10 minutes until a clear, colourless solution was obtained. The reaction mixture was cooled down to 0°C. After 10 min, POCl₃ (0.16 mL, 1.71 mmol, 2.00 eq) was added dropwise, whereupon a white solid precipitated. The reaction mixture was left to warm up to room temperature and was stirred overnight. Then, EtOAc was added and the reaction mixture was washed with saturated aqueous NaHCO₃ (x2), brine, dried over Na₂SO₄ and concentrated on the rotary evaporator. The crude product was diluted with a small amount of DCM and purified using Isolera One flash chromatography system (gradient: hexane/EtOAc 85/15 to 1/1 +1% of triethylamine) to give the product **26** in the form of a white foam (0.57 g; 88%).

¹H NMR (400 MHz, Acetonitrile-d_3): δ = 9.19 (s, 1H), 8.60 (d, J = 7.3 Hz, 1H), 8.15 (s, 1H), 7.43 (dd, J = 8.3, 1.2 Hz, 2H), 7.37 – 7.23 (m, 7H), 6.95 – 6.84 (m, 4H), 6.64 (d, J = 7.3 Hz, 1H), 5.83 (d, J = 2.2 Hz, 1H), 5.21 (dd, J = 7.8, 4.5 Hz, 1H), 4.59 (dd, J = 4.5, 2.2 Hz, 1H), 4.35 (dt, J = 7.8, 2.6 Hz, 1H), 3.78 (d, J = 0.7 Hz, 6H), 3.58 – 3.41 (m, 2H), 2.06 (s, 3H), 0.90 (s, 9H), 0.17 (s, 3H), 0.07 (s, 3H). **¹³C NMR** (101 MHz, Acetonitrile-d_3): δ = 154.9, 148.3, 144.2, 2 x 136.6, 2 x 131.2, 2 x 131.0, 4 x 129.1, 128.1, 6 x 118.3, 4 x 114.3, 95.4, 92.8, 88.0, 81.6, 75.5, 71.3, 62.1, 2 x 56.0, 3 x 26.0, 2 x 21.1, 2 x -4.5.

VII.3.2. Oligoribonucleotide synthesis, functionalization, deprotection and purification

VII.3.2.1. ORN-54a to ORN-54e and ORN-55a to ORN-55e

The synthesis, functionalization with use of the CLICK chemistry, deprotection and purification of the oligoribonucleotides were performed as described in sections VII.1.1 and VII.1.2 with small changes. 15 mg portions of the CPG with oligoribonucleotides modified with the 2'-*O*-propargyl cytidine (ORN-54 or ORN-55) were prepared and treated with the doubled amount of the reagents (azides **20a**-**20d** and 4'-azidomethyl-4,5',8-trimethylpsoralen were used). After the deprotection, oligoribonucleotides were dissolved in 200 µL of ultrapure water and purified by RP-HPLC in DMTr-on mode. Eluent A: 0.1 N aqueous triethylamine/acetic acid, pH 8.0; eluent D: 100% ACN; flow-rate: 5 mL/min. Gradient 10-50% D in 5 min. Fractions containing the DMTr-on product were collected, dried in a SpeedVac and treated with 40% aqueous acetic acid (200 µL) for 15 min at room temperature. Samples were concentrated in a SpeedVac, dissolved in 200 µL of ultrapure water and purified in DMTr-off mode by RP-HPLC. Eluent A: 0.1 N aqueous triethylamine/acetic acid, pH 8.0; eluent D: 100% ACN; flow-rate: 5 ml/min; gradient 2-20% D in 6 min. The gradient was changed for samples with PEG3 and PEG4 linkers (2-22% D in 8 min). Fractions containing the product were collected and analysed by LC-MS.

VII.3.2.2. ORN-56a to ORN-56e

The synthesis, deprotection and purification of the oligoribonucleotides were performed as described in section VII.1.1 with small changes.

The post-synthetic triazole substitution: 20 mg portions of the CPG with ORN-56 modified with the O^4-triazolyluridine were prepared in screw cap tubes. Amines (**21a-21d** or **22**; 7-20 mg/reaction, depending on the linker length) were dissolved in a mixture of 150 µL of dry ACN and 50 µL of DBU

and added to the CPG. The suspensions were then shaken for 6 h at 60°C and concentrated in the SpeedVac. Further deprotection and purification were performed according to the general procedure for the modified oligoribonucleotides. Gradient for the DMTr-on purification: 10-50% D in 5 min. Gradient for DMTr-off purification: 2-20% D in 6 min. The gradient was changed for samples with PEG3 and PEG4 linkers (2-22% D in 8 min). Some samples were re-purified multiple times. Fractions containing the products were collected and analysed by LC-MS.

VII.3.3. *In vitro* photo-cross-linking

The *in vitro* photo-cross-linking assay was performed as described in section VII.1.3.2. The concentration of each strand of the duplex was 3 µM. Samples were irradiated for 15 min.

The *in vitro* photo-cross-linking with a mixture of probes:
Samples were prepared according to the standard protocol with a small modification; calculated volumes (corresponding to the 2 µM concentrations) of the trioxsalen-modified miR-124-3p analogues (ORN-56a to ORN-56e) were taken and mixed with the calculated volume (corresponding to the 10 µM concentration) of the unmodified counter-strand (ORN-31). The mixture was concentrated, re-dissolved in 200 µL of the phosphate buffer, annealed and cooled down to room temperature. After irradiation, samples were analysed according to the standard procedure.

For some of the compounds, sodium and potassium adducts were observed during the LC-MS analysis.

VII.3.4. Thermal stability studies

UV melting was performed as described in section VII.2.4 with small changes: oligoribonucleotides were dissolved in 200 µL of the melting buffer so that the final concentration of each strand of the duplex was 2 µM. Additionally, 100 µL of mineral oil was added on the top of the samples to prevent the evaporation of the solutions during measurements.

VII.3.5. Chromatograms from the *in vitro* photo-cross-linking experiments

VII.3.5.1. miR-124-3p-C6-2′-*O*-prop-PEG

D42
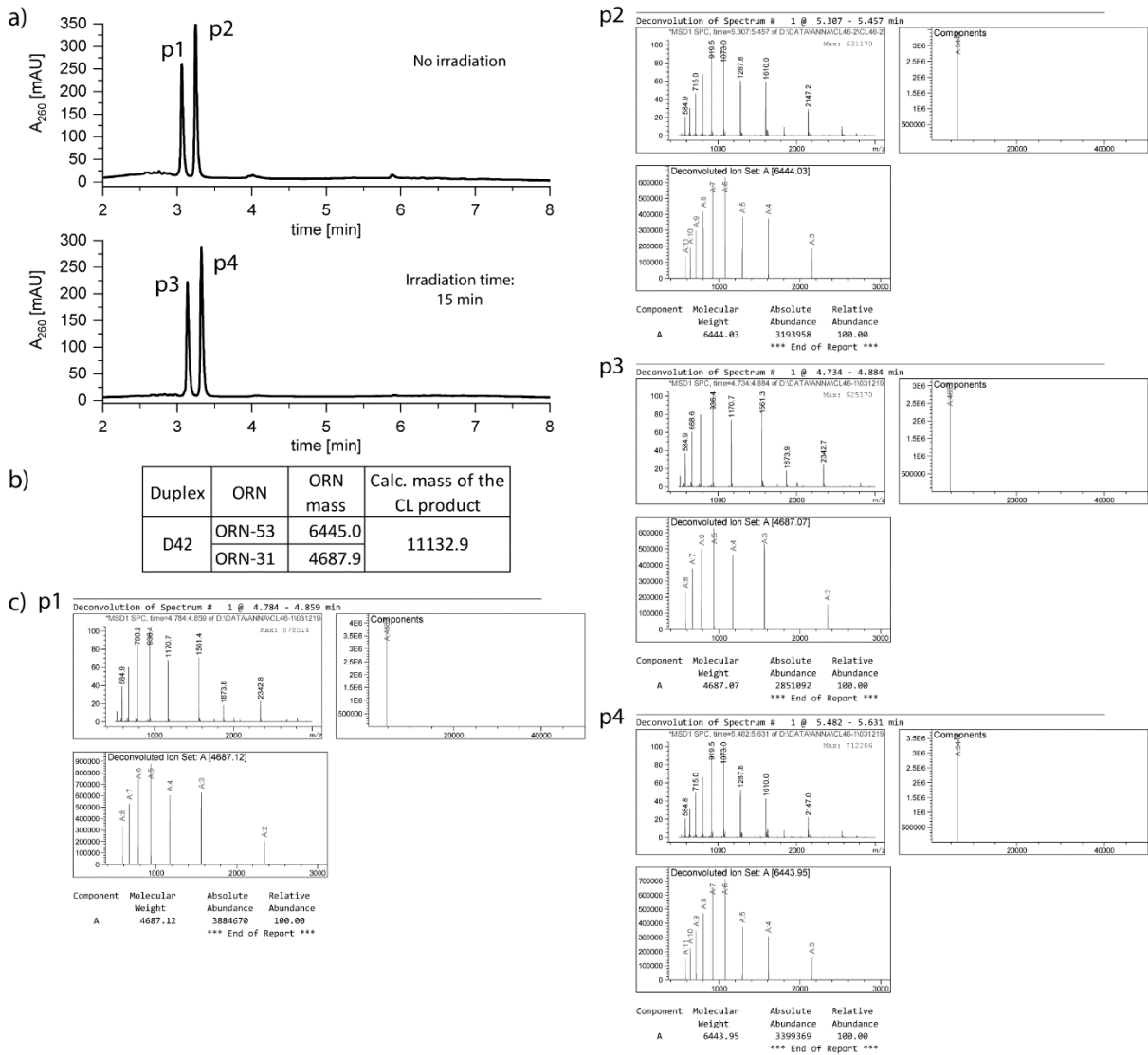
Figure 97: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D42; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D43

a)

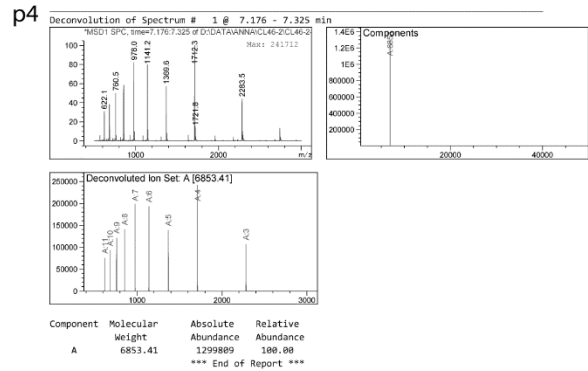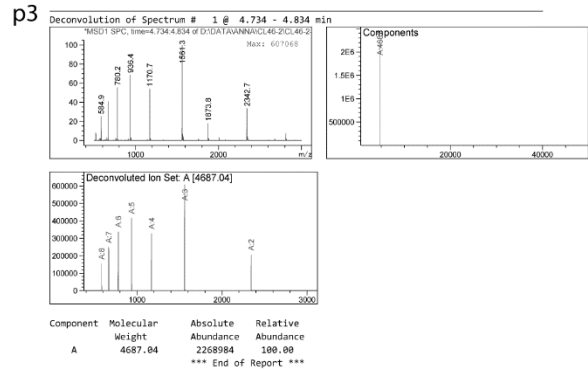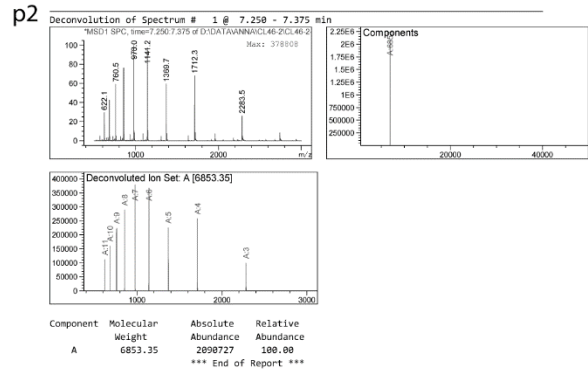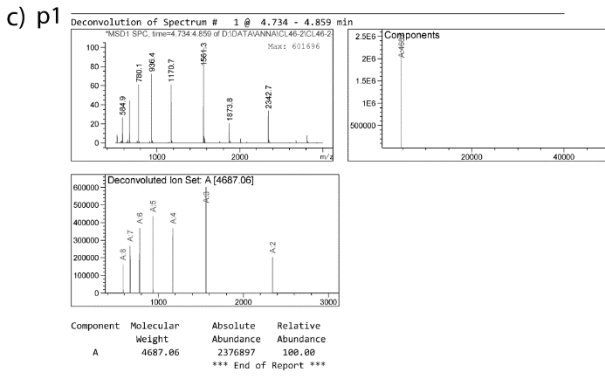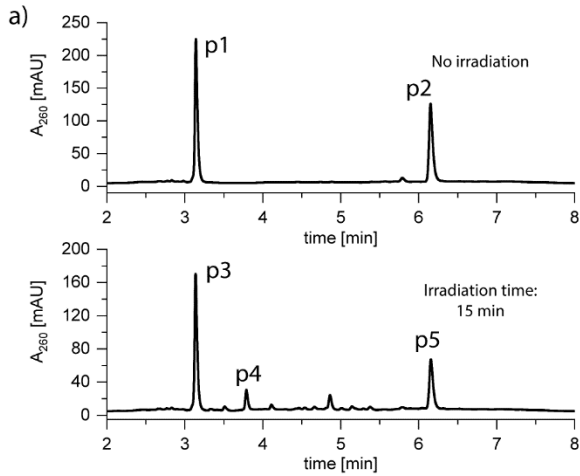b)
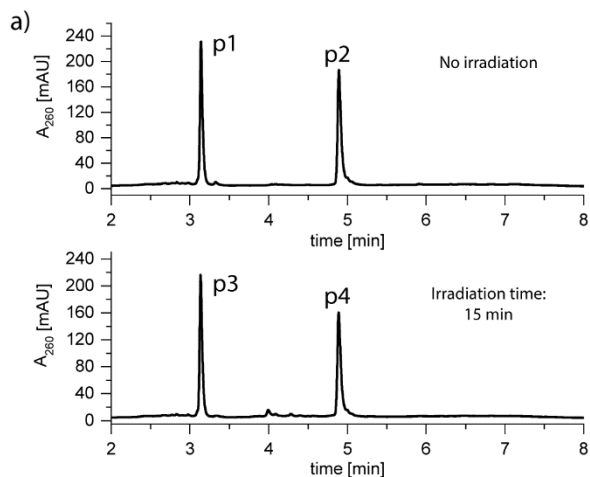
Duplex	ORN	ORN mass	Calc. mass of the CL product
D43	ORN-54a	6766.4	11454.3
	ORN-31	4687.9	

c) p1

p2

p3

p4
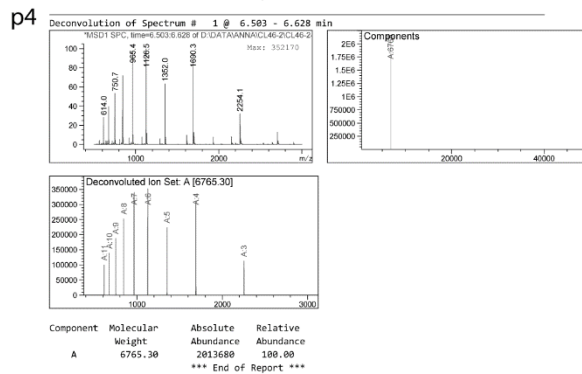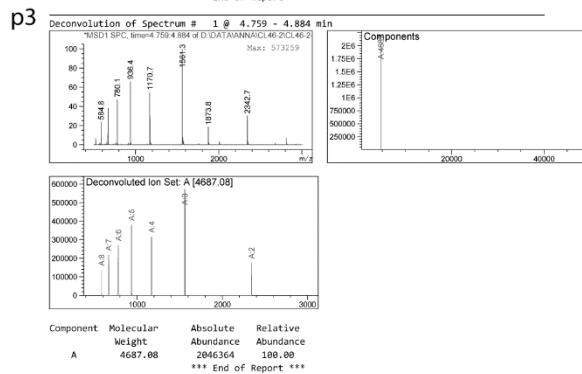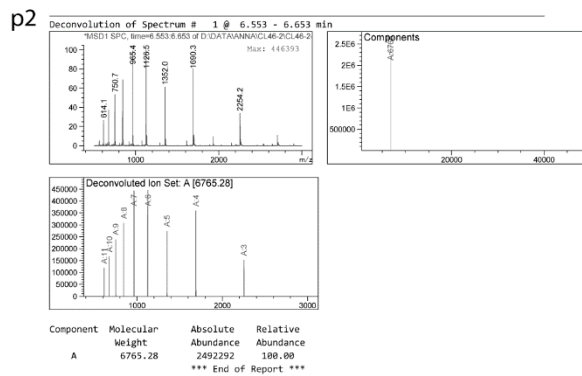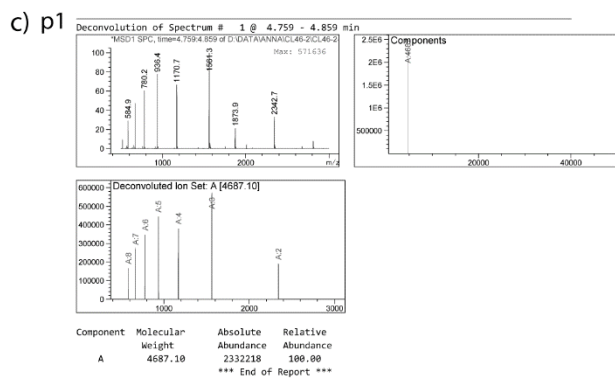
Figure 98: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D43; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

155

D44

a)

p3

b)
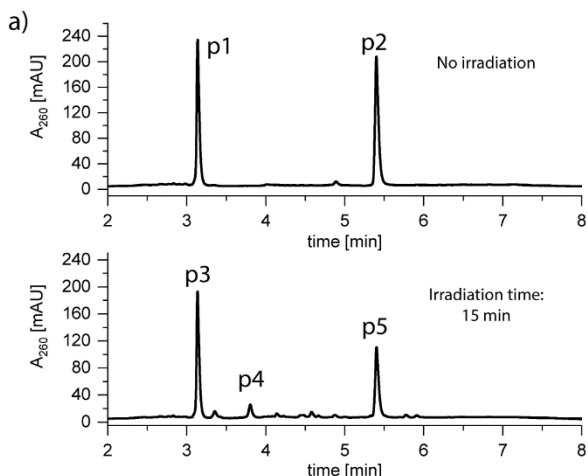
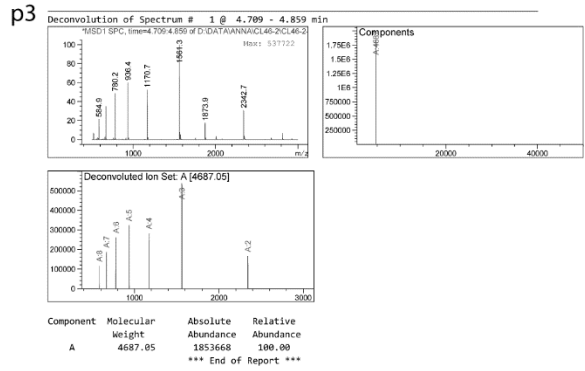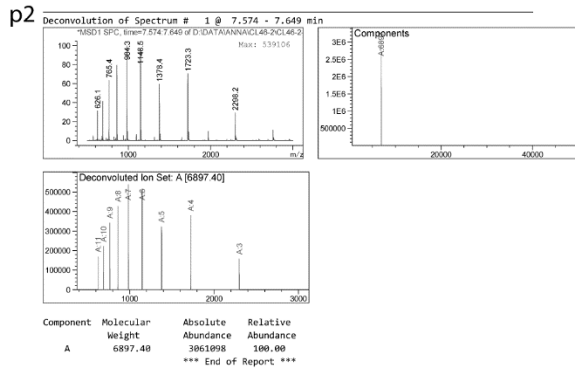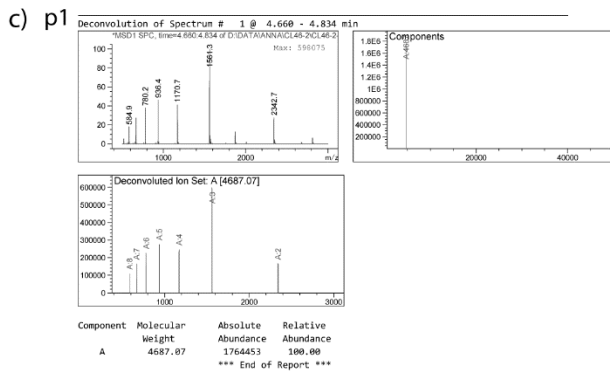| Duplex	ORN	ORN mass	Calc. mass of the CL product
D44	ORN-54b	6810.4	11498.3
	ORN-31	4687.9	

p4

c) p1

p2

p5
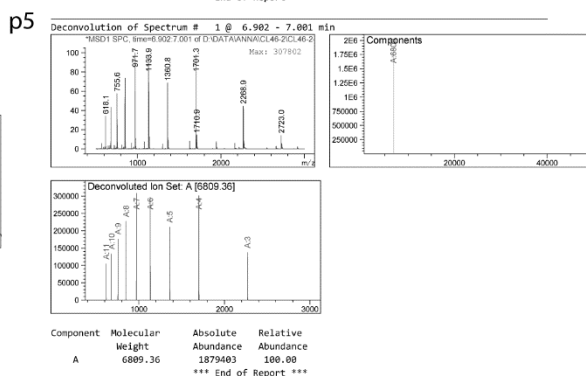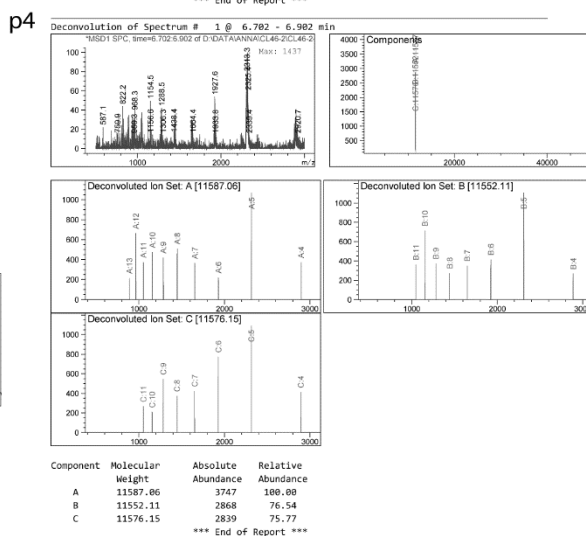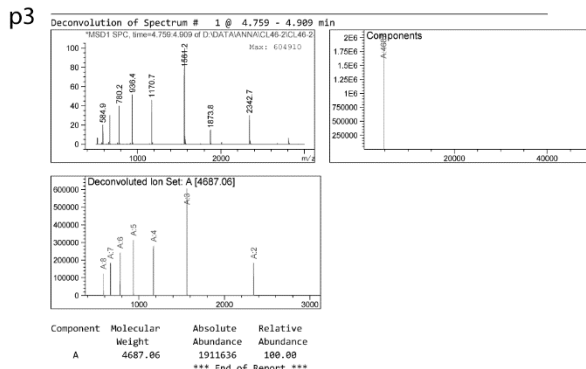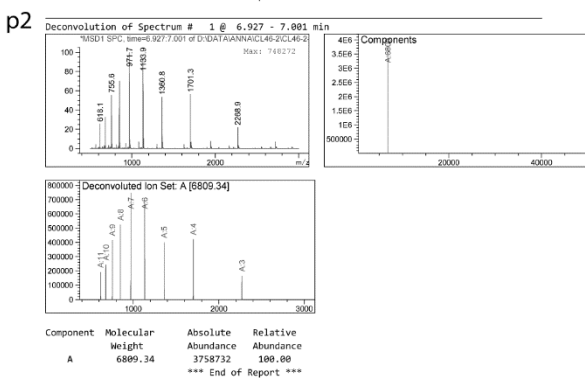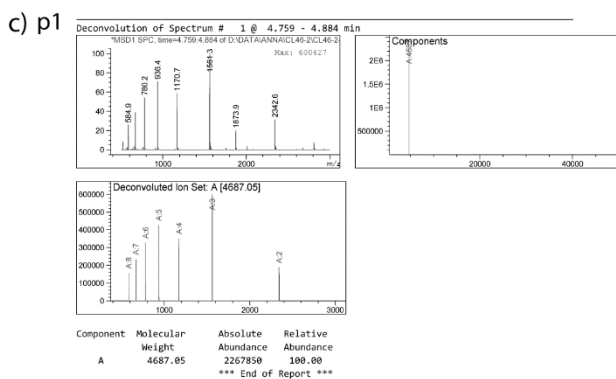
Figure 99: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D44; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D45

a)

b)
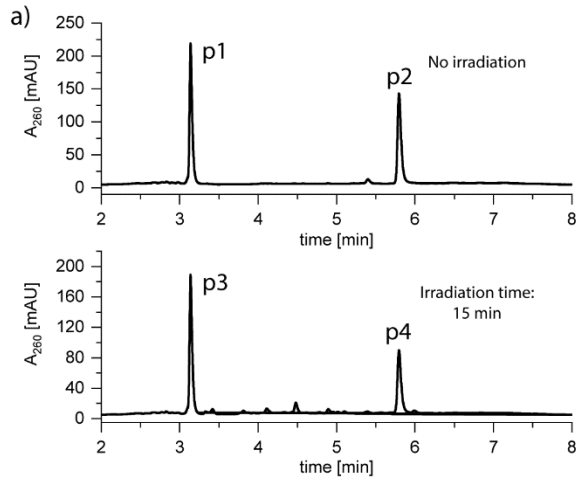
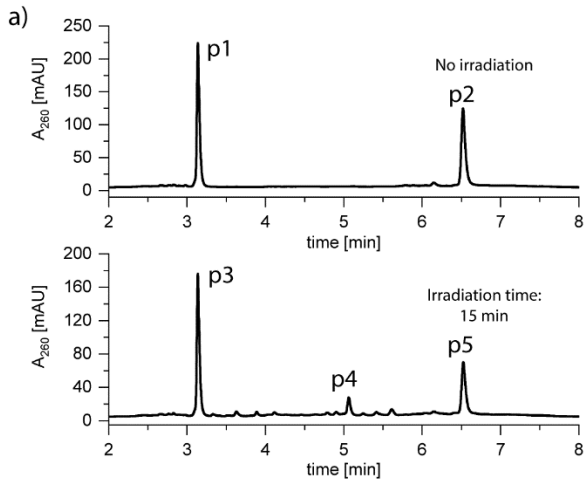| Duplex	ORN	ORN mass	Calc. mass of the CL product
D45	ORN-54c	6854.5	11542.4
	ORN-31	4687.9	

c) p1

p2

p3

p4

Figure 100: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D45; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
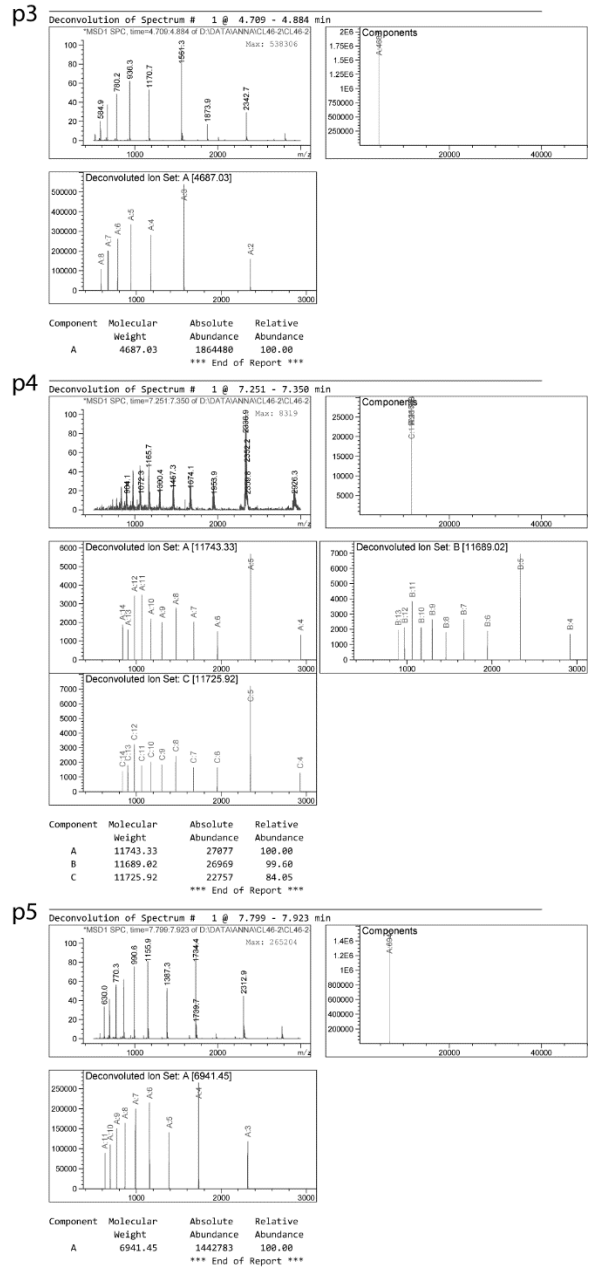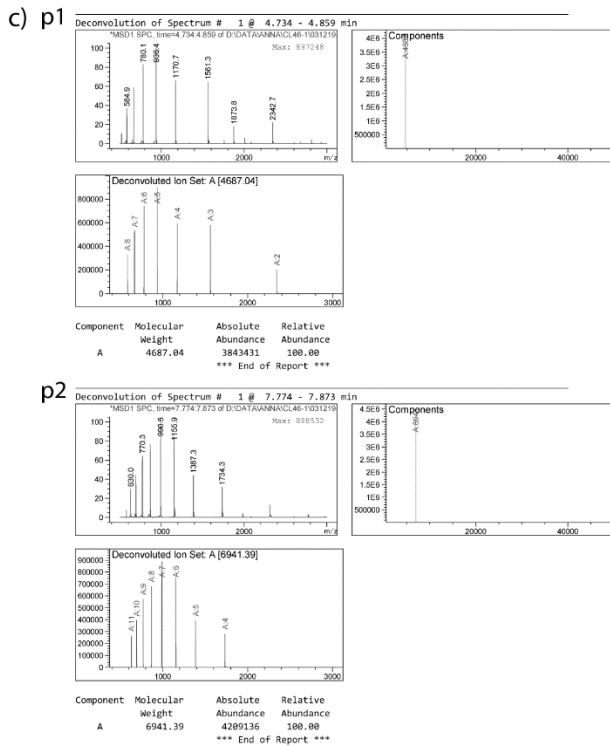
D46

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D46	ORN-54d	6898.5	11586.4
	ORN-31	4687.9	

c) p1
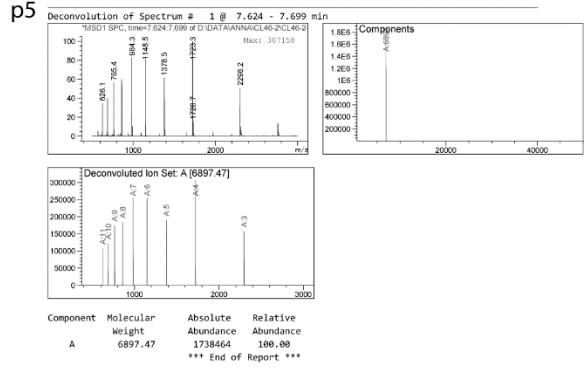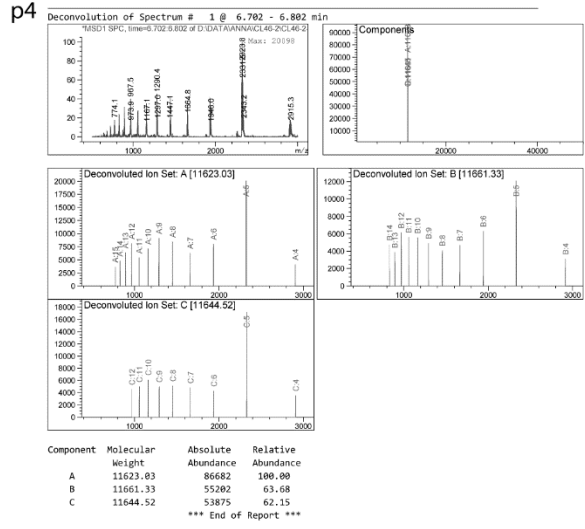
Figure 101: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D46; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D47

a)

b)

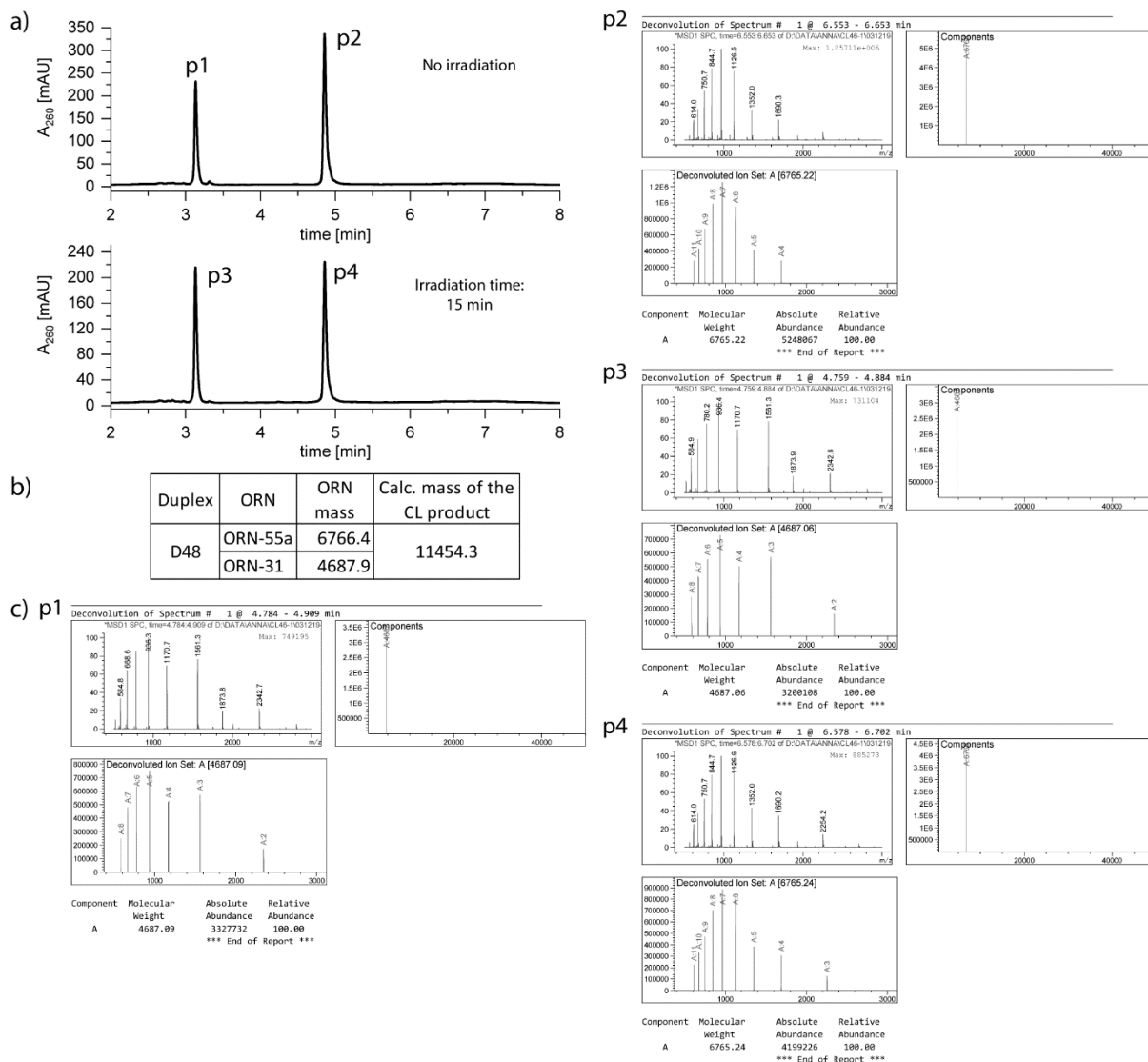| Duplex	ORN	ORN mass	Calc. mass of the CL product
D47	ORN-54e	6942.6	11630.5
	ORN-31	4687.9	

c) p1

p2

p3

p4

p5

Figure 102: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D47; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
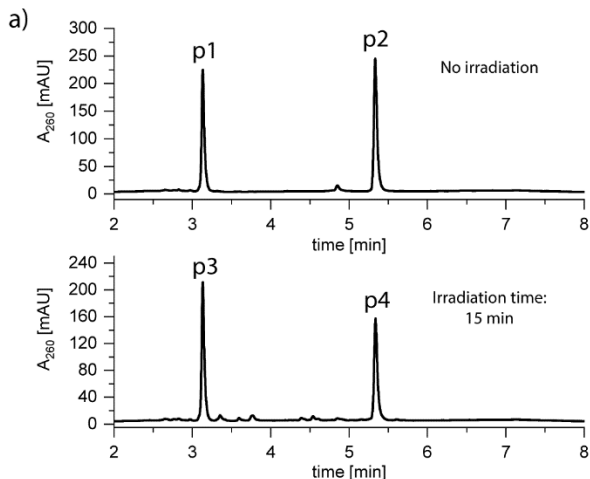
VII.3.5.2. miR-124-3p-C8-2′-O-prop-PEG

D48

a)

b)

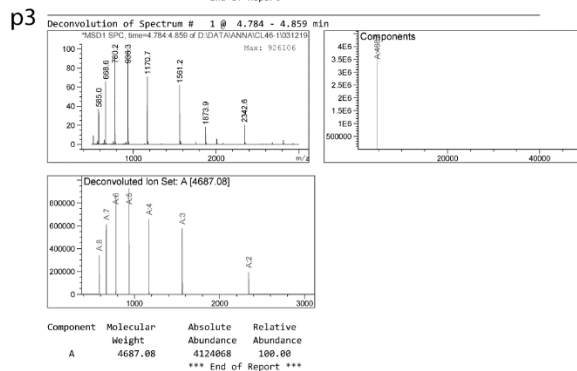| Duplex	ORN	ORN mass	Calc. mass of the CL product
D48	ORN-55a	6766.4	11454.3
	ORN-31	4687.9	

c) p1

Figure 103: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D48; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
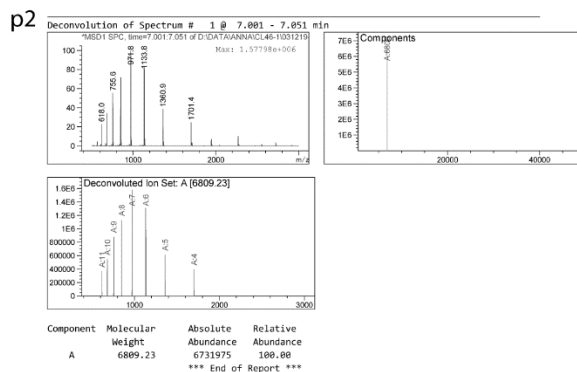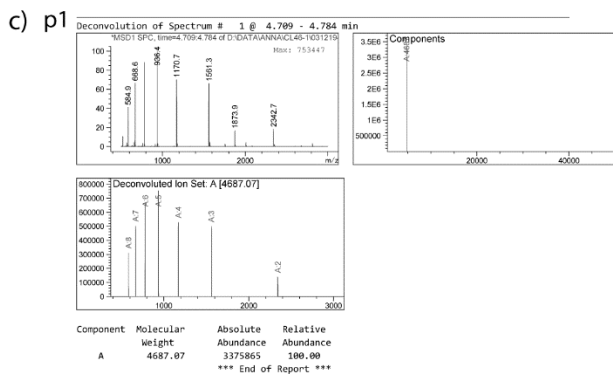
D49
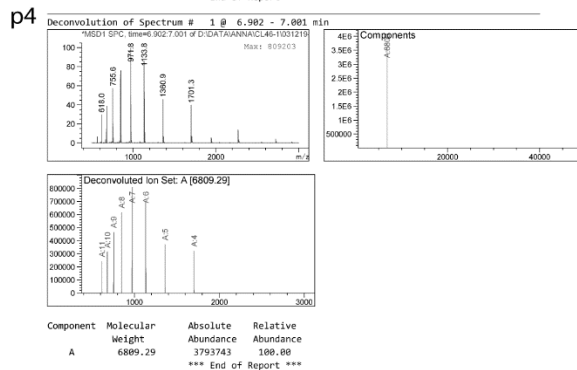
Figure 104: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D49; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
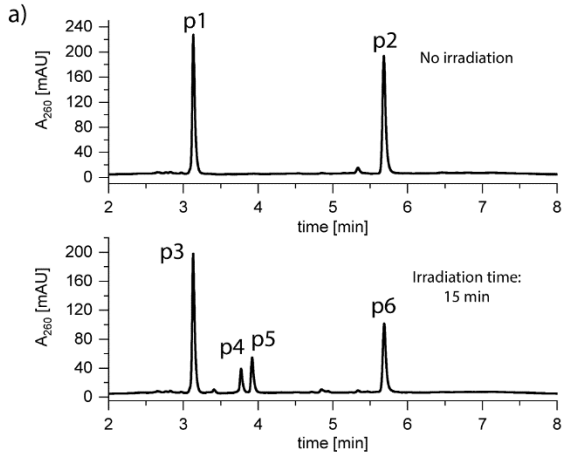
D50
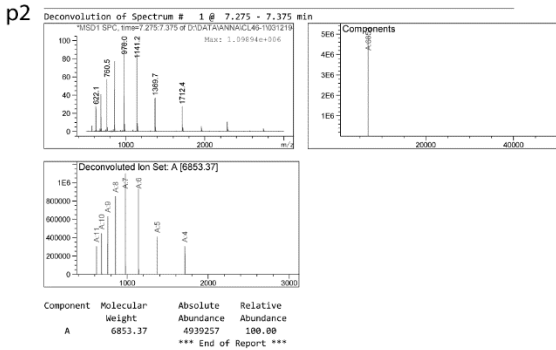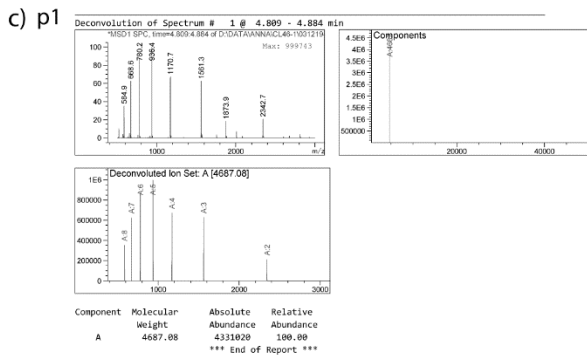
Figure 105: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D50; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
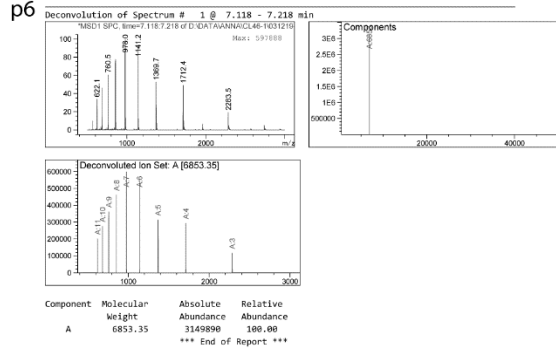
D51

a)

b)
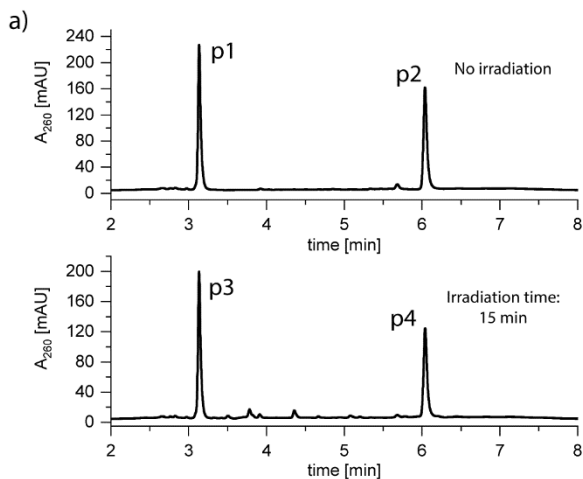
Duplex	ORN	ORN mass	Calc. mass of the CL product
D51	ORN-55d	6898.5	11586.4
	ORN-31	4687.9	

c) p1

Figure 106: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D51; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
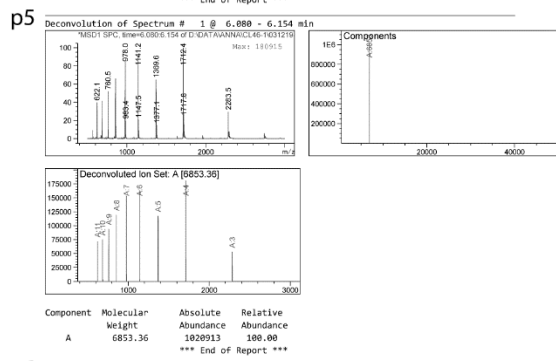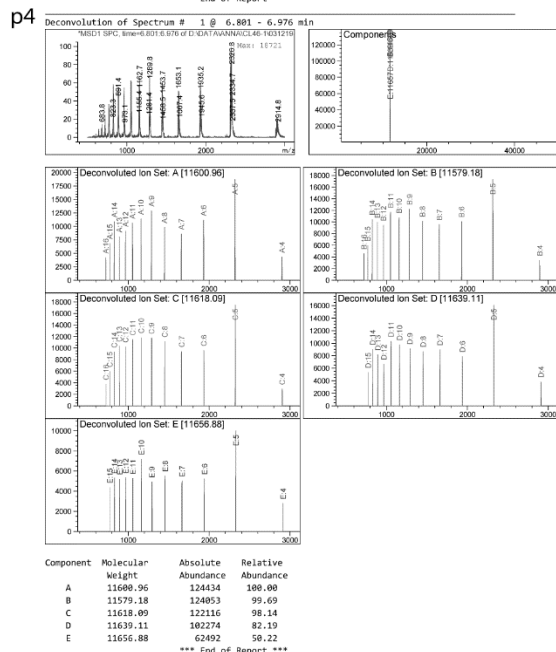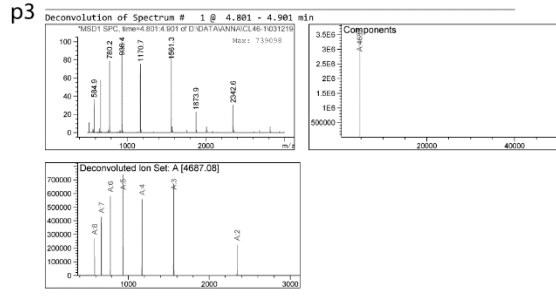
D52

a)

b)

Duplex	ORN	ORN mass	Calc. mass of the CL product
D52	ORN-55e	6942.6	11630.5
	ORN-31	4687.9	

c) p1

p2

p3

p4

Figure 107: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D52; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.3.5.3. miR-124-3p-C6-*N*[4]-PEG

D53
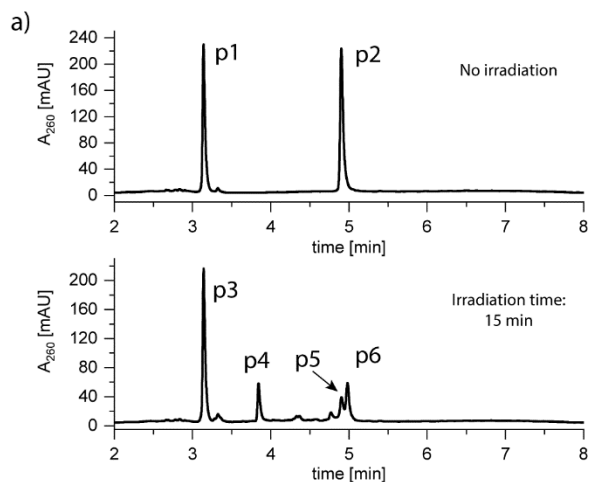
a)

b)

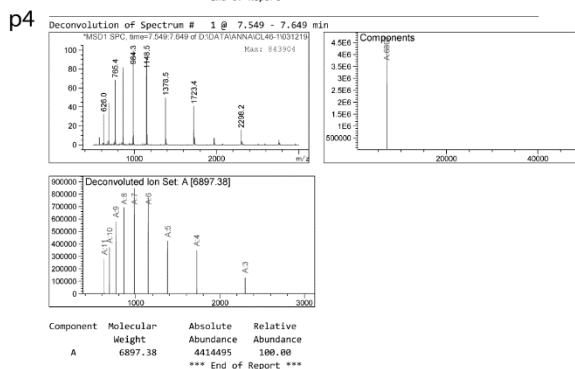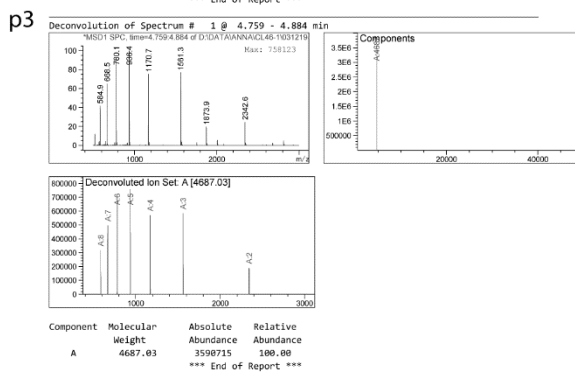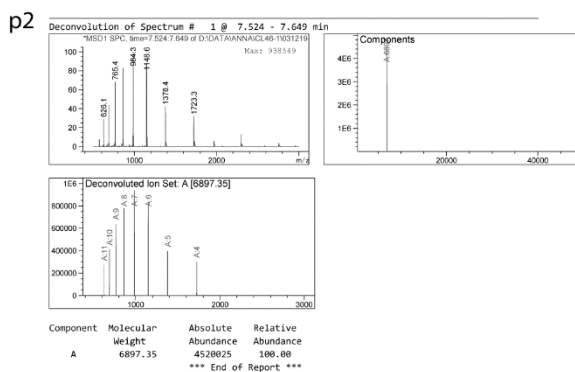| Duplex	ORN	ORN mass	Calc. mass of the CL product
D53	ORN-56a	6685.3	11373.2
	ORN-31	4687.9	

c) p1
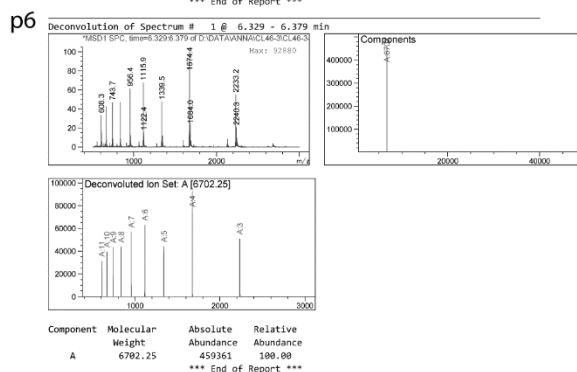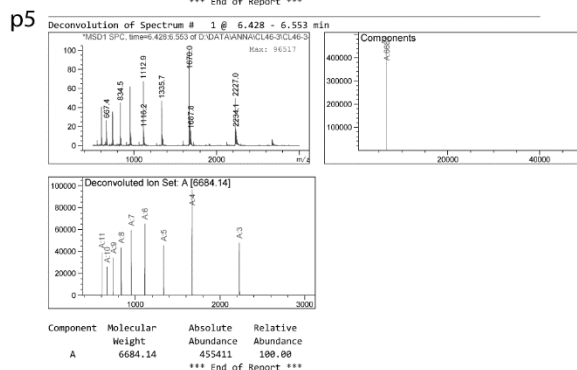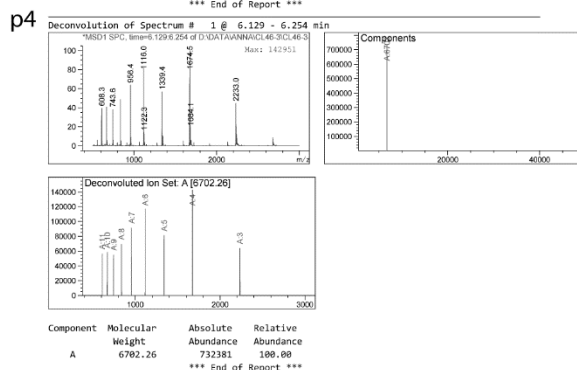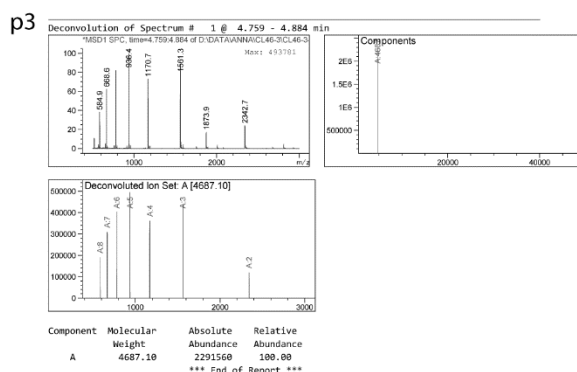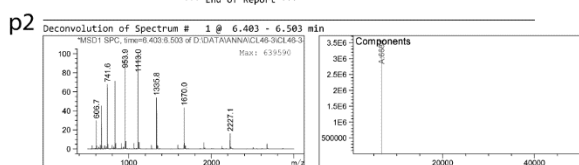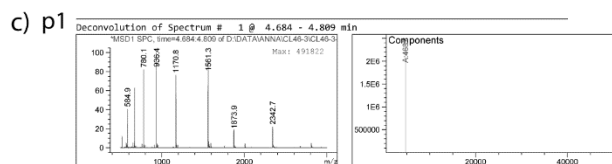
Figure 108: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D53; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D54

a)

b)
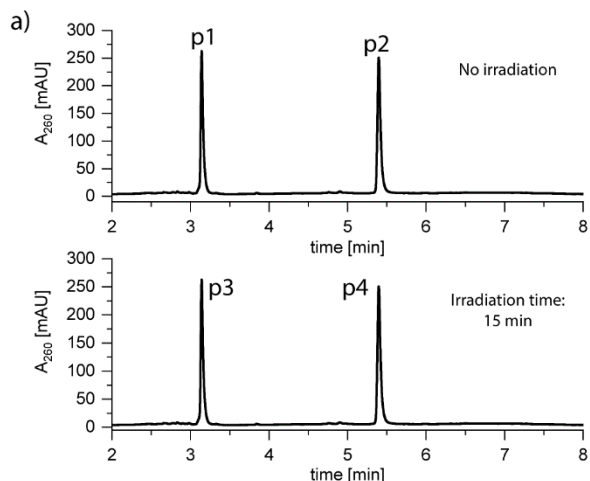
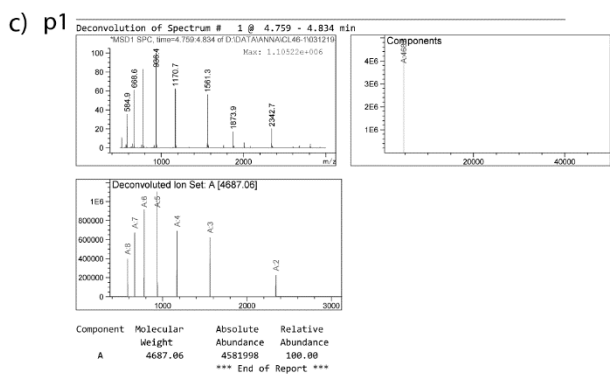| Duplex	ORN	ORN mass	Calc. mass of the CL product
D54	ORN-56b	6729.3	11417.2
	ORN-31	4687.9	

c) p1
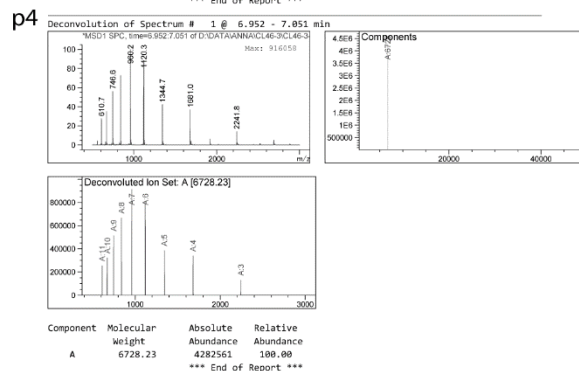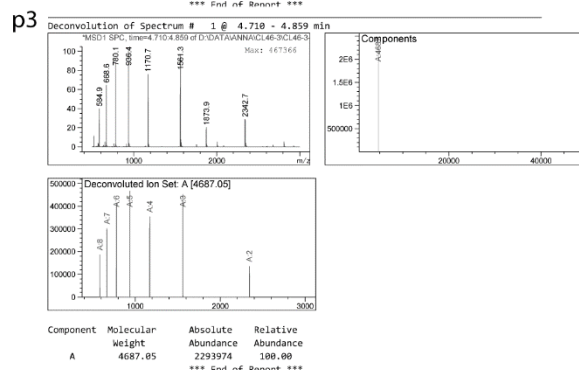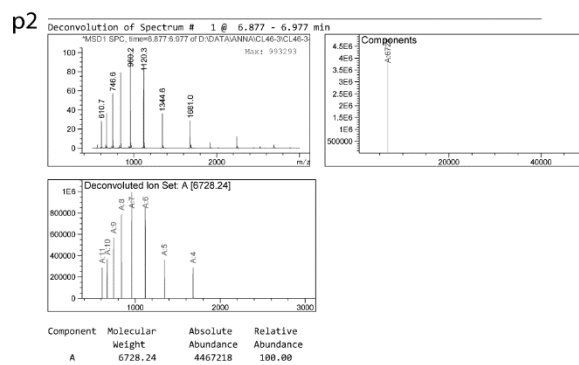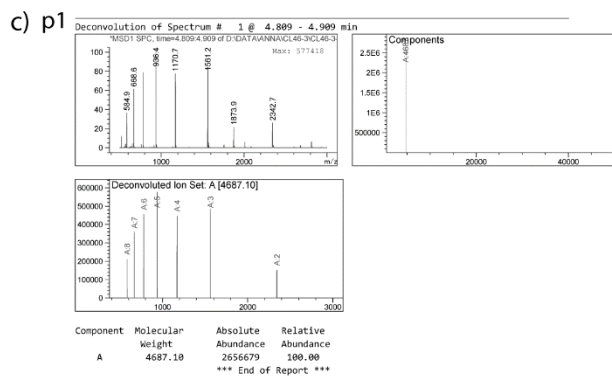
Figure 109: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D54; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.
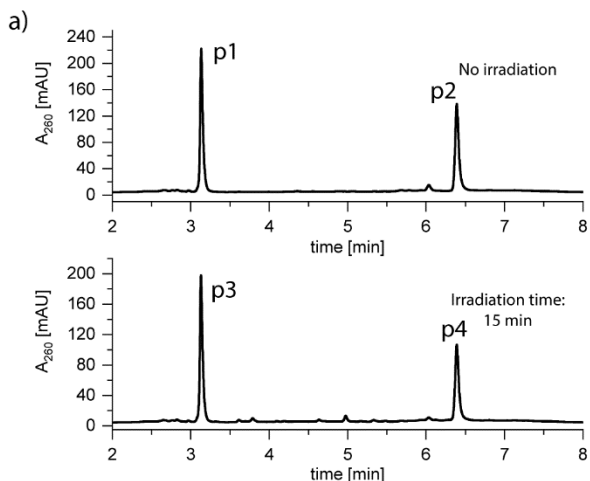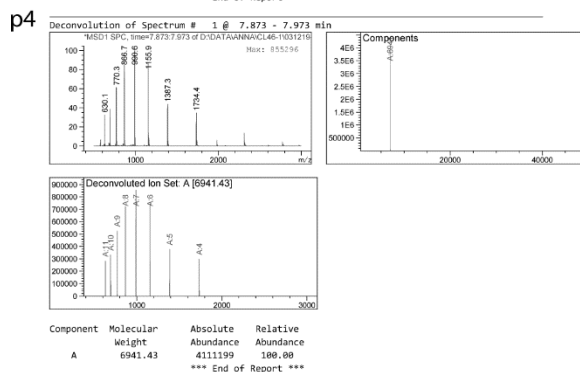
D55
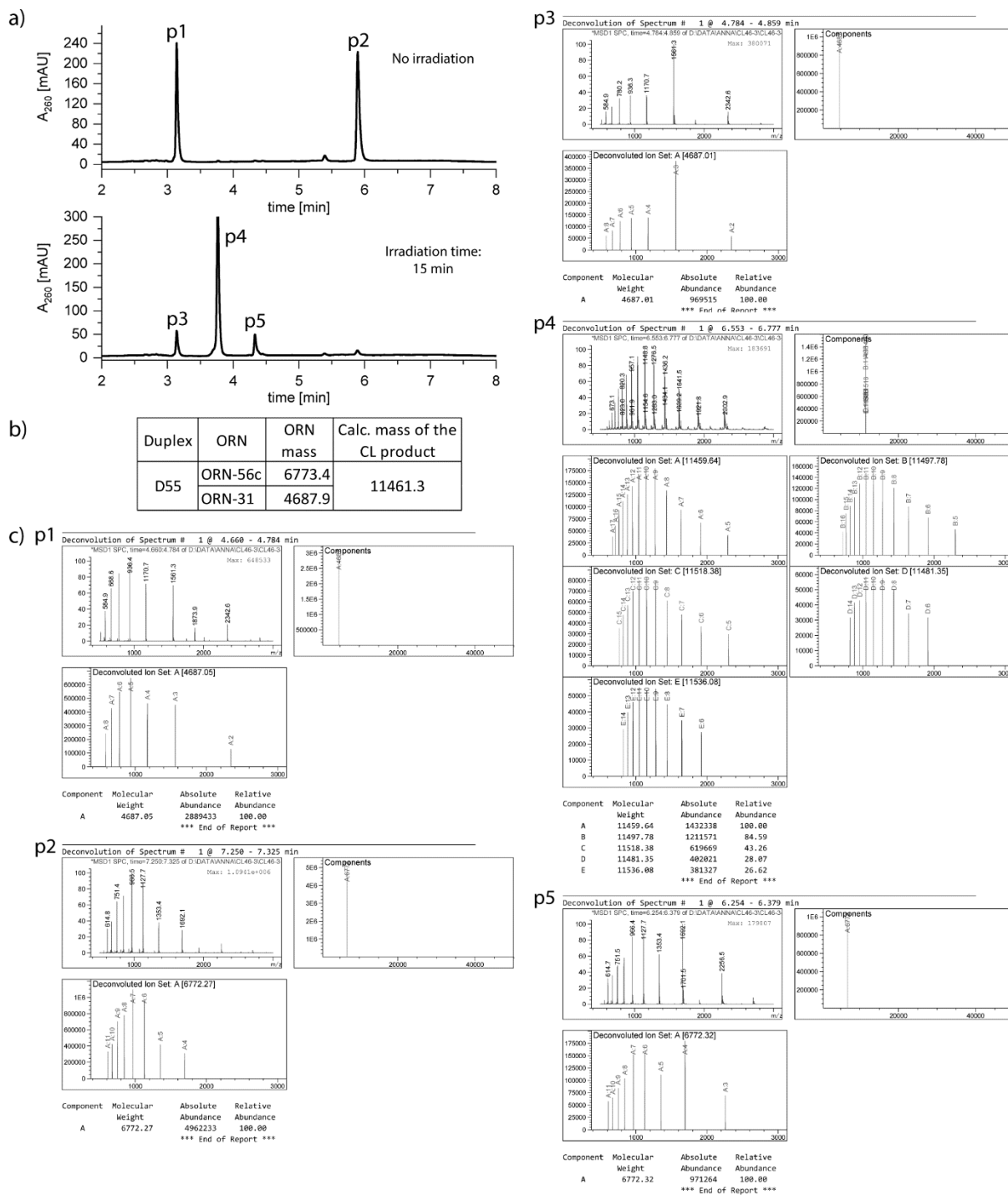
Figure 110: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D55; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D56

a)

b)
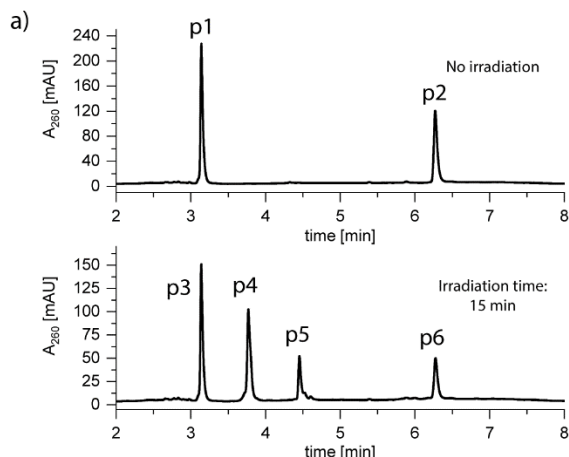
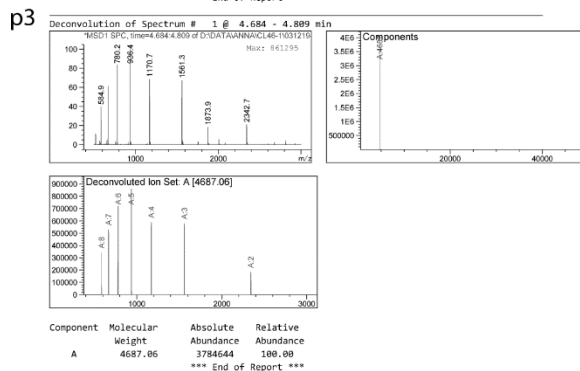| Duplex	ORN	ORN mass	Calc. mass of the CL product
D56	ORN-56d	6817.4	11505.3
	ORN-31	4687.9	

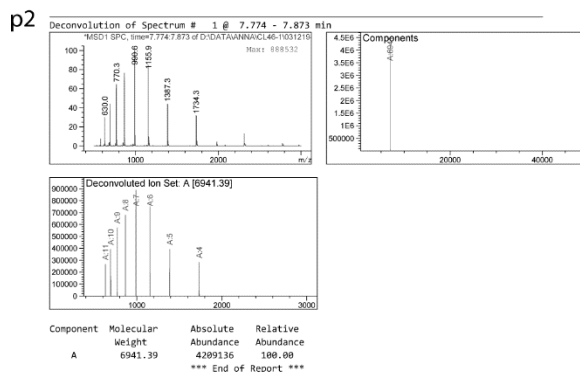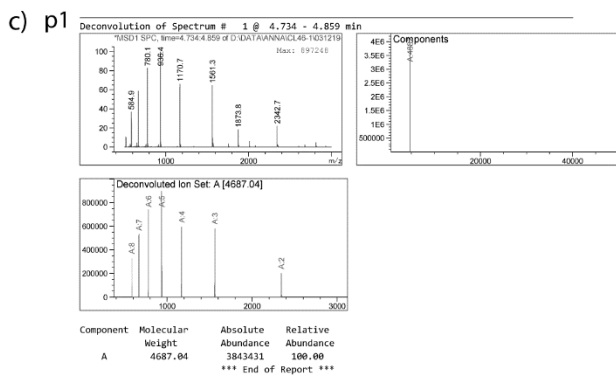c) p1

p2

p3

p4

p5

p6
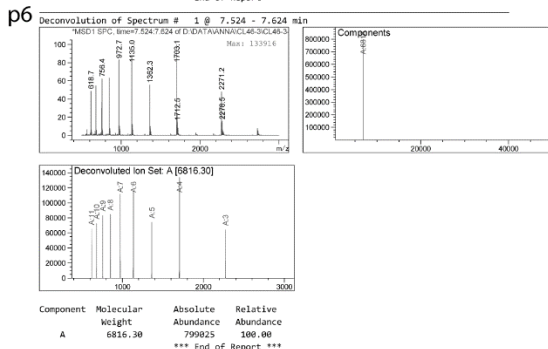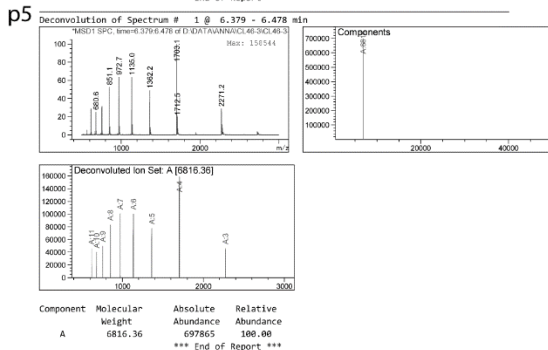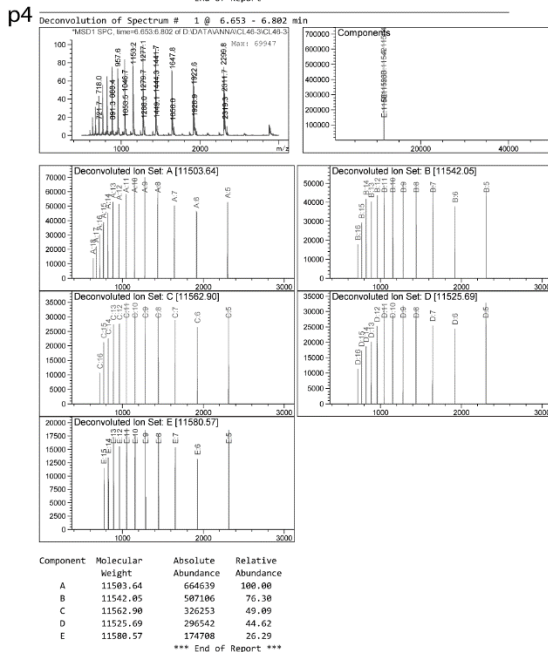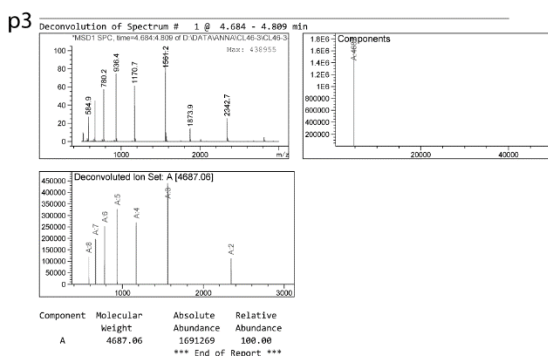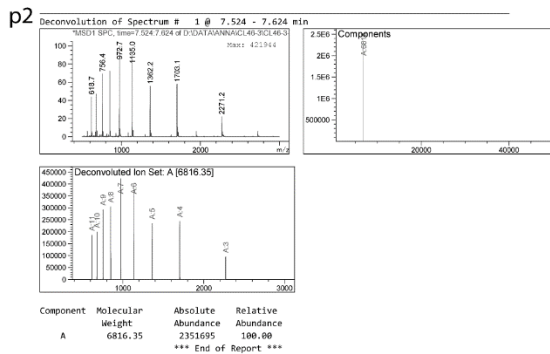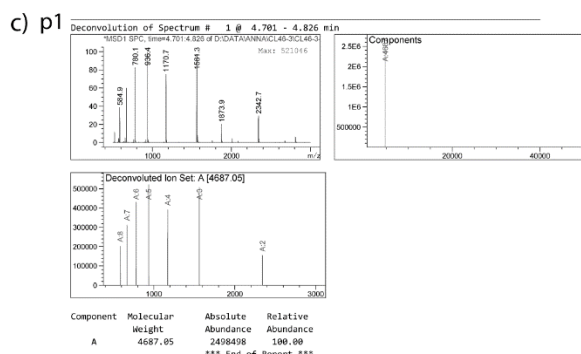
Figure 111: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D56; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D57

a)

b)
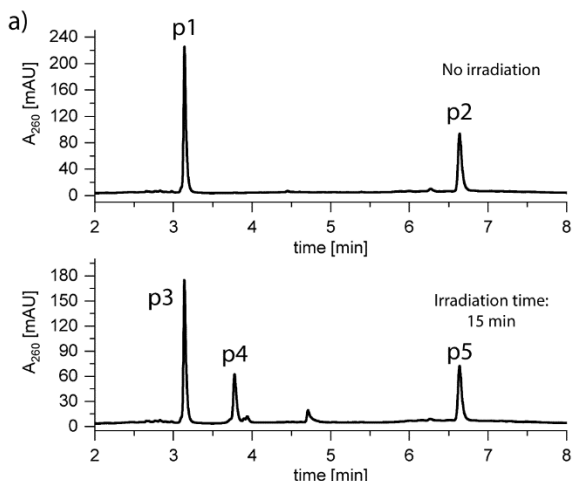
Duplex	ORN	ORN mass	Calc. mass of the CL product
D57	ORN-56e	6861.5	11549.4
	ORN-31	4687.9	

c)
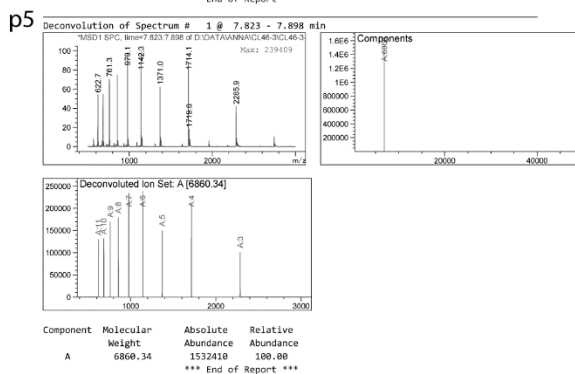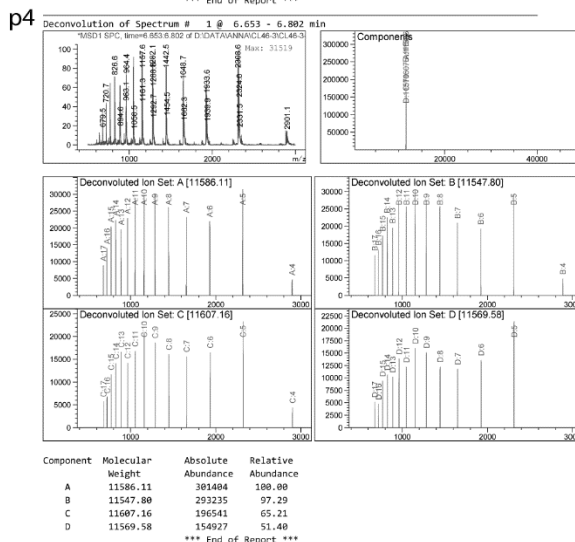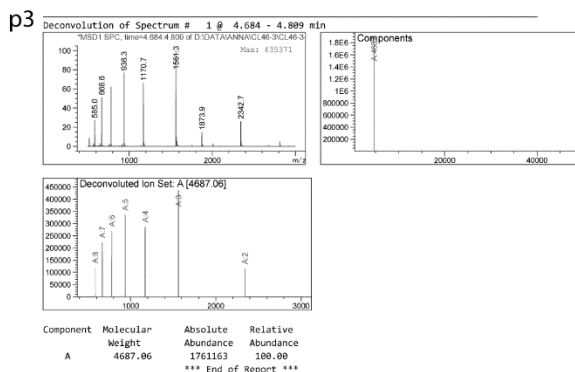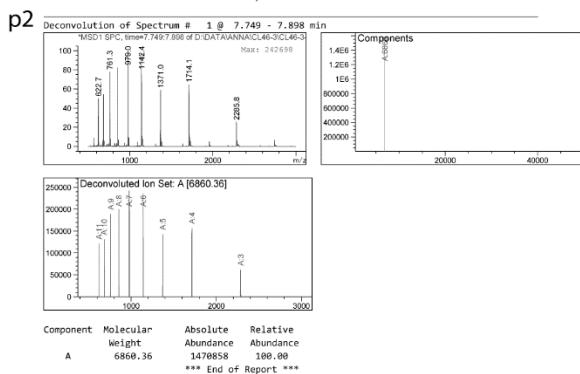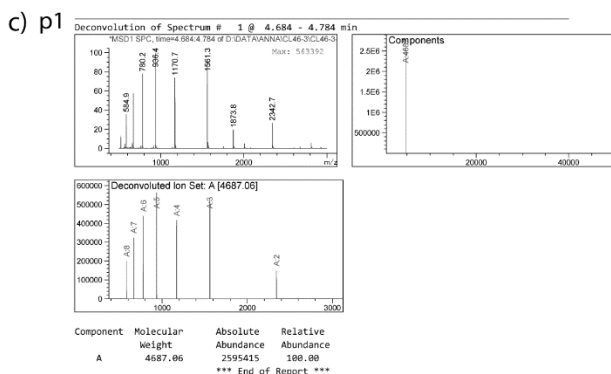
Figure 112: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D57; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.3.5.4. miR-124-3p-C6-N^4-PEG: a mixture of the probes

ORN-31 ORN56mix 2 uM

a)

No irradiation

Irradiation time: 15 min

b)
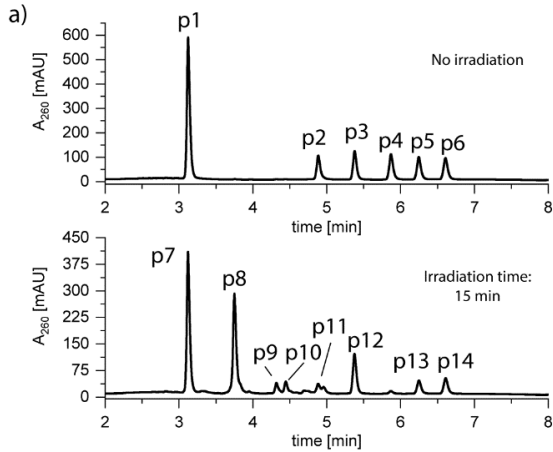
Duplex	ORN	ORN mass	Calc. mass of the CL product
D53	ORN-56a	6685.3	11373.2
	ORN-31	4687.9	
D54	ORN-56b	6729.3	11417.2
	ORN-31	4687.9	
D55	ORN-56c	6773.4	11461.3
	ORN-31	4687.9	
D56	ORN-56d	6817.4	11505.3
	ORN-31	4687.9	
D57	ORN-56e	6861.5	11549.4
	ORN-31	4687.9	

c) p1

p2

p3

p4

p5

p6

p7
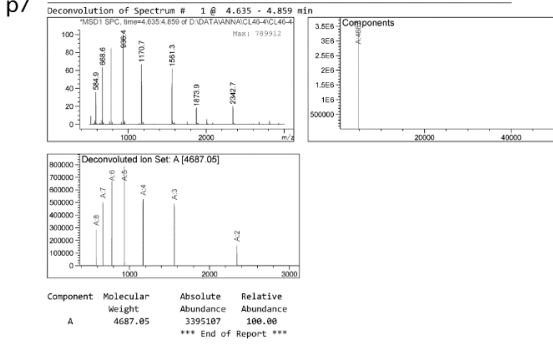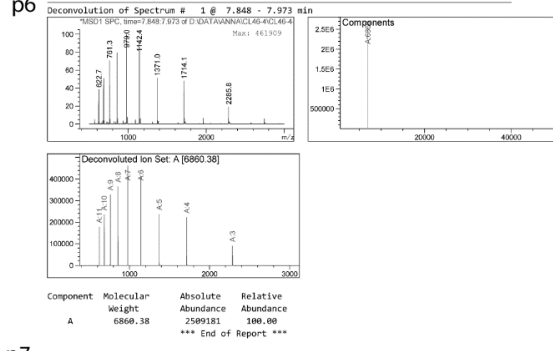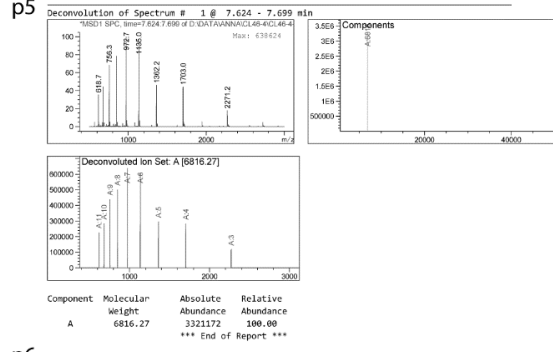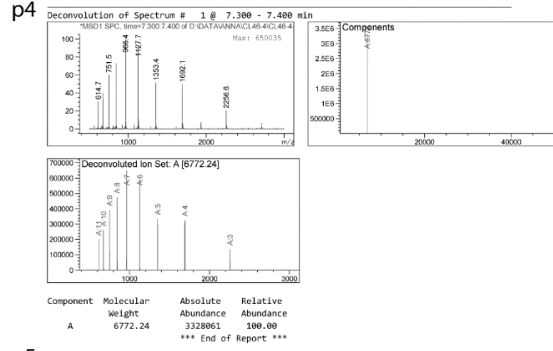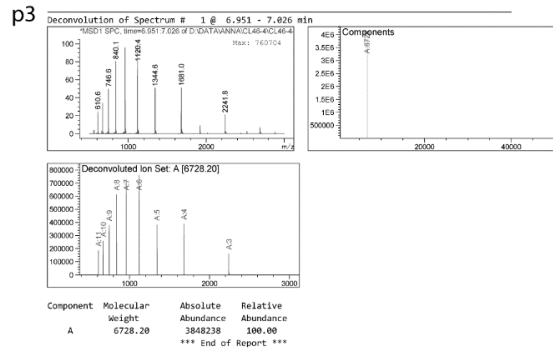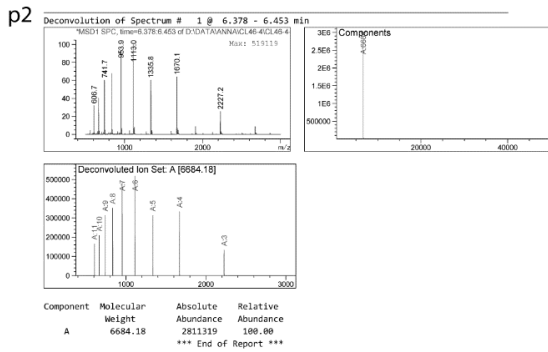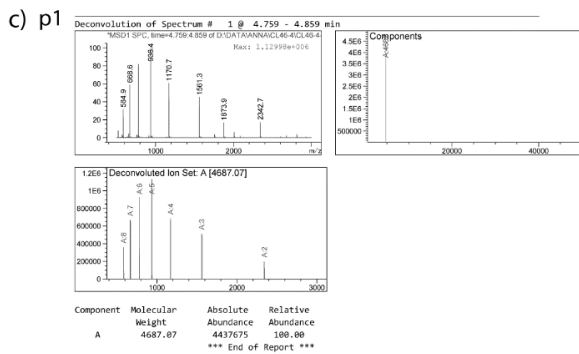
ORN-31 ORN56mix 2 uM
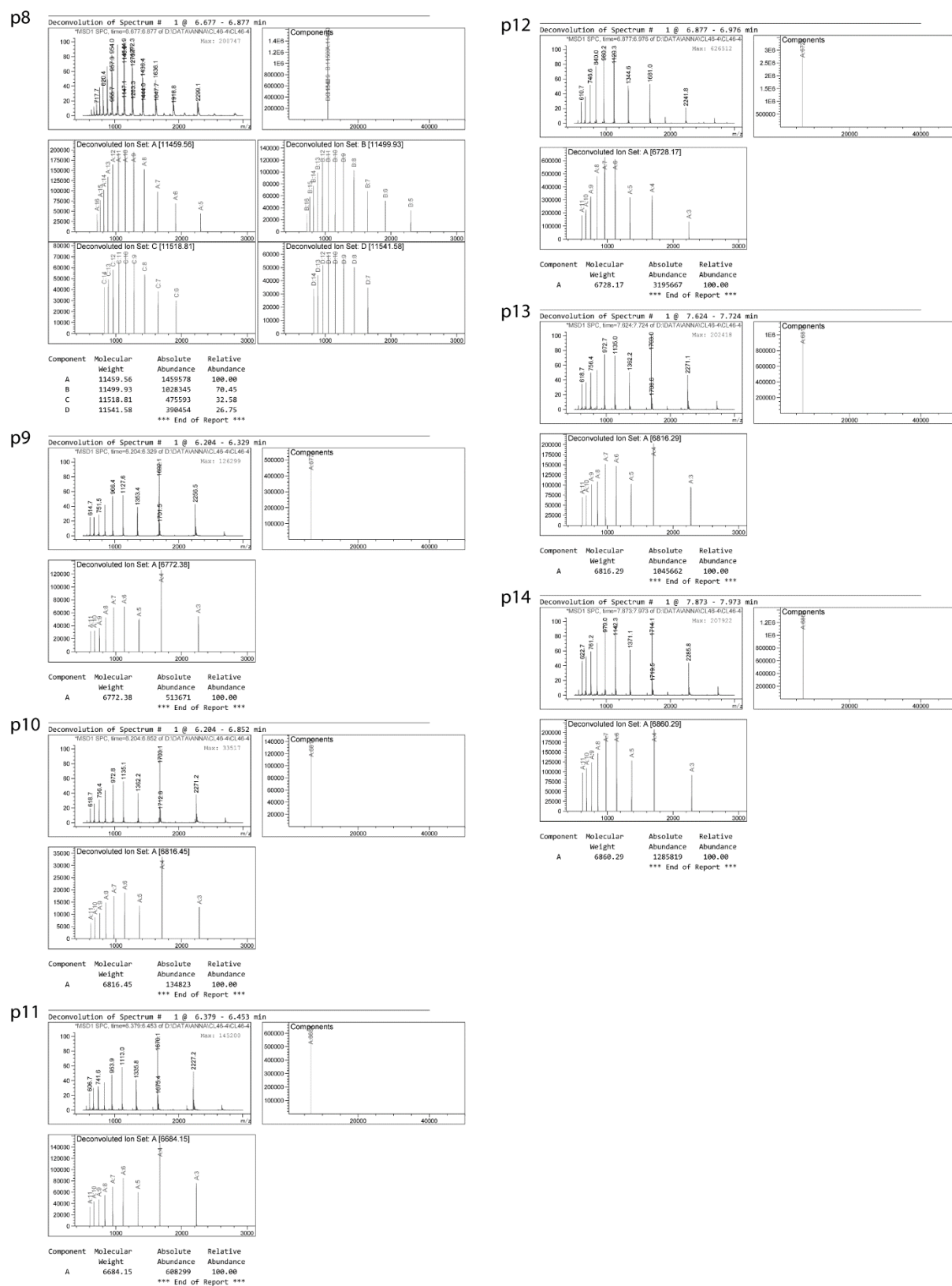
Figure 113: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with a mixture of probes (ORN-56a to ORN-56e + ORN-31); b) calculated masses of the ORNs forming the duplex and the potential cross-linked products; c) deconvoluted masses of the peaks indicated on the chromatograms.

VII.4. Chapter V

VII.4.1. Elucidation of the cross-linking site for the miR-106a-5p probe – pilot studies

VII.4.1.1. Screening of the conditions for the digestion assay

Calculated volumes of the ORN-2 (final concentration: 20 µL of 30 µM or 50 µM solutions, depending on the sample) were transferred to screw cap tubes and dried in the SpeedVac. Samples were re-dissolved in 10 µL of ultrapure water* and spun down. Then 10 µL of 0.25 M NaOH, 0.5 M NaOH or 25% aqueous ammonia were added to the samples, which were then subjected to the hydrolysis under different conditions. *For some of the samples, 20 µL of the base was added directly.

Hydrolysis with ammonia: after the hydrolysis, the samples were cooled down to the room temperature, evaporated to dryness, and re-dissolved in 30 µL of ultrapure water.

Hydrolysis with NaOH solutions: the reaction was quenched with 0.5 M HCl. Samples were then purified with Strata-X Polymeric Reversed-Phase columns (Strata-X 33u Polymeric Reversed Phase, 30 mg/1 mL, 200/Pk; Phenomenex).

Solid-phase extraction: to condition the columns, the Strata-X cartridges were washed with 2.0 mL of MeOH and 2 x 2.0 mL of 15 mM ammonium acetate. After application of the sample, cartridges were washed with 0.5 mL of 10 mM ammonium acetate, dried, wash with 0.2 mL of DCM and dried again. The analytes were eluted with 1.0 mL of MeOH, dried in the SpeedVac and re-dissolved in 30 µL of ultrapure water.

VII.4.1.2. Ion-exchange HPLC purification

The samples (volume: 30 µL) were injected on the IE-HPLC (Hitachi VWR LaChrom Elite HPLC fitted with a DNA Pac PA200 (4 x 250 mm) anion exchange column and a DNA Pac PA200 (4 x 50 mm) guard column at 25°C). Column temperature: 25°C, flow: 1 mL/min. Eluent E: 25 mM Tris HCl, pH 8.0; eluent F: 25 mM Tris HCl, pH 8.0 and 1.5 mM NaCl. Eluents were thoroughly degassed before use. Gradient: 0-100% F in 30 min. UV absorbance was monitored at λ=260 nm using a Hitachi VWR Elite LaChrom L-2420 UV detector.

VII.4.1.3. LC-MS analysis

Samples were analysed with LC-MS (Agilent 1200/6130 system) on an Acquity OST C-18 column (2.1 x 50 mm; Waters) or Synergi™ 4 µm Polar-RP 80 Å column (30 x 4.6 mm; Phenomenex). Different column oven temperatures, flow-rates, eluent systems and gradients were tested.

The masses of the 5'-monophosphates (concentrations in mM range) were detected for the following conditions: the Acquity OST C-18 column (2.1 x 50 mm; Waters), the temperature of the column oven: 40°C, flow-rate: 0.3 mL/min. Eluent C: ultrapure water containing 0.4 M HFIP and 15 mM triethylamine; eluent B: 100% MeOH; gradient 0.5% B in 0-3 min then 0.5-60% B in 3-5 min (total run time: 5 min).

VII.4.2. Synthesis of the miR-CLIP probes

VII.4.2.1. Chemical synthesis

Biotin dT phosphoramidite was purchased from ChemGenes (Catalog No: CLP-1502).

N-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-5-((4*S*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)pentanamide (27)
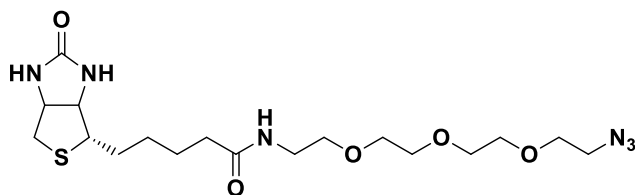
Chemical Formula: C₁₈H₃₂N₆O₅S
Molecular Weight: 444.55

The compound was synthesized according to the published procedure [284, 357]. D-Biotin (0.20 g, 0.82 mmol, 1.05 eq) was suspended in ACN/MeOH 3:1 mixture (8.0 mL). Subsequently, EDC x HCl (0.19 g, 1.01 mmol, 1.30 eq) and 11-azido-3,6,9-trioxaundecan-1-amine (0.15 mL, 0.78 mmol, 1.00 eq) were added and the reaction mixture was stirred overnight at room temperature under the argon atmosphere. The next day 0.65 eq (0.10 g, 0.52 mmol) of EDC x HCl and 0.50 eq (0.08 mL, 0.39 mmol) of 11-azido-3,6,9-trioxaundecan-1-amine were added and the reaction was stirred for few hours. Then, the reaction mixture was concentrated to give the transparent, slightly yellow oil. The crude was purified by flash column chromatography (gradient: DCM/MeOH 98/2 to 90/10) to give 0.27 g (78%) of the pure product in the form of a sticky solid.

1H NMR (400 MHz, DMSO-d_6): δ = 7.82 (t, *J* = 5.6 Hz, 1H), 6.39 (d, *J* = 24.4 Hz, 2H), 4.31 (dd, *J* = 7.6, 5.2 Hz, 1H), 4.16 – 4.10 (m, 1H), 3.61 (dd, *J* = 5.5, 4.4 Hz, 2H), 3.54 (dtt, *J* = 12.3, 4.7, 2.6 Hz, 8H), 3.40 (t, *J* = 5.1 Hz, 4H), 3.19 (q, *J* = 5.9 Hz, 2H), 3.10 (ddd, *J* = 8.5, 6.1, 4.6 Hz, 1H), 2.82 (dd, *J* = 12.4, 5.1 Hz, 1H), 2.58 (d, *J* = 12.4 Hz, 1H), 2.07 (t, *J* = 7.4 Hz, 2H), 1.62 (ddt, *J* = 12.4, 9.6, 6.1 Hz, 1H), 1.49 (tdd, *J* = 13.9, 9.6, 4.7 Hz, 3H), 1.30 (qt, *J* = 10.8, 5.3 Hz, 2H).

VII.4.3. Oligoribonucleotide synthesis, functionalization deprotection and purification

Oligoribonucleotide synthesis, functionalization, deprotection and purification was performed according to the standard procedure described in sections VII.1.1 and VII.1.2 with minor changes.

VII.4.3.1. miR-124-3p- and miR-132-3p-related probes

Oligoribonucleotides were synthesized on a 50 nmol or 1 μmol scale with the MM12 synthesizer (Bio Automation Inc.) using 2000 Å UnyLinker Controlled Pore Glass (CPG) (ChemGenes).

50 nmol scale: coupling times: 2 x 180 s for the standard RNA phosphoramidites and 2 x 240 s for the modified phosphoramidites.
1 μmol scale: coupling times: 1 x 240 s for the standard RNA phosphoramidites and 1 x 360 s for the modified phosphoramidites.

DMTr-on RP-HPLC purification: eluent A/eluent B system; 20-60% B in 5 min. DMTr-off RP-HPLC purification: eluent A/eluent B system; 5-30% B in 5 min. LC-MS analysis: eluent C/eluent B system; gradient 5-40% B in 14 min (ORN-57), 5-50% B in 14 min (ORN-58 and ORN-59), 5-35% B in 14 min (ORN-60).

The CLICK reaction with the biotin azide (**27**) was performed under argon atmosphere to prevent oxidation of the biotin.

VII.4.3.2. miR-155-related probes

Oligoribonucleotides were synthesized on a 50 nmol or 1 μmol scale with the MM12 synthesizer (Bio Automation Inc.) using 500 Å or 2000 Å UnyLinker Controlled Pore Glass (CPG) (ChemGenes).

50 nmol scale: coupling times: 2 x 180 s for the standard RNA phosphoramidites and 2 x 240 s for the modified phosphoramidites.

1 µmol scale: coupling times: 1 x 240 s for the standard RNA phosphoramidites and 1 x 360 s for the modified phosphoramidites.

DMTr-on RP-HPLC purification: eluent A/eluent D system; 10-50% D in 5 min. DMTr-off RP-HPLC purification: eluent A/eluent D system; 2-20% D in 6 min (ORN-64 to ORN-67) or 2-17% D in 8 min (ORN-61 to ORN-63). LC-MS analysis: eluent C/eluent B system; gradient 5-35% B in 14 min. The CLICK reaction with biotin azide (**27**) was performed under argon atmosphere to prevent oxidation of the biotin.

VII.4.3.3. miR-106a-5p- and miR-106b-5p-related probes

Oligoribonucleotides were synthesized on a 50 nmol scale with the MM12 synthesizer (Bio Automation Inc.) using 500 Å or 2000 Å UnyLinker Controlled Pore Glass (CPG) (ChemGenes). Coupling times: 2 x 180 s for the standard RNA phosphoramidites and 2 x 240 s for the modified phosphoramidites.

DMTr-on RP-HPLC purification: eluent A/eluent D system; 10-50% D in 5 min. DMTr-off RP-HPLC purification: eluent A/eluent D system; 2-20% D in 6 min. LC-MS analysis: eluent C/eluent B system; gradient 5-35% B in 14 min. The CLICK reaction with biotin azide (**27**) was performed under argon atmosphere to prevent oxidation of the biotin.

VII.4.4. Luciferase assay

The protocol for the luciferase assay was provided by Artur Łaski.

HEK293T cells (ATCC® CRL-3216™, Wesel, DE) were seeded in Dulbecco's Modified Eagle Medium (Gibco, Invitrogen, Basel, CH) supplemented with 10% Fetal Bovine Serum (Gibco, Invitrogen, Basel) in a 96-well plate. After 8 h, cells were transfected with RNA samples, in technical triplicates, using Lipofectamine 2000 (cat. no. 11668019, ThermoFisher Scientific, Basel, CH), according to the manufacturer protocol. RNA transfection was performed with increasing concentrations (0, 2.5, 10, 40 nM) of the miR-155-3p mimic (3pMIM), siRenilla acting against Renilla luciferase (siREN), synthetic pre-miR-155 (ORN-61), miR-155-3p pre-miR-CLIP probes (ORN-62 and ORN-63) or the negative control RNA (siRND2; double-stranded RNAs carrying randomized base pairs [358]). After 24 h from the RNA transfection, cells were transfected with 20 ng/well of the reporter plasmid, using JetPEI (10110N, Polyplus, Transfection, Illkirch, FR) according to the manufacturer's protocol. After 48 h from the plasmid transfection, luciferase readout was conducted according to the manufacturer's protocol (Dual-Glo Luciferase Assay System, Promega) with 30 µL/well Dual-Glo® Luciferase Reagent diluted in 1:1 ratio with water, and 15 µL/well of Dual-Glo® Stop & Glo® Reagent. Luminescence readout was performed on a microtiter plate reader (Mithras LB940, Berthold Technologies). Readout values were normalized to the firefly luciferase counts and additionally to the corresponding values obtained from 0 nM treatment.

VII.4.5. *In vitro* photo-cross-linking

The *in vitro* cross-linking assay was performed as described in section VII.1.3.2. 1 or 2 µM concentration of each strand was used. Samples were irradiated for 15 min.

VII.4.6. Chromatograms from the *in vitro* photo-cross-linking experiments

D58
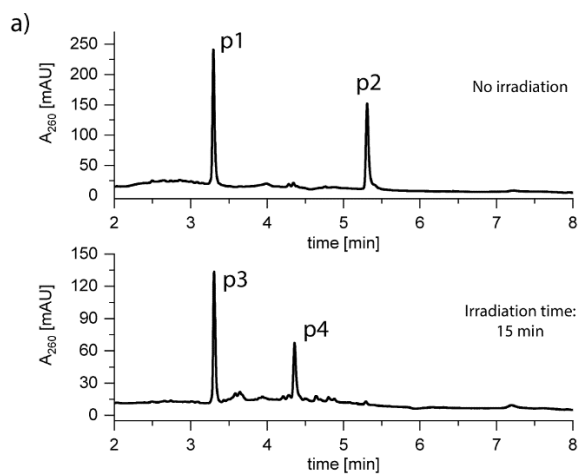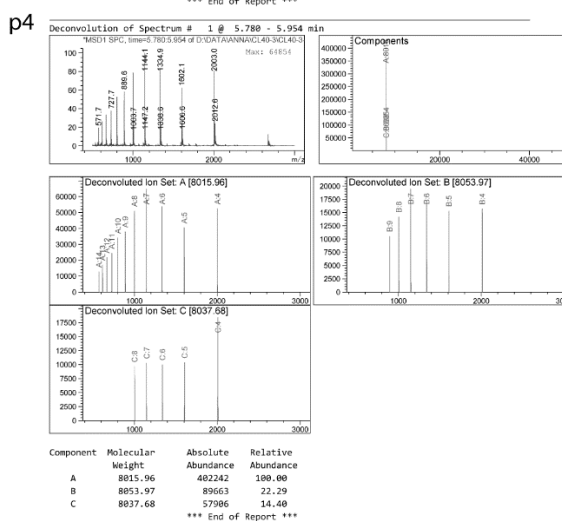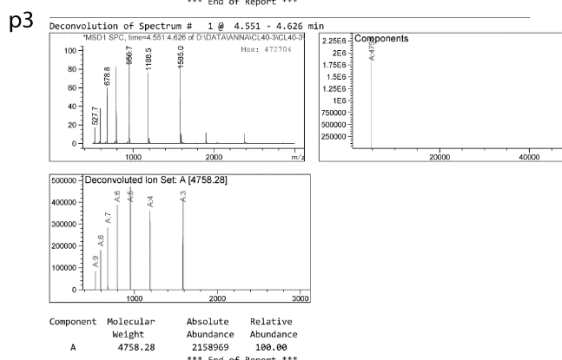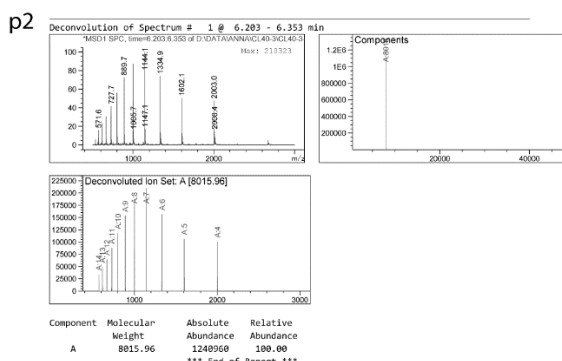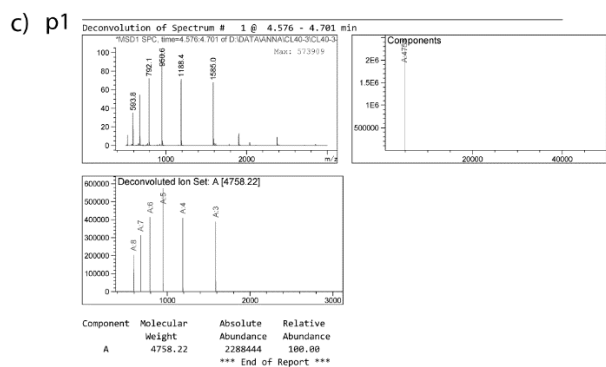
Figure 114: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D58; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

D59

a)

b)
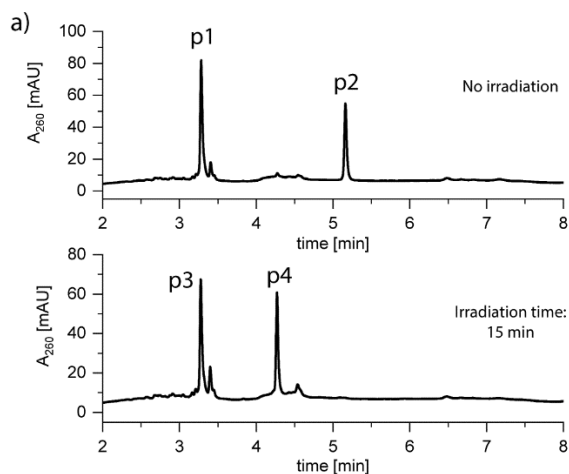
Duplex	ORN	ORN mass	Calc. mass of the CL product
D59	ORN-66	7212.6	12052.6
	ORN-67	4840.0	

c) p1

p2

p3

p4
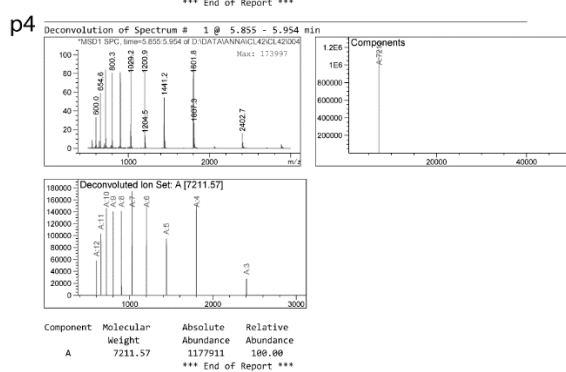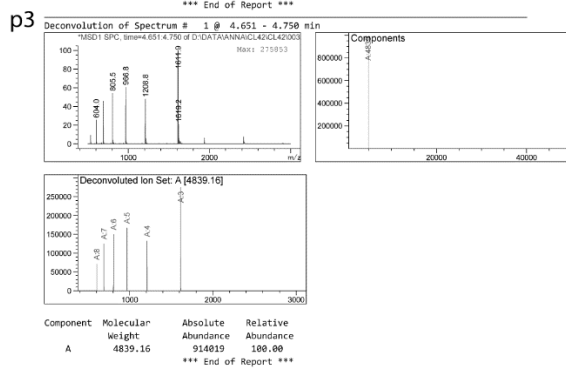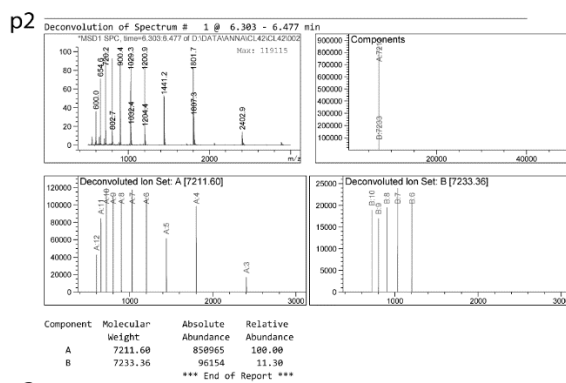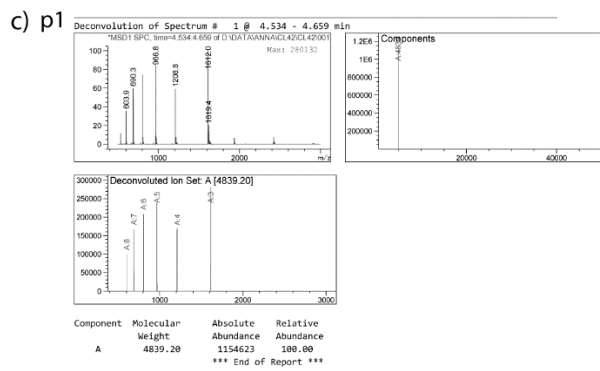
Figure 115: a) RP-HPLC chromatograms from the *in vitro* cross-linking experiment with duplex D59; b) calculated masses of the ORNs forming the duplex and the potential cross-linked product; c) deconvoluted masses of the peaks indicated on the chromatograms.

VIII. References

1.   Malinowska, A.L. and J. Hall, *CHAPTER 8 Research and Development of Oligonucleotides Targeting MicroRNAs (miRNAs)*, in *Advances in Nucleic Acid Therapeutics*. 2019. p. 151-180.

2.   Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-854.

3.   Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-862.

4.   Bartel, D.P., *Metazoan MicroRNAs.* Cell, 2018. **173**(1): p. 20-51.

5.   Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 509-24.

6.   Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* Embo j, 2002. **21**(17): p. 4663-70.

7.   Lagos-Quintana, M., et al., *Identification of novel genes coding for small expressed RNAs.* Science, 2001. **294**(5543): p. 853-858.

8.   Lin, S. and R.I. Gregory, *MicroRNA biogenesis pathways in cancer.* Nat Rev Cancer, 2015. **15**(6): p. 321-33.

9.   Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* Embo j, 2004. **23**(20): p. 4051-60.

10.   Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-5.

11.   Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.

12.   Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.

13.   Landthaler, M., A. Yalcin, and T. Tuschl, *The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis.* Current Biology, 2004. **14**(23): p. 2162-2167.

14.   Han, J., et al., *Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex.* Cell, 2006. **125**(5): p. 887-901.

15.   Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.

16.   Okada, C., et al., *A high-resolution structure of the pre-microRNA nuclear export machinery.* Science, 2009. **326**(5957): p. 1275-9.

17.   Wang, X., et al., *Dynamic mechanisms for pre-miRNA binding and export by Exportin-5.* RNA (New York, N.Y.), 2011. **17**(8): p. 1511-1528.

18.   Lei, E.P. and P.A. Silver, *Protein and RNA Export from the Nucleus.* Developmental Cell, 2002. **2**(3): p. 261-272.

19.   Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.* Genes Dev, 2003. **17**(24): p. 3011-6.

20.   Bohnsack, M.T., K. Czaplinski, and D. Gorlich, *Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.* Rna, 2004. **10**(2): p. 185-91.

21.   Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* Nature, 2001. **409**(6818): p. 363-366.

22.   Grishok, A., et al., *Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing.* Cell, 2001. **106**(1): p. 23-34.

23.   Hutvagner, G., et al., *A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.* Science, 2001. **293**(5531): p. 834-8.

24. Chendrimada, T., et al., *TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.* Nature, 2005. **436**: p. 740-4.

25. Haase, A.D., et al., *TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing.* EMBO reports, 2005. **6**(10): p. 961-967.

26. Wilson, R.C., et al., *Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis.* Mol Cell, 2015. **57**(3): p. 397-407.

27. Lee, Y., et al., *The role of PACT in the RNA silencing pathway.* Embo j, 2006. **25**(3): p. 522-32.

28. Macrae, I.J., et al., *Structural basis for double-stranded RNA processing by Dicer.* Science, 2006. **311**(5758): p. 195-8.

29. Park, J.E., et al., *Dicer recognizes the 5' end of RNA for efficient and accurate processing.* Nature, 2011. **475**(7355): p. 201-5.

30. Zhang, H., et al., *Single processing center models for human Dicer and bacterial RNase III.* Cell, 2004. **118**(1): p. 57-68.

31. Okamura, K., et al., *The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila.* Cell, 2007. **130**(1): p. 89-100.

32. Ruby, J.G., C.H. Jan, and D.P. Bartel, *Intronic microRNA precursors that bypass Drosha processing.* Nature, 2007. **448**: p. 83.

33. Cheloufi, S., et al., *A dicer-independent miRNA biogenesis pathway that requires Ago catalysis.* Nature, 2010. **465**(7298): p. 584-9.

34. Cifuentes, D., et al., *A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity.* Science, 2010. **328**(5986): p. 1694-8.

35. Diederichs, S. and D.A. Haber, *Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression.* Cell, 2007. **131**(6): p. 1097-108.

36. Kim, Y.K., B. Kim, and V.N. Kim, *Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. E1881-9.

37. Gregory, R.I., et al., *Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.* Cell, 2005. **123**(4): p. 631-40.

38. Liu, J., et al., *Argonaute2 is the catalytic engine of mammalian RNAi.* Science, 2004. **305**(5689): p. 1437-41.

39. Iwasaki, S., et al., *Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes.* Mol Cell, 2010. **39**(2): p. 292-9.

40. Kobayashi, H. and Y. Tomari, *RISC assembly: Coordination between small RNAs and Argonaute proteins.* Biochim Biophys Acta, 2016. **1859**(1): p. 71-81.

41. Khvorova, A., A. Reynolds, and S.D. Jayasena, *Functional siRNAs and miRNAs exhibit strand bias.* Cell, 2003. **115**(2): p. 209-16.

42. Schwarz, D.S., et al., *Asymmetry in the assembly of the RNAi enzyme complex.* Cell, 2003. **115**(2): p. 199-208.

43. Ro, S., et al., *Tissue-dependent paired expression of miRNAs.* Nucleic Acids Research, 2007. **35**(17): p. 5944-5953.

44. Nakanishi, K., et al., *Structure of yeast Argonaute with guide RNA.* Nature, 2012. **486**(7403): p. 368-74.

45. Schirle, N.T. and I.J. MacRae, *The crystal structure of human Argonaute2.* Science, 2012. **336**(6084): p. 1037-40.

46. Schirle, N.T., J. Sheu-Gruttadauria, and I.J. MacRae, *Structural basis for microRNA targeting.* Science, 2014. **346**(6209): p. 608-613.

47. Frank, F., N. Sonenberg, and B. Nagar, *Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2.* Nature, 2010. **465**(7299): p. 818-22.

48. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.

49. Lewis, B.P., et al., *Prediction of mammalian microRNA targets.* Cell, 2003. **115**(7): p. 787-98.

50.   Klum, S.M., et al., *Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition.* The EMBO journal, 2018. **37**(1): p. 75-88.

51.   Sheu-Gruttadauria, J., et al., *Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2.* Embo j, 2019. **38**(13): p. e101153.

52.   Schirle, N.T., et al., *Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets.* Elife, 2015. **4**: p. 07646.

53.   Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. **136**(2): p. 215-33.

54.   Brennecke, J., et al., *Principles of MicroRNA–Target Recognition.* PLOS Biology, 2005. **3**(3): p. e85.

55.   Didiano, D. and O. Hobert, *Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions.* Nature Structural &Amp; Molecular Biology, 2006. **13**: p. 849.

56.   Cloonan, N., *Re-thinking miRNA-mRNA interactions: intertwining issues confound target discovery.* Bioessays, 2015. **37**(4): p. 379-88.

57.   Shin, C., et al., *Expanding the microRNA targeting code: functional sites with centered pairing.* Mol Cell, 2010. **38**(6): p. 789-802.

58.   Martin, H.C., et al., *Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs.* Genome Biology, 2014. **15**(3): p. R51.

59.   Vella, M.C., et al., *The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR.* Genes & Development, 2004. **18**(2): p. 132-137.

60.   Tan, G.S., et al., *Small molecule inhibition of RISC loading.* ACS Chem Biol, 2012. **7**(2): p. 403-10.

61.   Llave, C., et al., *Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA.* Science, 2002. **297**(5589): p. 2053-6.

62.   Rhoades, M.W., et al., *Prediction of plant microRNA targets.* Cell, 2002. **110**(4): p. 513-20.

63.   Mathonnet, G., et al., *MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F.* Science, 2007. **317**(5845): p. 1764-7.

64.   Kiriakidou, M., et al., *An mRNA m7G cap binding-like motif within human Ago2 represses translation.* Cell, 2007. **129**(6): p. 1141-51.

65.   Pillai, R.S., et al., *Inhibition of translational initiation by Let-7 MicroRNA in human cells.* Science, 2005. **309**(5740): p. 1573-6.

66.   Guo, H., et al., *Mammalian microRNAs predominantly act to decrease target mRNA levels.* Nature, 2010. **466**(7308): p. 835-40.

67.   Eichhorn, S.W., et al., *mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues.* Mol Cell, 2014. **56**(1): p. 104-15.

68.   Jonas, S. and E. Izaurralde, *Towards a molecular understanding of microRNA-mediated gene silencing.* Nat Rev Genet, 2015. **16**(7): p. 421-33.

69.   Hutvagner, G. and P.D. Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex.* Science, 2002. **297**(5589): p. 2056-60.

70.   Yekta, S., I.H. Shih, and D.P. Bartel, *MicroRNA-directed cleavage of HOXB8 mRNA.* Science, 2004. **304**(5670): p. 594-6.

71.   Elbashir, S.M., W. Lendeckel, and T. Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs.* Genes Dev, 2001. **15**(2): p. 188-200.

72.   Schwarz, D.S., Y. Tomari, and P.D. Zamore, *The RNA-induced silencing complex is a Mg2+-dependent endonuclease.* Curr Biol, 2004. **14**(9): p. 787-91.

73.   Guo, Y., et al., *Characterization of the mammalian miRNA turnover landscape.* Nucleic Acids Res, 2015. **43**(4): p. 2326-41.

74.   Hajarnis, S., et al., *microRNA-17 family promotes polycystic kidney disease progression through modulation of mitochondrial metabolism.* Nature Communications, 2017. **8**: p. 14395.

75.   Kingston, E.R. and D.P. Bartel, *Global analyses of the dynamics of mammalian microRNA metabolism.* Genome Res, 2019. **29**(11): p. 1777-1790.

76. Ameres, S.L., et al., *Target RNA-directed trimming and tailing of small silencing RNAs.* Science (New York, N.Y.), 2010. **328**(5985): p. 1534-1539.

77. Baccarini, A., et al., *Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells.* Curr Biol, 2011. **21**(5): p. 369-76.

78. de la Mata, M., et al., *Potent degradation of neuronal miRNAs induced by highly complementary targets.* EMBO Reports, 2015. **16**(4): p. 500-511.

79. Lytle, J.R., T.A. Yario, and J.A. Steitz, *Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR.* Proc Natl Acad Sci U S A, 2007. **104**(23): p. 9667-72.

80. Tay, Y., et al., *MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation.* Nature, 2008. **455**(7216): p. 1124-8.

81. Brummer, A. and J. Hausser, *MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation.* Bioessays, 2014. **36**(6): p. 617-26.

82. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.* Nature, 2010. **465**: p. 1033.

83. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.

84. Imig, J., et al., *miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction.* Nat Chem Biol, 2015. **11**(2): p. 107-14.

85. Hansen, T.B., et al., *Natural RNA circles function as efficient microRNA sponges.* Nature, 2013. **495**: p. 384.

86. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**: p. 333.

87. Vasudevan, S., Y. Tong, and J.A. Steitz, *Switching from repression to activation: microRNAs can up-regulate translation.* Science, 2007. **318**(5858): p. 1931-4.

88. Orom, U.A., F.C. Nielsen, and A.H. Lund, *MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation.* Mol Cell, 2008. **30**(4): p. 460-71.

89. Krol, J., I. Loedige, and W. Filipowicz, *The widespread regulation of microRNA biogenesis, function and decay.* Nature Reviews Genetics, 2010. **11**(9): p. 597-610.

90. Michlewski, G. and J.F. Cáceres, *Post-transcriptional control of miRNA biogenesis.* RNA (New York, N.Y.), 2019. **25**(1): p. 1-16.

91. Gebert, L.F.R. and I.J. MacRae, *Regulation of microRNA function in animals.* Nat Rev Mol Cell Biol, 2018. **20**(1): p. 21-37.

92. Morin, R.D., et al., *Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.* Genome Res, 2008. **18**(4): p. 610-21.

93. Cloonan, N., et al., *MicroRNAs and their isomiRs function cooperatively to target common biological pathways.* Genome Biol, 2011. **12**(12): p. R126.

94. Neilsen, C.T., G.J. Goodall, and C.P. Bracken, *IsomiRs--the overlooked repertoire in the dynamic microRNAome.* Trends Genet, 2012. **28**(11): p. 544-9.

95. Ameres, S.L. and P.D. Zamore, *Diversifying microRNA sequence and function.* Nat Rev Mol Cell Biol, 2013. **14**(8): p. 475-88.

96. Lee, H.Y. and J.A. Doudna, *TRBP alters human precursor microRNA processing in vitro.* RNA (New York, N.Y.), 2012. **18**(11): p. 2012-2019.

97. Kawahara, Y., et al., *Redirection of silencing targets by adenosine-to-inosine editing of miRNAs.* Science, 2007. **315**(5815): p. 1137-40.

98. Tay, Y., et al., *Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs.* Cell, 2011. **147**(2): p. 344-57.

99. Miska, E.A., et al., *Most Caenorhabditis elegans microRNAs Are Individually Not Essential for Development or Viability.* PLoS Genetics, 2007. **3**(12): p. e215.

100. Bartel, D.P. and C.Z. Chen, *Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs.* Nat Rev Genet, 2004. **5**(5): p. 396-400.

101. Vidigal, J.A. and A. Ventura, *The biological functions of miRNAs: lessons from in vivo studies.* Trends Cell Biol, 2015. **25**(3): p. 137-47.

102. Mencia, A., et al., *Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss.* Nat Genet, 2009. **41**(5): p. 609-13.

103. Trajkovski, M., et al., *MicroRNAs 103 and 107 regulate insulin sensitivity.* Nature, 2011. **474**(7353): p. 649-53.

104. Dangwal, S. and T. Thum, *microRNA therapeutics in cardiovascular disease models.* Annu Rev Pharmacol Toxicol, 2014. **54**: p. 185-203.

105. Montgomery, R.L., et al., *Therapeutic inhibition of miR-208a improves cardiac function and survival during heart failure.* Circulation, 2011. **124**(14): p. 1537-47.

106. Hullinger, T.G., et al., *Inhibition of miR-15 protects against cardiac ischemic injury.* Circ Res, 2012. **110**(1): p. 71-81.

107. Small, E.M., R.J.A. Frost, and E.N. Olson, *MicroRNAs add a new dimension to cardiovascular disease.* Circulation, 2010. **121**(8): p. 1022-1032.

108. Rayner, K.J., et al., *MiR-33 Contributes to the Regulation of Cholesterol Homeostasis.* Science, 2010. **328**(5985): p. 1570-1573.

109. Najafi-Shoushtari, S.H., et al., *MicroRNA-33 and the SREBP Host Genes Cooperate to Control Cholesterol Homeostasis.* Science (New York, N.Y.), 2010. **328**(5985): p. 1566-1569.

110. Marquart, T.J., et al., *miR-33 links SREBP-2 induction to repression of sterol transporters.* Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12228-32.

111. Jopling, C.L., et al., *Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA.* Science, 2005. **309**(5740): p. 1577-1581.

112. Rupaimoole, R. and F.J. Slack, *MicroRNA therapeutics: towards a new era for the management of cancer and other diseases.* Nature Reviews Drug Discovery, 2017. **16**: p. 203.

113. Lekka, E. and J. Hall, *Noncoding RNAs in disease.* FEBS Letters. **592**(17): p. 2884-2900.

114. Gambari, R., et al., *Targeting oncomiRNAs and mimicking tumor suppressor miRNAs: New trends in the development of miRNA therapeutic strategies in oncology (Review).* International Journal of Oncology, 2016. **49**(1): p. 5-32.

115. Ling, H., M. Fabbri, and G.A. Calin, *MicroRNAs and other non-coding RNAs as targets for anticancer drug development.* Nature Reviews Drug Discovery, 2013. **12**: p. 847.

116. Di Leva, G., M. Garofalo, and C.M. Croce, *MicroRNAs in Cancer.* Annual Review of Pathology: Mechanisms of Disease, 2014. **9**(1): p. 287-314.

117. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-8.

118. Kuninty, P.R., et al., *MicroRNA Targeting to Modulate Tumor Microenvironment.* Frontiers in Oncology, 2016. **6**: p. 3.

119. Lou, W., et al., *MicroRNAs in cancer metastasis and angiogenesis.* Oncotarget, 2017. **8**(70): p. 115787-115802.

120. Abba, M.L., et al., *MicroRNA Regulation of Epithelial to Mesenchymal Transition.* J Clin Med, 2016. **5**(1), 8.

121. Kim, J., et al., *MicroRNAs and metastasis: small RNAs play big roles.* Cancer and Metastasis Reviews, 2018. **37**(1): p. 5-15.

122. Solé, C. and C.H. Lawrie, *MicroRNAs and metastasis.* Cancers, 2020. **12**(1), 96.

123. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer.* Nat Rev Cancer, 2006. **6**(4): p. 259-69.

124. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression.* Nature, 2005. **435**: p. 839.

125. Calin, G.A., et al., *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2999-3004.

126. Bracken, C.P., H.S. Scott, and G.J. Goodall, *A network-biology perspective of microRNA function and dysfunction in cancer.* Nat Rev Genet, 2016. **17**(12): p. 719-732.

127. Hurst, D.R., M.D. Edmonds, and D.R. Welch, *Metastamir: the field of metastasis-regulatory microRNA is spreading.* Cancer Res, 2009. **69**(19): p. 7495-8.

128. Abba, M.L., et al., *MicroRNAs as novel targets and tools in cancer therapy.* Cancer Lett, 2017. **387**: p. 84-94.

129. Cortez, M.A., et al., *MicroRNAs in body fluids--the mix of hormones and biomarkers.* Nat Rev Clin Oncol, 2011. **8**(8): p. 467-77.

130. Ebert, M.S., J.R. Neilson, and P.A. Sharp, *MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells.* Nat Methods, 2007. **4**(9): p. 721-6.

131. Ebert, M.S. and P.A. Sharp, *MicroRNA sponges: progress and possibilities.* Rna, 2010. **16**(11): p. 2043-50.

132. Tay, F.C., et al., *Using artificial microRNA sponges to achieve microRNA loss-of-function in cancer cells.* Adv Drug Deliv Rev, 2015. **81**: p. 117-27.

133. Lorenz, D.A. and A.L. Garner, *Approaches for the Discovery of Small Molecule Ligands Targeting microRNAs*, in *RNA Therapeutics*, A.L. Garner, Editor. 2018, Springer International Publishing: Cham. p. 79-110.

134. Wen, D., et al., *Small molecules targeting microRNA for cancer therapy: Promises and obstacles.* J Control Release, 2015. **219**: p. 237-247.

135. Roos, M., et al., *A Small-Molecule Inhibitor of Lin28.* ACS Chem Biol, 2016. **11**(10): p. 2773-2781.

136. Fernandez-Piñeiro, I., I. Badiola, and A. Sanchez, *Nanocarriers for microRNA delivery in cancer medicine.* Biotechnology Advances, 2017. **35**(3): p. 350-360.

137. Gebert, L.F., et al., *Miravirsen (SPC3649) can inhibit the biogenesis of miR-122.* Nucleic Acids Res, 2014. **42**(1): p. 609-21.

138. Slaby, O., R. Laga, and O. Sedlacek, *Therapeutic targeting of non-coding RNAs in cancer.* Biochem J, 2017. **474**(24): p. 4219-4251.

139. Weiler, J., J. Hunziker, and J. Hall, *Anti-miRNA oligonucleotides (AMOs): ammunition to target miRNAs implicated in human disease?* Gene Ther, 2005. **13**(6): p. 496-502.

140. Lennox, K.A. and M.A. Behlke, *Chemical modification and design of anti-miRNA oligonucleotides.* Gene Ther, 2011. **18**(12): p. 1111-1120.

141. Li, Z. and T.M. Rana, *Therapeutic targeting of microRNAs: current status and future challenges.* Nature Reviews Drug Discovery, 2014. **13**: p. 622.

142. Stenvang, J., et al., *Inhibition of microRNA function by antimiR oligonucleotides.* Silence, 2012. **3**: p. 1-1.

143. Boutla, A., C. Delidakis, and M. Tabler, *Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in Drosophila and the identification of putative target genes.* Nucleic Acids Research, 2003. **31**(17): p. 4973-4980.

144. Hutvágner, G., et al., *Sequence-Specific Inhibition of Small RNA Function.* PLOS Biology, 2004. **2**(4): p. e98.

145. Meister, G., et al., *Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing.* Rna, 2004. **10**(3): p. 544-50.

146. Eckstein, F., *Phosphorothioates, essential components of therapeutic oligonucleotides.* Nucleic Acid Ther, 2014. **24**(6): p. 374-87.

147. Spitzer, S. and F. Eckstein, *Inhibition of deoxyribonucleases by phosphorothioate groups in oligodeoxyribonucleotides.* Nucleic Acids Research, 1988. **16**(24): p. 11691-11704.

148. Brown, D.A., et al., *Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein binding.* J Biol Chem, 1994. **269**(43): p. 26801-5.

149. Geary, R.S., et al., *Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides.* Advanced Drug Delivery Reviews, 2015. **87**: p. 46-51.

150. Crooke, S.T., *Progress in antisense technology: the end of the beginning.* Methods Enzymol, 2000. **313**: p. 3-45.

151. Krutzfeldt, J., et al., *Silencing of microRNAs in vivo with /`antagomirs/'.* Nature, 2005. **438**(7068): p. 685-689.

152. Velu, C.S. and H.L. Grimes, *Utilizing antagomiR (antisense microRNA) to knock down microRNA in murine bone marrow cells.* Methods Mol Biol, 2012. **928**: p. 185-95.

153. Davis, S., et al., *Improved targeting of miRNA with antisense oligonucleotides.* Nucleic Acids Research, 2006. **34**(8): p. 2294-2304.

154. Esau, C., et al., *MicroRNA-143 Regulates Adipocyte Differentiation.* Journal of Biological Chemistry, 2004. **279**(50): p. 52361-52365.

155. Esau, C., et al., *miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting.* Cell Metab, 2006. **3**(2): p. 87-98.

156. Guo, F., Q. Li, and C. Zhou, *Synthesis and biological applications of fluoro-modified nucleic acids.* Organic & Biomolecular Chemistry, 2017. **15**(45): p. 9552-9565.

157. Guschlbauer, W. and K. Jankowski, *Nucleoside conformation is determined by the electronegativity of the sugar substituent.* Nucleic Acids Research, 1980. **8**(6): p. 1421-1433.

158. Altmann, K.-H., et al., *Second Generation of Antisense Oligonucleotides: From Nuclease Resistance to Biological Efficacy in Animals.* CHIMIA International Journal for Chemistry, 1996. **50**(4): p. 168-176.

159. Kawasaki, A.M., et al., *Uniformly modified 2'-deoxy-2'-fluoro phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for RNA targets.* J Med Chem, 1993. **36**(7): p. 831-41.

160. Davis, S., et al., *Potent inhibition of microRNA in vivo without degradation.* Nucleic Acids Res, 2009. **37**(1): p. 70-7.

161. Petersen, M., et al., *Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids.* J Am Chem Soc, 2002. **124**(21): p. 5974-82.

162. K. Singh, S., et al., *LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition.* Chemical Communications, 1998(4): p. 455-456.

163. Chan, J.A., A.M. Krichevsky, and K.S. Kosik, *MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells.* Cancer Res, 2005. **65**(14): p. 6029-33.

164. Orom, U.A., S. Kauppinen, and A.H. Lund, *LNA-modified oligonucleotides mediate specific inhibition of microRNA function.* Gene, 2006. **372**: p. 137-41.

165. Silahtaroglu, A.N., et al., *Detection of microRNAs in frozen tissue sections by fluorescence in situ hybridization using locked nucleic acid probes and tyramide signal amplification.* Nat Protoc, 2007. **2**(10): p. 2520-8.

166. Obernosterer, G., J. Martinez, and M. Alenius, *Locked nucleic acid-based in situ detection of microRNAs in mouse tissue sections.* Nat Protoc, 2007. **2**(6): p. 1508-14.

167. Fabani, M.M. and M.J. Gait, *miR-122 targeting with LNA/2'-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA–peptide conjugates.* RNA, 2008. **14**(2): p. 336-346.

168. Lennox, K.A. and M.A. Behlke, *A Direct Comparison of Anti-microRNA Oligonucleotide Potency.* Pharmaceutical Research, 2010. **27**(9): p. 1788-1799.

169. Elmen, J., et al., *Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver.* Nucleic Acids Res, 2008. **36**(4): p. 1153-62.

170. Obad, S., et al., *Silencing of microRNA families by seed-targeting tiny LNAs.* Nat Genet, 2011. **43**(4): p. 371-378.

171. Seth, P.P., et al., *Synthesis and biophysical evaluation of 2',4'-constrained 2'O-methoxyethyl and 2',4'-constrained 2'O-ethyl nucleic acid analogues.* J Org Chem, 2010. **75**(5): p. 1569-81.

172. Hogan, D.J., et al., *Anti-miRs Competitively Inhibit microRNAs in Argonaute Complexes.* PLOS ONE, 2014. **9**(7): p. e100951.

173. Quijano, E., et al., *Therapeutic Peptide Nucleic Acids: Principles, Limitations, and Opportunities.* Yale J Biol Med, 2017. **90**(4): p. 583-598.

174. Oh, S.Y., Y. Ju, and H. Park, *A highly effective and long-lasting inhibition of miRNAs with PNA-based antisense oligonucleotides.* Mol Cells, 2009. **28**(4): p. 341-5.

175. Fabani, M.M., et al., *Efficient inhibition of miR-155 function in vivo by peptide nucleic acids.* Nucleic Acids Res, 2010. **38**(13): p. 4466-75.

176. Torres, A.G., et al., *Chemical structure requirements and cellular targeting of microRNA-122 by peptide nucleic acids anti-miRs.* Nucleic Acids Res, 2012. **40**(5): p. 2152-67.

177. Summerton, J. and D. Weller, *Morpholino antisense oligomers: design, preparation, and properties.* Antisense Nucleic Acid Drug Dev, 1997. **7**(3): p. 187-95.

178. Moulton, J.D. and Y.L. Yan, *Using Morpholinos to control gene expression.* Curr Protoc Mol Biol, 2008. **Chapter 26**: p. 26.8.1-26.8.29.

179. Flynt, A.S., et al., *Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate.* Nat Genet, 2007. **39**(2): p. 259-63.

180. Kloosterman, W.P., et al., *Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development.* PLoS Biol, 2007. **5**(8): p. e203.

181. Cheng, C.J., et al., *MicroRNA silencing for cancer therapy targeted to the tumour microenvironment.* Nature, 2014. **518**: p. 107.

182. Andreev, O.A., D.M. Engelman, and Y.K. Reshetnyak, *pH-sensitive membrane peptides (pHLIPs) as a novel class of delivery agents.* Mol Membr Biol, 2010. **27**(7): p. 341-52.

183. Reshetnyak, Y.K., et al., *A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution, Attached to, and Inserted across Lipid Bilayers.* Biophysical Journal, 2007. **93**(7): p. 2363-2372.

184. Vaupel, P., F. Kallinowski, and P. Okunieff, *Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review.* Cancer Res, 1989. **49**(23): p. 6449-65.

185. Østergaard, H., C. Tachibana, and J.R. Winther, *Monitoring disulfide bond formation in the eukaryotic cytosol.* The Journal of Cell Biology, 2004. **166**(3): p. 337-345.

186. Sandhu, S.K., C.M. Croce, and R. Garzon, *Micro-RNA Expression and Function in Lymphomas.* Adv Hematol, 2011. **2011**: p. 347137.

187. Selcuklu, S.D., M.T. Donoghue, and C. Spillane, *miR-21 as a key regulator of oncogenic processes.* Biochem Soc Trans, 2009. **37**(Pt 4): p. 918-25.

188. Costa, P.M., et al., *MiRNA-21 silencing mediated by tumor-targeted nanoparticles combined with sunitinib: A new multimodal gene therapy approach for glioblastoma.* J Control Release, 2015. **207**: p. 31-9.

189. Deshane, J., C.C. Garner, and H. Sontheimer, *Chlorotoxin Inhibits Glioma Cell Invasion via Matrix Metalloproteinase-2.* Journal of Biological Chemistry, 2003. **278**(6): p. 4135-4144.

190. Zhang, Q., et al., *Simultaneous delivery of therapeutic antagomirs with paclitaxel for the management of metastatic tumors by a pH-responsive anti-microbial peptide-mediated liposomal delivery system.* J Control Release, 2015. **197**: p. 208-18.

191. Ma, L., J. Teruya-Feldstein, and R.A. Weinberg, *Tumour invasion and metastasis initiated by microRNA-10b in breast cancer.* Nature, 2007. **449**: p. 682.

192. Ma, L., et al., *Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model.* Nature Biotechnology, 2010. **28**: p. 341.

193. Makovitzki, A., A. Fink, and Y. Shai, *Suppression of human solid tumor growth in mice by intratumor and systemic inoculation of histidine-rich and pH-dependent host defense-like lytic peptides.* Cancer Res, 2009. **69**(8): p. 3458-63.

194. Papo, N., et al., *Inhibition of Tumor Growth and Elimination of Multiple Metastases in Human Prostate and Breast Xenografts by Systemic Inoculation of a Host Defense–Like Lytic Peptide.* Cancer Research, 2006. **66**(10): p. 5371-5378.

195. Matsui, M. and D.R. Corey, *Non-coding RNAs as drug targets.* Nature Reviews Drug Discovery, 2016. **16**: p. 167.

196. Chakraborty, C., et al., *Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine.* Molecular Therapy - Nucleic Acids, 2017. **8**: p. 132-143.

197. Hoofnagle, J.H., *Course and outcome of hepatitis C.* Hepatology, 2002. **36**(5 Suppl 1): p. S21-9.

198. Baek, J., S. Kang, and H. Min, *MicroRNA-targeting therapeutics for hepatitis C.* Arch Pharm Res, 2014. **37**(3): p. 299-305.

199. Jopling, C.L., S. Schutz, and P. Sarnow, *Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome.* Cell Host Microbe, 2008. **4**(1): p. 77-85.

200. Li, Y., D. Yamane, and S.M. Lemon, *Dissecting the roles of the 5' exoribonucleases Xrn1 and Xrn2 in restricting hepatitis C virus replication.* J Virol, 2015. **89**(9): p. 4857-65.

201. Sedano, C.D. and P. Sarnow, *Hepatitis C virus subverts liver-specific miR-122 to protect the viral genome from exoribonuclease Xrn2.* Cell Host Microbe, 2014. **16**(2): p. 257-264.

202. Machlin, E.S., P. Sarnow, and S.M. Sagan, *Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex.* Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3193-8.

203. Elmen, J., et al., *LNA-mediated microRNA silencing in non-human primates.* Nature, 2008. **452**(7189): p. 896-9.

204. Lanford, R.E., et al., *Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection.* Science, 2010. **327**(5962): p. 198-201.

205. Janssen, H.L., et al., *Treatment of HCV infection by targeting microRNA.* N Engl J Med, 2013. **368**(18): p. 1685-94.

206. van der Ree, M.H., et al., *Miravirsen dosing in chronic hepatitis C patients results in decreased microRNA-122 levels without affecting other microRNAs in plasma.* Aliment Pharmacol Ther, 2016. **43**(1): p. 102-13.

207. Ottosen, S., et al., *In vitro antiviral activity and preclinical and clinical resistance profile of miravirsen, a novel anti-hepatitis C virus therapeutic targeting the human factor miR-122.* Antimicrob Agents Chemother, 2015. **59**(1): p. 599-608.

208. van der Ree, M.H., et al., *Safety, tolerability, and antiviral effect of RG-101 in patients with chronic hepatitis C: a phase 1B, double-blind, randomised controlled trial.* Lancet, 2017. **389**(10070): p. 709-717.

209. Prakash, T.P., et al., *Targeted delivery of antisense oligonucleotides to hepatocytes using triantennary N-acetyl galactosamine improves potency 10-fold in mice.* Nucleic Acids Res, 2014. **42**(13): p. 8796-807.

210. Huang, Y., *Preclinical and Clinical Advances of GalNAc-Decorated Nucleic Acid Therapeutics.* Molecular Therapy - Nucleic Acids, 2017. **6**: p. 116-132.

211. Spiess, M., *The asialoglycoprotein receptor: a model for endocytic transport receptors.* Biochemistry, 1990. **29**(43): p. 10009-18.

212. Bhat, B., Neben, S., Tay, J., Liu, K., Chau, N., Hogan, D., MacKenna, D., Gibson, N., *Late-breaking abstracts: RG-101, a GalNAC-conjugated anti-miR employing a unique mechanism of action by targeting host factor microRNA-122 (miR-122), demonstrates potent activity and reduction of HCV in preclinical studies.* Hepatology, 2013. **58**(6): p. 1377A-1395A.

213. Regulus Therapeutics, *Regulus Reports Clinical Hold of RG-101.* 2016, cited 30.03.2020. Available from: http://ir.regulusrx.com/news-releases/news-release-details/regulus-reports-clinical-hold-rg-101.

214. Regulus Therapeutics, *Regulus Announces Pipeline Updates and Advancements.* 2017, cited 30.03.2020. Available from: http://ir.regulusrx.com/news-releases/news-release-details/regulus-announces-pipeline-updates-and-advancements.

215. Geary, R.S., et al., *Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides.* Adv Drug Deliv Rev, 2015. **87**: p. 46-51.

216. Gomez, I.G., et al., *Anti-microRNA-21 oligonucleotides prevent Alport nephropathy progression by stimulating metabolic pathways.* J Clin Invest, 2015. **125**(1): p. 141-56.

217. Duffield, J., B. Bhat, and D. MACKENNA (Regulus Therapeutics inc), *Methods for treatment of alport syndrome*. WO2014058881A1, 2014, Google Patents.

218. ClinicalTrials.gov database, *Study of SAR339375 in Patients With Alport Syndrome (HERA).* Cited 30.03.2020. Available from: https://clinicaltrials.gov/ct2/show/NCT02855268?term=NCT02855268&draw=2&rank=1.

219. National Library of Medicine (US), *ClinicalTrials.gov database.* Available from: https://clinicaltrials.gov/ct2/home.

220. European Union Clinical Trials Register, EU Clinical Trials Register version 2.2. Available from: https://www.clinicaltrialsregister.eu/.

221. Lee, E.C., et al., *Discovery and preclinical evaluation of anti-miR-17 oligonucleotide RGLS4326 for the treatment of polycystic kidney disease.* Nature Communications, 2019. **10**(1): p. 4148.

222. Regulus Therapeutics, *Regulus Announces Strategic Update and Corporate Restructuring.* 2018, cited 30.03.2020. Available from: http://ir.regulusrx.com/news-releases/news-release-details/regulus-announces-strategic-update-and-corporate-restructuring.

223. Regulus Therapeutics, *Regulus Therapeutics Announces FDA Removal of Partial Clinical Hold for Multiple Ascending Dose Study of RGLS4326.* 2019, cited 30.03.2020. Available from: http://ir.regulusrx.com/news-releases/news-release-details/regulus-therapeutics-announces-fda-removal-partial-clinical-hold.

224. ClinicalTrials.gov database, *SOLAR: Efficacy and Safety of Cobomarsen (MRG-106) vs. Active Comparator in Subjects With Mycosis Fungoides (SOLAR).* Cited 30.03.2020. Available from: https://clinicaltrials.gov/ct2/show/NCT03713320?term=NCT03713320&draw=2&rank=1.

225. Querfeld, C., et al., *Phase 1 Study of the Safety and Efficacy of MRG-106, a Synthetic Inhibitor of microRNA-155, in CTCL Patients.* Blood, 2017. **130**(Suppl 1): p. 820-820.

226. Querfeld, C., et al., *Preliminary Results of a Phase 1 Trial Evaluating MRG-106, a Synthetic microRNA Antagonist (LNA antimiR) of microRNA-155, in Patients with CTCL.* Blood, 2016. **128**(22): p. 1829-1829.

227. Seto, A.G., et al., *Cobomarsen, an oligonucleotide inhibitor of miR-155, co-ordinately regulates multiple survival pathways to reduce cellular proliferation and survival in cutaneous T-cell lymphoma.* Br J Haematol, 2018. **183**(3): p. 428-444.

228. Catela Ivkovic, T., et al., *microRNAs as cancer therapeutics: A step closer to clinical application.* Cancer Lett, 2017. **407**: p. 113-122.

229. Lei, C., et al., *Up-regulated miR155 Reverses the Epithelial-mesenchymal Transition Induced by EGF and Increases Chemo-sensitivity to Cisplatin in Human Caski Cervical Cancer Cells.* PLOS ONE, 2012. **7**(12): p. e52310.

230. Bader, A.G., *miR-34 - a microRNA replacement therapy is headed to the clinic.* Front Genet, 2012. **3**(120): p. 120.

231. Beg, M.S., et al., *Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors.* Invest New Drugs, 2017. **35**(2): p. 180-188.

232. ClinicalTrials.gov database, *A Multicenter Phase I Study of MRX34, MicroRNA miR-RX34 Liposomal Injection.* Cited 30.03.2020. Available from: https://clinicaltrials.gov/ct2/show/NCT01829971?term=NCT01829971&draw=2&rank=1.

233. Gallant-Behm, C.L., et al., *A MicroRNA-29 Mimic (Remlarsen) Represses Extracellular Matrix Expression and Fibroplasia in the Skin.* J Invest Dermatol, 2019. **139**(5): p. 1073-1081.

234. ClinicalTrials.gov database, *Efficacy, Safety, and Tolerability of Remlarsen (MRG-201) Following Intradermal Injection in Subjects With a History of Keloids.* Cited 30.03.2020. Available from: https://clinicaltrials.gov/ct2/show/NCT03601052?term=NCT03601052&draw=2&rank=1.

235. Reid, G., et al., *Clinical development of TargomiRs, a miRNA mimic-based treatment for patients with recurrent thoracic cancer.* Epigenomics, 2016. **8**(8): p. 1079-1085.

236. van Zandwijk, N., et al., *Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study.* Lancet Oncol, 2017. **18**(10): p. 1386-1396.

237. Lee, T., et al., *Dosage and Temporal Thresholds in microRNA Proteomics.* Molecular & Cellular Proteomics, 2015. **14**(2): p. 289-302.

238. Steinkraus, B.R., M. Toegel, and T.A. Fulga, *Tiny giants of gene regulation: experimental strategies for microRNA functional studies.* Wiley Interdiscip Rev Dev Biol, 2016. **5**(3): p. 311-62.

239. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-486.

240. Ule, J., H.W. Hwang, and R.B. Darnell, *The Future of Cross-Linking and Immunoprecipitation (CLIP).* Cold Spring Harb Perspect Biol, 2018. **10**(8): a032243.

241. Greenberg, J.R., *Ultraviolet light-induced crosslinking of mRNA to proteins.* Nucleic Acids Research, 1979. **6**(2): p. 715-732.

242. Hockensmith, J.W., et al., *Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system.* J Biol Chem, 1986. **261**(8): p. 3512-8.

243. WAGENMAKERS, A.J.M., R.J. REINDERS, and W.J. VAN VENROOIJ, *Cross-linking of mRNA to Proteins by Irradiation of Intact Cells with Ultraviolet Light.* European Journal of Biochemistry, 1980. **112**(2): p. 323-330.

244. Smith, K.C., *Photochemical addition of amino acids to 14C-uracil.* Biochemical and Biophysical Research Communications, 1969. **34**(3): p. 354-357.

245. Hafner, M., et al., *Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.* Cell, 2010. **141**(1): p. 129-41.

246. Hearst, J.E., *Psoralen photochemistry.* Annu Rev Biophys Bioeng, 1981. **10**: p. 69-86.

247. Lipson, S.E. and J.E. Hearst, *Psoralen cross-linking of ribosomal RNA.* Methods Enzymol, 1988. **164**: p. 330-41.

248. Cimino, G.D., et al., *Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry.* Annu Rev Biochem, 1985. **54**: p. 1151-93.

249. Shi, Y. and J.E. Hearst, *Thermostability of double-stranded deoxyribonucleic acids: effects of covalent additions of a psoralen.* Biochemistry, 1986. **25**(20): p. 5895-902.

250. Spielmann, H.P., et al., *DNA structural reorganization upon conversion of a psoralen furan-side monoadduct to an interstrand cross-link: implications for DNA repair.* Proc Natl Acad Sci U S A, 1995. **92**(6): p. 2345-9.

251. Dewick, P.M., *The Shikimate Pathway: Aromatic Amino Acids and Phenylpropanoids*, in *Medicinal Natural Products*. 2009, John Wiley & Sons, Ltd. p. 137-186.

252. Isaacs, S.T., et al., *Synthesis and characterization of new psoralen derivatives with superior photoreactivity with DNA and RNA.* Biochemistry, 1977. **16**(6): p. 1058-1064.

253. Bachellerie, J.P. and J.E. Hearst, *Specificity of the photoreaction of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen with RNA. Identification of reactive sites in E. coli tRNAPhe.* Biochemistry, 1982. **21**(6): p. 1357-1363.

254. Thompson, J.F., et al., *Dependence of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen photoaddition on the conformation of ribonucleic acid.* Biochemistry, 1982. **21**(6): p. 1363-8.

255. Dall'Acqua, F., D. Vedaldi, and M. Recher, *THE PHOTOREACTION BETWEEN FUROCOUMARINS AND VARIOUS DNA WITH DIFFERENT BASE COMPOSITIONS.* Photochemistry and Photobiology, 1978. **27**(1): p. 33-36.

256. Stern, R.S., *Psoralen and ultraviolet a light therapy for psoriasis.* N Engl J Med, 2007. **357**(7): p. 682-90.

257. Sastry, S.S., B.M. Ross, and A. P'Arraga, *Cross-linking of DNA-binding proteins to DNA with psoralen and psoralen furan-side monoadducts. Comparison of action spectra with DNA-DNA cross-linking.* J Biol Chem, 1997. **272**(6): p. 3715-23.

258. Wassarman, D.A., *Psoralen crosslinking of small RNAs in vitro.* Mol Biol Rep, 1993. **17**(2): p. 143-51.

259. Toussaint, M., et al., *Psoralen photocrosslinking, a tool to study the chromatin structure of RNA polymerase I - transcribed ribosomal genes.* Biochemistry and Cell Biology, 2005. **83**(4): p. 449-459.

260. Thompson, J.F. and J.E. Hearst, *Structure of E. coli 16S RNA elucidated by psoralen crosslinking.* Cell, 1983. **32**(4): p. 1355-1365.

261. Wassarman, D. and J. Steitz, *Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing.* Science, 1992. **257**(5078): p. 1918-1925.

262. Wang, Z. and T.M. Rana, *Probing RNA-protein interactions by psoralen photocrosslinking.* Methods Mol Biol, 1999. **118**: p. 49-62.

263. Kobori, A., A. Yamayoshi, and A. Murakami, *Synthesis of Oligonucleotides Containing 4,5',8-Trimethylpsoralen at the 2'-O Position and Their Cross-Linking Properties with RNAs.* Curr Protoc Nucleic Acid Chem, 2014. **58**: p. 5.15.1-15.

264. Lee, B.L., et al., *Interaction of psoralen-derivatized oligodeoxyribonucleoside methylphosphonates with single-stranded DNA.* Biochemistry, 1988. **27**(9): p. 3197-3203.

265. Kean, J.M., et al., *Photochemical cross-linking of psoralen-derivatized oligonucleoside methylphosphonates to rabbit globin messenger RNA.* Biochemistry, 1988. **27**(26): p. 9113-21.

266. Pieles, U. and U. Englisch, *Psoralen covalently linked to oligodeoxyribonucleotides: synthesis, sequence specific recognition of DNA and photo-cross-linking to pyrimidine residues of DNA.* Nucleic Acids Res, 1989. **17**(1): p. 285-99.

267. Murakami, A., et al., *Photodynamic antisense regulation of human cervical carcinoma cell growth using psoralen-conjugated oligo(nucleoside phosphorothioate).* Eur J Pharm Sci, 2001. **13**(1): p. 25-34.

268. Yamayoshi, A., et al., *Psoralen-conjugated oligonucleotide with hairpin structure as a novel photo-sensitive antisense molecule.* Chem Commun (Camb), 2003(12): p. 1370-1.

269. Kean, J.M. and P.S. Miller, *Effect of Target Structure on Crosslinking by Psoralen-Derivatized Oligonucleoside Methylphosphonates.* Biochemistry, 1994. **33**(31): p. 9178-9186.

270. Higuchi, M., et al., *Selective photo-cross-linking of 2'-O-psoralen-conjugated oligonucleotide with RNAs having point mutations.* Nucleosides Nucleotides Nucleic Acids, 2007. **26**(3): p. 277-90.

271. Higuchi, M., et al., *Synthesis of antisense oligonucleotides containing 2'-O-psoralenylmethoxyalkyl adenosine for photodynamic regulation of point mutations in RNA.* Bioorg Med Chem, 2009. **17**(2): p. 475-83.

272. Baigude, H., et al., *miR-TRAP: a benchtop chemical biology strategy to identify microRNA targets.* Angew Chem Int Ed Engl, 2012. **51**(24): p. 5880-3.

273. Matsuyama, Y., et al., *Functional regulation of RNA-induced silencing complex by photoreactive oligonucleotides.* Bioorg Med Chem, 2014. **22**(3): p. 1003-7.

274. Kobori, A., et al., *Synthesis and Photoinduced Cross-linking Reactions of 4,5',8-Trimethylpsoralen-incorporated Oligodeoxyribonucleotide.* Chemistry Letters, 2009. **38**(3): p. 272-273.

275. Kobori, A., et al., *Synthesis and Crosslinking Activity of 4-N-(4, 5', 8-Trimethylpsoralen-4'-ylmethyl)-2'-deoxycytidine-containing Oligodeoxyribonucleotides.* Chemistry letters, 2012. **41**(8): p. 804-805.

276. Sun, H., H. Fan, and X. Peng, *Quantitative DNA interstrand cross-link formation by coumarin and thymine: structure determination, sequence effect, and fluorescence detection.* J Org Chem, 2014. **79**(23): p. 11359-69.

277. Shigeno, A., et al., *Quick regulation of mRNA functions by a few seconds of photoirradiation.* Organic & Biomolecular Chemistry, 2012. **10**(38): p. 7820-7825.

278. Nakamoto, K. and Y. Ueno, *Diazirine-Containing RNA Photo-Cross-Linking Probes for Capturing microRNA Targets.* The Journal of Organic Chemistry, 2014. **79**(6): p. 2463-2472.

279. He, L., et al., *A microRNA polycistron as a potential human oncogene.* Nature, 2005. **435**(7043): p. 828-833.

280. Rostovtsev, V.V., et al., *A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes.* Angewandte Chemie International Edition, 2002. **41**(14): p. 2596-2599.

281. Tornøe, C.W., C. Christensen, and M. Meldal, *Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides.* The Journal of Organic Chemistry, 2002. **67**(9): p. 3057-3064.

282. El-Sagheer, A.H. and T. Brown, *Click chemistry with DNA.* Chemical Society Reviews, 2010. **39**(4): p. 1388-1405.

283. Astakhova, I.K. and J. Wengel, *Interfacing click chemistry with automated oligonucleotide synthesis for the preparation of fluorescent DNA probes containing internal xanthene and cyanine dyes.* Chemistry, 2013. **19**(3): p. 1112-22.

284. Pradere, U., et al., *Chemical synthesis of mono- and bis-labeled pre-microRNAs.* Angew Chem Int Ed Engl, 2013. **52**(46): p. 12028-32.

285. Porschke, D. and M. Eigen, *Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH.* J Mol Biol, 1971. **62**(2): p. 361-81.

286. Patel, P.C., et al., *Duplex end breathing determines serum stability and intracellular potency of siRNA-Au NPs.* Molecular pharmaceutics, 2011. **8**(4): p. 1285-1291.

287. Wang, Y., *Tailoring CLIP-Based Methods for Exploring the miRNA Targetome.* ETH Zurich, 2019.

288. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

289. Dubinsky, L., B.P. Krom, and M.M. Meijler, *Diazirine based photoaffinity labeling.* Bioorganic & Medicinal Chemistry, 2012. **20**(2): p. 554-570.

290. Hashimoto, M. and Y. Hatanaka, *Recent Progress in Diazirine-Based Photoaffinity Labeling.* European Journal of Organic Chemistry, 2008. **2008**(15): p. 2513-2523.

291. Mackinnon, A.L. and J. Taunton, *Target Identification by Diazirine Photo-Cross-linking and Click Chemistry.* Curr Protoc Chem Biol, 2009. **1**: p. 55-73.

292. Yamaguchi, T. and M. Saneyoshi, *A Photolabile 2', 3'-Dideoxyuridylate Analog Bearing an Aryl(trifluoromethyl)diazirine Moiety: Photoaffinity Labeling of HIV-1 Reverse Transcriptase.* Nucleic Acids Research, 1996. **24**(17): p. 3364-3369.

293. Pattison, D.I. and M.J. Davies, *Actions of ultraviolet light on cellular structures*, in *Cancer: Cell Structures, Carcinogens and Genomic Instability*. 2006, Birkhäuser Basel: Basel. p. 131-157.

294. Masuma, R., et al., *Effects of UV wavelength on cell damages caused by UV irradiation in PC12 cells.* J Photochem Photobiol B, 2013. **125**: p. 202-8.

295. Frey, H.M., *The Photolysis of the Diazirines*, in *Advances in Photochemistry*. 2007, John Wiley & Sons, Inc. p. 225-256.

296. Brunner, J., H. Senn, and F.M. Richards, *3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents.* J Biol Chem, 1980. **255**(8): p. 3313-8.

297. Das, J., *Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments.* Chemical Reviews, 2011. **111**(8): p. 4405-4417.

298. Bonneau, R. and M.T.H. Liu, *Quantum Yield of Formation of Diazo Compounds from the Photolysis of Diazirines.* Journal of the American Chemical Society, 1996. **118**(30): p. 7229-7230.

299. Babu Kumar, A., J.M. Anderson, and R. Manetsch, *Design, synthesis and photoactivation studies of fluorous photolabels.* Organic & Biomolecular Chemistry, 2011. **9**(18): p. 6284-6292.

300. Yamamoto, N., et al., *Mechanism of Carbene Formation from the Excited States of Diazirine and Diazomethane: An MC-SCF Study.* Journal of the American Chemical Society, 1994. **116**(5): p. 2064-2074.

301. Zhang, Y., et al., *Direct Observation of Carbene and Diazo Formation from Aryldiazirines by Ultrafast Infrared Spectroscopy.* Journal of the American Chemical Society, 2008. **130**(48): p. 16134-16135.

302. Mueller, P.H., et al., *Carbene singlet-triplet gaps. Linear correlations with substituent .pi.-donation.* Journal of the American Chemical Society, 1981. **103**(17): p. 5049-5052.

303. Chee, G.L., et al., *A diazirine-based photoaffinity etoposide probe for labeling topoisomerase II.* Bioorg Med Chem, 2010. **18**(2): p. 830-8.

304. Sergiev, P.V., et al., *The path of mRNA through the bacterial ribosome: a site-directed crosslinking study using new photoreactive derivatives of guanosine and uridine.* RNA, 1997. **3**(5): p. 464-475.

305. Wang, J., et al., *Influence of Solvent on Carbene Intersystem Crossing Rates.* Journal of the American Chemical Society, 2008. **130**(20): p. 6604-6609.

306. Winnacker, M., et al., *Novel Diazirine-Containing DNA Photoaffinity Probes for the Investigation of DNA-Protein-Interactions.* ChemBioChem, 2009. **10**(1): p. 109-118.

307. Liebmann, M., F. Di Pasquale, and A. Marx, *A New Photoactive Building Block for Investigation of DNA Backbone Interactions: Photoaffinity Labeling of Human DNA Polymerase β.* ChemBioChem, 2006. **7**(12): p. 1965-1969.

308. Shigdel, U.K., J. Zhang, and C. He, *Diazirine-Based DNA Photo-Cross-Linking Probes for the Study of Protein–DNA Interactions.* Angewandte Chemie International Edition, 2008. **47**(1): p. 90-93.

309. Kuboe, S., et al., *Diazirine-containing RNA photocrosslinking probes for the study of siRNA-protein interactions.* Chem Commun (Camb), 2010. **46**(39): p. 7367-9.

310. Qiu, Z., et al., *A Diazirine-Based Nucleoside Analogue for Efficient DNA Interstrand Photocross-Linking.* Journal of the American Chemical Society, 2008. **130**(44): p. 14398-14399.

311. Nakamoto, K., et al., *Labeling of target mRNAs using a photo-reactive microRNA probe.* Chemical Communications, 2016. **52**(40): p. 6720-6722.

312. Nakamoto, K., Y. Akao, and Y. Ueno, *Diazirine-containing tag-free RNA probes for efficient RISC-loading and photoaffinity labeling of microRNA targets.* Bioorganic & Medicinal Chemistry Letters, 2018. **28**(17): p. 2906-2909.

313. Kan, T., et al., *Convenient Synthesis of Photoaffinity Probes and Evaluation of Their Labeling Abilities.* Organic Letters, 2007. **9**(11): p. 2055-2058.

314. Savechenkov, P.Y., et al., *Allyl m-Trifluoromethyldiazirine Mephobarbital: An Unusually Potent Enantioselective and Photoreactive Barbiturate General Anesthetic.* Journal of Medicinal Chemistry, 2012. **55**(14): p. 6554-6565.

315. Cameron, M.A., S.B. Cush, and R.P. Hammer, *Facile Preparation of Protected Furanoid Glycals from Thymidine.* The Journal of Organic Chemistry, 1997. **62**(26): p. 9065-9069.

316. Macmillan, A.M. and G.L. Verdine, *Engineering tethered DNA molecules by the convertible nucleoside approach.* Tetrahedron, 1991. **47**(14): p. 2603-2616.

317. MacMillan, A.M. and G.L. Verdine, *Synthesis of functionally tethered oligodeoxynucleotides by the convertible nucleoside approach.* The Journal of Organic Chemistry, 1990. **55**(24): p. 5931-5933.

318. Brzezinska, J., et al., *Polyaminooligonucleotide: NMR structure of duplex DNA containing a nucleoside with spermine residue, N-[4,9,13-triazatridecan-1-yl]-2'-deoxycytidine.* Biochim Biophys Acta, 2014. **1840**(3): p. 1163-70.

319. Decuypere, E., et al., *Increased Affinity of 2'-O-(2-Methoxyethyl)-Modified Oligonucleotides to RNA through Conjugation of Spermine at Cytidines.* Helvetica Chimica Acta, 2019. **102**(12): p. e1900222.

320. Domingo, O., et al., *Intermolecular 'cross-torque': the N4-cytosine propargyl residue is rotated to the 'CH'-edge as a result of Watson-Crick interaction.* Nucleic Acids Res, 2015. **43**(11): p. 5275-83.

321. Cervi, A.R., et al., *The crystal structure of N4-methylcytosine.guanosine base-pairs in the synthetic hexanucleotide d(CGCGm4CG).* Nucleic acids research, 1993. **21**(24): p. 5623-5629.

322. Grasby, J.A., et al., *Synthesis and Applications of Oligoribonucleotides Containing N 4-Methylcytidine.* Nucleosides and Nucleotides, 1995. **14**(3-5): p. 1129-1132.

323. Engel, J.D. and P.H. von Hippel, *Effects of methylation on the stability of nucleic acid conformations: studies at the monomer level.* Biochemistry, 1974. **13**(20): p. 4143-58.

324. Shah, K., H. Wu, and T.M. Rana, *Synthesis of Uridine Phosphoramidite Analogs: Reagents for Site-Specific Incorporation of Photoreactive Sites into RNA Sequences.* Bioconjugate Chemistry, 1994. **5**(6): p. 508-512.

325. Guennewig, B., et al., *Properties of N4-Methylated Cytidines in miRNA Mimics.* Nucleic Acid Therapeutics, 2012. **22**(2): p. 109-116.

326. Berk, C., *Alternative scaffolds for systemic delivery of small interfering RNAs.* ETH Zurich, 2019.

327. Hei, D., *Methods and devices for the removal of psoralens from blood products.* WO9640857A1, 1996.

328. Lu, X. and K. Zhang, *PEGylation of therapeutic oligonucletides: From linear to highly branched PEG architectures.* Nano research, 2018. **11**(10): p. 5519-5534.

329. Ng, E.W., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease.* Nat Rev Drug Discov, 2006. **5**(2): p. 123-32.

330. Vinores, S.A., *Pegaptanib in the treatment of wet, age-related macular degeneration.* International journal of nanomedicine, 2006. **1**(3): p. 263-268.

331. Drolet, D.W., et al., *Pharmacokinetics and safety of an anti-vascular endothelial growth factor aptamer (NX1838) following injection into the vitreous humor of rhesus monkeys.* Pharm Res, 2000. **17**(12): p. 1503-10.

332. Manoharan, M., *2'-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation.* Biochim Biophys Acta, 1999. **1489**(1): p. 117-30.

333. Martin, P., *Ein neuer Zugang zu 2'-O-Alkylribonucleosiden und Eigenschaften deren Oligonucleotide.* Helvetica Chimica Acta, 1995. **78**(2): p. 486-504.

334. Gao, Y., et al., *Synthesis of Uridine Derivatives Containing Amino Acid Residues.* Synthetic Communications, 2003. **33**(15): p. 2635-2641.

335. Pradere, U. and J. Hall, *Site-Specific Difunctionalization of Structured RNAs Yields Probes for microRNA Maturation.* Bioconjug Chem, 2016. **27**(3): p. 681-7.

336. Husken, D., et al., *Creating RNA bulges: cleavage of RNA in RNA/DNA duplexes by metal ion catalysis.* Biochemistry, 1996. **35**(51): p. 16591-600.

337. Menzi, M., et al., *Site-Specific Labeling of MicroRNA Precursors: A Structure-Activity Relationship Study.* Chembiochem, 2016. **17**(21): p. 2012-2017.

338. Guo, Y.E. and J.A. Steitz, *3'-Biotin-tagged microRNA-27 does not associate with Argonaute proteins in cells.* Rna, 2014. **20**(7): p. 985-8.

339. Allard, S.T.M.K., K., *Luciferase reporter assays: powerful, adaptable tools for cell biology research.* Cell Notes 21, p. 23–26 Promega Corporation, 2008.

340. Smale, S.T., *Luciferase assay.* Cold Spring Harb Protoc, 2010. **2010**(5): p. pdb.prot5421.

341. Zhou, H., et al., *miR-155 and its star-form partner miR-155* cooperatively regulate type I interferon production by human plasmacytoid dendritic cells.* Blood, 2010. **116**(26): p. 5885-94.

342. Pathak, S., et al., *MiR-155 modulates the inflammatory phenotype of intestinal myofibroblasts by targeting SOCS1 in ulcerative colitis.* Experimental & Molecular Medicine, 2015. **47**(5): p. e164-e164.

343. Bhattacharyya, D., G. Mirihana Arachchilage, and S. Basu, *Metal Cations in G-Quadruplex Folding and Stability.* Frontiers in Chemistry, 2016. **4**: p. 38.

344. Rhodes, D. and H.J. Lipps, *G-quadruplexes and their regulatory roles in biology.* Nucleic Acids Res, 2015. **43**(18): p. 8627-37.

345. Lightfoot, H.L., et al., *The diverse structural landscape of quadruplexes.* FEBS Lett, 2019. **593**(16): p. 2083-2102.

346. Norman, D.G., et al., *Location of Cyanine-3 on Double-Stranded DNA: Importance for Fluorescence Resonance Energy Transfer Studies.* Biochemistry, 2000. **39**(21): p. 6317-6324.

347. Moreira, B.G., Y. You, and R. Owczarzy, *Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model.* Biophysical Chemistry, 2015. **198**: p. 36-44.

348. Iqbal, A., et al., *The Structure of Cyanine 5 Terminally Attached to Double-Stranded DNA: Implications for FRET Studies.* Biochemistry, 2008. **47**(30): p. 7857-7862.

349. Hagen, T., et al., *Site-Specific Fluorophore Labeling of Guanosines in RNA G-Quadruplexes.* ACS Omega, 2019. **4**(5): p. 8472-8479.

350. Masliah, G., et al., *Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains.* Embo j, 2018. **37**(6): p. e97089.

351. Menzi, M., et al., *Towards Improved Oligonucleotide Therapeutics Through Faster Target Binding Kinetics.* Chemistry – A European Journal, 2017. **23**(57): p. 14221-14230.

352. Li, M., et al., *Synthesis and cellular activity of stereochemically-pure 2[prime or minute]-O-(2-methoxyethyl)-phosphorothioate oligonucleotides.* Chemical Communications, 2017. **53**(3): p. 541-544.

353. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Research, 2006. **34**(suppl_1): p. D140-D144.

354. Xu, Y., et al., *A U-Tetrad Stabilizes Human Telomeric RNA G-Quadruplex Structure.* Journal of the American Chemical Society, 2010. **132**(21): p. 7231-7233.

355. Hassler, M., et al., *RNA synthesis via dimer and trimer phosphoramidite block coupling.* Tetrahedron Letters, 2011. **52**(20): p. 2575-2578.

356. Wiesmayr, A., P. Fournier, and A. Jaschke, *An on-bead tailing/ligation approach for sequencing resin-bound RNA libraries.* Nucleic Acids Res, 2012. **40**(9): p. e68.

357. Fusz, S., et al., *Photocleavable Initiator Nucleotide Substrates for an Aldolase Ribozyme*. J. Org. Chem., 2008. **73**(13): 5069-5077.

358. Zagalak, J.A., et al., *Properties of short double-stranded RNAs carrying randomized base pairs: toward better controls for RNAi experiments.* RNA, 2015. **21**(12): p. 2132-42.

IX. Appendix

IX.1. LC-MS chromatograms of the synthesized oligoribonucleotides

IX.1.1. Chapter II

IX.1.1.1. ORN-1 to ORN-28

Table 25: Summary of the synthesized oligoribonucleotides (ORN-1 to ORN-28). **X**, **Y**, **Z**, **W** designate adenosine, guanosine, uridine or cytidine with the trioxsalen attached at the 2'-*O*-position of the ribose, respectively.

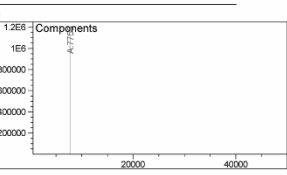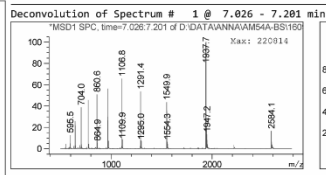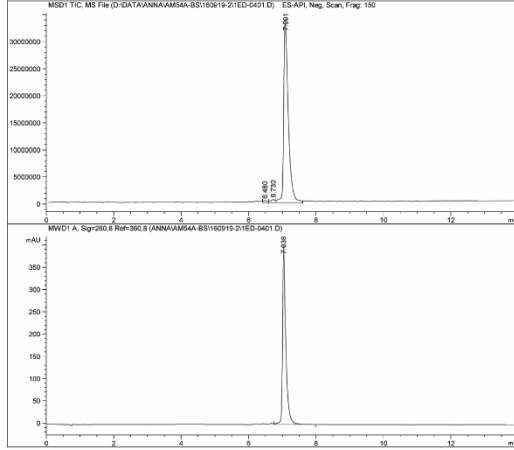| Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-1	AAA**X**GUGCUUACAGUGCAGGUAG	23	7754.0	7754.9	0.01	>98	50 nmol
2	ORN-2	ACUGUAAGCACUUUU	15	4698.0	4697.1	0.02	>95	50 nmol
3	ORN-3	CAA**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7733.1	0.01	>90	50 nmol
4	ORN-4	ACUGUAAGCACUUUG	15	4736.9	4736.9	0.00	>80	50 nmol
5	ORN-5	ACA**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7730.2	0.02	>90	50 nmol
6	ORN-6	ACUGUAAGCACUUGU	15	4736.9	4736.1	0.02	>90	50 nmol
7	ORN-7	AAC**X**GUGCUUACAGUGCAGGUAG	23	7732.0	7730.9	0.01	>98	50 nmol
8	ORN-8	ACUGUAAGCACUGUU	15	4736.9	4736.0	0.02	>95	50 nmol
9	ORN-9	CCA**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7706.8	0.01	>90	50 nmol
10	ORN-10	ACUGUAAGCACUUGG	15	4776.0	4775.1	0.02	>70	50 nmol
11	ORN-11	CAC**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7707.0	0.01	>98	50 nmol
12	ORN-12	ACUGUAAGCACUGUG	15	4776.0	4774.9	0.02	>90	50 nmol
13	ORN-13	ACC**X**GUGCUUACAGUGCAGGUAG	23	7707.8	7706.8	0.01	>85	50 nmol
14	ORN-14	ACUGUAAGCACUGGU	15	4776.0	4775.1	0.02	>90	50 nmol
15	ORN-15	CCC**X**GUGCUUACAGUGCAGGUAG	23	7683.8	7682.5	0.02	>98	50 nmol
16	ORN-16	ACUGUAAGCACUGGG	15	4815.0	4814.2	0.02	>85	50 nmol
17	ORN-17	ACUGUAAGCACCGGU	15	4775.0	4774.0	0.02	>90	50 nmol
18	ORN-18	ACUGUAAGCACCGUG	15	4775.0	4771.1*	-	>80	50 nmol
19	ORN-19	AAAA**Y**UGCUUACAGUGCAGGUAG	23	7756.0	7755.4	0.01	>98	50 nmol
20	ORN-20	CACC**Y**UGCUUACAGUGCAGGUAG	23	7683.9	7682.8	0.01	>98	50 nmol
21	ORN-21	ACUGUAAGCACGGUG	15	4815.0	4813.9	0.02	>98	50 nmol
22	ORN-22	AAAAG**Z**GCUUACAGUGCAGGUAG	23	7756.0	7755.0	0.01	>98	50 nmol
23	ORN-23	ACACG**Z**GCUUACAGUGCAGGUAG	23	7707.9	7706.5	0.02	>98	50 nmol
24	ORN-24	ACUGUAAGCACGUGU	15	4775.9	4775.1	0.02	>95	50 nmol
25	ORN-25	AAAAGU**Y**CUUACAGUGCAGGUAG	23	7756.0	7755.1	0.01	>85	50 nmol
26	ORN-26	AAAAGUG**W**UUACAGUGCAGGUAG	23	7756.0	7755.1	0.01	>98	50 nmol
27	ORN-27	GCCAGUGACC**X**CUGUGCAGGUAG	23	7722.9	7722.0	0.01	>98	50 nmol
28	ORN-28	ACAGUGGUCACUGGC	15	4791.0	4790.4	0.01	>95	50 nmol

*mass calculated as (m*z)-z = (1591.4*3)-3 = 4771.2

ORN-1

Additional Info : Peak(s) manually integrated
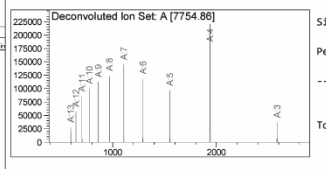
MSD1 TIC, MS File (D:\DATA\ANNA\AM54A-BS\160919-21ED-0401.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM54A-BS\160919-21ED-0401.D)

Deconvolution of Spectrum #   1 @ 7.026 - 7.201 min

*MSD1 SPC, time=7.026:7.201 of D:\DATA\ANNA\AM54A-BS\160

Deconvoluted Ion Set: A [7754.86]

Components

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.038	VB	0.0947	2718.68652	402.87503	100.0000
Totals :				2718.68652	402.87503	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7754.86	1117108	100.00

*** End of Report ***

ORN-2

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\L-246-013-AM27-BS-AM54A-BS\050519-\1AE-0501.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\L-246-013-AM27-BS-AM54A-BS\050519-\1AE-0501.D)

Deconvolution of Spectrum #   1 @ 6.355 - 6.529 min

*MSD1 SPC, time=6.355:6.529 of D:\DATA\ANNA\L-246-013-AM2

Deconvoluted Ion Set: A [4697.05]

Components
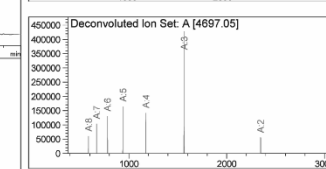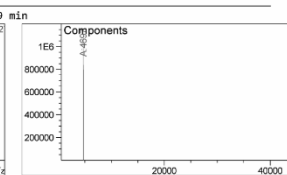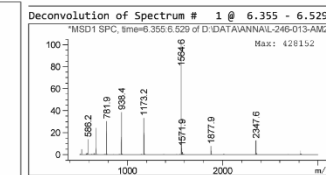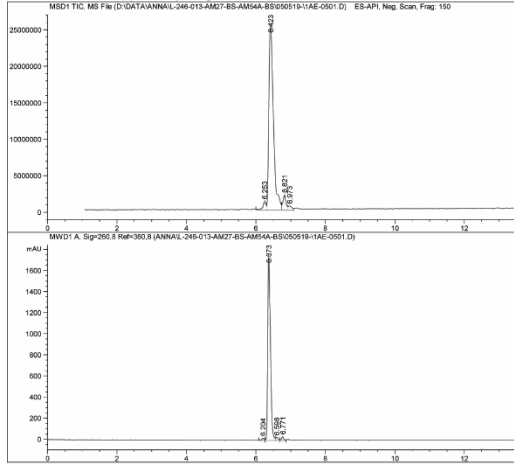
Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	6.204	VV	0.0648	96.91198	22.23708	0.9935
2	6.373	VV	0.0816	9339.25195	1769.37292	95.7465
3	6.598	VV	0.0734	154.61998	29.36279	1.5852
4	6.771	VV	0.0667	163.36398	36.19304	1.6748
Totals :				9754.14790	1857.16583	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4697.05	1072575	100.00

*** End of Report ***

ORN-3

Additional Info : Peak(s) manually integrated
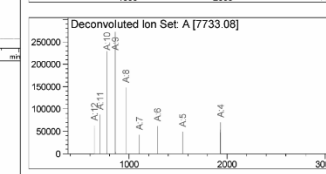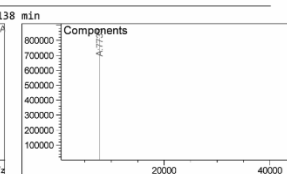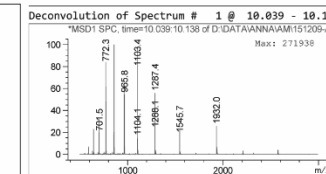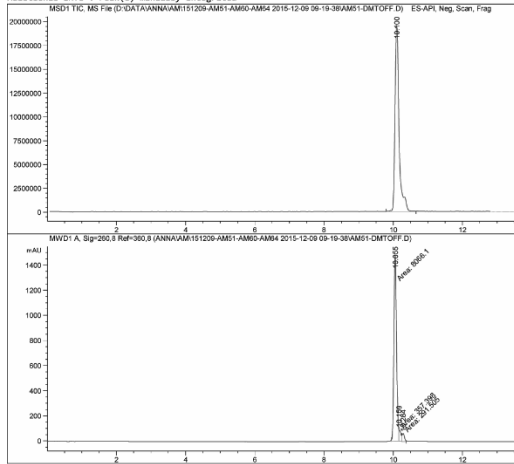
MSD1 TIC, MS File (D:\DATA\ANNA\AM\151209-AM51-AM60-AM64 2015-12-09 09-19-38\AM51-DMTOFF.D)   ES-API, Neg, Scan, Frag

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\151209-AM51-AM60-AM64 2015-12-09 09-19-38\AM51-DMTOFF.D)

Deconvolution of Spectrum #   1 @ 10.039 - 10.138 min

*MSD1 SPC, time=10.039:10.138 of D:\DATA\ANNA\AM\151209-A

Deconvoluted Ion Set: A [7733.08]

Components

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.055	MM	0.0906	8066.09521	1483.88159	92.5542
2	10.159	MM	0.0481	357.39801	92.44298	4.1010
3	10.284	MM	0.0778	291.50534	62.42254	3.3449
Totals :				8714.99857	1638.74711	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7733.08	814388	100.00

*** End of Report ***

ORN-4

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\151209-AM61-AM66 2015-12-09 20-02-38\AM60-2-OFF.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\151209-AM61-AM66 2015-12-09 20-02-38\AM60-2-OFF.D)
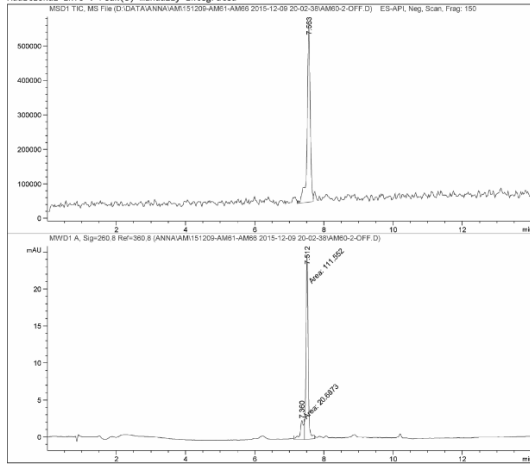
Deconvolution of Spectrum #   1 @ 7.490 - 7.614 min

*MSD1 SPC, time=7.490:7.614 of D:\DATA\ANNA\AM\151209-AM Max: 7809
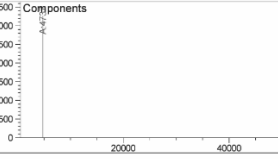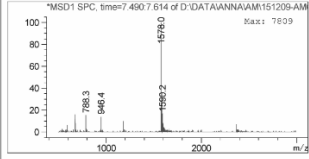
Deconvoluted Ion Set: A [4736.88]

Components

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	7.360	MM	0.1330	20.68729	2.59206	15.6438
2	7.512	MM	0.0746	111.55235	24.92814	84.3562

Totals :   132.23964   27.52021

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4736.88	3261	100.00

*** End of Report ***

ORN-5

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM\151209-AM51-AM60-AM64 2015-12-09 09-19-38\AM52-DMTOFF.D)   ES-API, Neg, Scan, Frag

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\151209-AM51-AM60-AM64 2015-12-09 09-19-38\AM52-DMTOFF.D)
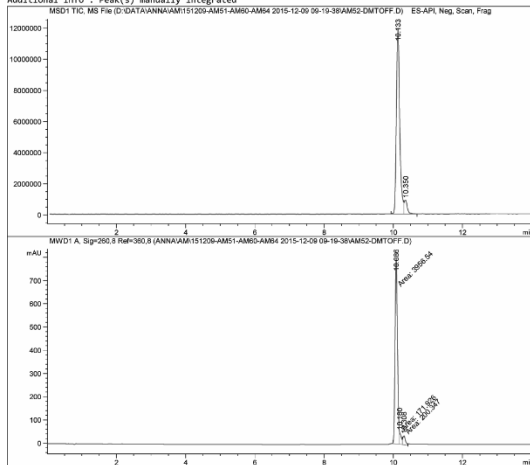
Deconvolution of Spectrum #   1 @ 10.089 - 10.163 min

*MSD1 SPC, time=10.089:10.163 of D:\DATA\ANNA\AM\151209-A Max: 153790
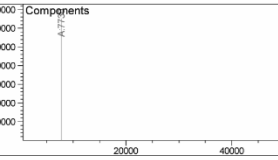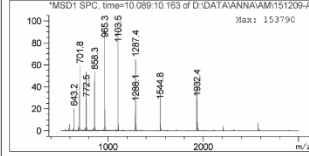
Deconvoluted Ion Set: A [7730.15]

Components

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	10.086	MM	0.0824	3956.54053	800.29095	91.4001
2	10.180	MM	0.0413	171.92566	53.20490	3.9717
3	10.308	MM	0.0861	200.34697	38.78242	4.6282

Totals :   4328.81316   892.27828

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7730.15	652244	100.00

*** End of Report ***

ORN-6

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM61-AM114\170819000003.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM61-AM114\170819000003.D)

Deconvolution of Spectrum #   1 @ 8.472 - 8.596 min

*MSD1 SPC, time=8.472:8.596 of D:\DATA\ANNA\AM61-AM114\1 Max: 726837
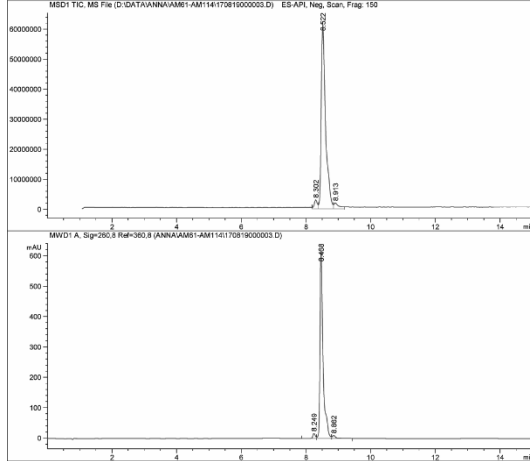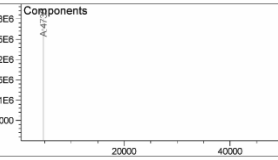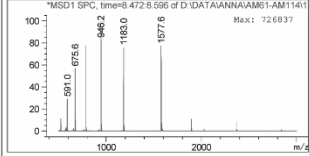
Deconvoluted Ion Set: A [4736.06]
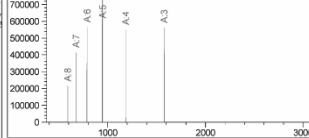
Components

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	8.249	BV	0.0823	84.85616	15.88531	1.9798
2	8.468	VV	0.1006	4127.60742	612.88196	96.3031
3	8.862	VB	0.1067	73.59500	9.46755	1.7171

Totals :   4286.05858   638.23481

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4736.06	3011635	100.00

*** End of Report ***

ORN-7

MSD1 TIC, MS File (D:\DATA\ANNA\AM\151126-AM32-45 DMT-OFF2 2015-11-27 12-45-08\AM34-MAIN-10UL.D)   ES-API, Neg, Scan

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\151126-AM32-45 DMT-OFF2 2015-11-27 12-45-08\AM34-MAIN-10UL.D)

Deconvolution of Spectrum #   1 @ 10.138 - 10.238 min

*MSD1 SPC, time=10.138:10.238 of D:\DATA\ANNA\AM\151126-A   Max: 13601

Components

Deconvoluted Ion Set: A [7730.90]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.133	BV	0.0600	203.36925	51.54964	100.0000
Totals :				203.36925	51.54964	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7730.90	27593	100.00

*** End of Report ***

ORN-8

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM61-AM114\170819000004.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM61-AM114\170819000004.D)

Deconvolution of Spectrum #   1 @ 8.422 - 8.522 min

*MSD1 SPC, time=8.422:8.522 of D:\DATA\ANNA\AM61-AM114\1   Max: 753766

Components

Deconvoluted Ion Set: A [4736.04]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.164	BV	0.0899	35.95938	5.99800	0.8417
2	8.390	VV	0.0993	4166.37451	628.99139	97.5212
3	8.773	VB	0.1215	69.94080	7.73612	1.6371
Totals :				4272.27469	642.72552	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4736.04	3073866	100.00

*** End of Report ***

ORN-9

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM\160203-AM73-AM79- DMT-OFF 2016-02-03 11-36-44\AM73-DMTOFF-1STPEAK.D)   ES-API, Neg

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\16...3-AM73-AM79- DMT-OFF 2016-02-03 11-36-44\AM73-DMTOFF-1STPEAK.D)

Deconvolution of Spectrum #   1 @ 10.213 - 10.288 min

*MSD1 SPC, time=10.213:10.288 of D:\DATA\ANNA\AM\160203-A   Max: 125362

Components

Deconvoluted Ion Set: A [7706.81]

Deconvoluted Ion Set: B [7728.82]

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7706.81	728985	100.00
B	7728.82	241240	33.09

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.213	BV	0.0704	2819.36597	604.33765	91.5304
2	10.408	VB	0.0712	260.88556	53.21116	8.4696
Totals :				3080.25153	657.54881	

ORN-10

ORN-11

ORN-12

ORN-13

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM\160203-AM73-AM79- DMT -OFF 2016-02-03 11-36-44\AM75-DMTOFF.D)   ES-API, Neg, Scan,

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\160203-AM73-AM79- DMT- DMT-OFF 2016-02-03 11-36-44\AM75-DMTOFF.D)

Deconvolution of Spectrum #   1 @ 10.213 - 10.313 min

*MSD1 SPC, time=10.213:10.313 of D:\DATA\ANNA\AM\160203-A   Max: 136466

Components

Deconvoluted Ion Set: A [7706.76]

Deconvoluted Ion Set: B [7728.46]

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7706.76	806390	100.00
B	7728.46	265719	32.95

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.220	VV	0.0742	3584.61060	744.03729	87.7367
2	10.421	VB	0.0702	501.03372	104.02650	12.2633

Totals :   4085.64432 848.06379

ORN-14

Additional Info : Peak(s) manually integrated
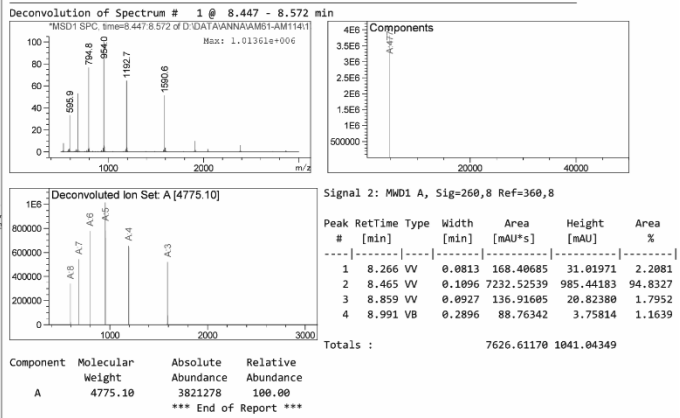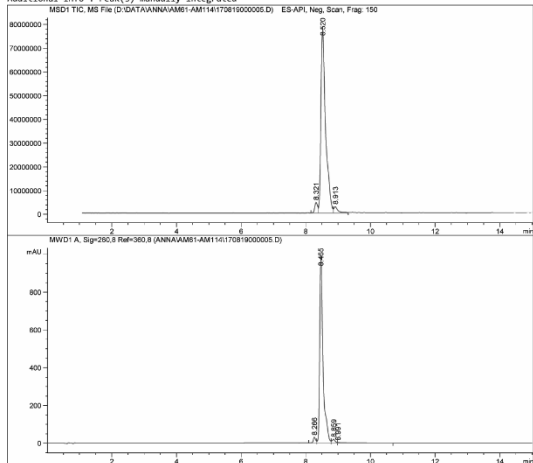
MSD1 TIC, MS File (D:\DATA\ANNA\AM\AM61-AM114\170819000005.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM\AM61-AM114\170819000005.D)

Deconvolution of Spectrum #   1 @ 8.447 - 8.572 min

*MSD1 SPC, time=8.447:8.572 of D:\DATA\ANNA\AM\AM114\1   Max: 1.01361e+006

Components

Deconvoluted Ion Set: A [4775.10]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.266	VV	0.0813	168.40685	31.01971	2.2081
2	8.465	VV	0.1096	7232.52539	985.44183	94.8327
3	8.859	VV	0.0927	136.91605	20.82380	1.7952
4	8.991	VB	0.2896	88.76342	3.75814	1.1639

Totals :   7626.61170 1041.04349

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4775.10	3821278	100.00

*** End of Report ***

ORN-15

MSD1 TIC, MS File (D:\DATA\ANNA\AM67-AM86 REPEAT\160205-AM67-AM86- DMT -OFF-2 2016-02-05 13-26-51\AM76-DMTOFF-2.D)   E

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM67-...60205-AM67-AM86- DMT -OFF-2 2016-02-05 13-26-51\AM76-DMTOFF-2.D)

Deconvolution of Spectrum #   1 @ 10.188 - 10.263 min

*MSD1 SPC, time=10.188:10.263 of D:\DATA\ANNA\AM67-AM86   Max: 32925

Components

Deconvoluted Ion Set: A [7682.53]

Deconvoluted Ion Set: B [7704.06]

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7682.53	185918	100.00
B	7704.06	71672	38.55

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.187	BV	0.0667	1045.47876	240.57066	100.0000

Totals :   1045.47876 240.57066

ORN-16

ORN-17

ORN-18

ORN-19

MSD1 TIC, MS File (D:\DATA\ANNA\AM109-AM118\160425-AM109-AM118-1 2016-04-25 15-26-40\AM109-DMTOFF.D)   ES-API, Neg, S

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM109-AM118\160425-AM109-AM118-1 2016-04-25 15-26-40\AM109-DMTOFF.D)

Deconvolution of Spectrum #   1 @ 10.014 - 10.064 min

Deconvoluted Ion Set: A [7755.37]

Deconvoluted Ion Set: B [7792.70]

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7755.37	64854	100.00
B	7792.70	25051	38.63

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	9.991	BV	0.0578	263.48761	70.20113	100.0000

Totals :   263.48761   70.20113

ORN-20

MSD1 TIC, MS File (D:\DATA\ANNA\AM109-AM116\160426_AM109-AM116-OFF 2016-04-26 11-54-39\AM110-DMTOFF.D)   ES-API, Neg,

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM109-AM116\160426_AM109-AM116-OFF 2016-04-26 11-54-39\AM110-DMTOFF.D)

Deconvolution of Spectrum #   1 @ 9.931 - 10.006 min

Deconvoluted Ion Set: A [7682.81]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	9.929	VV	0.0624	1588.13281	399.05032	98.2819
2	10.153	VB	0.0843	27.76203	4.88550	1.7181

Totals :   1615.89485   403.93583

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7682.81	355595	100.00

*** End of Report ***

ORN-21

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM109-AM118\160425-AM109-AM118-1 2016-04-25 15-26-40\AM111-DMTOFF-10UL.D)   ES-API, N

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM109...8\160425-AM109-AM118-1 2016-04-25 15-26-40\AM111-DMTOFF-10UL.D)

Deconvolution of Spectrum #   1 @ 7.374 - 7.473 min

Deconvoluted Ion Set: A [4813.92]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.386	VV	0.0669	1706.16077	391.05478	100.0000

Totals :   1706.16077   391.05478

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4813.92	333099	100.00

*** End of Report ***

ORN-22

MSD1 TIC, MS File (D:\DATA\ANNA\AM109-AM118\160425-AM109-AM118-1 2016-04-25 15-26-40\AM112-DMTOFF.D)   ES-API, Neg, S

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM109-AM118\160425-AM109-AM118-1 2016-04-25 15-26-40\AM112-DMTOFF.D)

Deconvolution of Spectrum #   1 @ 10.163 - 10.238 min

*MSD1 SPC, time=10.163:10.238 of D:\DATA\ANNA\AM109-AM11   Max: 8100

Components

Deconvoluted Ion Set: A [7755.04]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	10.133	BV	0.0588	164.60254	42.88432	100.0000

Totals :   164.60254   42.88432

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7755.04	32545	100.00

*** End of Report ***

ORN-23
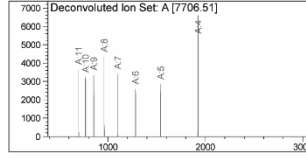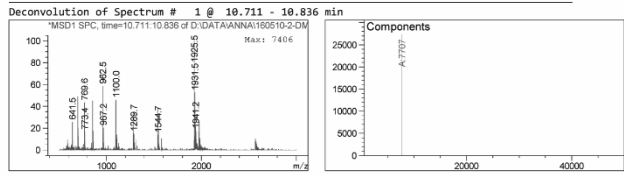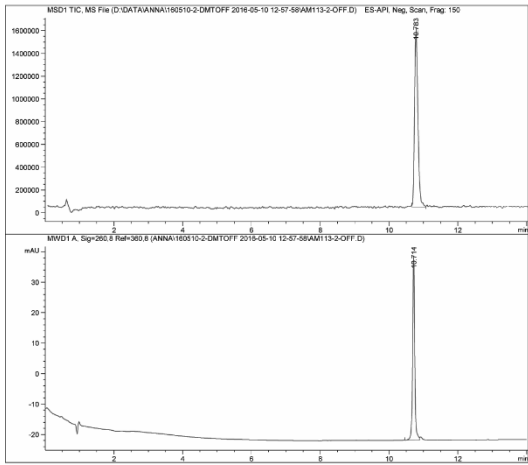
MSD1 TIC, MS File (D:\DATA\ANNA\160510-2-DMTOFF 2016-05-10 12-57-58\AM113-2-OFF.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\160510-2-DMTOFF 2016-05-10 12-57-58\AM113-2-OFF.D)

Deconvolution of Spectrum #   1 @ 10.711 - 10.836 min

*MSD1 SPC, time=10.711:10.836 of D:\DATA\ANNA\160510-2-DM   Max: 7406

Components

Deconvoluted Ion Set: A [7706.51]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	10.714	BV	0.0671	265.42978	60.62601	100.0000

Totals :   265.42978   60.62601

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7706.51	27321	100.00

*** End of Report ***

ORN-24
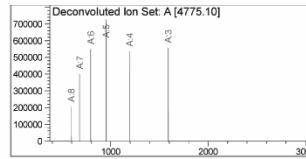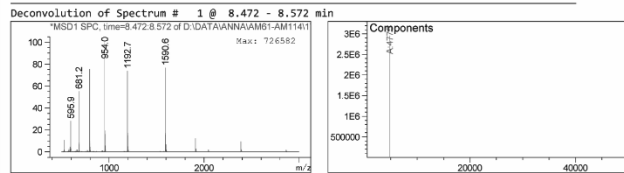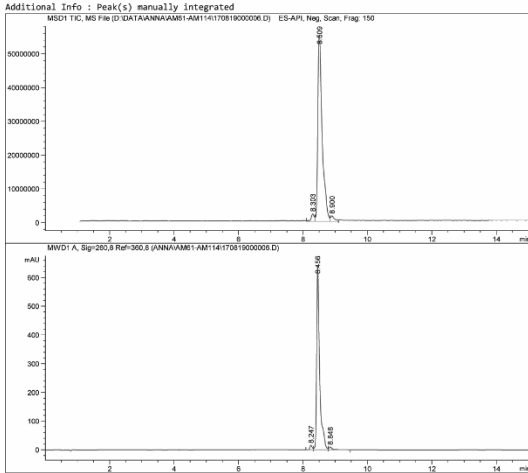
Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM114\170819000006.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM61-AM114\170819000006.D)

Deconvolution of Spectrum #   1 @ 8.472 - 8.572 min

*MSD1 SPC, time=8.472:8.572 of D:\DATA\ANNA\AM61-AM114\1   Max: 726582

Components

Deconvoluted Ion Set: A [4775.10]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

| Peak | RetTime | Type | Width | Area | Height | Area |
#	[min]		[min]	[mAU*s]	[mAU]	%
1	8.247	VV	0.0784	77.24522	14.92512	1.7380
2	8.456	VV	0.0998	4291.82715	643.50177	96.5669
3	8.848	VB	0.1076	75.33495	9.81018	1.6951

Totals :   4444.40733   668.23708

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4775.10	2952491	100.00

*** End of Report ***

ORN-25
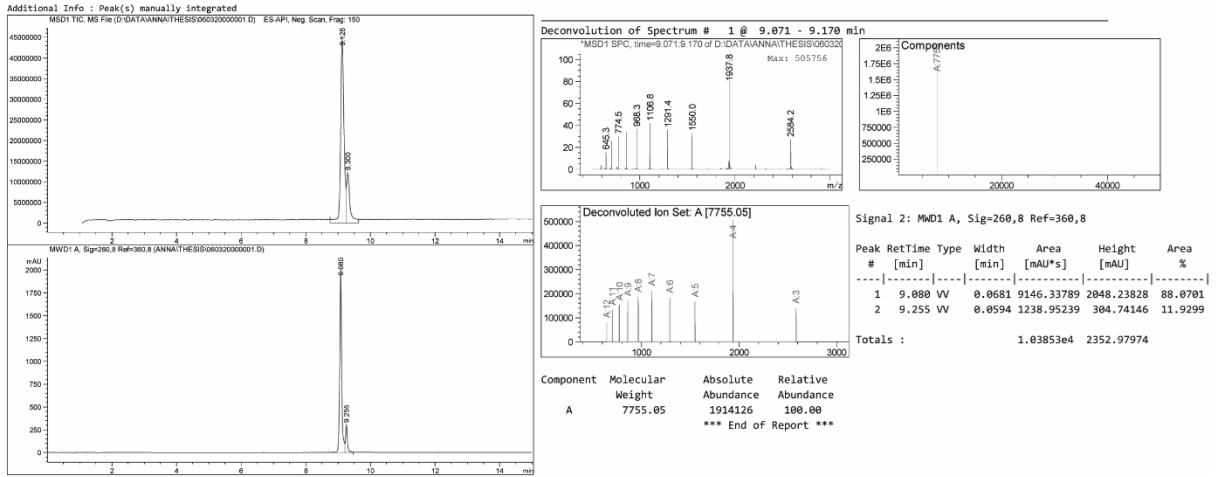
ORN-26
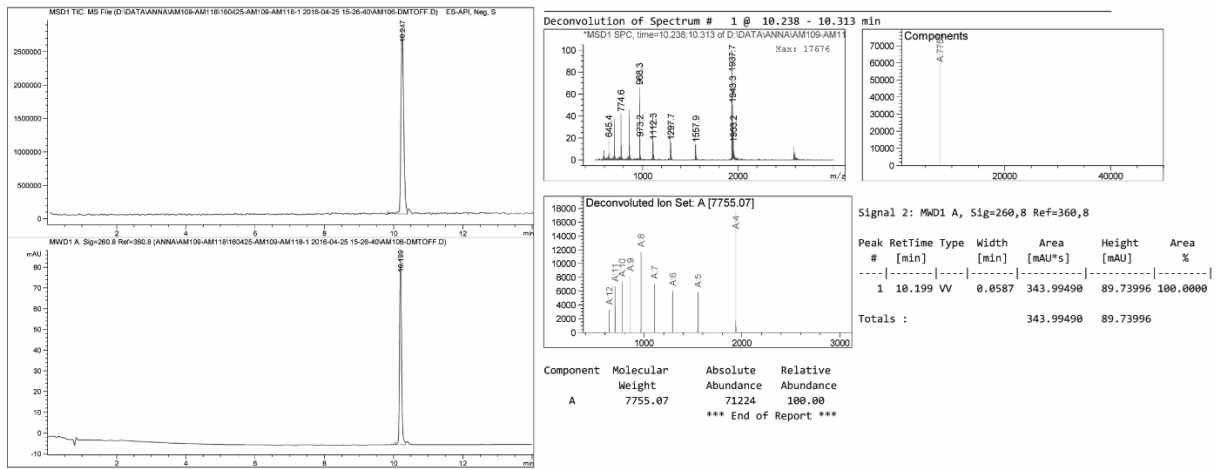
ORN-27
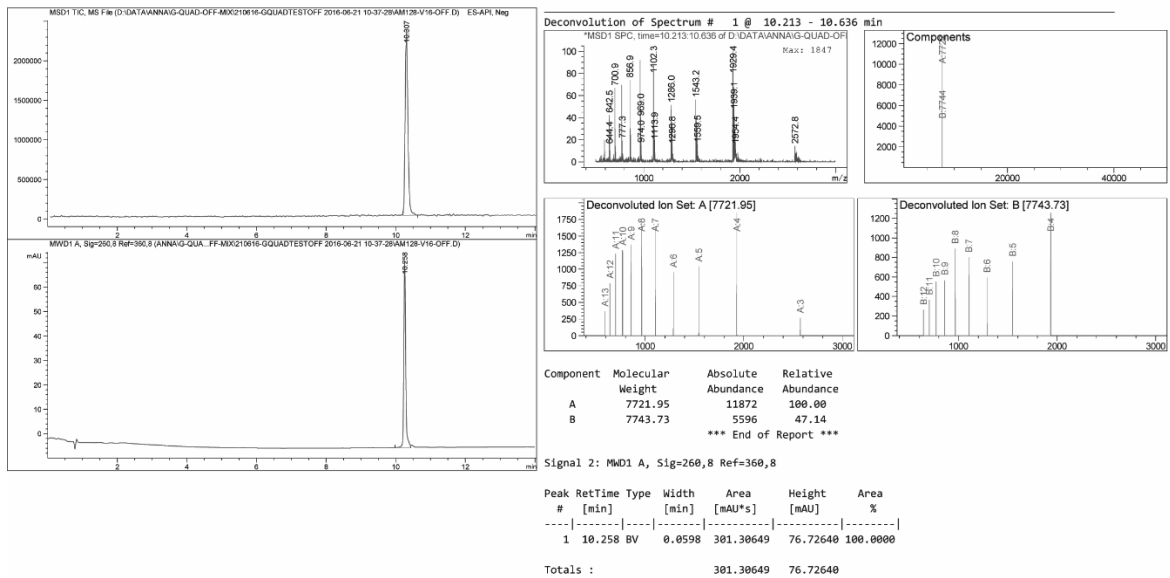
ORN-28
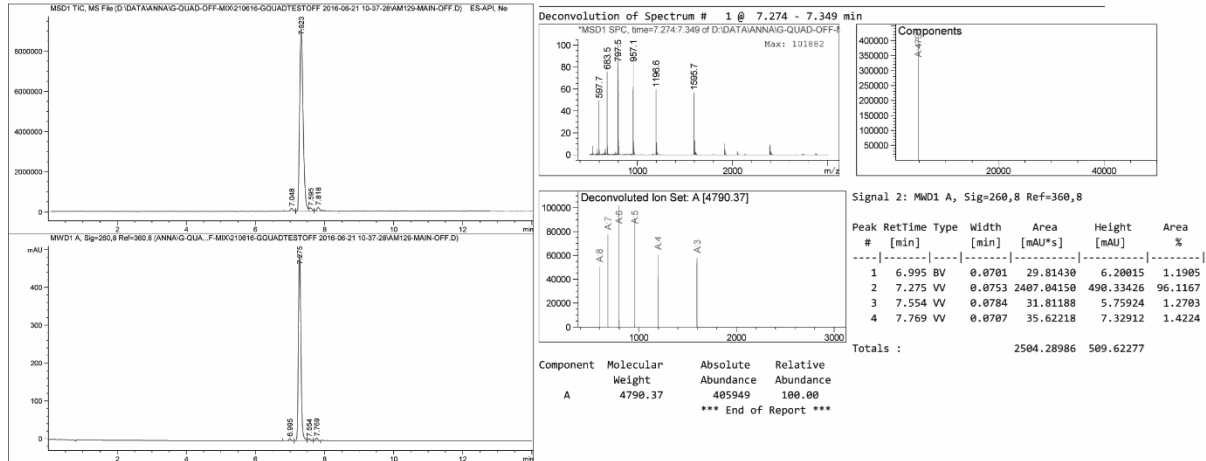
IX.1.1.2. ORN-29 to ORN-39

Table 26: Summary of the synthesized oligoribonucleotides (ORN-29 to ORN-39). **X** designates adenosine with the trioxsalen attached at the 2'-*O*-position of the ribose.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-29	UAAGGCACGCGGUG**X**AUGCC	20	6766.4	6765.5	0.01	>95	50 nmol and 1 µmol
2	ORN-30	UA**X**GGCACGCGGUGAAUGCC	20	6766.4	6765.4	0.01	>98	50 nmol
3	ORN-31	UCACCGCGUGCCUUA	15	4687.9	4687.1	0.02	>98	50 nmol
4	ORN-32	ACAGCGGACCUUGAU	15	4775.0	4774.0	0.02	>98	50 nmol
5	ORN-33	AGUGUUAGCAUCGUGUGUCUUG	22	7013.2	7012.4	0.01	>98	50 nmol and 1 µmol
6	ORN-34	UAUUCACUCUGUGCCUUAA	19	5920.6	5919.8	0.01	>98	50 nmol and 1 µmol
7	ORN-35	CAAUGUAGUGGCCUUG	16	5098.1	5097.2	0.02	>85	50 nmol and 1 µmol
8	ORN-36	UAAC**X**GUCUACAGCCAUGGUCG	22	7322.7	7321.7	0.01	>98	50 nmol
9	ORN-37	GGCUGUAGACUGUUA	15	4792.9	4792.2	0.02	>98	50 nmol
10	ORN-38	CUUUCGAUUGUUACU	15	4651.8	4651.1	0.02	>95	50 nmol
11	ORN-39	GCCAUUGGCUGUGGACUGUU	20	4785.9	4785.1	0.02	>95	50 nmol

ORN-29
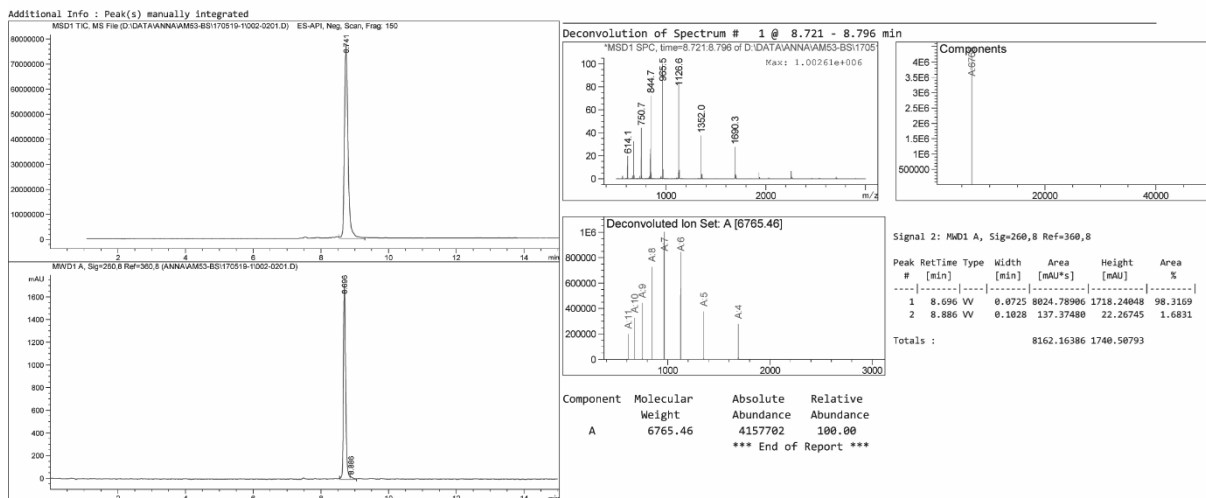
ORN-30

ORN-31

ORN-32

ORN-33

ORN-34

ORN-35

ORN-36

ORN-37

ORN-38

The mass was not deconvoluted

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	6.544	VV	0.1488	9121.45215	895.96588	97.0164
2	6.810	VV	0.0872	148.68414	23.71111	1.5814
3	7.151	VB	0.1060	131.83525	17.48309	1.4022
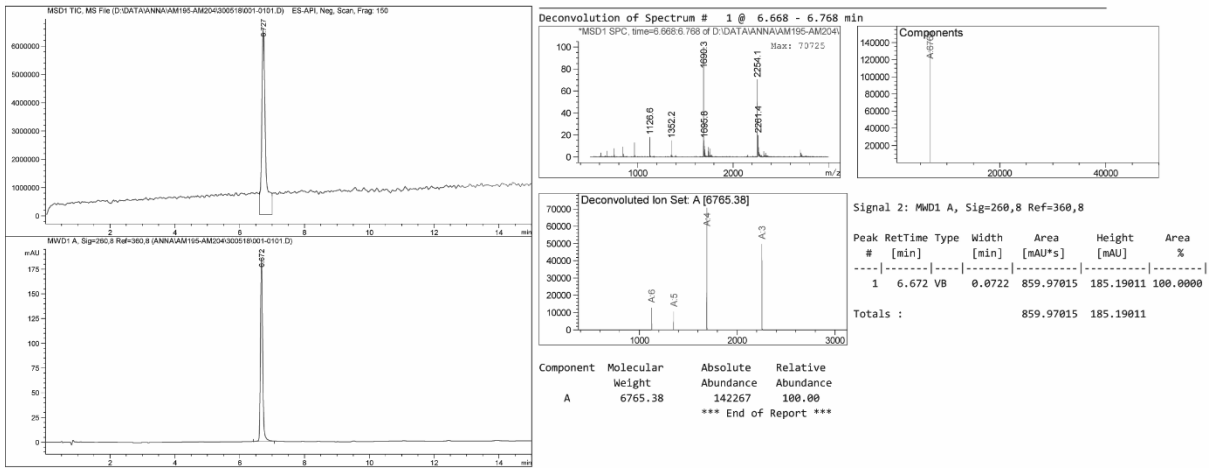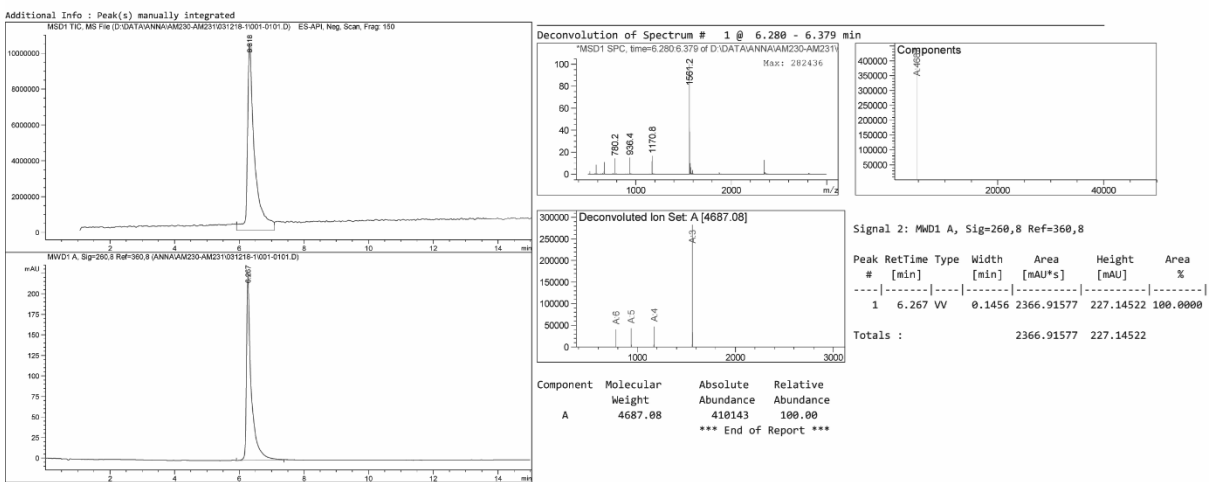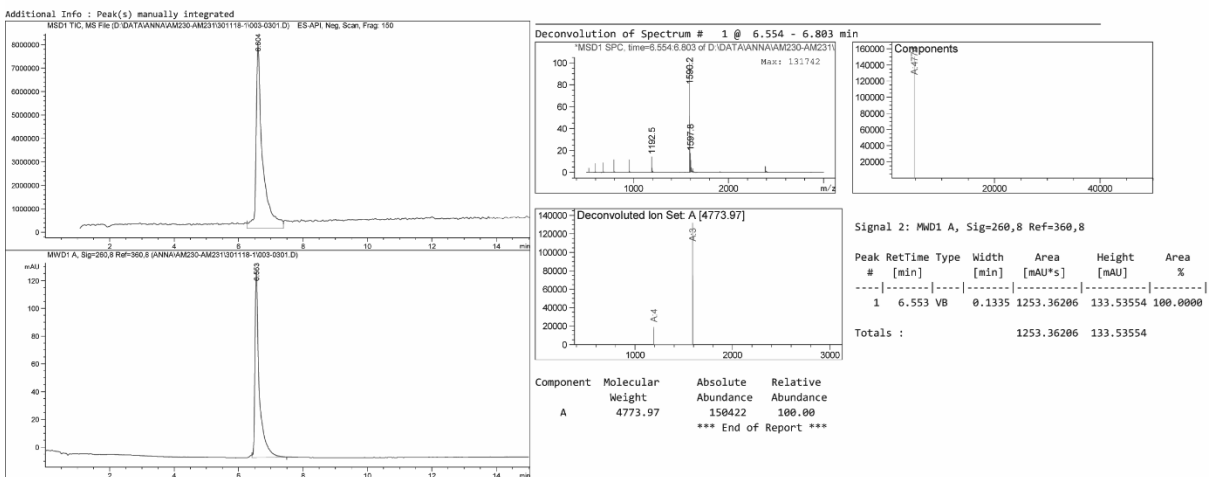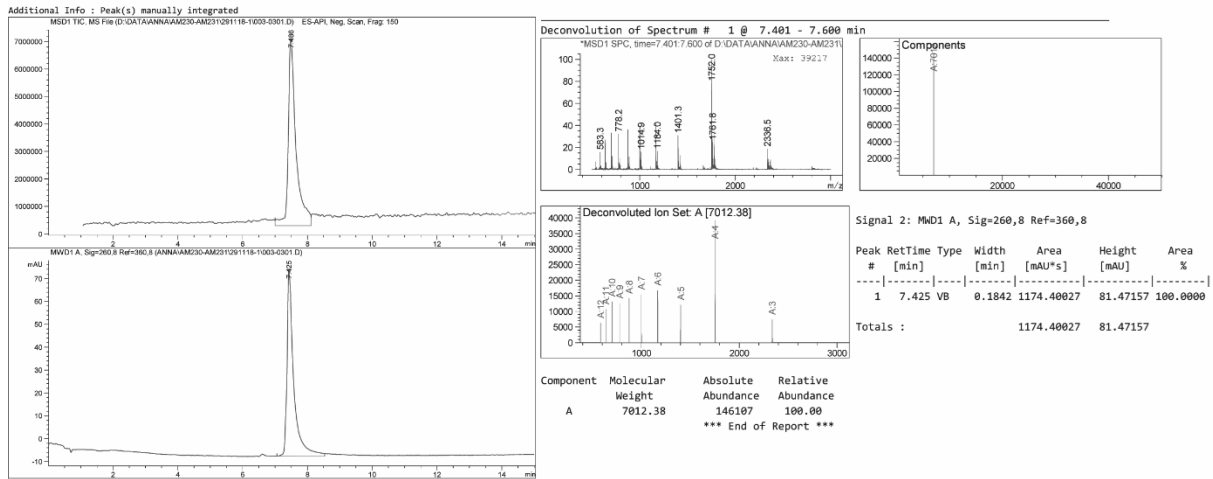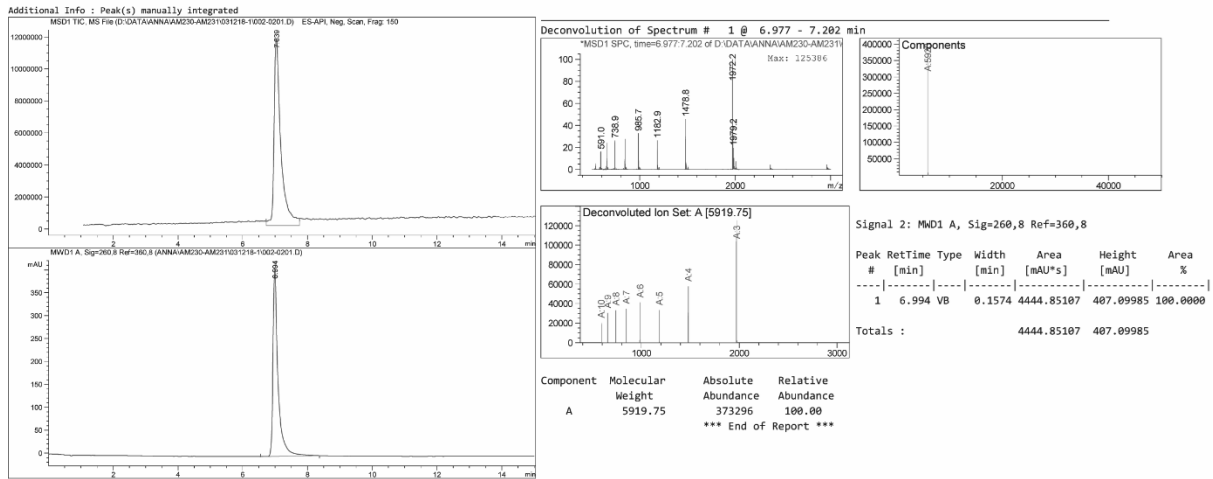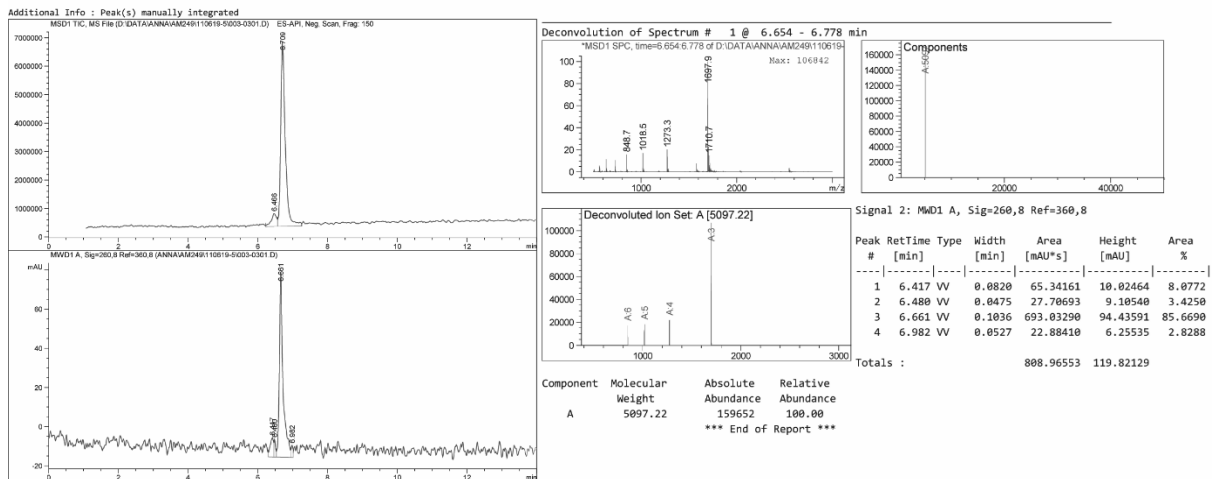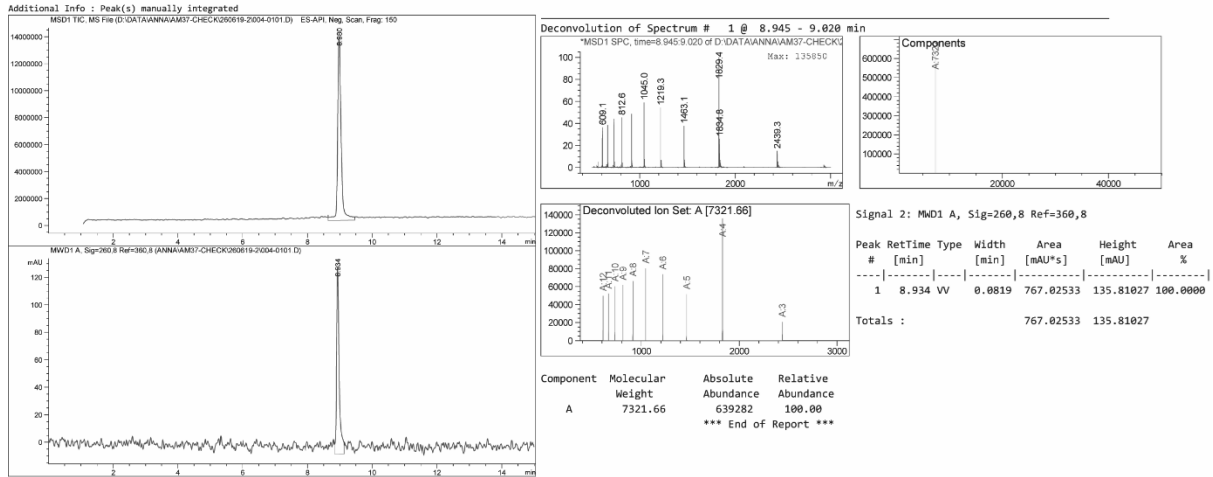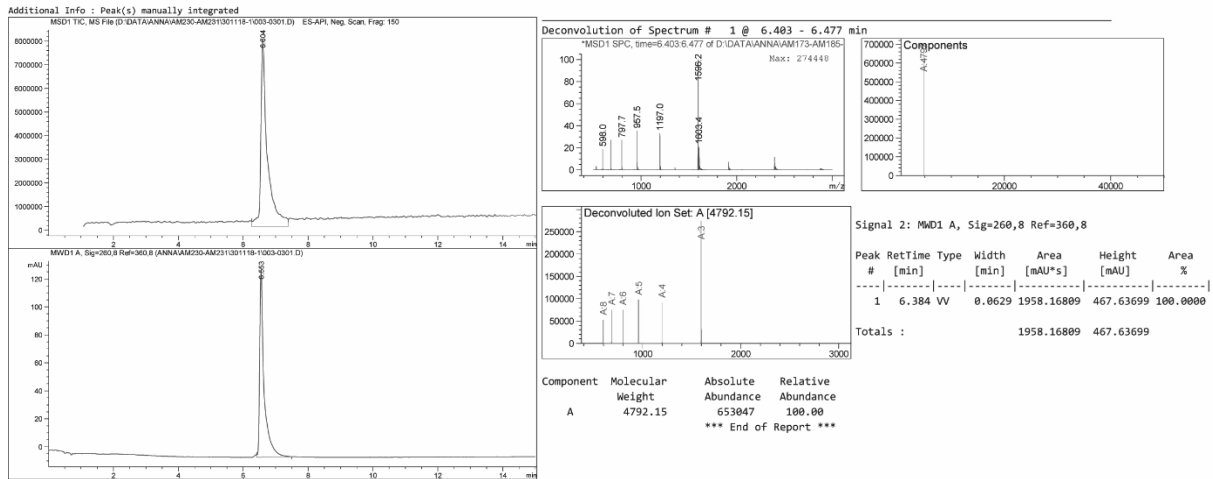
Totals : 9401.97154 937.16008
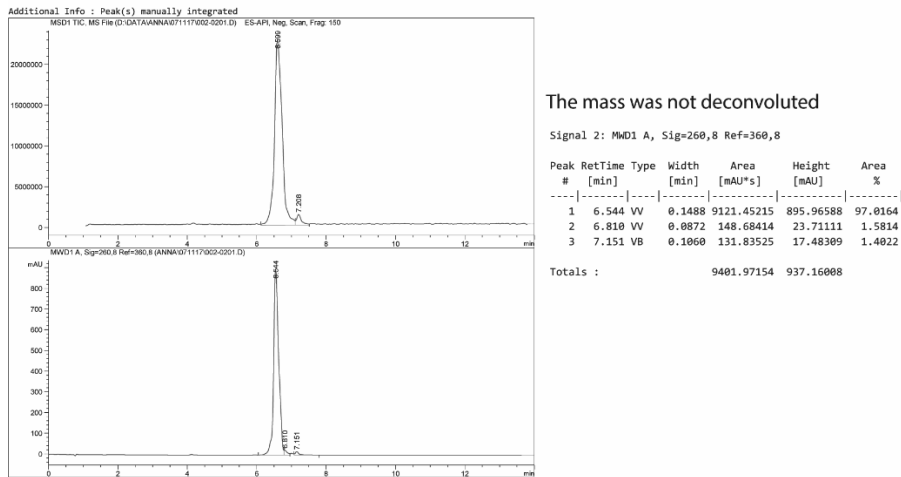
ORN-39
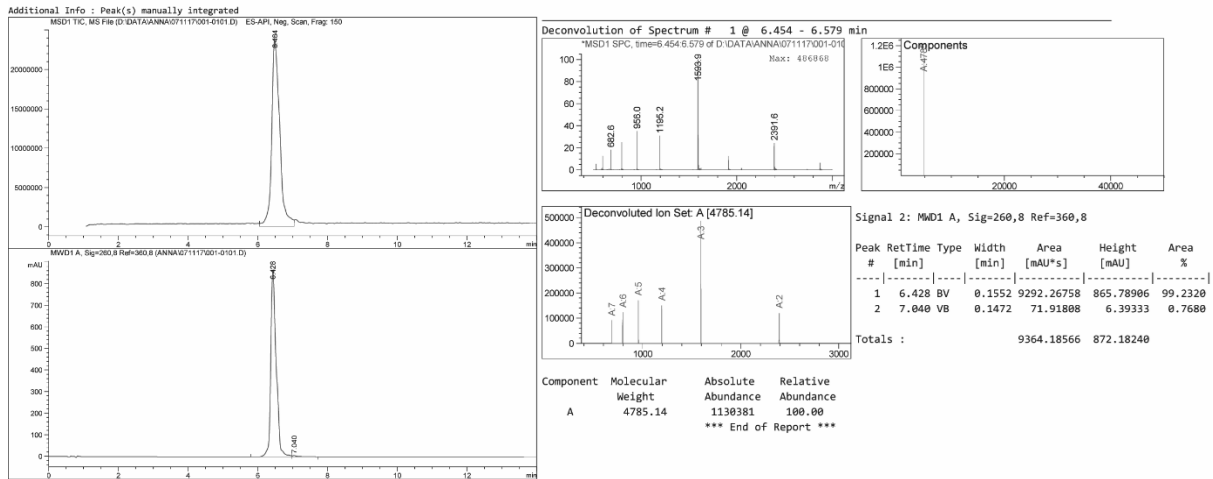
IX.1.1.3. ORN-40 to ORN-47

Table 27: Summary of the synthesized oligoribonucleotides (ORN-40 to ORN-47). **X** and **Y** designate adenosine and cytidine with the trioxsalen attached at the 2'-*O*-position of the ribose, respectively.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-40	UGAGGUAGU**X**GUUUGUACAGUU	22	7382.7	7381.2	0.02	>98	50 nmol
2	ORN-41	AACUGUACAAACUACUACCUCA	22	6913.3	6913.5	0.00	>98	50 nmol
3	ORN-42	UAAGGCACGCGGAGA**X**UGCC	20	6789.4	6788.5	0.01	>95	50 nmol
4	ORN-43	GGCAUUCUCCGCGUG	15	4743.9	4743.1	0.02	>95	50 nmol
5	ORN-44	UAC**Y**CUGUAGAACCGAAUUUGUG	23	7629.8	7628.9	0.01	>90	50 nmol
6	ORN-45	CGGUUCUACAGGGUA	15	4791.9	4791.2	0.01	>90	50 nmol
7	ORN-46	UAG**Y**UUAUCAGACUGAUGUUGA	22	7325.7	7324.7	0.01	>95	50 nmol
8	ORN-47	CAGUCUGAUAAGCUA	15	4759.9	4759.2	0.01	>90	50 nmol

ORN-40
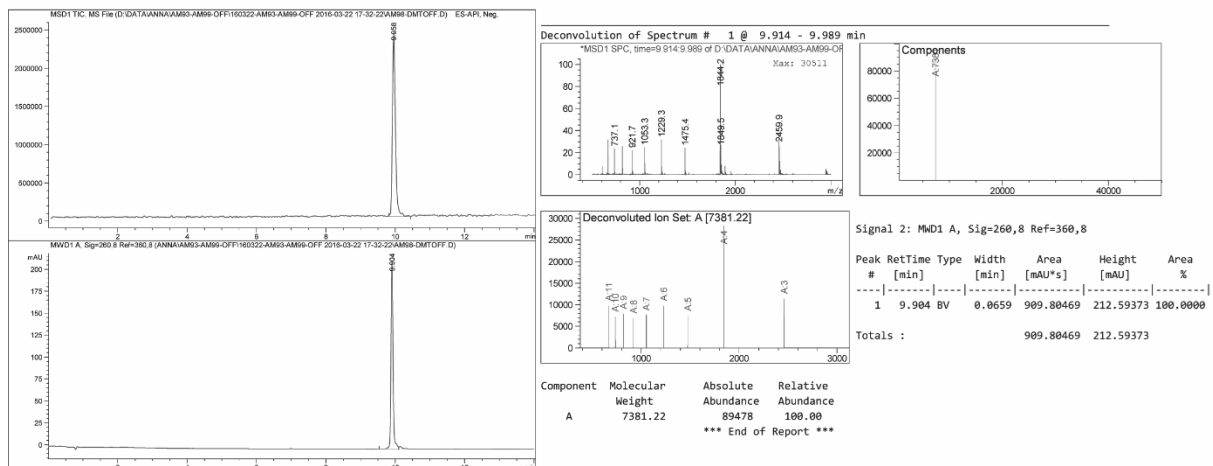
ORN-41
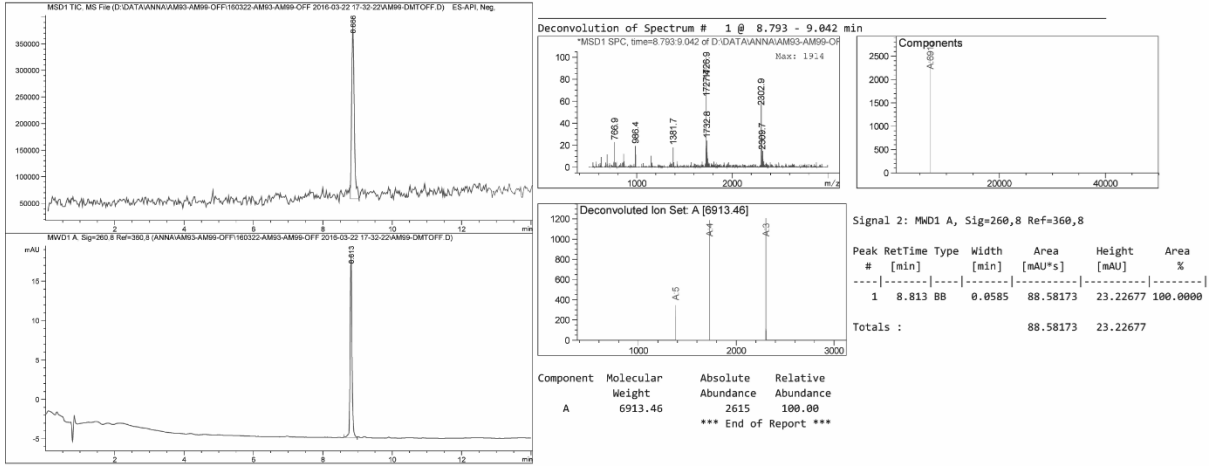
ORN-42
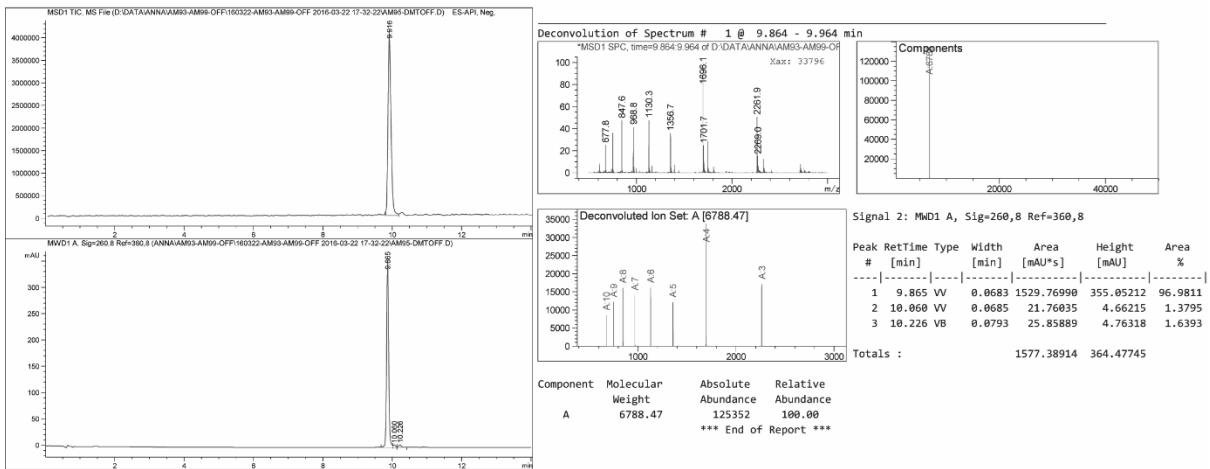
ORN-43
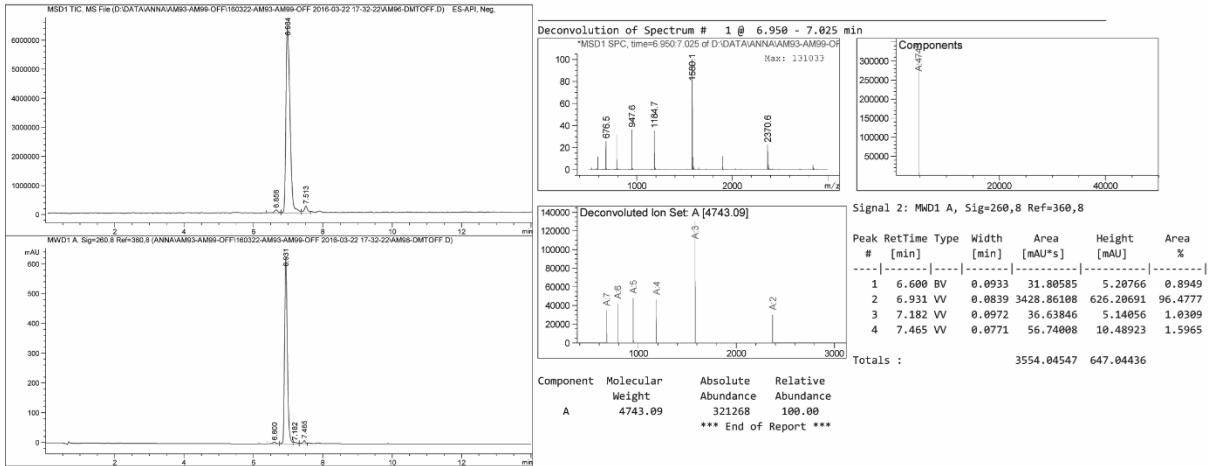
ORN-44
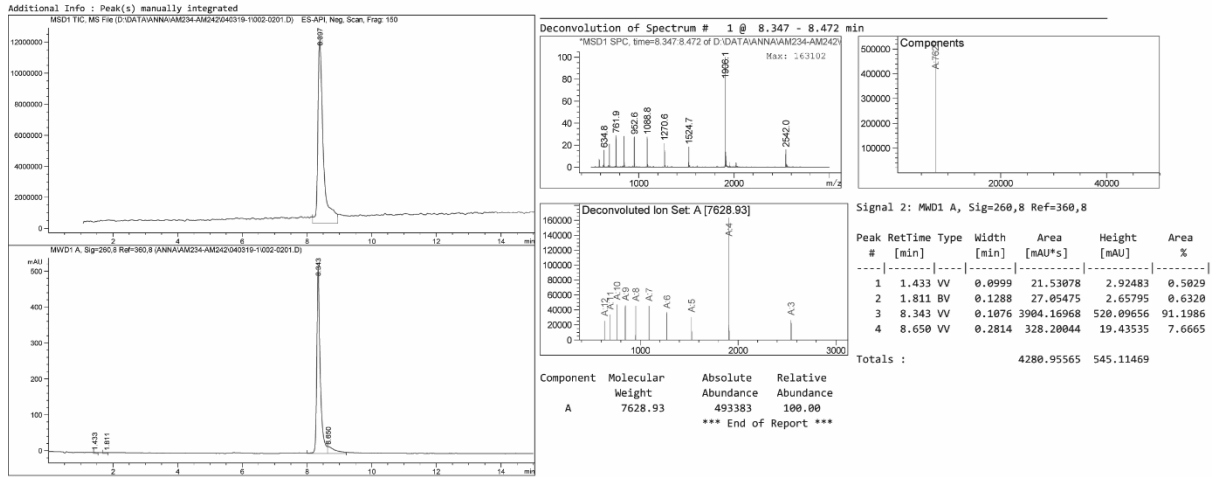
ORN-45
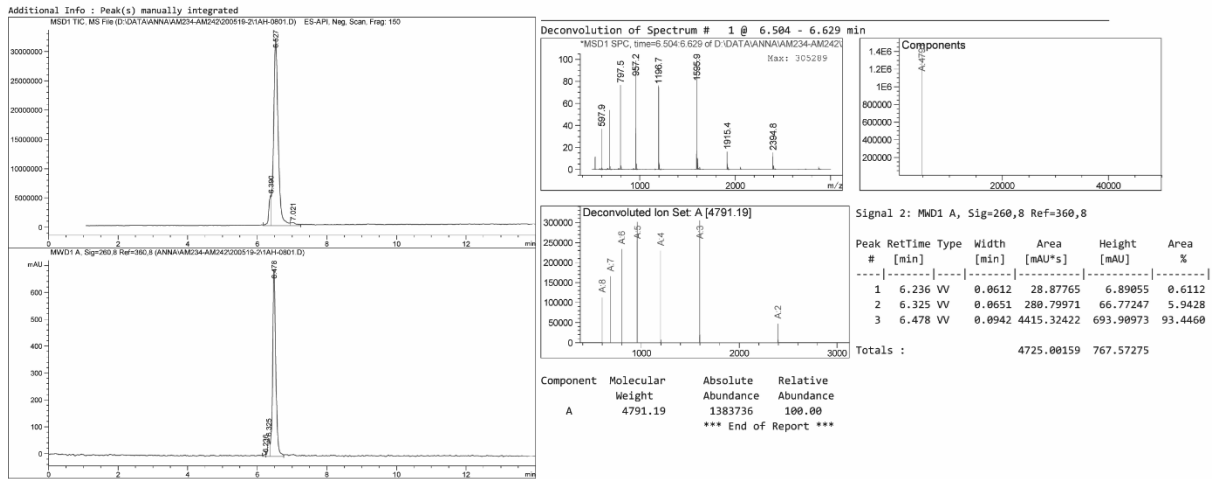
ORN-46
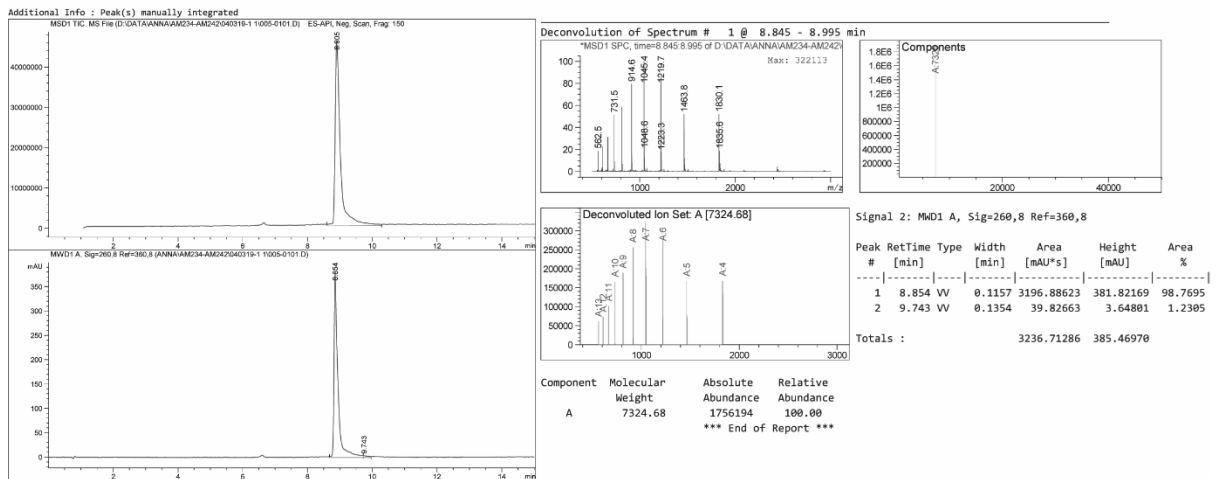
ORN-47
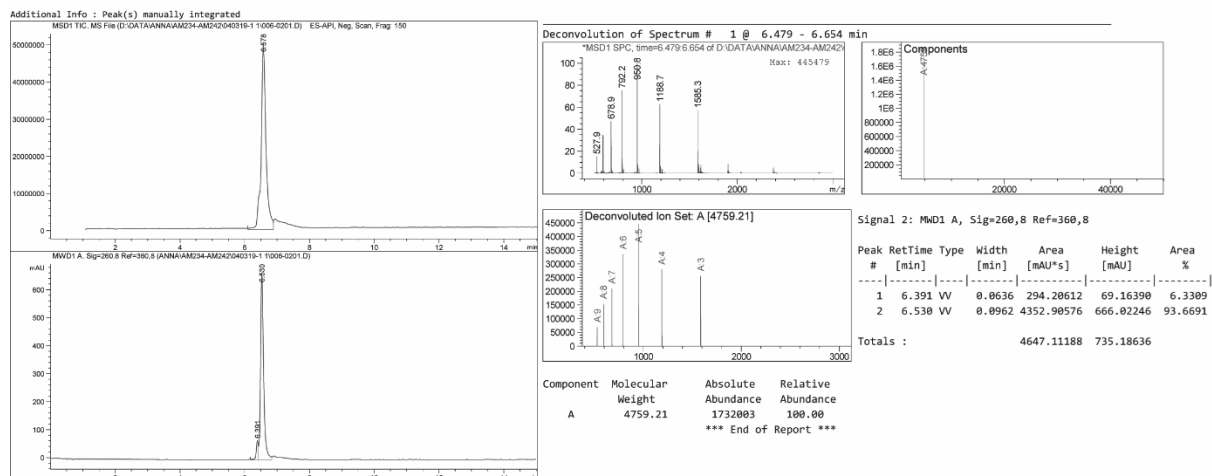
IX.1.2. Chapter III

IX.1.2.1. ORN-48 to ORN-52

Table 28: Summary of the synthesized oligoribonucleotides (ORN-48 to ORN-52). **Y**, **Z** and **W** designate diazirine attached at the 2'-*O*-position of adenosine, uridine or at the C^5-position of uridine, respectively.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-48	AAA**Y**GUGCUUACAGUGCAGGUAG	23	7713.8	7713.0	0.01	>98	50 nmol
2	ORN-49	AAAAG**Z**GCUUACAGUGCAGGUAG	23	7713.8	7713.1	0.01	>98	50 nmol
3	ORN-50	AAAAG**W**GCUUACAGUGCAGGUAG	23	7699.8	7698.9	0.01	>98	50 nmol
4	ORN-51	GUAAGCACUUCUUAC	15	4696.9	4696.2	0.02	>98	50 nmol
5	ORN-52	AAAAGUGCUUACAGUGCAGGUAG	23	7434.6	7433.6	0.01	>98	50 nmol

ORN-48

ORN-49

Additional Info : Peak(s) manually integrated

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.495	VB	0.0596	2877.46436	736.44000	100.0000

Totals :   2877.46436   736.44000

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7713.08	3362445	100.00

*** End of Report ***

ORN-50

Additional Info : Peak(s) manually integrated

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.542	VB	0.0609	3609.87866	898.79553	100.0000

Totals :   3609.87866   898.79553

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7698.92	3376022	100.00

*** End of Report ***

ORN-51

Additional Info : Peak(s) manually integrated

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	6.072	BB	0.0720	1856.35474	386.88232	100.0000

Totals :   1856.35474   386.88232

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4696.18	1356752	100.00

*** End of Report ***

ORN-52

IX.1.3. Chapter IV

IX.1.3.1. ORN-53 to ORN-56

Table 29: Summary of the synthesized oligoribonucleotides (ORN-53 to ORN-56). **X**: cytidine with the trioxsalen at the 2'-*O*-position of the ribose; **Y**: cytidine with the trioxsalen at the N^4-position of the nucleobase. PEG0-PEG4 indicates how many ethylene glycol units are present within the linker.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-53	UAAGGCACGCGGUGAAUGCC	20	6445.0	6444.7	0.00	>98	50 nmol
2	ORN-54a	UAAGG**X**$_{PEG0}$ACGCGGUGAAUGCC	20	6766.4	6765.2	0.02	>98	1 µmol
3	ORN-54b	UAAGG**X**$_{PEG1}$ACGCGGUGAAUGCC	20	6810.4	6809.3	0.02	>98	1 µmol
4	ORN-54c	UAAGG**X**$_{PEG2}$ACGCGGUGAAUGCC	20	6854.5	6853.4	0.02	>98	1 µmol
5	ORN-54d	UAAGG**X**$_{PEG3}$ACGCGGUGAAUGCC	20	6898.5	6897.4	0.02	>98	1 µmol
6	ORN-54e	UAAGG**X**$_{PEG4}$ACGCGGUGAAUGCC	20	6942.6	6941.5	0.02	>98	1 µmol
7	ORN-55a	UAAGGCA**X**$_{PEG0}$GCGGUGAAUGCC	20	6766.4	6765.3	0.02	>98	1 µmol
8	ORN-55b	UAAGGCA**X**$_{PEG1}$GCGGUGAAUGCC	20	6810.4	6809.4	0.01	>98	1 µmol
9	ORN-55c	UAAGGCA**X**$_{PEG2}$GCGGUGAAUGCC	20	6854.5	6853.4	0.02	>98	1 µmol
10	ORN-55d	UAAGGCA**X**$_{PEG3}$GCGGUGAAUGCC	20	6898.5	6897.4	0.02	>98	1 µmol
11	ORN-55e	UAAGGCA**X**$_{PEG4}$GCGGUGAAUGCC	20	6942.6	6941.6	0.01	>98	1 µmol
12	ORN-56a	UAAGG**Y**$_{PEG0}$ACGCGGUGAAUGCC	20	6685.3	6684.2	0.02	>98	1 µmol
13	ORN-56b	UAAGG**Y**$_{PEG1}$ACGCGGUGAAUGCC	20	6729.3	6728.3	0.01	>98	1 µmol
14	ORN-56c	UAAGG**Y**$_{PEG2}$ACGCGGUGAAUGCC	20	6773.4	6772.3	0.02	>98	1 µmol
15	ORN-56d	UAAGG**Y**$_{PEG3}$ACGCGGUGAAUGCC	20	6817.4	6816.4	0.01	>98	1 µmol
16	ORN-56e	UAAGG**Y**$_{PEG4}$ACGCGGUGAAUGCC	20	6861.5	6860.4	0.02	>98	1 µmol

ORN-53

ORN-54a

ORN-54b

ORN-54c

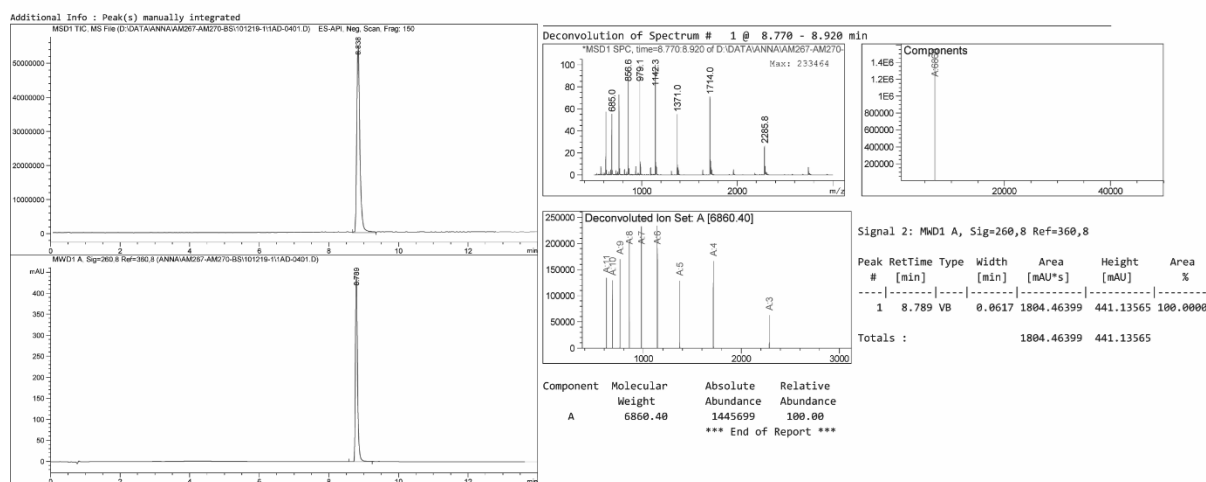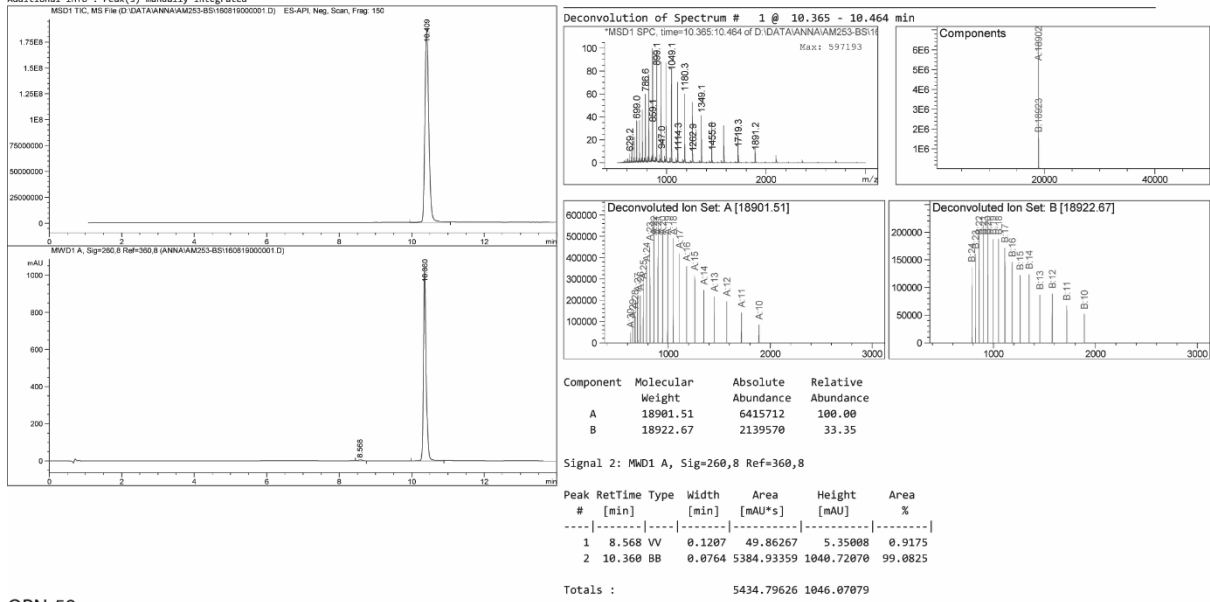Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 8.097 - 8.197 min

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.094	VB	0.0687	6785.69482	1501.16394	100.0000

Totals :   6785.69482 1501.16394

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	6853.39	5638943	100.00

*** End of Report ***

ORN-54d

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 8.446 - 8.546 min

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	6897.39	4321139	100.00
B	6918.98	380294	8.80

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.141	VV	0.0783	21.33633	3.86980	0.4689
2	8.443	VV	0.0676	4528.90430	1023.01953	99.5311

Totals :   4550.24063 1026.88933

ORN-54e

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 8.895 - 8.969 min

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.864	VB	0.0688	6158.58154	1361.52734	100.0000

Totals :   6158.58154 1361.52734

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	6941.48	5741473	100.00

*** End of Report ***

ORN-55a

ORN-55b

ORN-55c

ORN-55d

ORN-55e

ORN-56a

ORN-56b

ORN-56c

ORN-56d

ORN-56e

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM267-AM270-BS\101219-1\1AD-0401.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM267-AM270-BS\101219-1\1AD-0401.D)

Deconvolution of Spectrum #   1 @ 8.770 - 8.920 min

*MSD1 SPC, time=8.770:8.920 of D:\DATA\ANNA\AM267-AM270-   Max: 233464

Components

Deconvoluted Ion Set: A [6860.40]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

```
Peak RetTime Type  Width     Area      Height     Area
 #   [min]        [min]    [mAU*s]     [mAU]       %
---|-------|----|-------|----------|----------|--------|
 1   8.789 VB   0.0617 1804.46399  441.13565 100.0000

Totals :                  1804.46399  441.13565
```

```
Component  Molecular    Absolute    Relative
            Weight      Abundance   Abundance
  A         6860.40     1445699     100.00
          *** End of Report ***
```

IX.1.4. Chapter V

IX.1.4.1. ORN-57 to ORN-75

Table 30: Summary of the synthesized oligoribonucleotides (ORN-57 to ORN-75). **X:** adenosine with the trioxsalen attached at the 2'-*O*-position of the ribose. **V, Y, W**: biotin introduced at the 2'-*O*-position of cytidine, uridine or adenosine, respectively. **Z**: biotin dT phosphoramidite.

Entry	ORN	Sequence (5' to 3')	Length	Mass calc.	Mass found	Δm (%)	UV Purity [%]	Synthesis scale
1	ORN-57	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCA UACAAUUAAGGCACGCGGUGAAUGCC	59	18903.4	18901.5	0.01	>95	1 µmol
2	ORN-58	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCA UACAAUUAAGG**V**ACGCGGUG**X**AUGCC	59	19707.4	19706.1	0.01	>90	1 µmol
3	ORN-59	CGUGUUCACAGCGGACCUUGAUUUAAAUGUCCA UACAAUUA**X**GGCACG**V**GGUGAAUGCC	59	19707.4	19707.4	0.00	>95	1 µmol
4	ORN-60	ACCGUGGCUUUCGAUUGUUACUGUGGGAACUG GAGGUAAC**X**GUCUACAGC**V**AUGGUCG	58	19467.2	19464.5	0.01	>95	50 nmol
5	ORN-61	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUC CAACUGACUCCUACAUAUUAGCAUUAACA	61	19383.6	19382.9	0.00	>95	1 µmol
6	ORN-62	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUC CAACUGACUCCUAC**X**UAUUAGCAU**Y**AACA	61	20187.6	20186.3	0.01	>95	1 µmol
7	ORN-63	UUAAUGCUAAUCGUGAUAGGGGUUUUUGCCUC CAACUGACUCCUAC**X**UAUUAGCAU**Z**AACA	61	20083.5	20082.5	0.00	>90	1 µmol
8	ORN-64	UUA**X**UGCUAAUCGUGAUAGGGGUU	24	8017.1	8016.1	0.01	>85	50 nmol
9	ORN-65	UGCACGCAGCAUUAA	15	4759.0	4758.3	0.02	>95	50 nmol
10	ORN-66	CUCCUAC**X**UAUUAGCAUUAACA	22	7212.6	7211.7	0.01	>85	50 nmol
11	ORN-67	GCUAAUAUGUAGGAG	15	4840.0	4839.3	0.01	>85	50 nmol
12	ORN-68	AAA**X**GUGCUUACAGUGC**W**GGUAGCUUUUUGAG AUCUACUGCAAUGUAAGCACUUCUUAC	59	19655.3	19654.0	0.01	>85	50 nmol
13	ORN-69	AAA**X**GUGCUUACAGUGC**W**GGUAG	23	8238.6	8237.5	0.01	>80	50 nmol
14	ORN-70	UAA**X**GUGCUGACAGUGC**W**GAUAGUGGUCCUCU CCGUGCUACCGCACUGUGGGUACUUGCUGC	62	20664.9	20663.7	0.01	>85	50 nmol
15	ORN-71	UAA**X**GUGCUGACAGUGC**W**GAU	21	7564.2	7563.1	0.01	>80	50 nmol
16	ORN-72	CUGCACUGUCAGCACUUUAUG	22	6610.0	6608.8	0.02	>95	50 nmol
17	ORN-73	CCGCACUGUGGGUACUUGCUG	21	6681.0	6680.0	0.02	>95	50 nmol
18	ORN-74	ACCUGCACUGUAAGCACUUUUAC	23	7228.4	7227.4	0.01	>95	50 nmol
19	ORN-75	CUGCAAUGUAAGCACUUCUUAC	22	6923.2	6922.2	0.01	>95	50 nmol

ORN-57

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	18901.51	6415712	100.00
B	18922.67	2139570	33.35

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.568	VV	0.1207	49.86267	5.35008	0.9175
2	10.360	BB	0.0764	5384.93359	1040.72070	99.0825

Totals : 5434.79626 1046.07079

ORN-58

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	19706.10	336302	100.00
B	19727.64	155328	46.19

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.253	BV	0.0822	28.68416	4.90342	7.7253
2	8.359	VB	0.0745	342.61847	68.35164	92.2747

Totals : 371.30263 73.25506

ORN-59

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.299	VV	0.0555	26.21266	7.02226	2.7685
2	8.395	VB	0.0629	920.62122	228.99434	97.2315

Totals : 946.83387 236.01660

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	19707.39	1240225	100.00

*** End of Report ***

ORN-60

Deconvolution of Spectrum #   1 @ 10.902 - 11.002 min

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	10.916	VB	0.0747	1845.57593	367.09244	100.0000
Totals :				1845.57593	367.09244	

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	19464.52	135454	100.00

*** End of Report ***

ORN-61

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 8.945 - 9.070 min

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	19382.86	948251	100.00
B	19403.82	297208	31.34

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	8.940	VV	0.0775	1848.11011	350.83829	96.6752
2	9.230	VV	0.0626	36.40510	8.39545	1.9044
3	9.343	VV	0.0593	27.15487	6.17220	1.4205
Totals :				1911.67008	365.40594	

ORN-62

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 9.593 - 9.692 min

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	20186.29	1488587	100.00
B	20207.48	552224	37.10

*** End of Report ***

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	9.598	VV	0.0731	1780.32129	351.56985	97.6500
2	9.880	VV	0.0515	42.84434	12.65671	2.3500
Totals :				1823.16563	364.22656	

ORN-63

ORN-64

ORN-65

ORN-66

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM234-AM242\200519-2\1AC-0301.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM234-AM242\200519-2\1AC-0301.D)

Deconvolution of Spectrum #   1 @ 9.095 - 9.169 min

*MSD1 SPC, time=9.095:9.169 of D:\DATA\ANNA\AM234-AM242\   Max: 101969

Components

Deconvoluted Ion Set: A [7211.68]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	2.074	VV	0.1157	24.19696	2.62093	2.8955
2	5.269	VV	0.0853	23.36159	3.52373	2.7955
3	8.799	VV	0.0674	26.73889	5.63516	3.1997
4	9.075	VV	0.0690	741.04065	163.20477	88.6761
5	9.340	VV	0.0586	20.33270	4.88585	2.4331

Totals :   835.67078 179.87045

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7211.68	554889	100.00

*** End of Report ***

ORN-67

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\L--246-026-AM250-BS-AM251-BS\220519-2\1CA-1001.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\L--246-026-AM250-BS-AM251-BS\220519-2\1CA-1001.D)
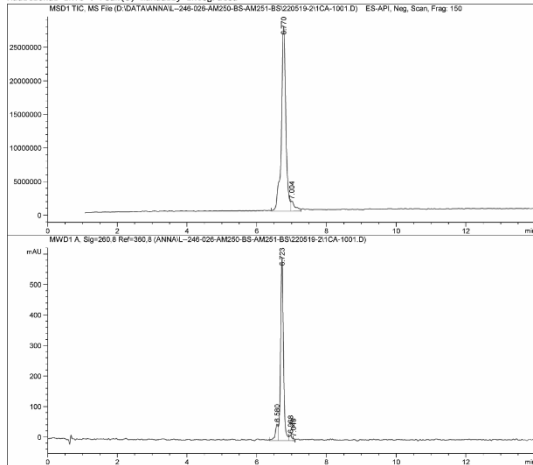
Deconvolution of Spectrum #   1 @ 6.728 - 6.878 min
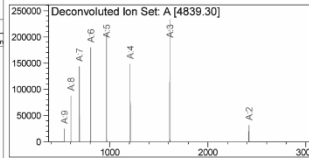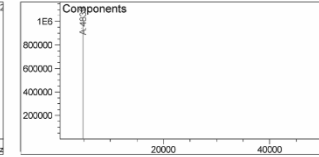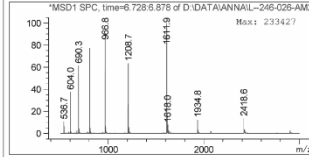
*MSD1 SPC, time=6.728:6.878 of D:\DATA\ANNA\L--246-026-AM2   Max: 233427

Components

Deconvoluted Ion Set: A [4839.30]

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	6.580	VV	0.0834	305.87042	54.54379	8.0385
2	6.723	VV	0.0853	3384.28784	604.32507	88.9417
3	6.968	VV	0.0780	93.30985	17.55908	2.4523
4	7.049	VV	0.0486	21.59431	7.40259	0.5675

Totals :   3805.06243 683.83053

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	4839.30	1040167	100.00

*** End of Report ***

ORN-68

Additional Info : Peak(s) manually integrated

MSD1 TIC, MS File (D:\DATA\ANNA\AM205-AM210 FINAL\100718-1\001-0101.D)   ES-API, Neg, Scan, Frag: 150

MWD1 A, Sig=260,8 Ref=360,8 (ANNA\AM205-AM210 FINAL\100718-1\001-0101.D)

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.306	BV	0.0684	533.27399	114.46214	12.8155
2	7.401	VB	0.0830	3627.88159	631.25720	87.1845

Totals :   4161.15558 745.71935
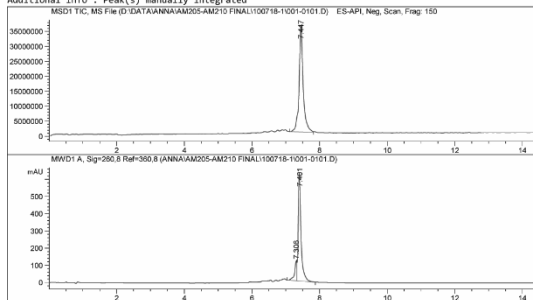
Deconvolution of Spectrum #   1 @ 7.391 - 7.490 min

*MSD1 SPC, time=7.391:7.490 of D:\DATA\ANNA\AM205-AM210   Max: 67187

Components

Deconvoluted Ion Set: A [19653.99]

Deconvoluted Ion Set: B [19683.97]

Deconvoluted Ion Set: C [19712.42]

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	19653.99	562808	100.00
B	19683.97	296162	52.62
C	19712.42	95245	16.92

*** End of Report ***

ORN-69

Additional Info : Peak(s) manually integrated
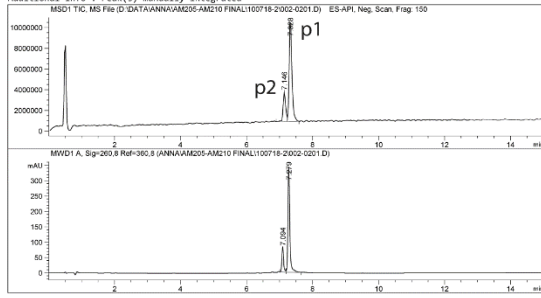
Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.094	BV	0.0578	298.94809	79.59460	18.1036
2	7.279	VV	0.0601	1352.36707	342.46896	81.8964

Totals :   1651.31516   422.06356

p1 Deconvolution of Spectrum # 1 @ 7.291 - 7.391 min
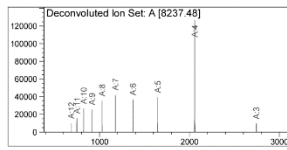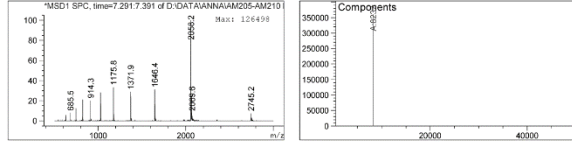
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	8237.48	360665	100.00

*** End of Report ***

p2 Deconvolution of Spectrum # 1 @ 7.117 - 7.167 min
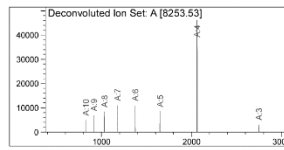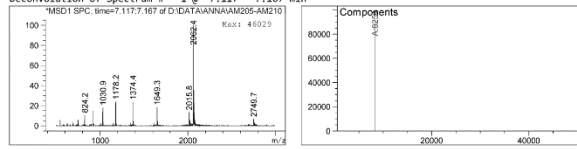
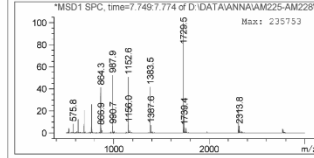| Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	8253.53	92256	100.00

*** End of Report ***

ORN-70

Additional Info : Peak(s) manually integrated
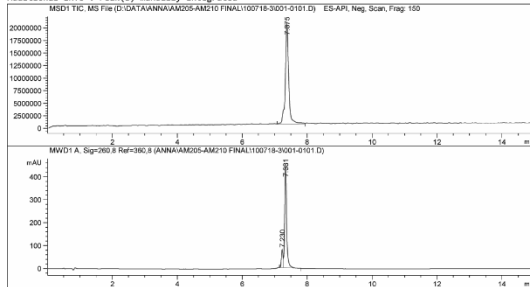
Deconvolution of Spectrum # 1 @ 7.316 - 7.440 min
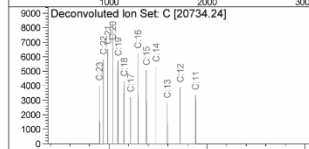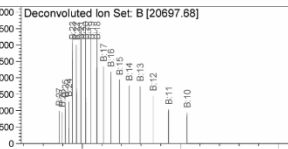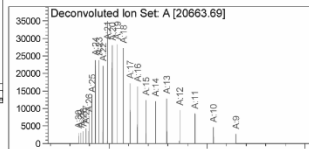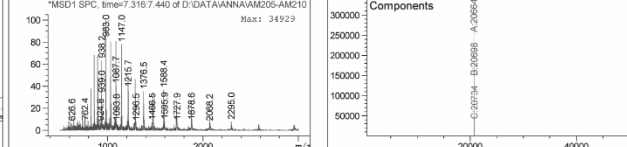
Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.230	BV	0.0520	268.67047	78.45155	11.9345
2	7.331	VB	0.0651	1982.53345	452.40854	88.0655

Totals :   2251.20392   530.86009

Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	20663.69	303686	100.00
B	20697.68	176156	58.01
C	20734.24	48879	16.10

*** End of Report ***

ORN-71

Additional Info : Peak(s) manually integrated
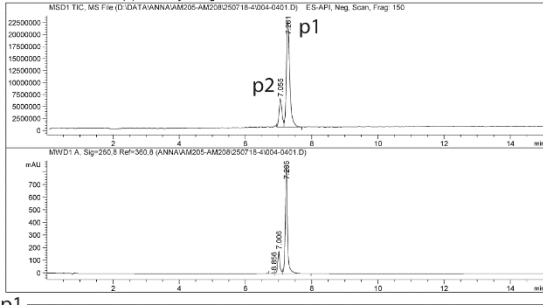
Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	6.856	BV	0.0608	34.66503	8.28440	0.7008
2	7.006	VV	0.0700	818.52240	176.81244	16.5466
3	7.235	VB	0.0719	4093.57153	854.29773	82.7526

Totals :   4946.75896 1039.39457

p1

Deconvolution of Spectrum #   1 @ 7.249 - 7.324 min
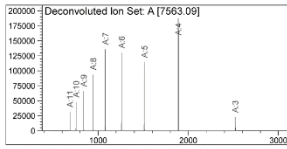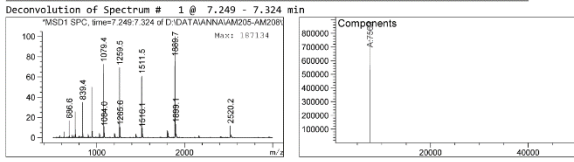
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7563.09	818929	100.00

*** End of Report ***

p2

Deconvolution of Spectrum #   1 @ 7.025 - 7.075 min
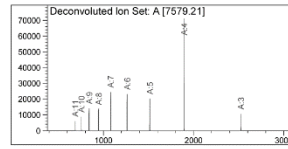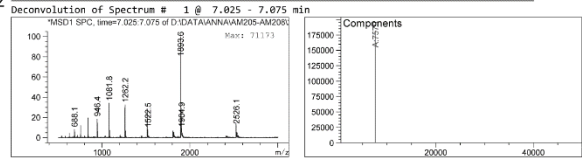
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	7579.21	182062	100.00

*** End of Report ***

ORN-72

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 7.575 - 7.675 min

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.586	VV	0.0571	187.79953	50.88990	100.0000

Totals :   187.79953   50.88990
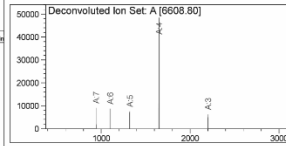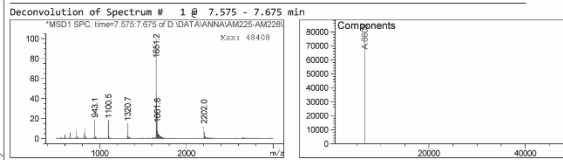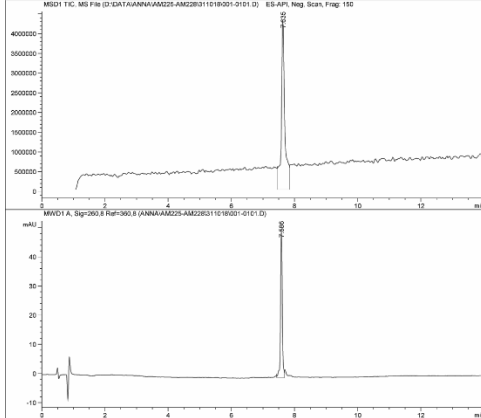
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	6608.80	79341	100.00

*** End of Report ***

ORN-73

Additional Info : Peak(s) manually integrated

Deconvolution of Spectrum #   1 @ 7.451 - 7.525 min

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.441	VV	0.0561	462.75455	128.32333	100.0000

Totals :   462.75455 128.32333
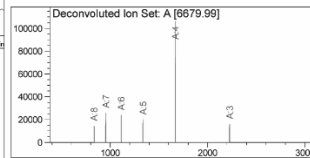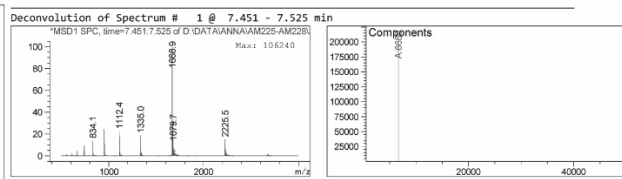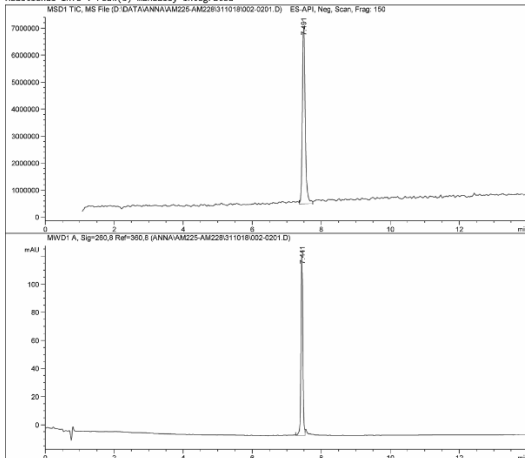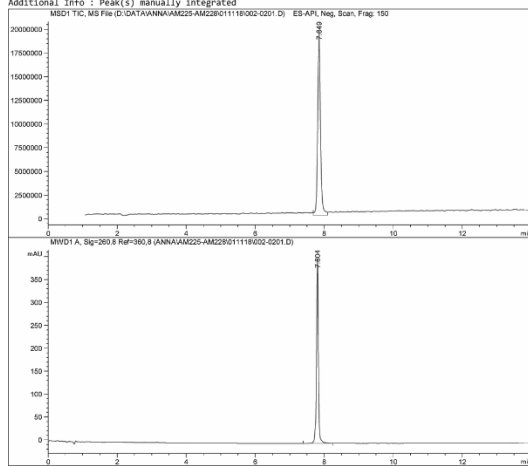
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	6679.99	202684	100.00

*** End of Report ***

ORN-74

Additional Info : Peak(s) manually integrated

Component | Molecular Weight | Absolute Abundance | Relative Abundance
A | 7227.41 | 637093 | 100.00
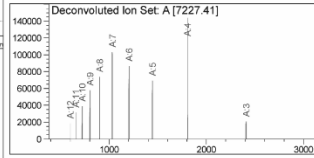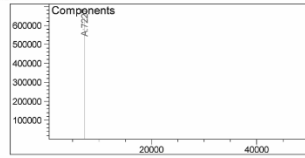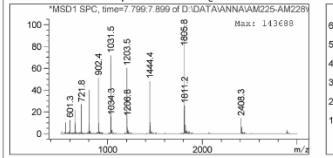
Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak RetTime Type Width Area Height Area
[min] [min] [mAU*s] [mAU] %
----|-------|----|-------|----------|----------|--------|
1 7.804 VB 0.0577 1581.76538 404.05362 100.0000

Totals : 1581.76538 404.05362

*** End of Report ***

ORN-75

Additional Info : Peak(s) manually integrated
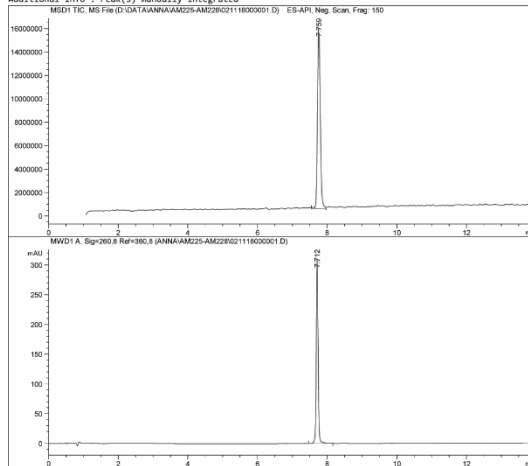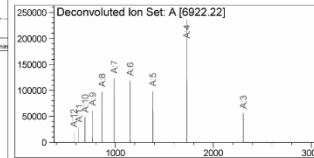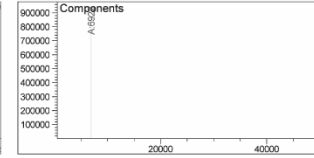
Component | Molecular Weight | Absolute Abundance | Relative Abundance
A | 6922.22 | 875433 | 100.00

Signal 2: MWD1 A, Sig=260,8 Ref=360,8

Peak RetTime Type Width Area Height Area
[min] [min] [mAU*s] [mAU] %
----|-------|----|-------|----------|----------|--------|
1 7.712 VB 0.0562 1124.50806 310.87842 100.0000

Totals : 1124.50806 310.87842

*** End of Report ***

225

IX.2. NMR spectra of the synthesized compounds

IX.2.1. Chapter IV
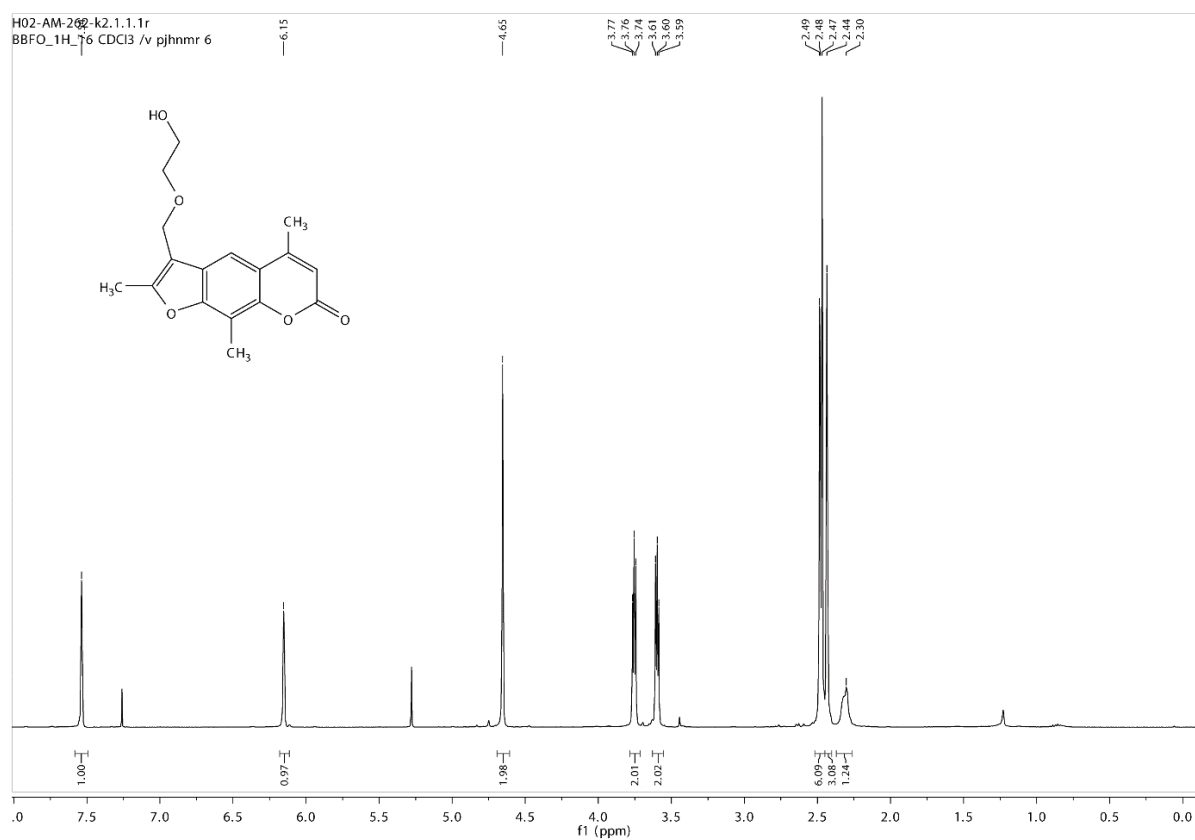
Figure 116: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **18a.**
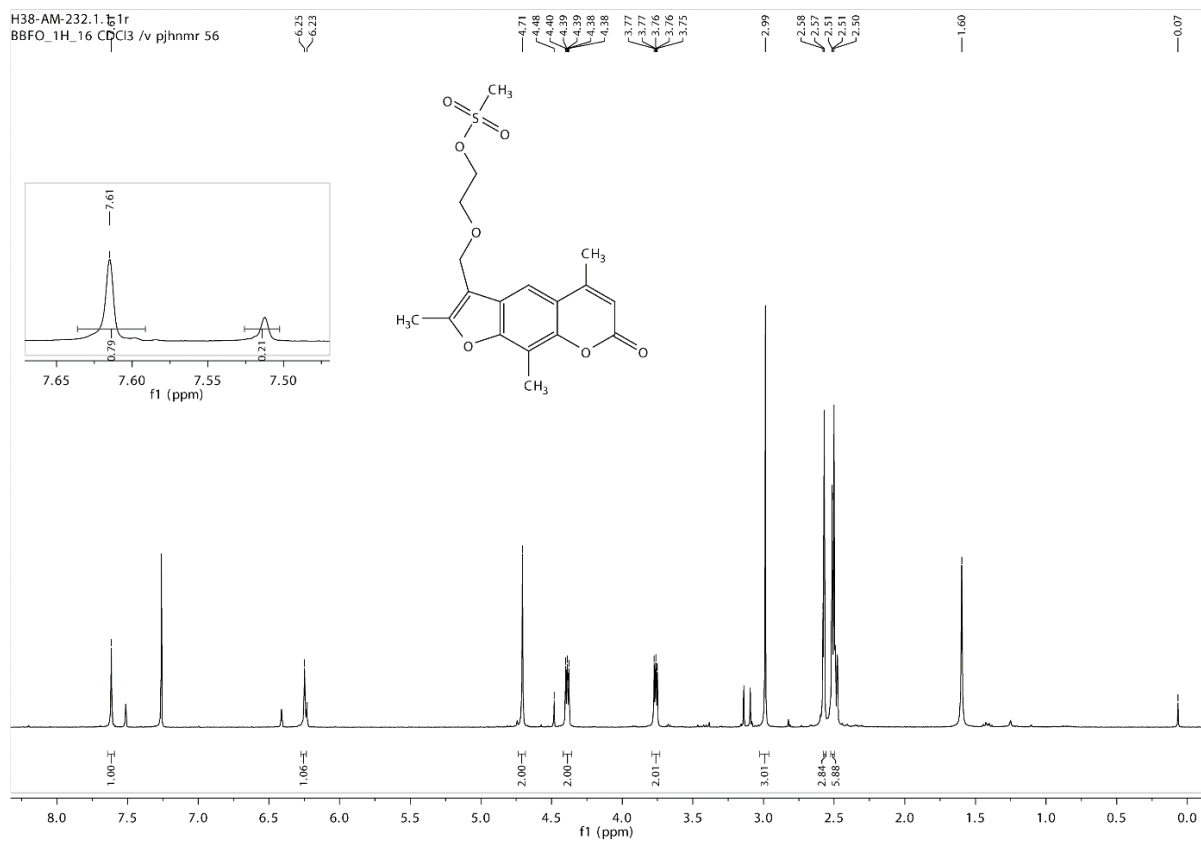
Figure 117: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **19a.**
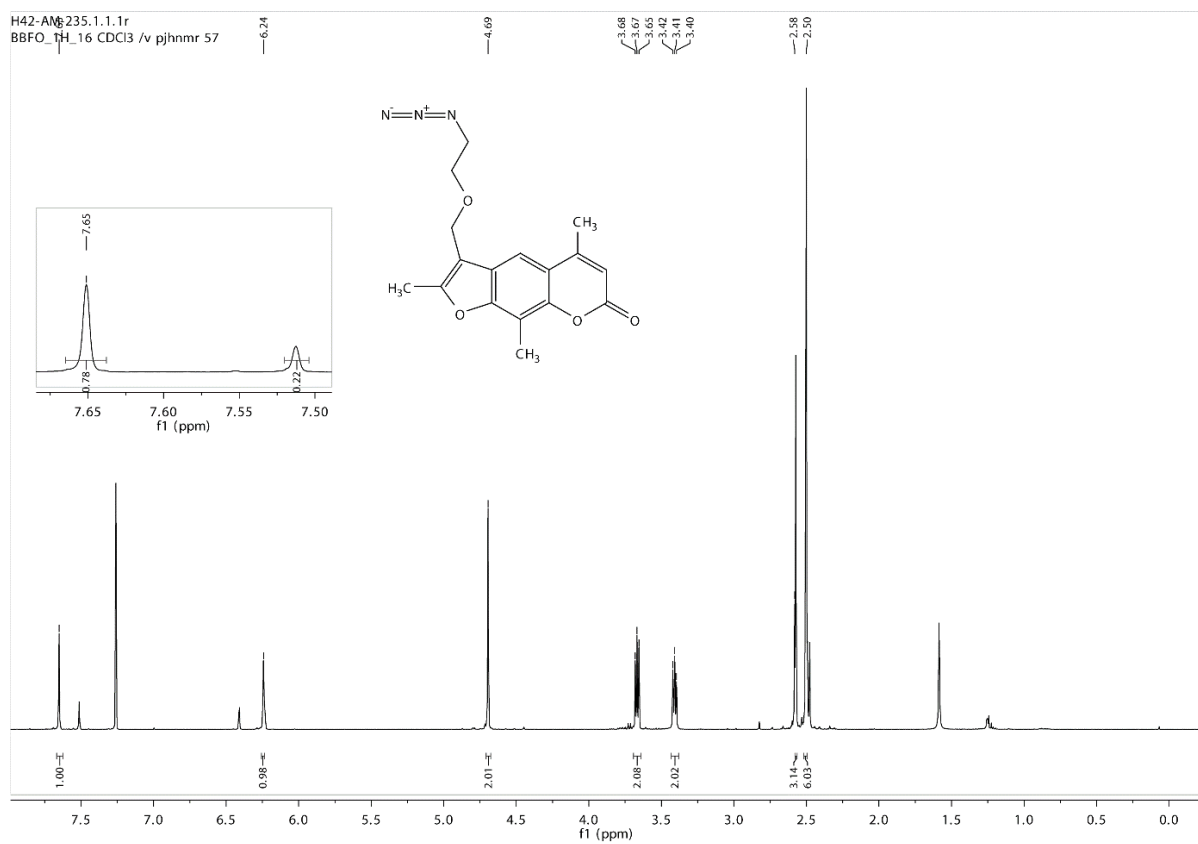
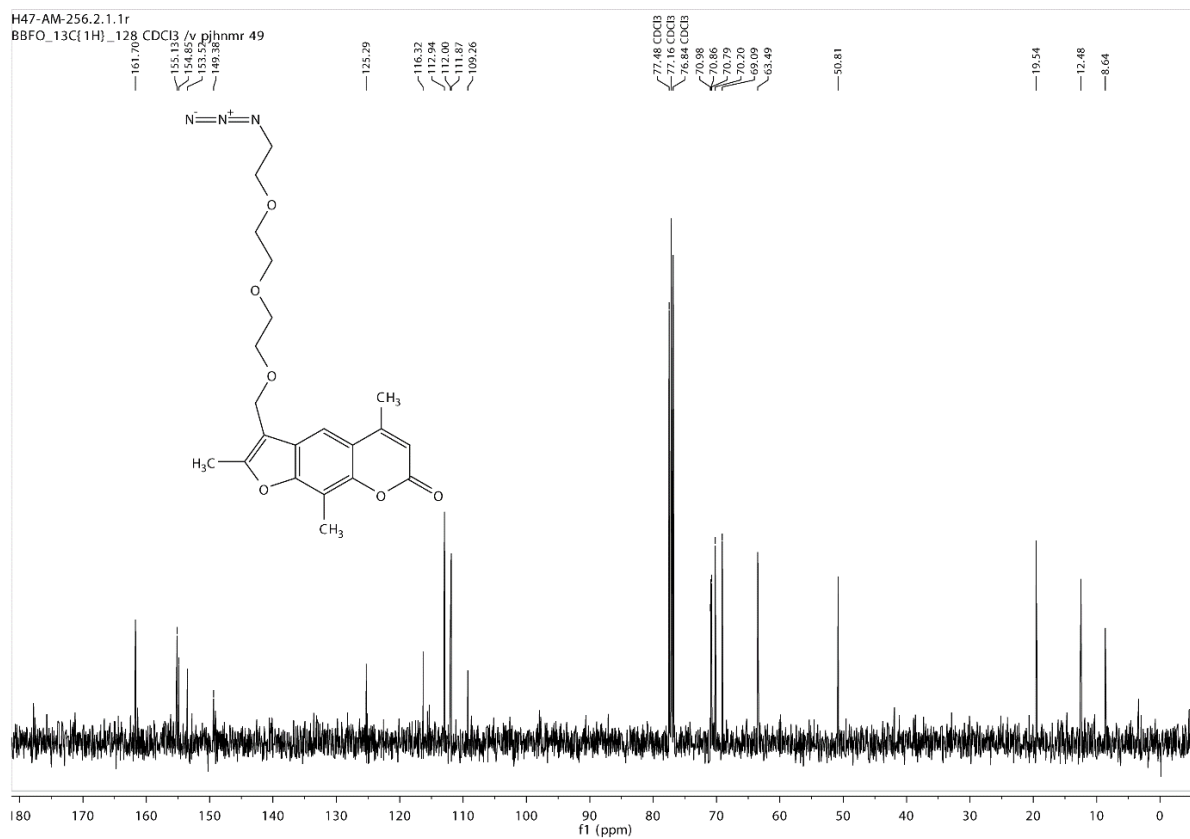**Figure 118:** ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **20a.**
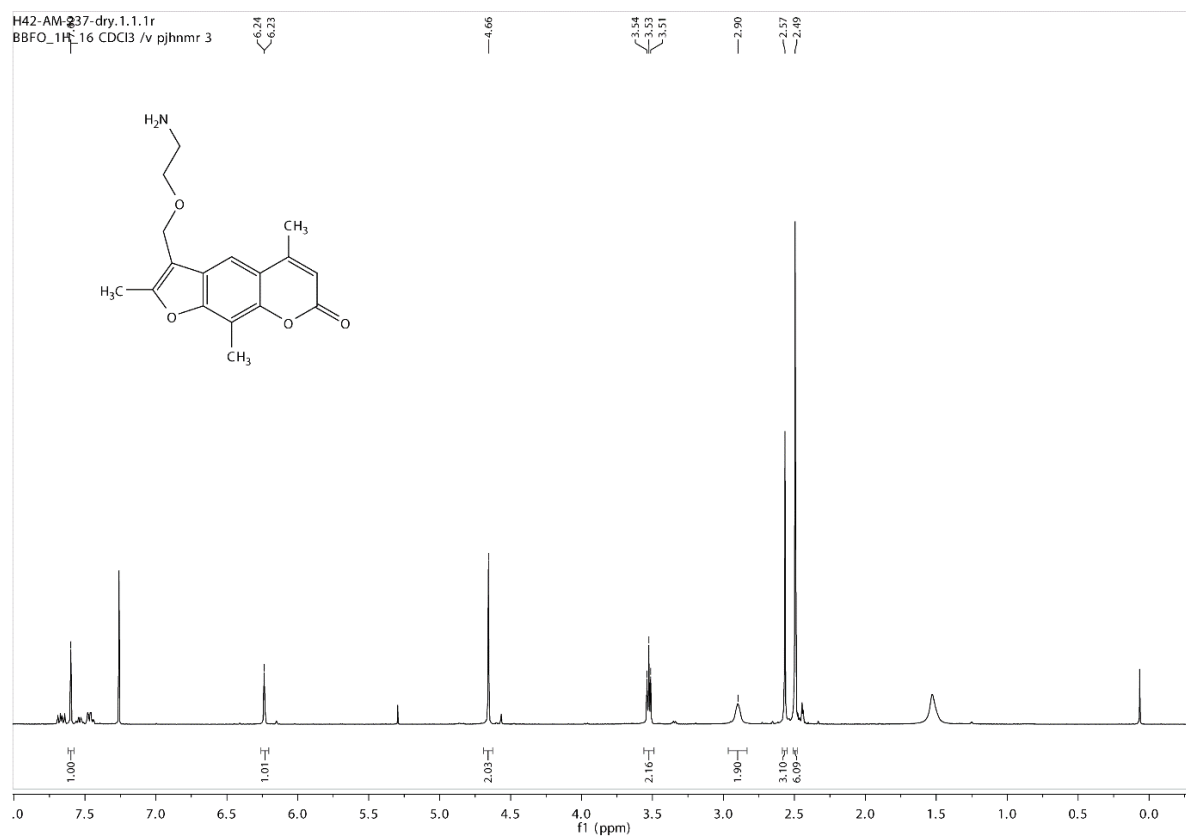
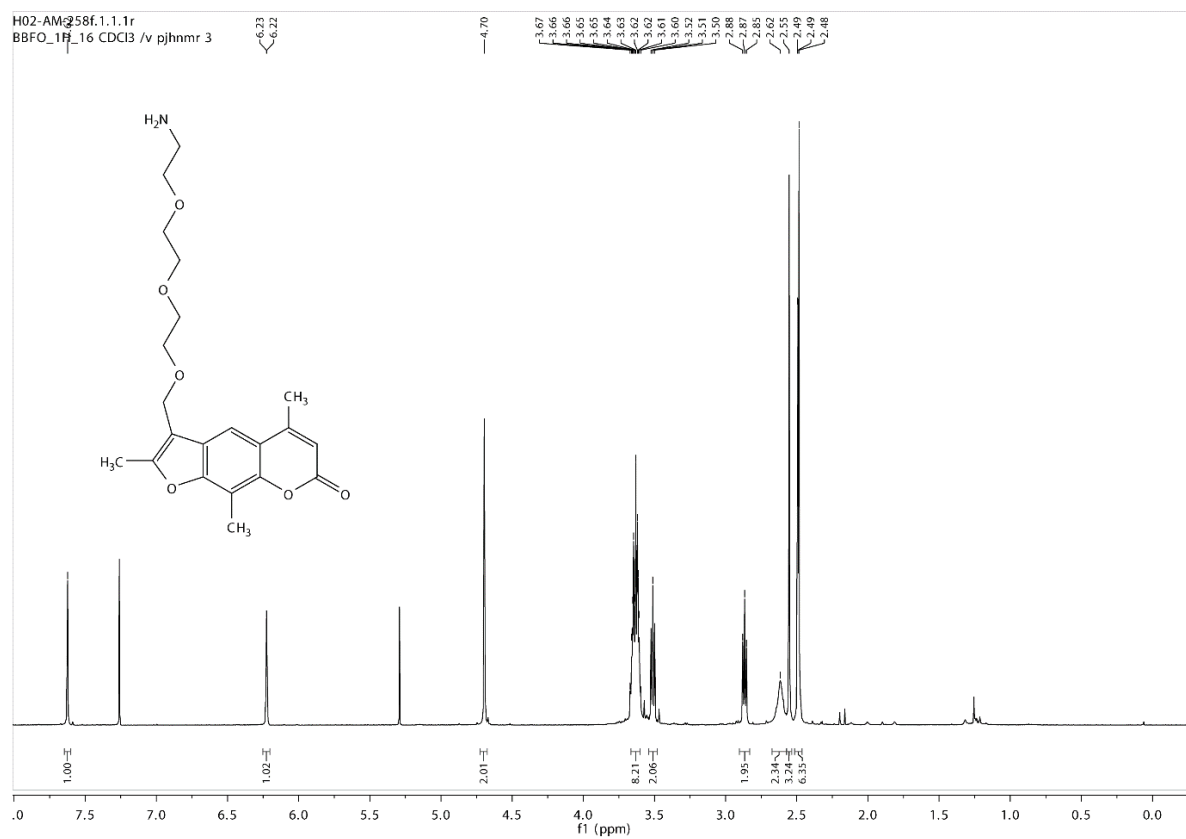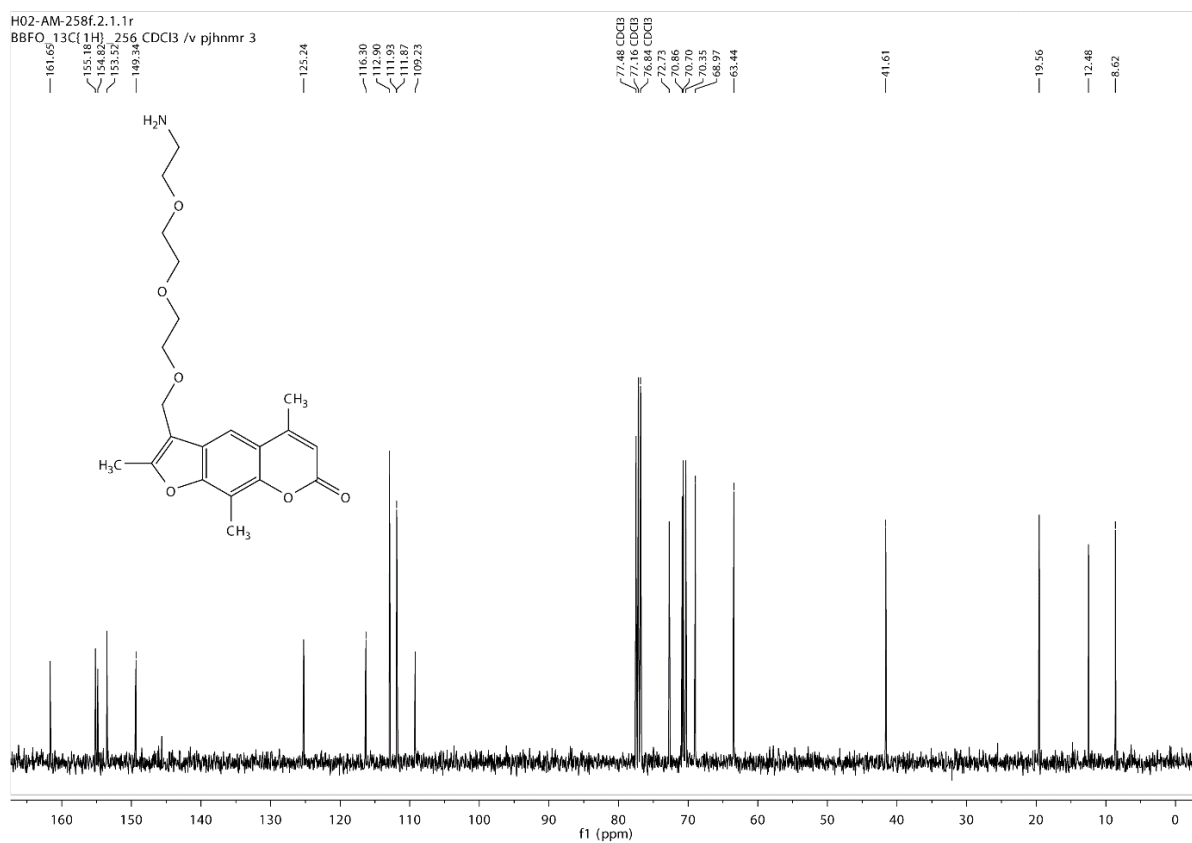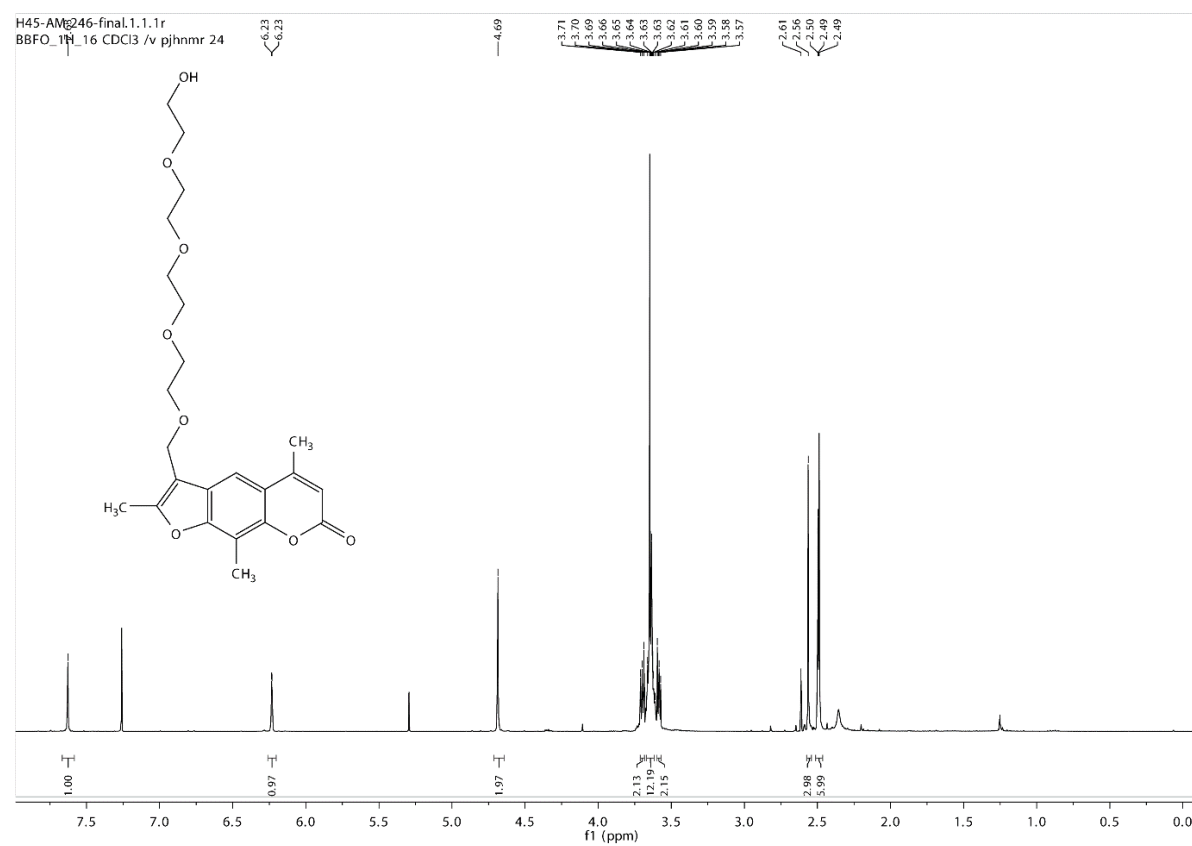**Figure 119:** ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **21a.**
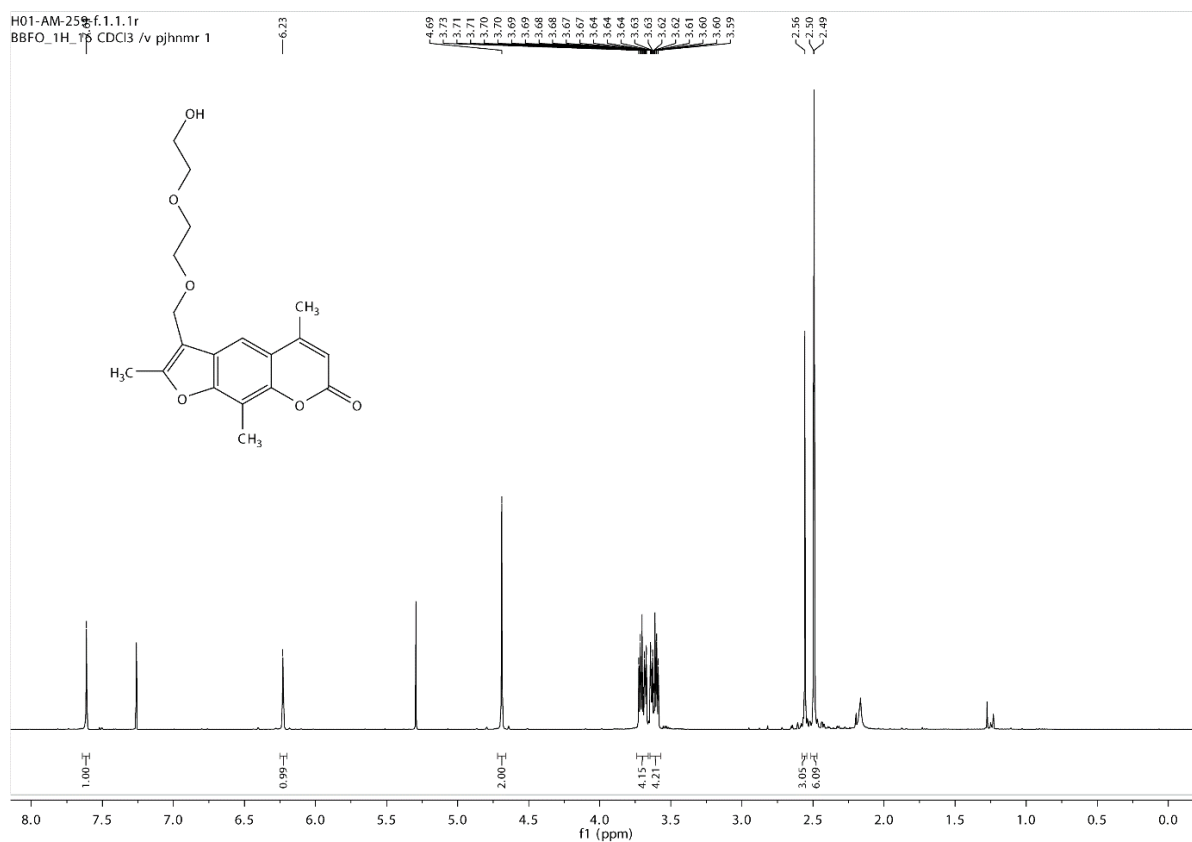
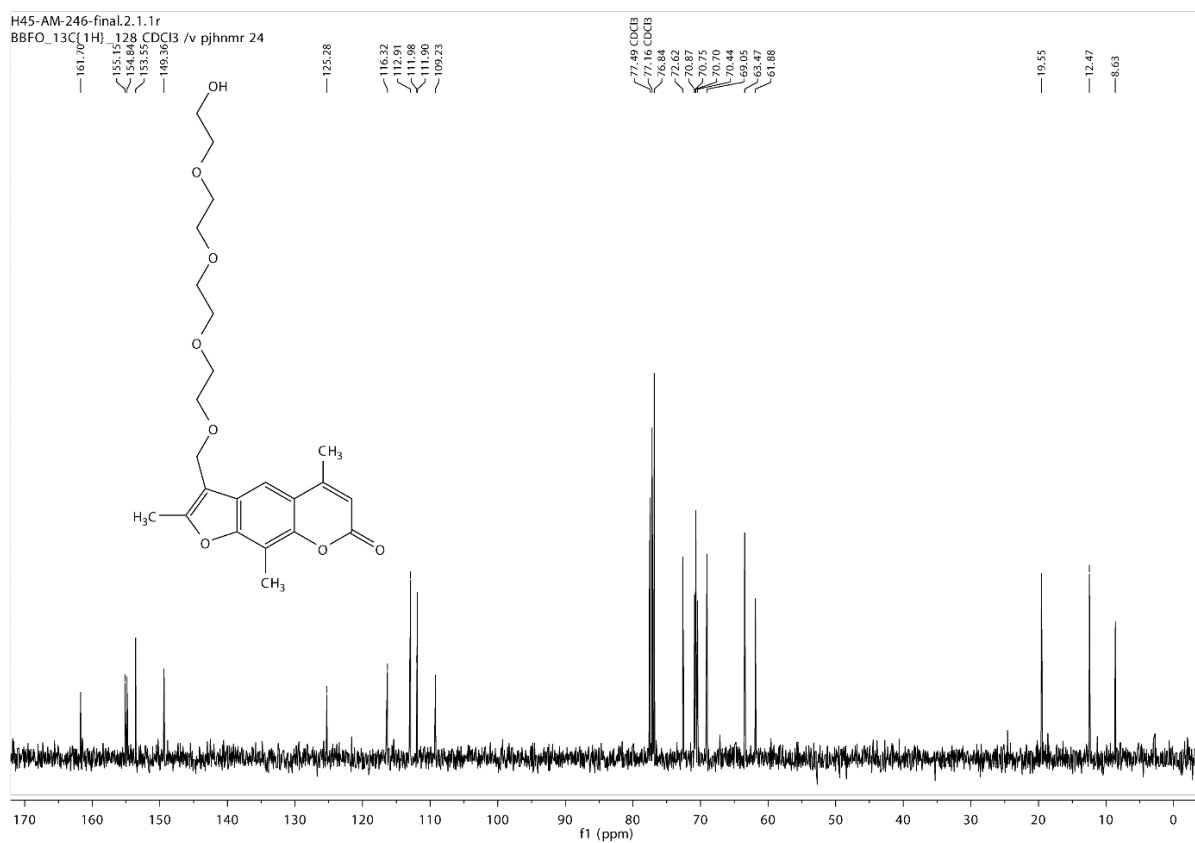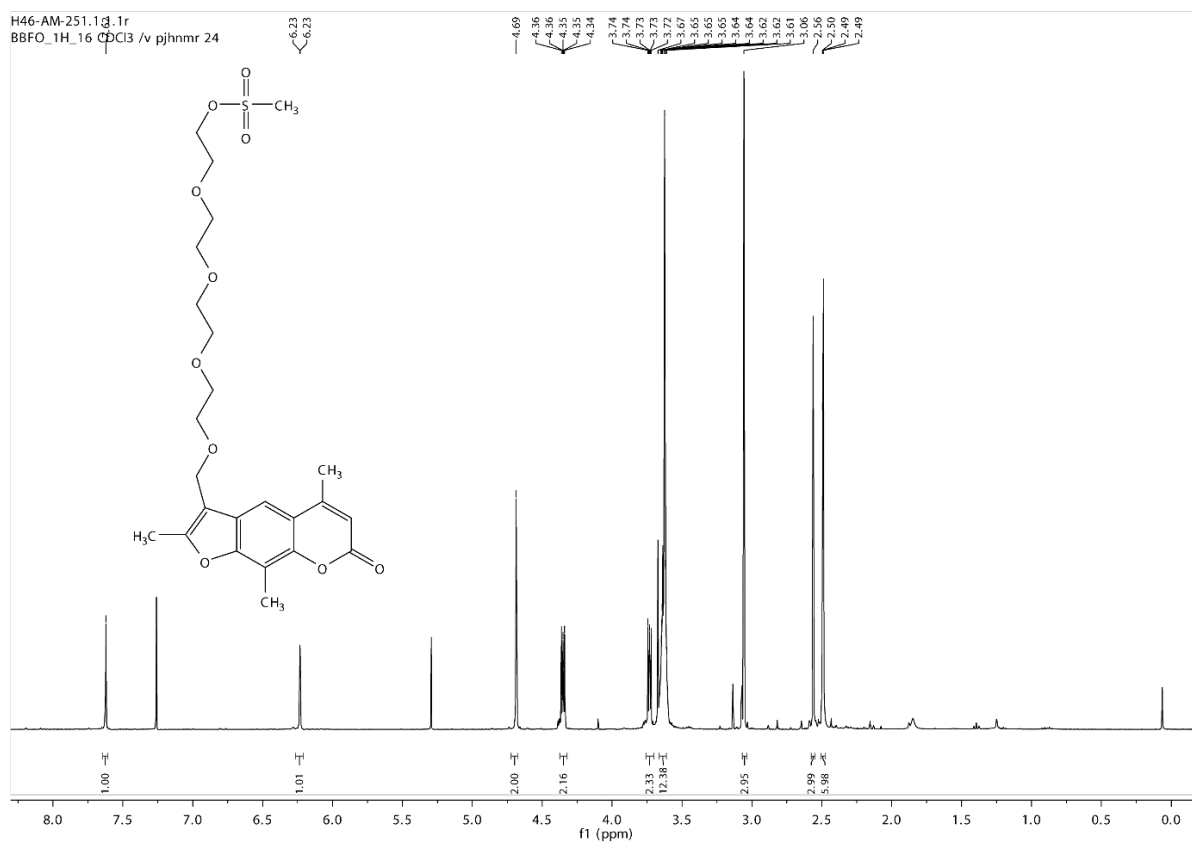**Figure 120:** ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **18b.**
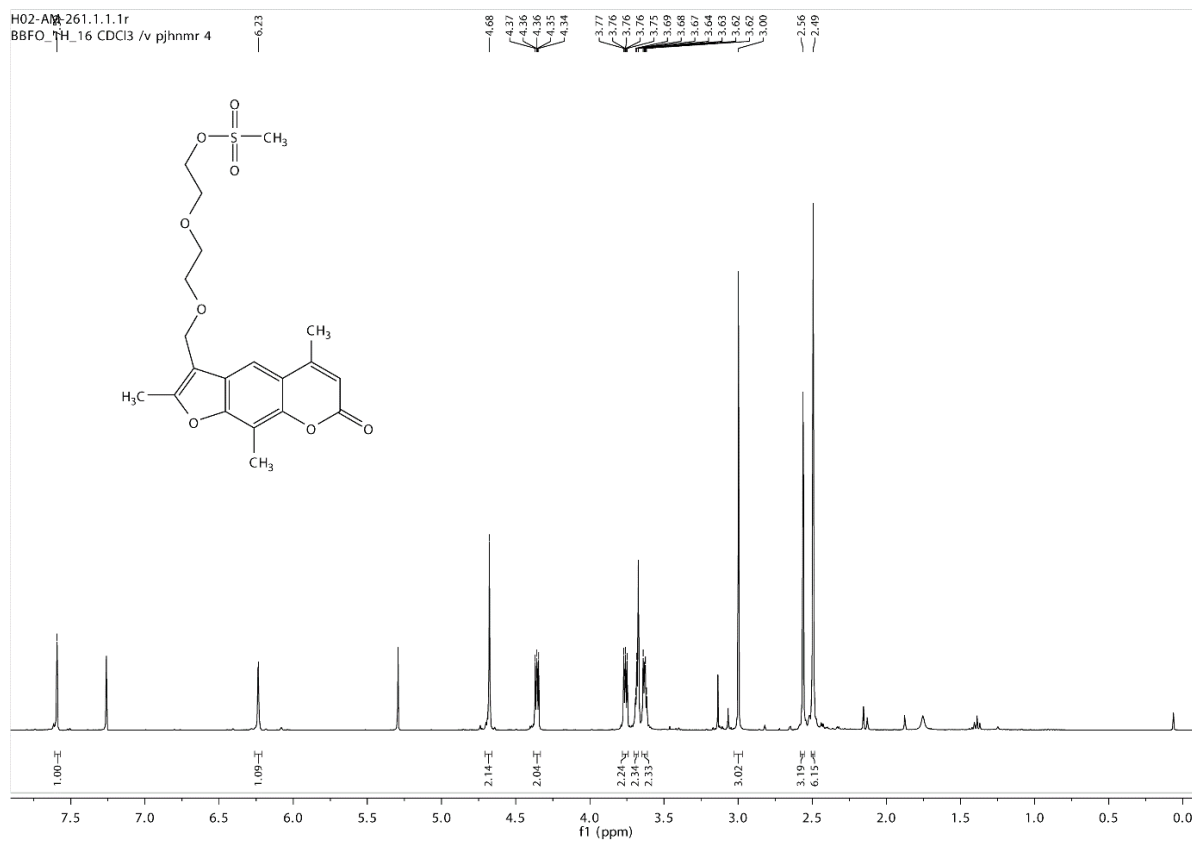
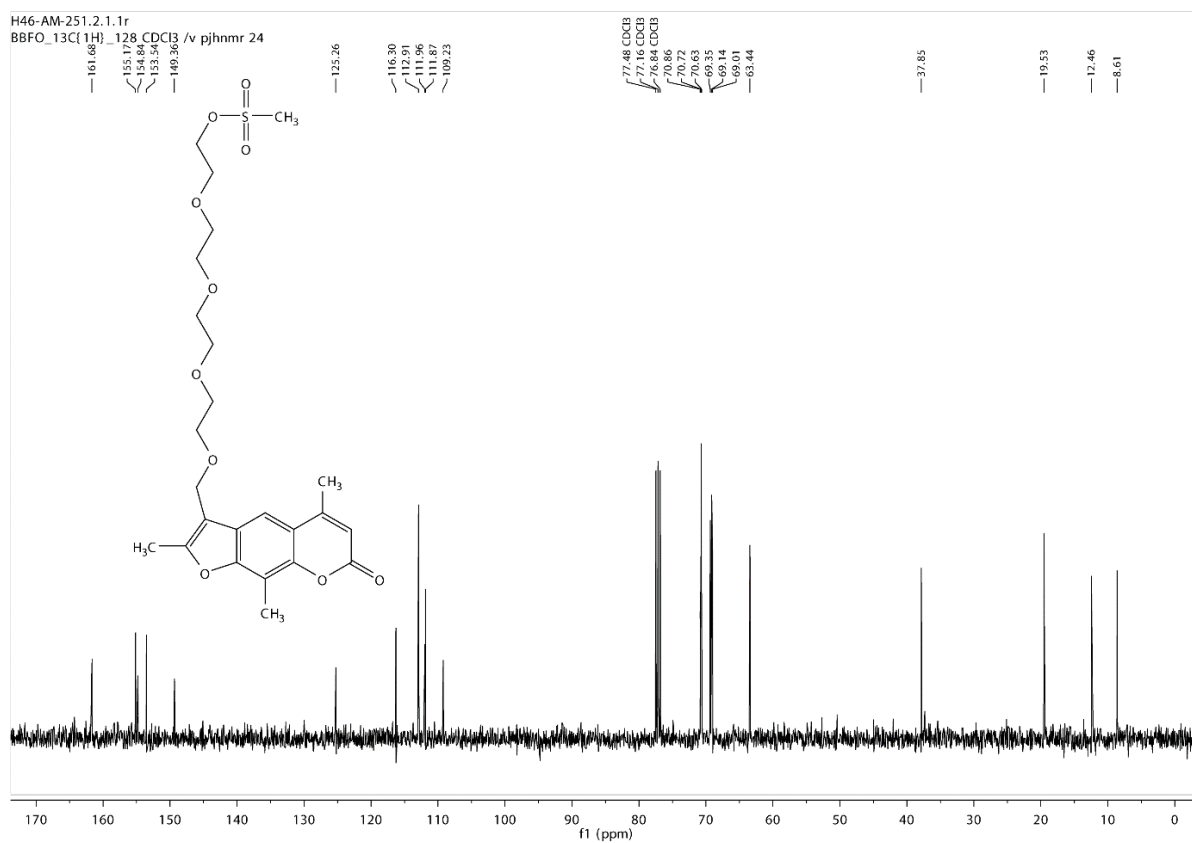**Figure 121:** ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **19b.**
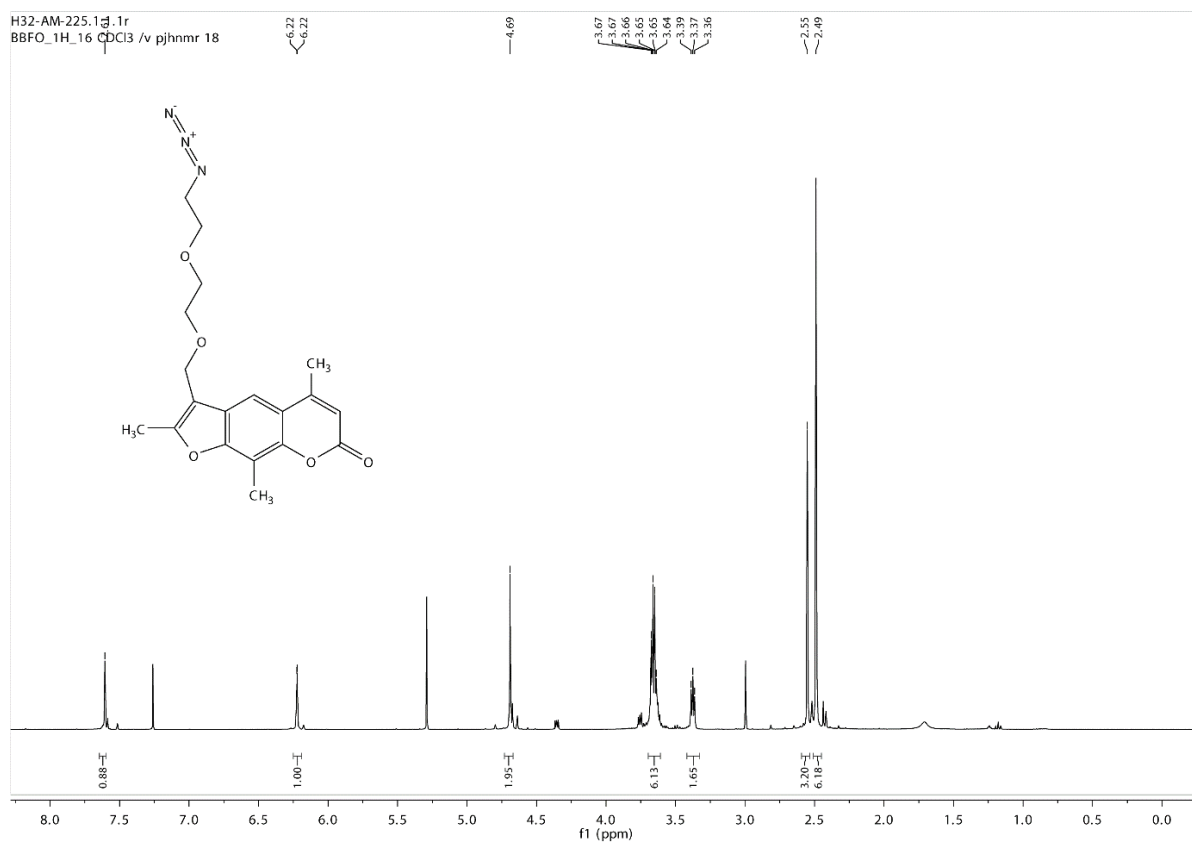
Figure 122: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **20b.**

Figure 123: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **21b.**

Figure 124: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **18c.**
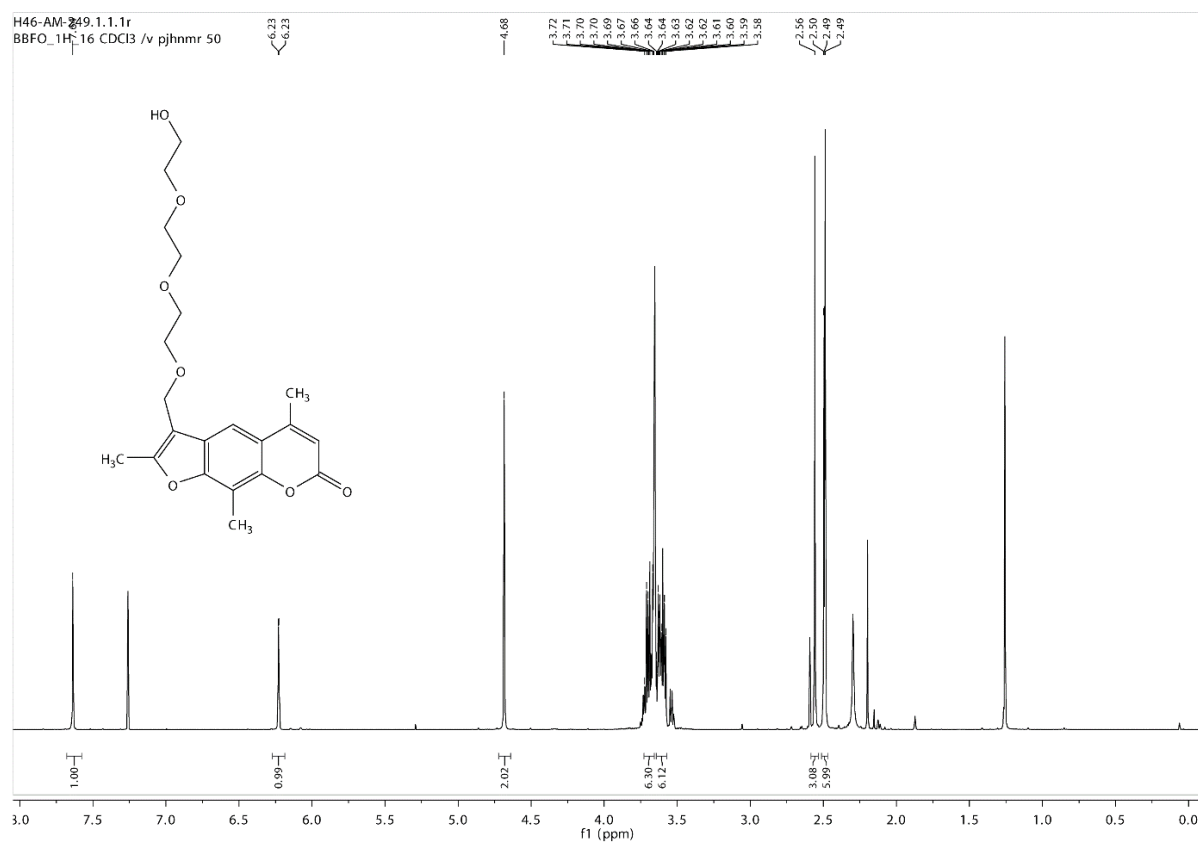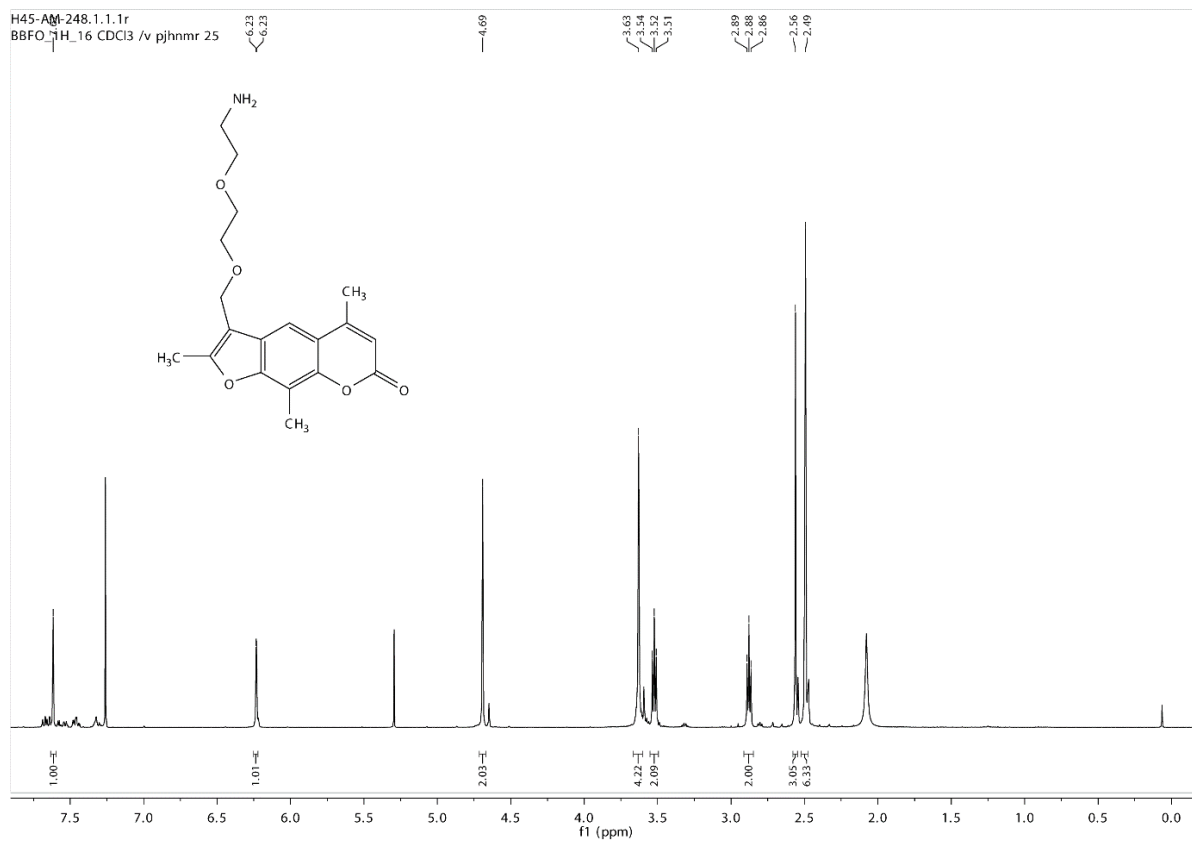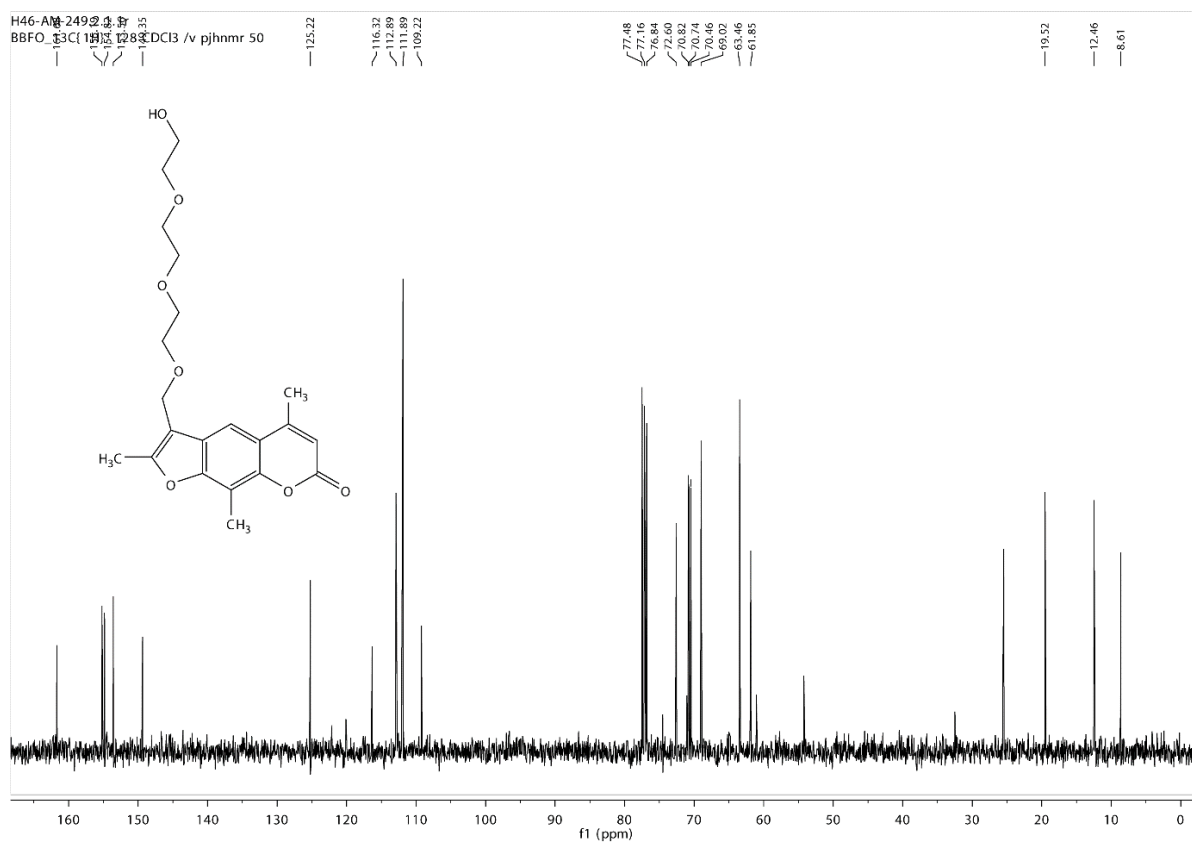
Figure 125: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **18c**.
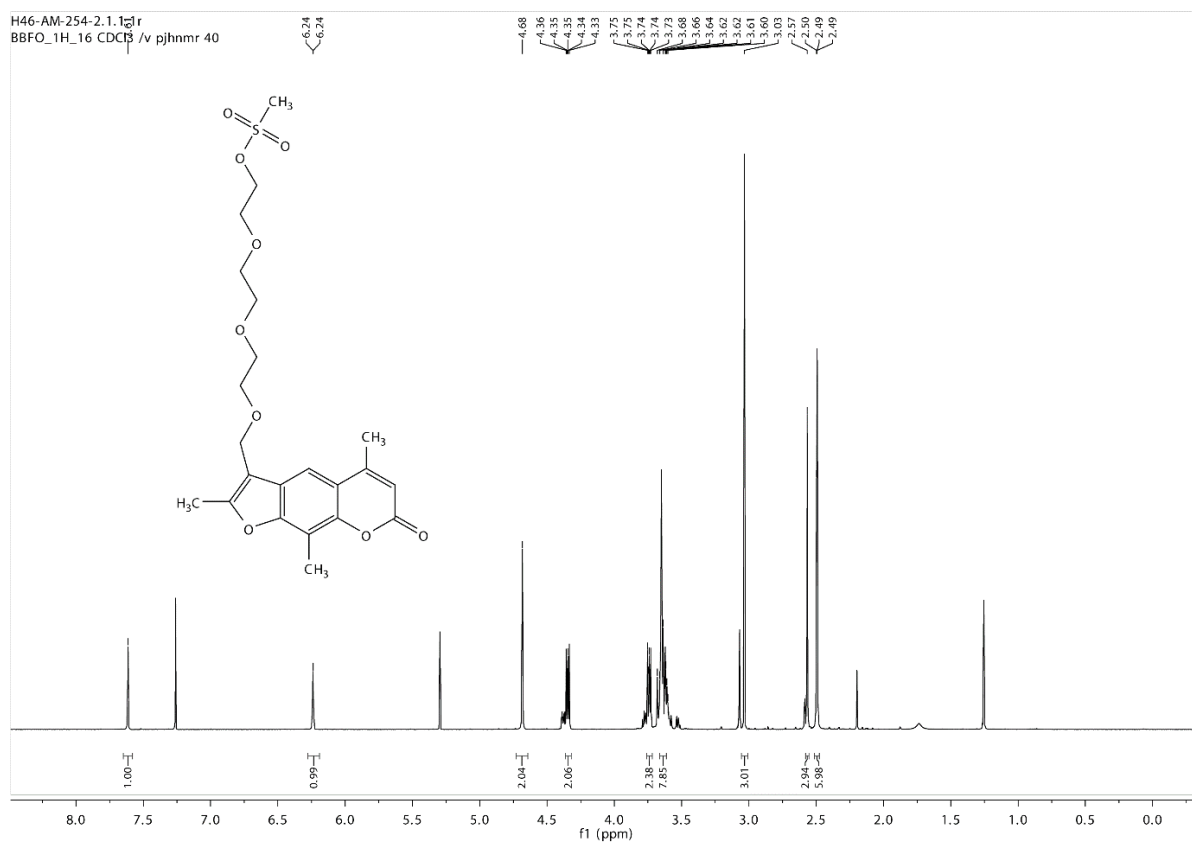
Figure 126: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **19c.**
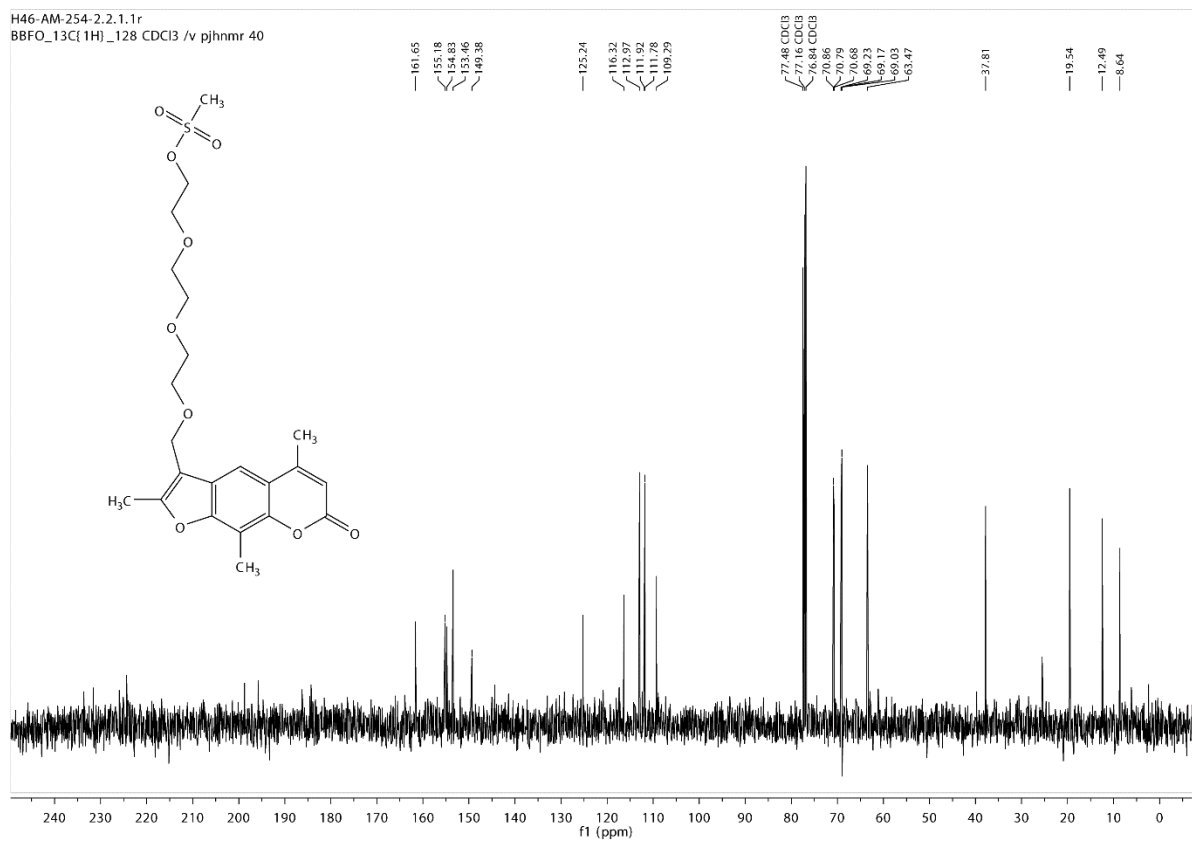
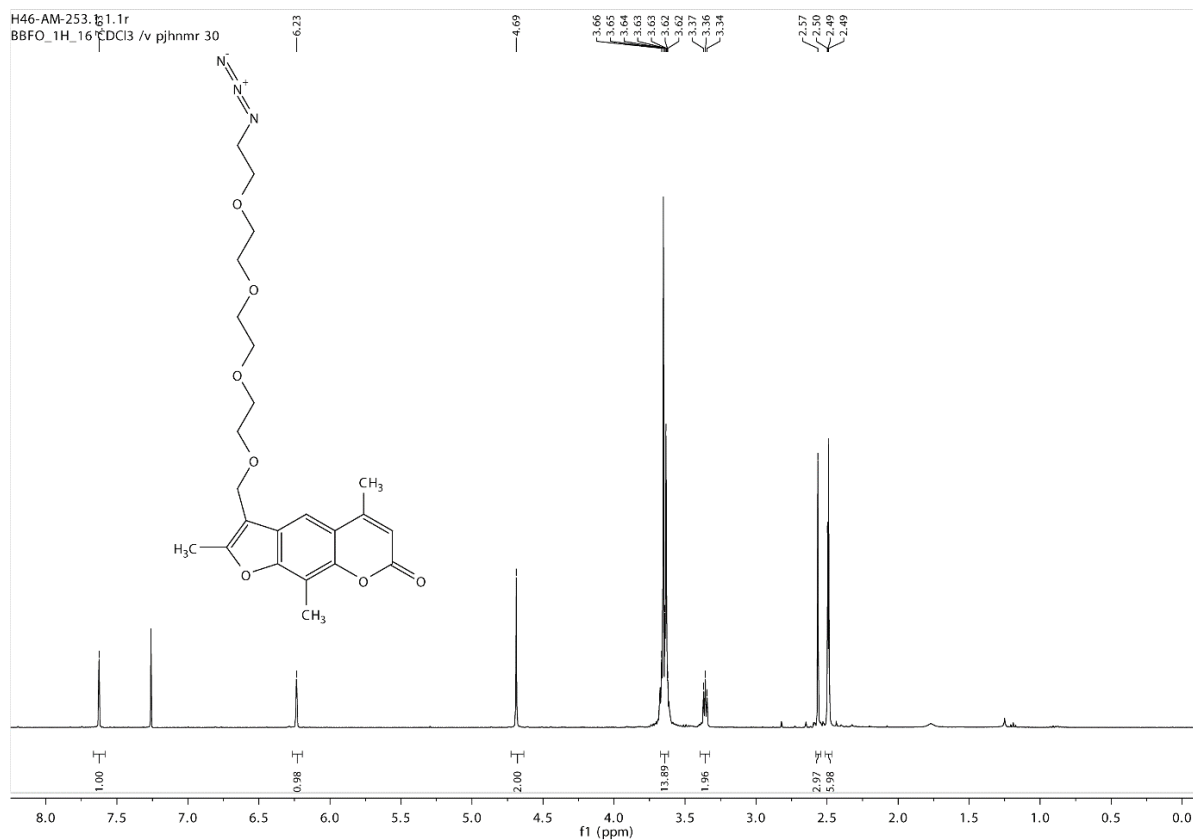**Figure 127:** ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **19c**.
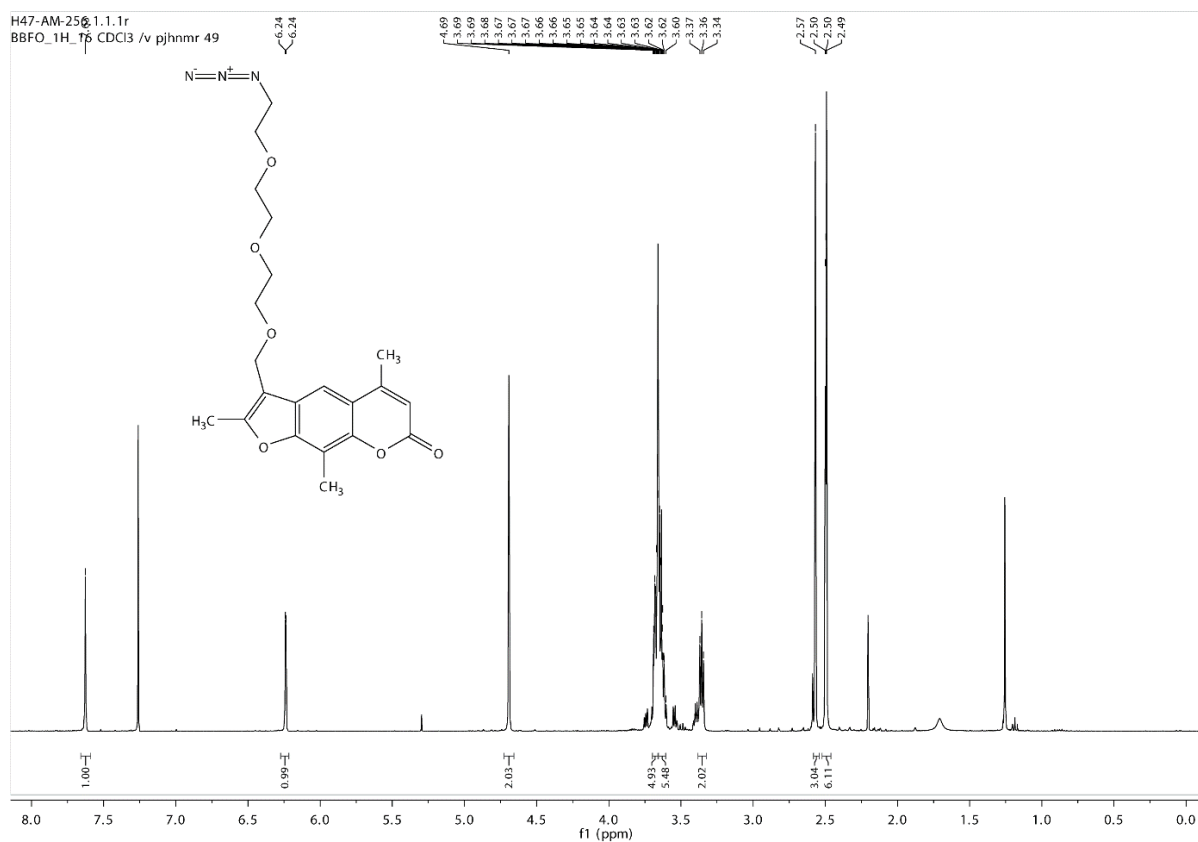
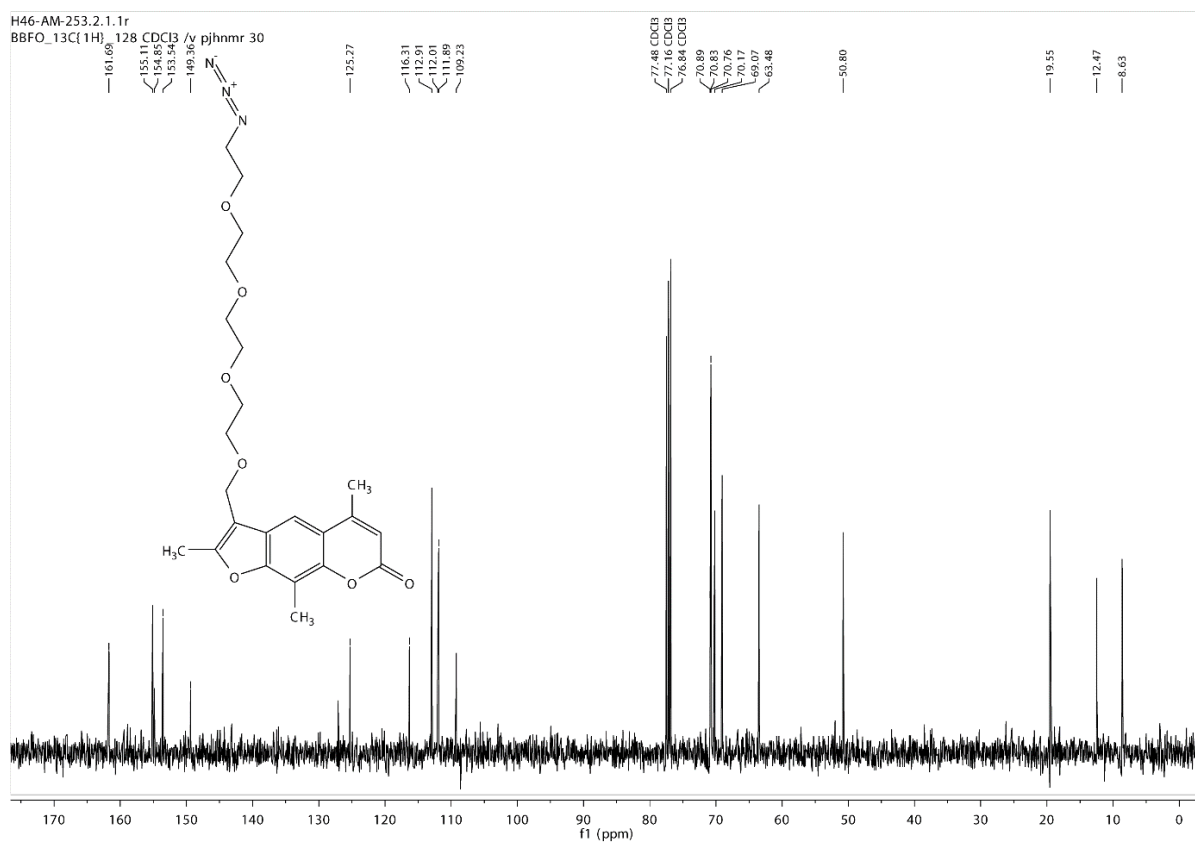**Figure 128:** ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **20c.**

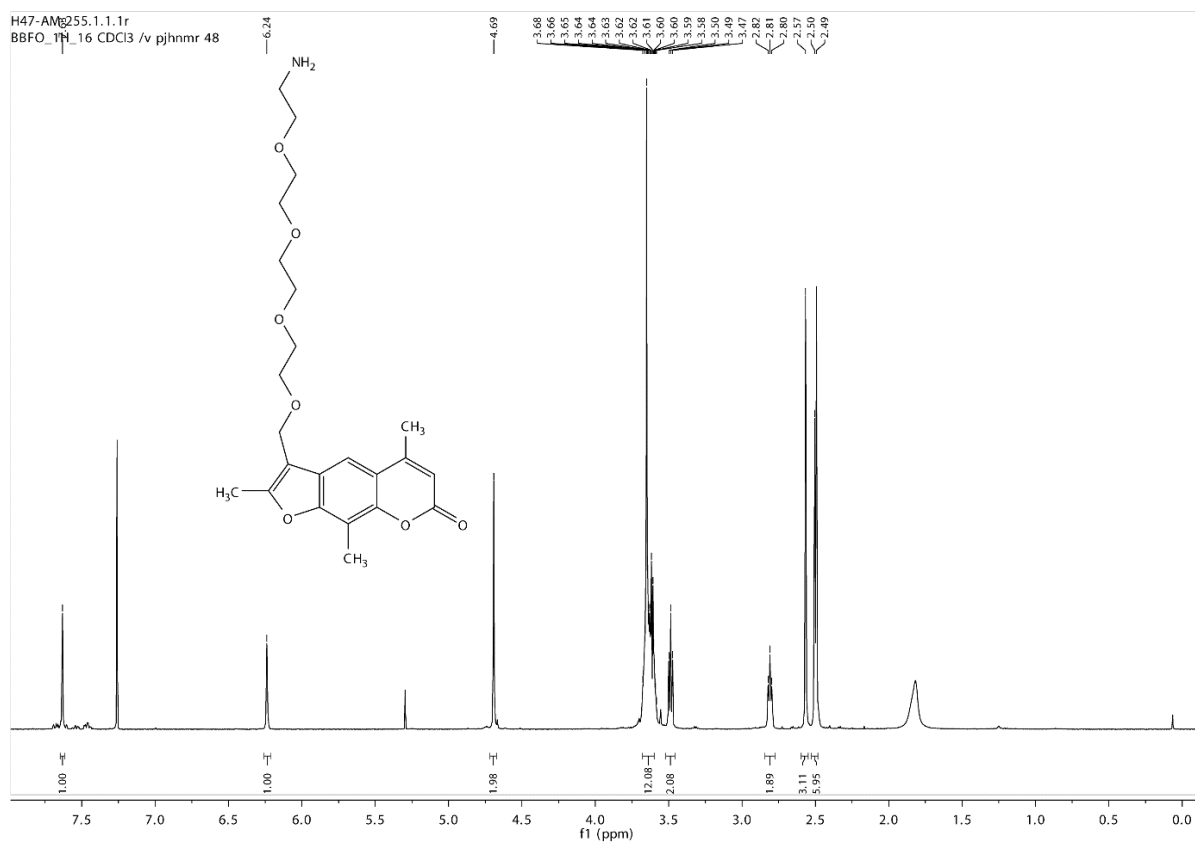**Figure 129:** ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **20c**.

Figure 130: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **21c.**

Figure 131: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **21c**.

Figure 132: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **18d.**
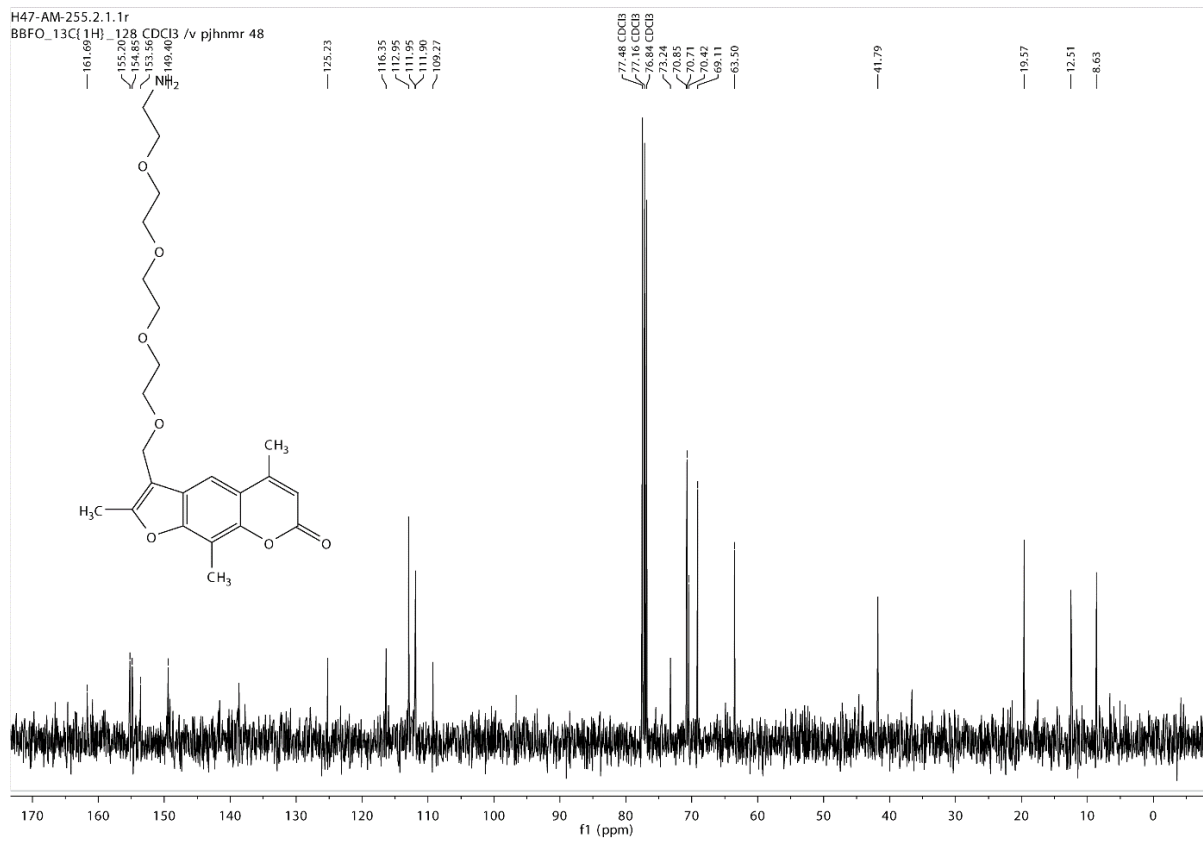
Figure 133: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **18d**.

Figure 134: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **19d.**

Figure 135: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **19d**.

Figure 136: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **20d.**

Figure 137: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **20d**.

Figure 138: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **21d.**

H47-AM-255.2.1.1r
BBFO_13C{1H}_128 CDCl3 /v pjhnmr 48

Figure 139: ^{13}C NMR (101 MHz, Chloroform-*d*) spectrum of the compound **21d**.

H49-AM-179.1.1.1r
BBFO_1H_16 MeOD /v pjhnmr 50

Figure 140: ^{1}H NMR (400 MHz, Methanol-*d*$_4$) spectrum of the compound **22**.
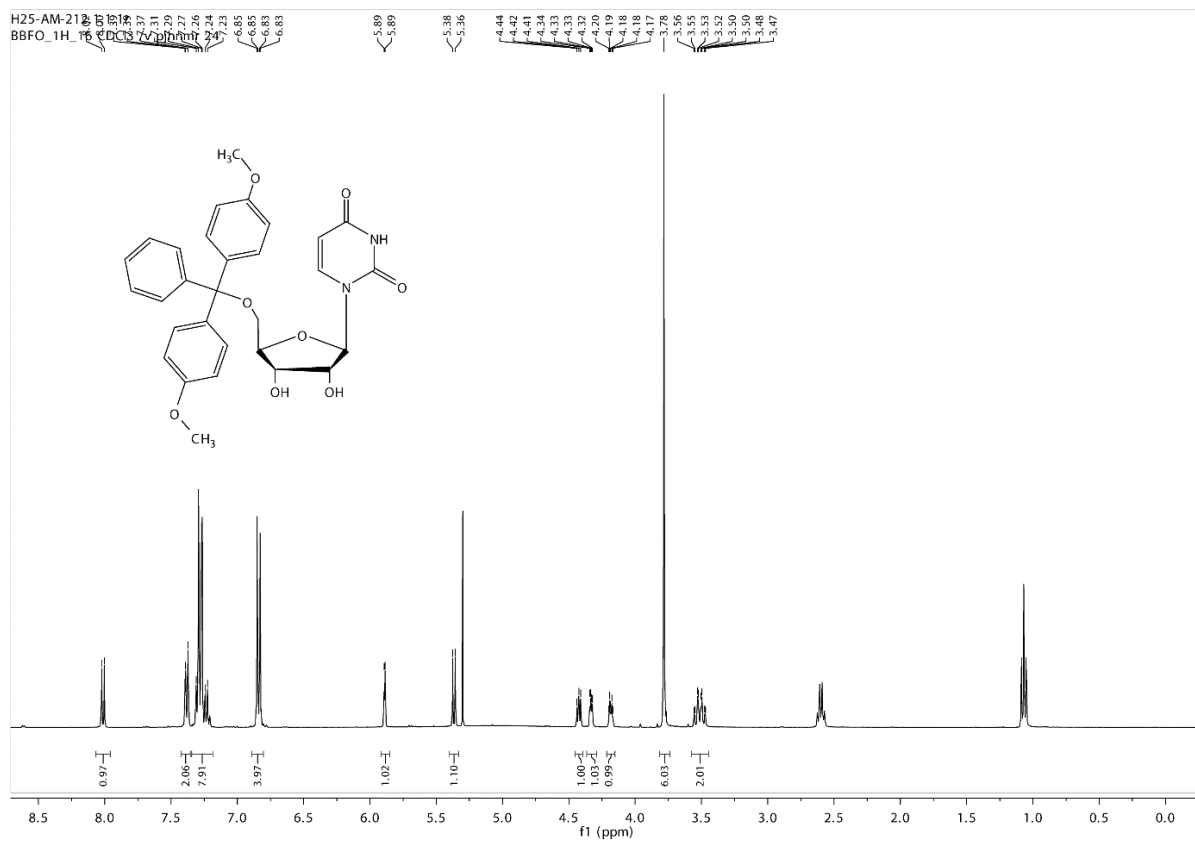
Figure 141: ^{1}H NMR (400 MHz, Chloroform-*d*) spectrum of the compound **23.**
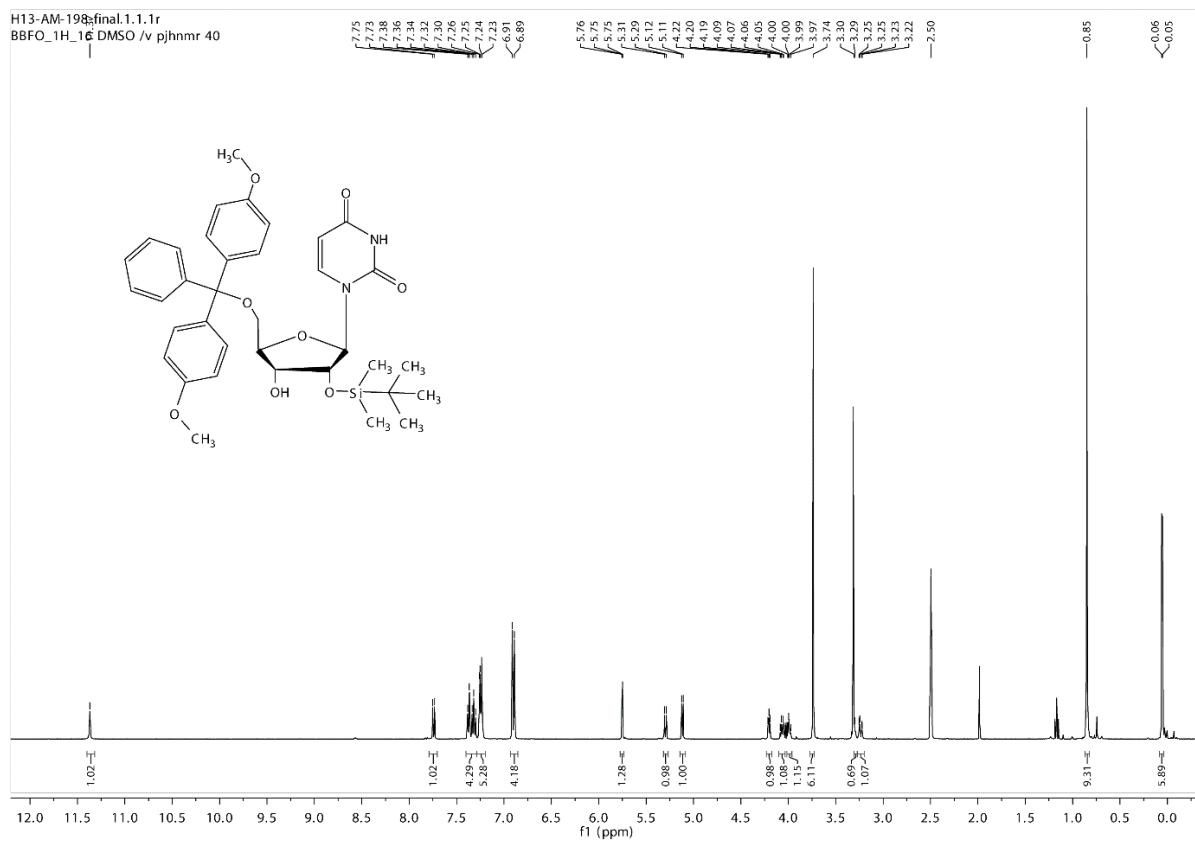
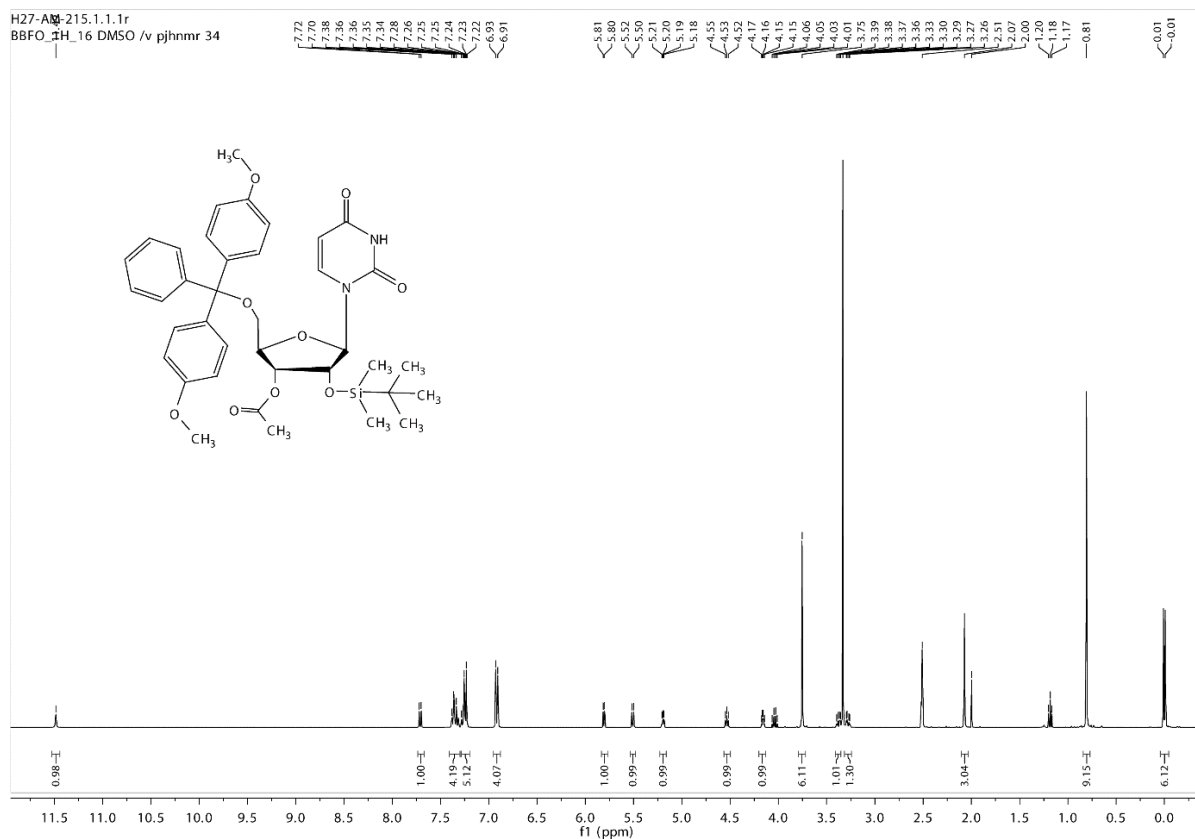**Figure 142:** ^{1}H NMR (400 MHz, DMSO-*d*$_6$) spectrum of the compound **24.**

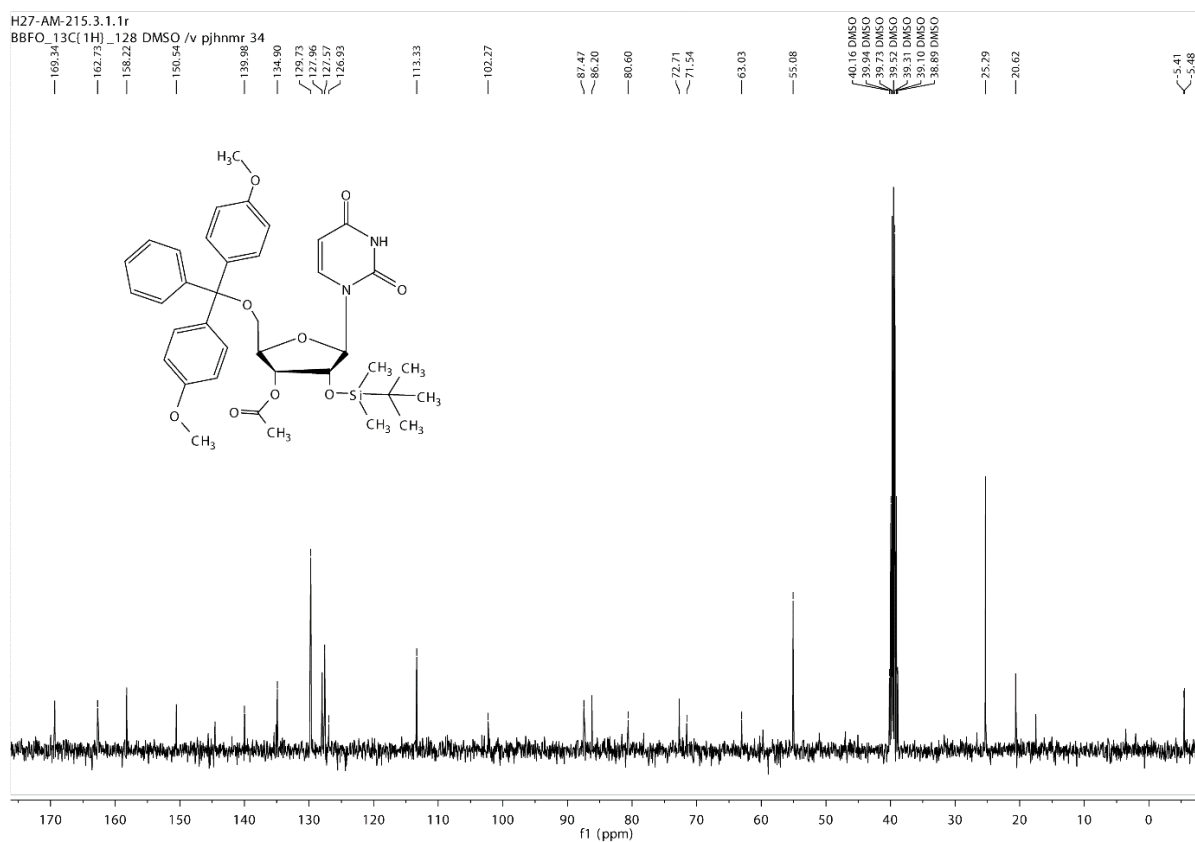**Figure 143:** ^{1}H NMR (400 MHz, DMSO-d_6) spectrum of the compound **25.**

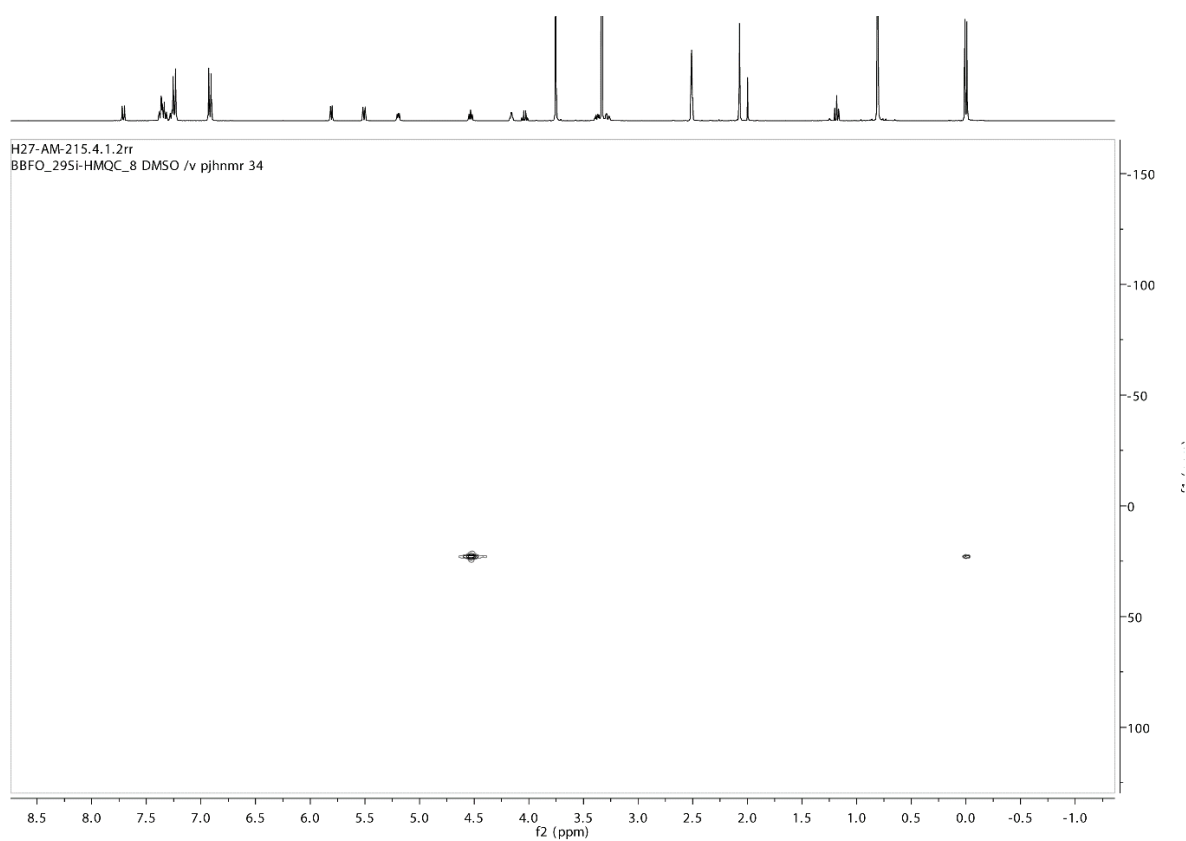**Figure 144:** ^{13}C NMR (101 MHz, DMSO-d_6) spectrum of the compound **25.**

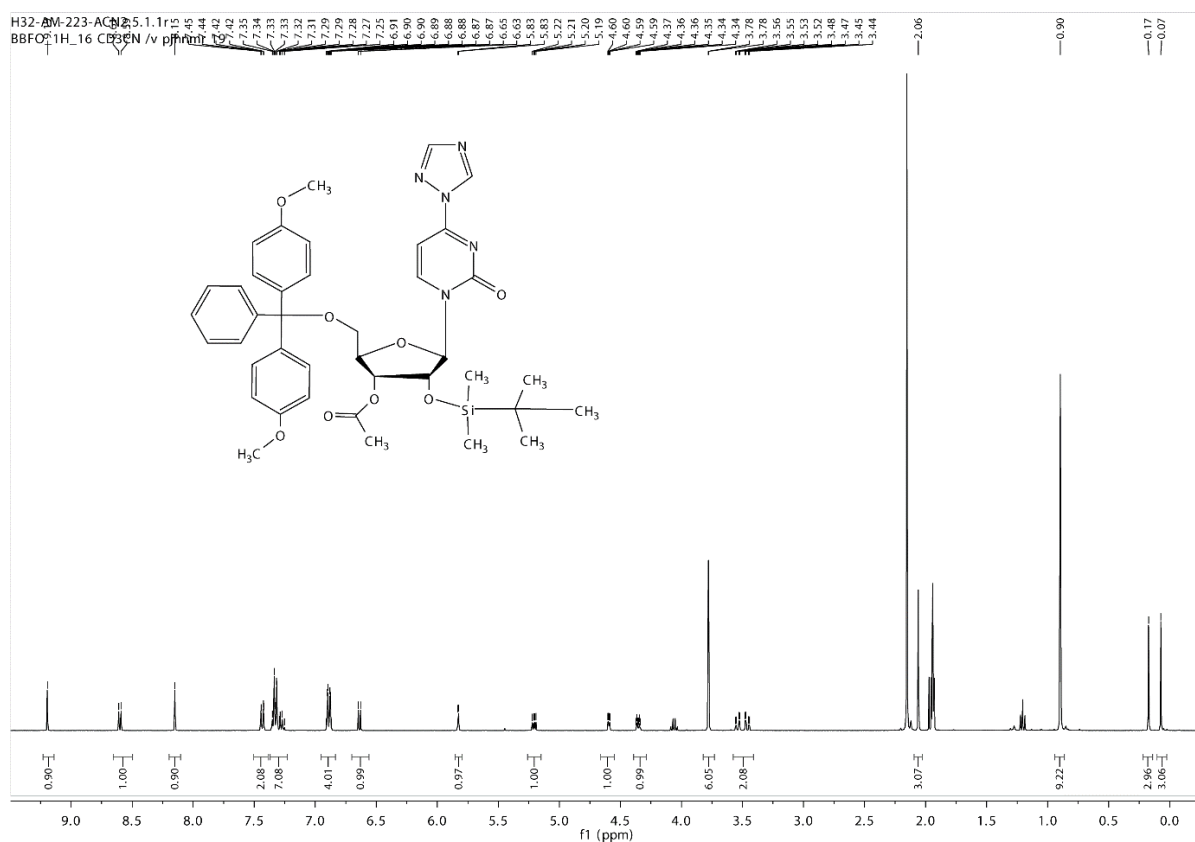**Figure 145:** ^{29}Si HMQC (DMSO-d_6) spectrum of the compound **25**.

Figure 146: ^{1}H NMR (400 MHz, Acetonitrile-d_3) spectrum of the compound **26.**

Figure 147: ^{13}C NMR (101 MHz, Acetonitrile-d_3) spectrum of the compound **26**.