FACULTY OF SCIENCES

DEVELOPING FURAN AND TRIAZOLINEDIONE CHEMISTRIES FOR SITE-SELECTIVE PROTEIN MODIFICATION

Klaas Decoene

Promotor: Prof. Dr. Annemieke Madder
Co-promotor: Prof. Dr. Jan Gettemans

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Science: Chemistry

2021

GHENT UNIVERSITY

Acknowledgements

In a certain way this part is easy to write since it is so obvious that this entire PhD chapter of my life was only possible with the help and support of many people. However, it is also hard to find the right words to thank everyone and express the gratitude I feel. The first person I need to thank is my promotor prof. Annemieke Madder, who was there since the very beginning of my academic adventures. Already as a bachelor student I was welcomed in the OBCR group to do my "bachelorproject". Two years later, after I had re-joined the group for a joint master thesis together with the Carell group in Munich, I decided to ask Annemieke if I could stay and start a PhD in her group. I really could not have wished for a better promotor. Annemieke provided funding, challenging research topics, interesting opportunities for (international) collaboration and a never ending stream of positivity and new ideas if I got stuck. So it really is safe to say that none of this would have been possible without Annemieke.

I want to thank all the past and current members of the OBCR-group and by extension also the OS-group for the nice working atmosphere, the many funny moments in the lab, the official team buildings and the less official activities such as the "OBCR night out" which were so much fun. It was very comforting that Joke, Yentl and I joined the group at the same time so we were basically in the same boat when we had to face the struggles of a beginning PhD student like grant applications. I want to thank Willem, who showed me around in the world of peptide and furan chemistry already during my master thesis. During my time as a PhD student, I have had the unbelievable luck that there was always an experienced post-doc present in the lab to ask all my questions to. First, Smita was there to help me during the beginning of my PhD thesis and after Smita left the group to go back to the UK, Kurt came to Ghent to take over. Also Jos deserves a special mention as he was always there to help and assist with all technical aspects of equipment, which is of crucial importance. I also want to thank Ewout, who joined the lab a couple years after me, for all the laughter we shared in lab A. Fun was also guaranteed when Vic would drop by to tell jokes and complain about his daily struggles. Benz, with your positivity and infectious smile you can always make me laugh. Dorien and Laure, you are doing a great job by keeping the spirit high in lab A, I wish you all the very best in the future.

I also need to thank my co-promotor prof. Jan Gettemans, Tim and the rest of the nanobody lab at the Rommelaere institute for the nice collaboration. Organic chemistry and biochemistry seemed like different worlds at first, and indeed we had to overcome quite a few challenges, but overall I think we did a great job together.

I also want to thank Toby for guiding me during my stay in the Suga lab in Tokyo. This was definitely one of the highlights of my PhD and an enriching experience both on a scientific and personal level. I still remember the great Christmas party of the Suga lab.

I want to thank Kamil and Prof. Johan Winne for the very nice collaboration on the TAD project. This project was a nice example of how a bad result can lead to a new application. The flipping of this unfortunate side reaction to a new method was only possible with our combined efforts and fruitful discussions. For the last part of this project we were able to collaborate with the lab of prof. Kris Gevaert, also thanks to An Staes for the help with these experiments, which really lift this project to a higher level. This collaboration was in fact so successful that I did not want to stop, so another thanks to Kris for hiring me and also the entire Gevaert group at VIB for welcoming me into the lab.

I would like to thank all the jury members for reading through the manuscript and for the valuable comments and suggestions which lifted the work to a higher level.

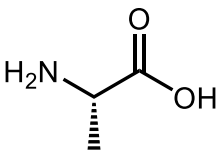
I would like to thank all my friends for the good times and great memories we shared and I look forward to keep doing that as much as possible! Lastly, I would also like to thank my parents and brothers for the continuous love and support with every step I make along the way in my scientific career and life. And of course, I can't forget my lovely girlfriend Sanne in this acknowledgement section. Without your love and support this would have been so much harder. With you by my side I feel confident that we will be able to overcome the challenges and grasp the opportunities that lie ahead. I can't wait to see what the future holds for us!

Klaas Decoene

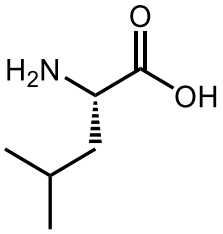
Ghent, August 2021

Amino acids

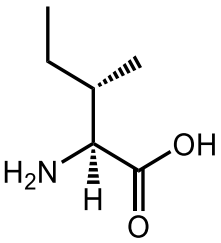
Non-polar



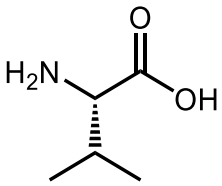
alanine
Ala
A



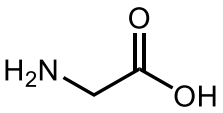
leucine
Leu
L



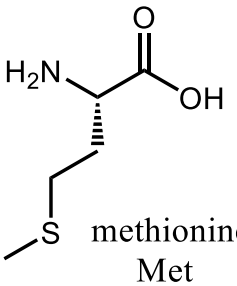
isoleucine
Ile
I



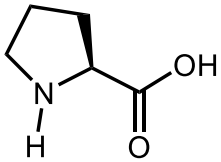
valine
Val
V



glycine
Gly
G

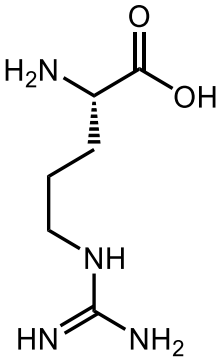


methionine
Met
M

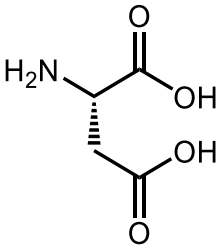


proline
Pro
P

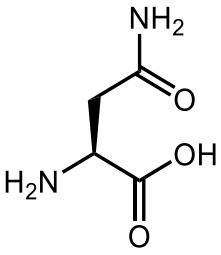
Polar



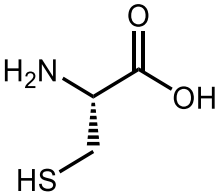
arginine
Arg
R



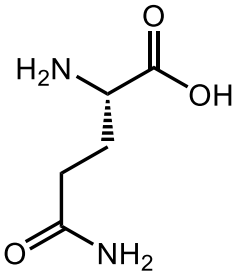
aspartic acid
Asp
D



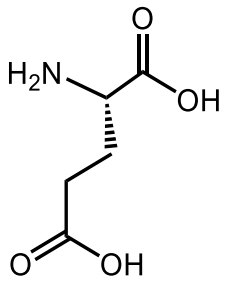
asparagine
Asn
N



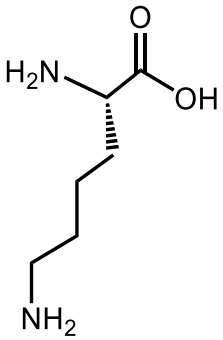
cysteine
Cys
C



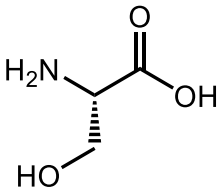
glutamine
Gln
Q



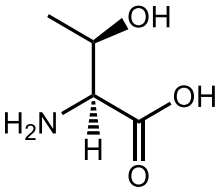
glutamic acid
Glu
E



lysine
Lys
K

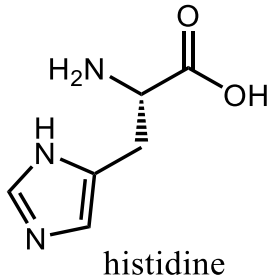


serine
Ser
S

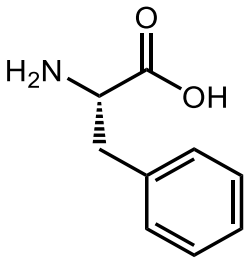


threonine
Thr
T

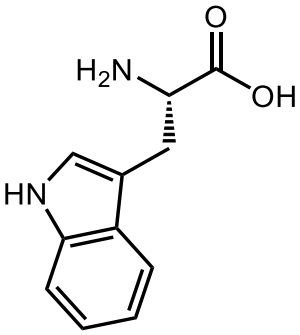
Aromatic



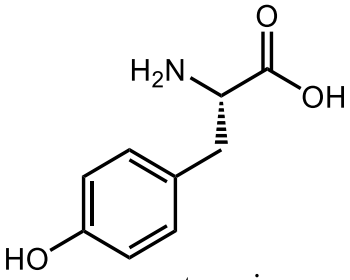
histidine
His
H



phenylalanine
Phe
F



tryptophan
Trp
W



tyrosine
Tyr
Y

Abbreviations

AA	amino acid
Acm	acetamidomethyl
ADC	antibody drug conjugate
ATP	adenosine triphosphate
Boc	tert-butyloxycarbonyl
BSA	bovine serum albumin
CBD	chitin binding domain
CD	circular dichroism
CID	collision induced dissociation
Cp	cyclopentadiene
Dha	dehydroalanine
DBCO	aza-dibenzocyclooctyne
DLS	dynamic light scattering
DMEQ-TAD	[2-(3,4-Dihydro-6,7-dimethoxy-4-methyl-3-oxo-2-quinoxalinyl)ethyl]-3*H*-1,2,4-triazole-3,5(4*H*)-dione
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EAS	electrophilic aromatic substitution
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ESI	electrospray ionisation
FAF	familial amyloidosis-Finnish type
FGE	formylglycine generating enzyme
FGly	formylglycine
FIT	flexible *in-vitro* translation
FPP	farnesyl pyrophosphate
FTase	farnesyl transferase
GGTase	geranylgeranyltransferase

GFP	green fluorescent protein
HCD	higher energy collision induced dissociation
HMPB	4-(4-hydroxymethyl-3-methoxyphenoxy) butyric acid
HPLC	high pressure liquid chromatography
HRP	horse radish peroxidase
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kd	dissociation constant
LC-MS	liquid chromatography mass spectrometry
LpIA	lipoic acid ligase
LYCH	Lucifer Yellow carbohydrazide
MALDI-TOF	matrix assisted laser desorption ionisation-time of filght
MeCN	acetonitrile
MeOH	methanol
MT1-MMP	membrane type 1-matrix metalloproteinase
NBS	*N*-bromo succinimide
ncAA	non canonical amino acid
NCL	native chemical ligation
NHS	*N*-hydroxysuccinimide
NMP	*N*-methyl-2-pyrrolidone
nt	nucleotide
pAzF	para azido phenylalanine
PBS	phosphate buffered saline
PCP	peptide carrier protein
PCR	polymerase chain reaction
PEG	polyethylene glycol
PNA	peptide nucleic acid
POI	protein of interest
Pyl	pyrrolysine
RaPID	random non-standard peptide integrated discovery
RB	Rose Bengal
RF1	release factor 1

RNA	ribonucleic acid
ROS	reactive oxygen species
Scm	*S*-carbomethoxysulfenyl
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate poly acrylamide gel electrophoresis
Sec	selenocysteine
sEGFR	soluble fraction EGFR
SPPS	solid phase peptide synthesis
SrtA	Sortase A
TAD	triazolinedione
TFA	trifluoroacetic acid
TTL	tubulin tyrosine ligase
Tris	trisaminomethane
VNB	vapor nano bubble
XIC	extracted ion chromatogram

Table of contents

1. Preface

Site-selective protein modification is of utmost importance for many applications from fundamental biology (fluorescent tagging) to therapeutic development (antibody-drug conjugates) and everything in between[1][2][3]. Owing to the growing interest of many researchers in academia and industry, this resulted in the booming of protein modification research and has not stopped since. A vast array of new strategies and methods for site-selective protein modification has been developed in recent years (figure 1.1).

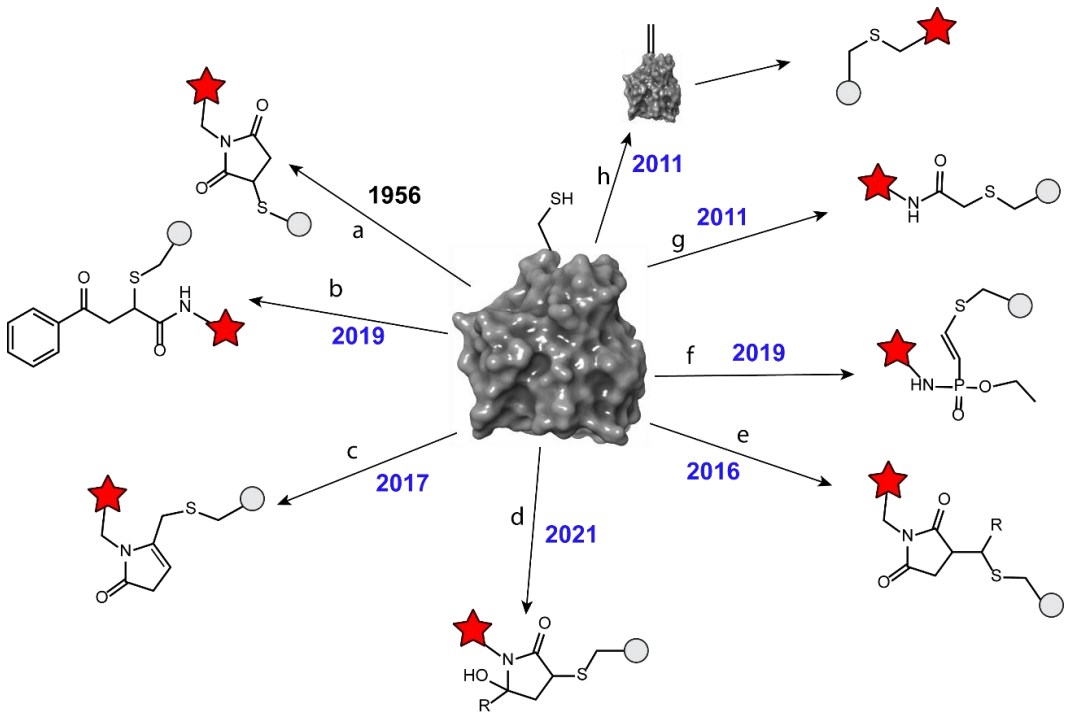


Figure 1.1 Schematic representation of cysteine modification methods. While the classical maleimide chemistry (a) was developed in the 1950's, methods b-h were reported in the last decade alone.

It is hard to overestimate the role proteins play in all living matter including the human body. Proteins are often rightfully described as the workhorses in biology since they are instrumental for an enormous variety of functions and processes. If we make a small simplification we could say that proteins in all living organisms are made up from the same 20 natural amino acids. It is remarkable how nature has harnessed these 20 simple building blocks to do so much and so many complex tasks. Proteins come in all sizes from only a dozen amino acids to proteins encompassing tens of thousands amino acids. Besides the amino acid building blocks also the way a protein is three-dimensionally folded, is of crucial importance for the function it is meant to carry out. The number of possible ways to fold a typical protein is astronomical, it was estimated in 1969 by Cyrus Levinthal to be in the range of 10^{300} for a protein of 150 amino acids. The inability to predict the protein folding from a known amino acid sequence is known among biologists as the folding problem. At the time of writing a team of DeepMind, an artificial intelligence company acquired by Google in 2014, recently made a huge breakthrough. They were able to correctly predict the 3D structure of proteins based on their amino acid sequence with their AI system AlphaFold[1]. This unexpected and ground breaking development will surely lead to another surge in protein research in the years to come.

[1] https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

The study of proteins in biological processes is a challenging endeavour. Proteins carry out their tasks often together with other proteins and/or other large or smaller organic molecules in a highly complex biological environment. Due to the overwhelming number of proteins and other compounds it is extremely complicated to investigate any individual protein. To overcome this issue scientists have developed many different strategies to make a modification to one specific protein to make it stand out from the crowd. This modification can consist in a fluorescent label to allow protein localisation via microscopy. Another modification can encompass the addition of an extra functionality for example to form an antibody drug conjugate. Other proteins have been modified to allow them to make a chemical connection to other proteins they interact with so as to map protein interaction patterns and networks. Although there are already many methods available for protein modification, there remains a powerful drive to enlarge this toolbox and to create better and more selective strategies to investigate, adapt and control the function of proteins.

2. Introduction and state of the art

2.1 Site selective protein modification

Proteins are a combination of 20 amino acid building blocks, which are very similar except that they all contain a different side chain. Some of these side chains contain a functional group that can be important for the task of the protein; other side chains are unreactive but can for example be important for the folding. Through the specific folding of the amino acid chain, nature manages to allow these functional groups to work together and to perform a whole variety of tasks. However, at some point in evolution the functionalities of the side chains were not sufficient for the countless tasks that proteins need to carry out. To overcome this problem nature developed enzymes, which are also proteins, to recognize a part of the sequence of a protein and site-selectively modify that protein with a functional group that is not present in the naturally available amino acid building blocks. By doing so the number of available functional groups can be expanded beyond the functionalities of amino acid building blocks rendering them better equipped for their complex tasks. This process is called post translational modification. Some modifications are included during the protein translation process, these modifications are called co-translational modifications. Like nature, scientists are also interested in expanding the structural diversity of proteins for all kind of modification purposes, as we will see later. Since proteins are often present in biological samples containing many different proteins we will refer to the specific protein under investigation as the protein of interest (POI) in what follows. At this stage a distinction between two types of POI's needs to be introduced.

On the one hand, there are pre-engineered proteins; these are proteins of which the amino acid sequence is engineered specifically for the selective modification of this protein. These proteins are typically produced by administering the DNA encoding for the POI to a host organism, which then transcribes the DNA into messenger RNA, which in turn is translated into a protein, a process that is called recombinant protein production. The host organism produces the POI as it produces the many thousands of proteins the host needs for its own vital functions and growth. A typical example of a pre-engineered protein is the combination of a POI and the green fluorescent protein (GFP). GFP was discovered in 1962 and the boundless use of GFP and related proteins as fusion constructs was recognized with the 2008 Nobel Prize in Chemistry.[4][5][6] A huge advantage of this POI-fluorescent protein approach is that the resulting fusion protein is fluorescent in itself and there is no need for additional modification. However the fluorescent protein part of such a fusion construct is large in size (ca. 200 amino acids) and this large cargo attached to the POI can have an impact on its biological function.[7]

This prompted scientists to investigate other strategies, making use of smaller labels to avoid interfering with the biological function of the POI. In one of the approaches, the pre-engineered POI sequence is adapted so that a part of the sequence is recognized by a specific enzyme which can modify the POI (part 2.1.1). Alternatively, the amino acid side chains of the engineered part of the sequence can be recognized by a chemical modification agent (part 2.1.2). Due to the increased biological understanding of protein translation, in the last decades scientists were able to develop customized protein expression methodologies. With these methods, it is possible to specifically introduce one or more noncanonical amino acids (ncAA) into a POI. These noncanonical amino acids allow decoration of the POI with bio-orthogonal functionalities, meaning that they are not present in any of the functionalities required to sustain life.[8] These new

functionalities can then be used for various purposes (part 2.1.3). Lastly, fully functional proteins can also be prepared by chemical synthesis of short peptide segments followed by linking these segments together. New functionalities can easily be introduced in the chemical synthesis however the completion of the finally envisaged protein can be laborious (part 2.1.4).

On the other hand, one can consider native proteins of which the amino acid is sequence is not altered. It is often particularly challenging to site selectively modify native proteins and even more to do so in biological samples (part 2.1.5).

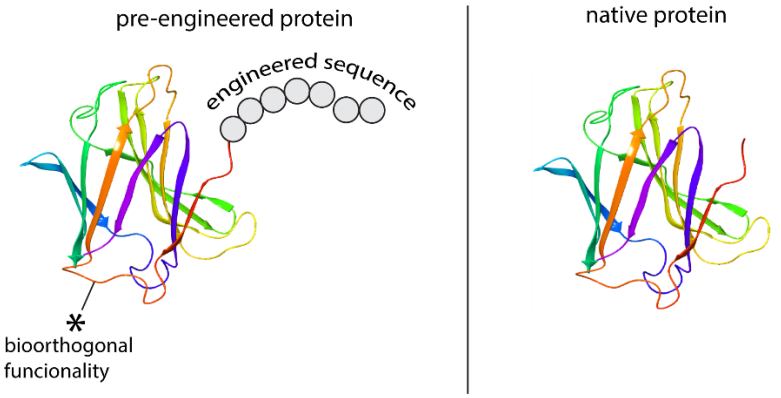


Figure 2.1 Schematic representation of a pre-engineered protein with a bioorthogonal functionality or an engineered sequence (left panel) or of a native protein (right panel).

2.1.1 Enzymatic protein modification

The enzymes responsible for post-translational modifications are extraordinarily specific since they are able to recognize only a specific sequence within the entire amino acid chain of proteins. This remarkable feature has inspired scientists to try aiming for the same level of site-selectivity by adaptation and use of such enzymes for site-selective protein conjugation. For this purpose, a series of amino acids are typically added to the N- or C-terminus of the protein as a recognition sequence. Enzymes are particularly suited for protein modification due to their ability to function in aqueous media, very high specificity and often fast reaction rates[9].

In this section, a selection of several widely used enzymes is described as well as some more recent approaches. Enzymatic reactions are typically high yielding but can often be reversible. The efficiency is very substrate- and protein-dependent and can vary a lot. For example tubulin tyrosine ligase (TTL) achieves a conversion of 90 % for the substrate meta-azido tyrosine but only 27 % conversion for para-azido phenylalanine.

In 2003, formylglycine generating enzyme (FGE) was discovered as the enzyme responsible for type I sulfatase activation in eukaryotes[10]. This FGE enzyme transforms a cysteine residue embedded in a 13 amino acid conserved sequence into a formylglycine (FGly) residue containing an aldehyde functionality. Several years later, Bertozzi and co-workers harnessed this enzyme for the modification of proteins[11]. The recognition motif with the target cysteine residue was optimized and reduced to 6 amino acids (CXPXR). This sequence was introduced in a POI and addition of FGE resulted in the aldehyde modified POI. The aldehyde functionality is a very interesting chemical functionality since it can selectively react with complementary aminooxy- or hydrazide functionalized reagents (figure 2.2).

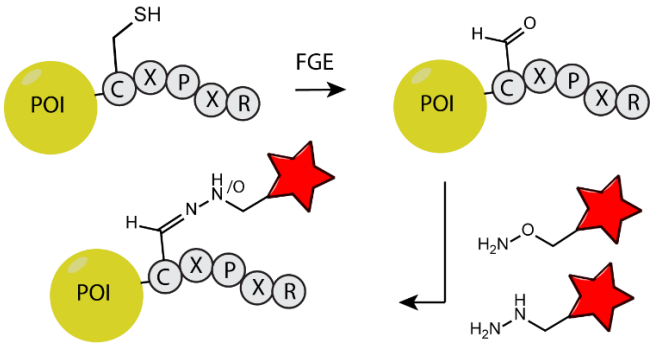


Figure 2.2 Schematic representation of a POI with a CXPXR motif followed by the enzymatic generation of formylglycine residue. Subsequently, an external reagent bearing an aminooxy or hydrazide functionality allows for site-selective modification.

Similar FGE enzymes are present in both *E. Coli* and mammalian cells, and the expression of a POI (containing the recognition motif) in these hosts allows for *in vivo* modification of the POI[12][13].

Another interesting approach is based on the mechanism of action of the Sortase A (SrtA) enzyme derived from *S. Aureus*.[14] SrtA is involved in a process called cell wall sorting. The SrtA recognition sequence (LPXTG) of a specific protein is cleaved in between the threonine and glycine amino acids. In a following step, the N-terminal part of the protein is linked via the carboxyl group of threonine to the pentaglycine cross-bridge in the peptidoglycan cell wall of *S. Aureus*.[15] Ploegh and co-workers have reported this method for protein modification in 2007.[16] The method works with a short recognition sequence of 5 amino acids and the only requirement for the label is that is should contain three N-terminal glycine residues (figure 2.3). Because of these very desirable properties sortagging was adopted widely for *in vitro* and *in vivo* protein labeling[17][18][19].

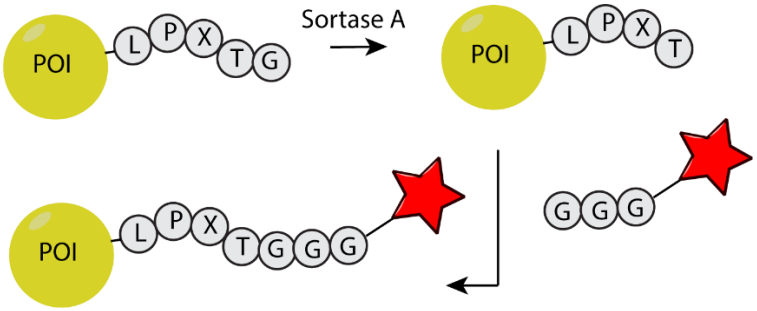


Figure 2.3 Schematic representation of a POI with a LPXTG motif and the enzymatic cleavage of the peptide bond between threonine and glycine, mediated by sortase A. In a second step also mediated by sortase A, a label agent containing a poly glycine motif is ligated to the carboxy terminus of threonine.

The polyvalence of this approach was further demonstrated by the ligation of very large molecules. Tanaka *et al.* were able to ligate an entire green fluorescent protein (27 kDa) to a cell surface protein using sortase.[17] Additionally also protein immobilization onto glass slides, derivatized with triglycine peptides, was reported.[20] Finally, also protein-protein conjugates were constructed using the sortase methodology.[21]

Biotin is a small ligand forming one of the strongest protein-ligand interactions known in nature (Kd = 10^{-15}) with the (strept)avidin protein.[22] Because of its small size and strong interaction, the biotin-(strept)avidin system is very often used in biotechnology and related fields. Typically, one component of a biological system of interest is biotinylated, this can be a peptide, a protein, DNA, RNA, etc. On the other hand, also large collections of functionalized (strept)avidin are available: fluorophores, horse radish peroxidase (HRP), solid phases among other functionalities allowing for an efficient experimental setup. It is often overlooked that biotinylation is also a rare post translational modification with only between 1 and 5 biotinylated protein species in certain organisms.[23] The enzyme responsible for protein biotinylation in *E. Coli* is BirA. The lysine in a conserved sequence of 15 amino acids (GLNDIFEAQKIEWHE) is selectively biotinylated by BirA. Ting and co-workers used this 15 amino acid sequence to biotinylate a cell surface protein and these biotinylated proteins were connected to streptavidin containing quantum dots.[24] Additionally they have demonstrated that BirA also accepts a ketone isostere of biotin (see figure 2.4 (b)), this structure was attached to lysine by BirA and the ketone functionality was used for reaction with aminooxy- and hydrazide labels.[25]

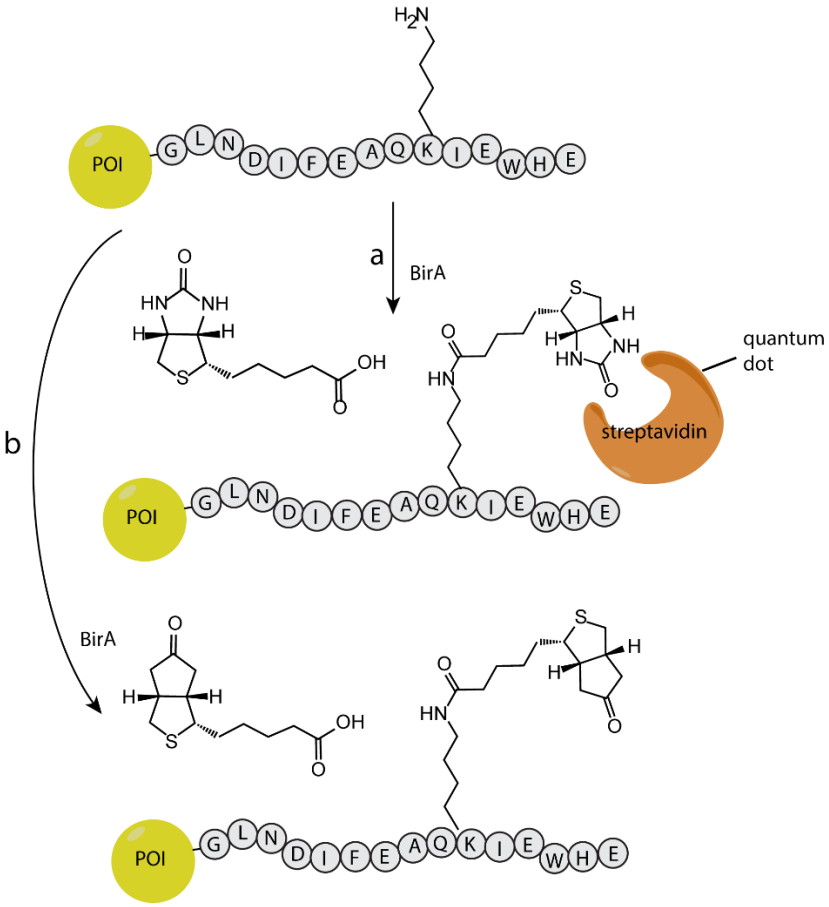


Figure 2.4 Schematic representation of the biotinylation process of a lysine residue in a conserved 15 amino acid sequence, catalysed by the enzyme BirA (a). This biotinylated POI can be used for decoration of quantum dots via the interaction with streptavidin. Additionally, an isostere of biotin with a ketone functionality can also be recognized by BirA and can be used for aminooxy- and hydrazide labelling after attachment to lysine (b).

Prenylation can be subdivided in farnesylation and geranylgeranylation corresponding to the covalent addition of a 15-carbon or 20-carbon poly-isoprene unit respectively. This post translational modification (PTM) occurs at a specific conserved cysteine residue near the C-terminus of proteins.[26][27] The farnesyltransferase (FTase) enzyme recognizes a conserved sequence often referred to as the CaaX-box ("C" is cysteine, "a" can be any of the aliphatic amino acids and "X" can be any amino acid). The attachment of isoprenoids to a protein promotes the membrane association and this effect can be attributed to hydrophobic interactions between the lipids in the membrane and the prenyl groups. FTase is the enzyme responsible for the farnesyl post translational modification using farnesyl pyrophosphate (FPP) as a substrate.[28][29][30] It was found that FTase is promiscuous and therefore farnesyl pyrophosphate analogues bearing bio-orthogonal functionalities for second step labelling have been developed. The main applications of these FPP analogues can be divided into two groups:

- In the first group, a protein of interest is engineered to have the Caax sequence at the C-terminus. After recombinant expression the protein is purified and in a second step the protein is farnesylated by FTase with a FPP analogue added to the mixture. This results in a protein with a unique functional group.

- In the second group, researchers are interested in native proteins that have the CaaX sequence. Upon addition of a FPP analogue to the cell medium the promiscuous FTase modifies native proteins with this analogue. Since regular FPP is naturally present in the cells, both the regular FPP and the FPP analogue will compete as a substrate for FTase this procedure results in a mixture of CaaX-proteins, some of which are modified with regular FPP while others bear a FPP analogue (Figure 2.5). When the cells are lysed, bio-orthogonal chemistry is then used to enrich for the proteins containing the FPP analogue. Afterwards, this enriched sample is analysed to identify CaaX-containing proteins. Zhao and co-workers developed a FPP-azide analogue for subsequent Staudinger ligation with a biotin probe for an enrichment step using streptavidin conjugated beads.[31] Distefano and co-workers used a similar FPP-azide for a copper mediated azide alkyne click reaction (CuAAC) for attaching proteins to alkyne-functionalized beads.[32] Shortly thereafter the Distefano group also reported the use of an FPP-alkyne for CuAAC chemistry with azide modified beads.[33] The FTase was even able to modify short pentapeptides with this FPP analogue. These short sequences were used to investigate the sequence specificity of FTase. Another approach developed by the same group is the use of an FPP-aldehyde for aminooxy and hydrazide labeling/immobilization purposes.[34][35]

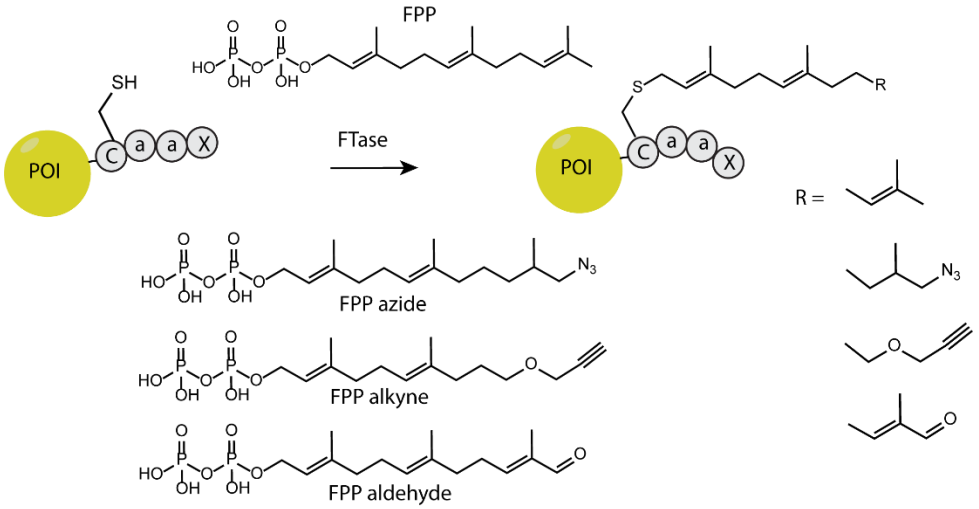


Figure 2.5 Schematic representation of protein farnesylation using FPP (above arrow). Structural representation of several bioorthogonal FPP analogues (below arrow).

In 2015, a new enzymatic protein labelling approach was reported by Hackenberger and co-workers. They use the tubulin tyrosine ligase enzyme (TTL) to attach tyrosine analogues containing several bio-orthogonal handles (Figure 2.6).[36][37] The function of TTL in nature involves attaching a tyrosine residue to the C-terminus of α-tubulin.[38] The recognition sequence for TTL is the fourteen amino acid sequence VDSVEGEGEEEGEE, this tag can be engineered at the C-terminus of a POI. A wide variety of bio-orthogonal handles were shown to be accepted by TTL through the use of azide, alkyne, fluorescent and biotin containing tyrosine analogues among others.

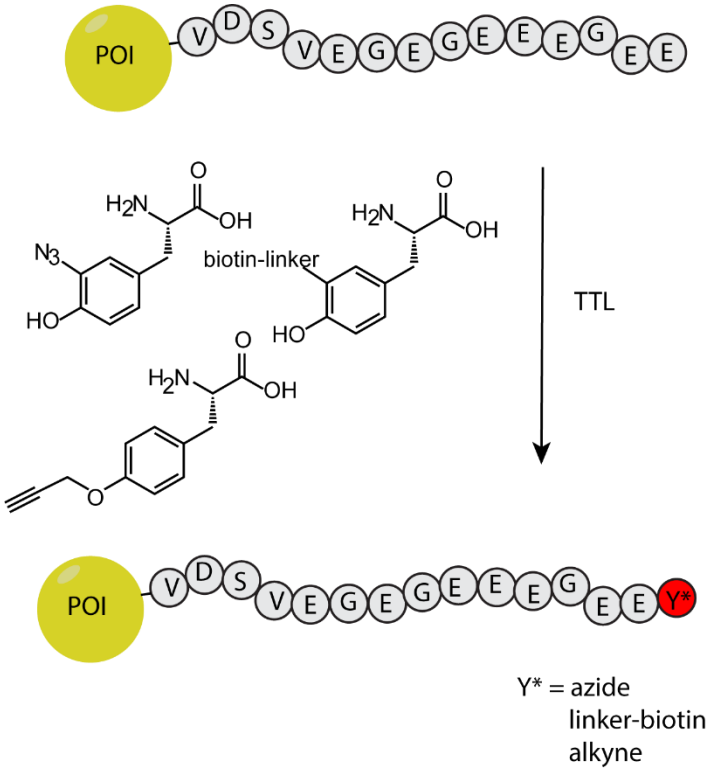


Figure 2.6 Schematic representation of protein modification using tubulin tyrosine ligase (TTL) with several tyrosine analogues (azide, biotin, alkyne).

Another interesting enzymatic protein labelling approach was introduced by Ting and co-workers who employ lipoic acid ligase(LpIA).[39] The LpIA enzyme originates from *E. Coli* where it ligates lipoic acid to proteins in a Mg^{2+} and ATP dependent process. The promiscuity of LpIA was harnessed for the ligation of an alkyne analogue of lipoic acid. In this example, the LpIA recognition sequence was engineered through iterative cycles of rational design to find a minimally invasive tag that can be fused to any POI. A 22-amino acid long sequence was identified and shown to allow efficient LpIA ligation.

An approach utilizing the Sfp phosphophantetheinyl transferase enzyme was first reported in 2004.[40] In this approach a POI is expressed as a fusion construct with peptide carrier protein (PCP). PCP's are domains or subunits in the assembly line of large nonribosomal peptide synthetase complexes and they can be as small as 80 amino acids. These PCP domains each have one specific serine residue, which acts as the substrate for phosphopantetheinyl modification. Walsh and co-workers used the remarkable substrate promiscuity of Sfp by using a phosphopantetheinyl biotin conjugate for the labeling of a POI-PCP fusion construct. A year later the same group reported that a new much shorter peptide recognition sequence (11 amino acids) can be used with the Sfp phosphopantetheinyl transferase enzyme and several fluorescent moieties could be attached to a POI.[41]

Another method makes use of the human enzyme DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT), this protein labeling method was developed by Johnsson and co-workers.[42] hAGT is involved in DNA repair by removing an alkylgroup on the O^6 of an alkylated guanine. Similar as in other examples of enzymatic protein modification described in this section, scientists have taken advantage of the enzyme promiscuity. Johnsson *et al.* showed that an alkylated nucleobase (O^6-benzylguanine) is readily accepted as a substrate for hAGT meaning that the alkylated guanine should not necessarily be incorporated in DNA. The alkyl group on the substrate is transferred to a cysteine residue of hAGT. This remarkable property of hAGT was harnessed for protein labelling via the generation of POI-hAGT fusion constructs. When alkylated guanine variants containing a biotin moiety or fluorescein were added to this fusion protein the functionalized alkyl chain was transferred to the fusion construct yielding a labelled protein.[43] The name SNAP-tag was coined for this method (Figure 2.7). A similar method was developed for the enzyme responsible for DNA repair of modified cytosine residues and termed CLIP-tag.[44]

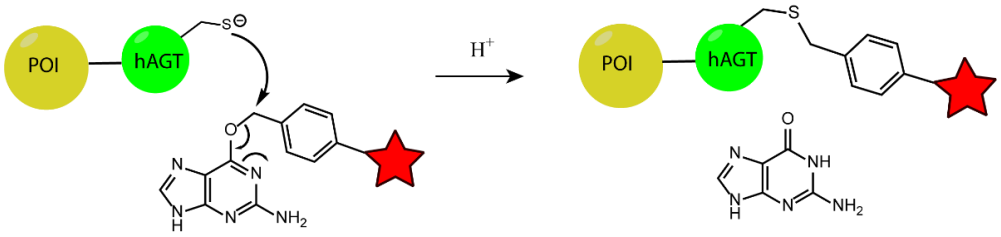


Figure 2.7 Schematic representation of the SNAP-tag self labeling protein tag. The functionalized alkyl group on a guanine is transferred to a POI-hAGT fusion protein.

The mechanism of these SNAP and CLIP tags is somewhat different from the other enzymatic methods described above. In this case, the POI is expressed as a fusion construct with the hAGT enzyme (SNAP-tag). Upon addition of the labelled guanine, the hAGT tag labels itself, which is why these are often called self-labelling protein

tags. Similar to the example of GFP fusion proteins an important downside of this approach is the size of the tag. The SNAP tag by itself is 182 amino acids long and this can perturb the function of the POI it is fused to. In the other methods presented in this section, POI's with a specific enzyme recognition sequence are used. A specific enzyme recognizes this sequence and is able to modify the POI. In these examples the POI and the enzyme are two different protein species.

The SpyTag/SpyCatcher system (Figure 2.8) was reported in 2010 by Zakeri *et al.*, and the system is based on the bacterial adhesin fibronectin binding protein.[45][46] A specific domain in this protein contains a spontaneous isopeptide bond between the ε-amine of a specific Lys and a specific Asn. By dividing this domain in two parts and rational design of the fragments, a 13-residue peptide tag (SpyTag) was obtained, capable of covalently attaching to a 116-residue protein partner (SpyCatcher). Expression of a POI-SpyCatcher fusion construct allows for covalent modification with the SpyTag. This approach has many applications, for example protein hydrogels were prepared using the SpyTag/SpyCatcher chemistry.[47]

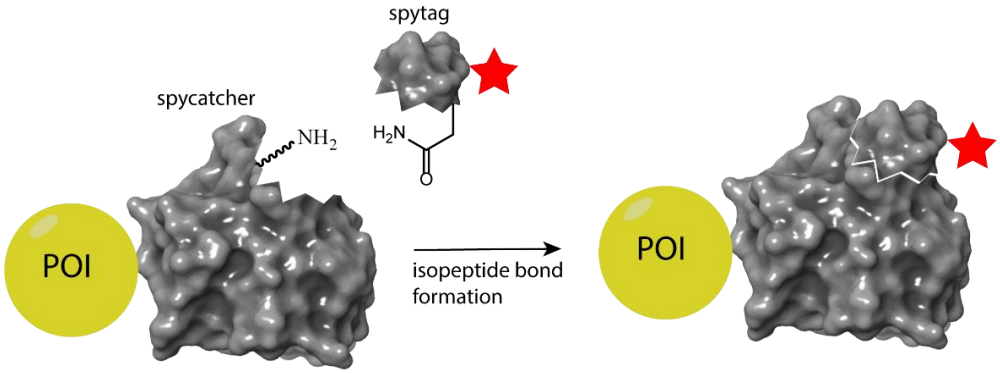


Figure 2.8 Schematic representation of the spytag/spycatcher approach.

2.1.2 Sequence engineered chemical modification

The amino acid, which is by far the most involved in sequence engineered chemical modification of natural amino acids is cysteine. This amino acid with an abundance of 1.3 %[48] is the second least abundant of all. This low abundance means that in proteins there will not be many cysteine residues present. Moreover cysteine residues are often involved in disulphide bridges, meaning that they are unlikely to participate in modification reactions. This rareness in combination with its high nucleophilicity (pKa of 8.2) makes cysteine an ideal target for many chemical modifications.[49] Moreover this pKa is highly influenced by the nature and relative positions of neighbouring amino acid side chains.[50] By carefully engineering the sequence of a POI it can be possible to create a uniquely reactive cysteine residue on a specific location within the protein.

Tsien and co-workers have discovered a small genetically encoded tag consisting of 6 amino acids CCXXCC that can selectively react with biarsenic fluorescent dyes (figure 2.9). Selective recognition occurs when this motif is situated in a helical or hairpin structure.[51][52] It was shown that these dyes are fluorogenic so they are much more fluorescent when protein bound. This feature ensures a low background and remediates the need for excessive washing steps. However, these molecules also have their drawbacks since they suffer from off-target hydrophobic interactions and competition with other natural thiols among other things.

In an effort inspired by the work of Tsien, Schepartz and co-workers have reported a rhodamine based dye that can bind to a SSPGSS sequence when present on a hairpin structure.[53] This bisboronic dye is also fluorogenic but has less drawbacks compared to the bisarsenic alternatives, there is less off-target binding and there is no need for the cytotoxic arsenic.

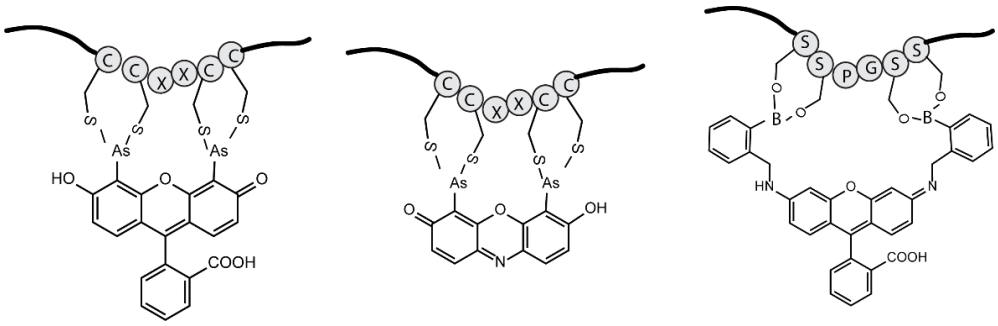


Figure 2.9 Schematic representation of two bisarsenic fluorescent dyes bound to a tetracysteine CCXXCC sequence (left, middle). Schematic representation of a bisboronic fluorescent dye bound to a tetraserine SSPGSS sequence (right).

Pentelute and co-workers developed another example of engineered sequence cysteine conjugation. They reported a very short FCPF sequence for site-specific modification of the cysteine. It is hypothesized that the aromatic phenylalanine residues activate the cysteine thiol and interact with the perfluoroaryl moiety. The proline ensures a conformation where the two phenylalanines “sandwich” the perfluoroaryl on both sides with π-π interactions. This so-called π-clamp sequence is typically inserted at the C-terminus of a POI. Perfluoroaromatic reagents can selectively react with the cysteine of the π-clamp in presence of other endogenous cysteines and even with competing free thiols (figure 2.10).[54] In water, amines do not react with these perfluoro-aromatic reagents and amino acid selectivity was shown via MS/MS analysis. Additionally, the same group reported on another tag for regioselective cysteine modification. The aza-dibenzocyclooctyne-tag DBCO-tag was selected from a library screen for peptides that increase the reaction rate with DBCO linked probes. The optimized DBCO-tag allows for a 220 fold increased thiol-cyclooctyne modification rate in comparison to a cysteine control peptide.[55] The DBCO-tag is seven residues long (LCYPWVY) and can be engineered both at the N- and C-terminus of a POI. Both the π-clamp and DBCO-tag methods have been used for the generation of antibody drug conjugates (ADC’s).

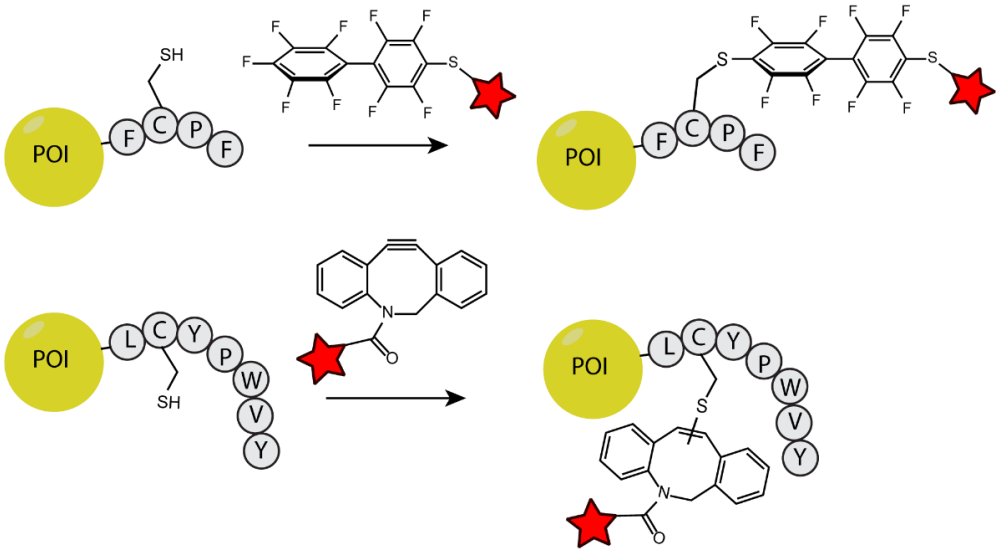


Figure 2.10 Schematic representation of chemical protein modification of a specific cysteine residue in an engineered π-clamp sequence with a perfluoro aromatic reagent (top). Schematic representation of chemical protein modification of a specific cysteine residue in an engineered DBCO-tag with a DBCO.

In a different approach, the Jensen group uses a short N-terminal sequence (GHHHHHH) for selective acylation. The method is based on N-terminal gluconoylation, a side reaction that occurs during protein expression of proteins with an N-terminal His-tag. This side reaction was harnessed for N-terminal acylation of a protein with the GH_6-tag.[56] As labeling reagents, 4-methoxyphenyl esters are used with different functionalities such as azides, biotin or fluorescent moieties.

2.1.3 Customized protein expression systems

2.1.3.1 Genetic code expansion

Genetic code *expansion* is based on the natural incorporation of pyrrolysine, the 22nd natural amino acid. In 2002, pyrrolysine was discovered by Srinivasan and co-workers at the active site of the methyltransferase enzyme in a methane producing archaea species named *Methanosarcina barkeri.*[57] This noncanonical amino acid is incorporated into the growing protein in response to an amber stop codon. This means that instead of being interpreted as a termination signal after the last sense codon on the mRNA, this amber stopcodon (UAG) is read as a regular codon for incorporation of pyrrolysine. For each amino acid, the organism requires an orthogonal tRNA/tRNA synthetase pair, the tRNA synthetase attaches the amino acid of interest to the corresponding tRNA and this tRNA has an anticodon that can hybridize with a specific codon on the mRNA. This ensures that only the amino acid corresponding to the codon on the mRNA is incorporated in the growing protein chain. The species that are able to incorporate pyrrolysine have developed an additional tRNA/tRNA synthetase pair for incorporation of pyrrolysine in response to the amber stop codon. To avoid unintentional read through of the UAG amber stop codon elsewhere in the *methanosarcinaceae* proteome it was shown that the

UAG stop codon is used at a much lower frequency and if used, it is often immediately followed by a different stop codon.[58]

Scientists have adopted this pyrrolysine system for the incorporation of many noncanonical amino acids. An orthogonal tRNA/tRNA synthetase pair was engineered for each noncanonical amino acid. Additionally, also a POI with an amber stop codon was engineered. A host organism having the orthogonal tRNA/tRNA synthetase pair and the DNA code for the POI is capable of expressing this POI with the noncanonical amino acid included at the location defined by the amber codon. Since the host organism cannot synthesize this new amino acid, the new amino acid must be added to the host cells for this approach to work. This method was first demonstrated with *O*-methyl-L-tyrosine in *E. Coli* by Schultz and co-workers.[59] Shortly thereafter the same group expanded the genetic code of eukaryotes with noncanonical amino acids containing a series of interesting chemically reactive functionalities including a ketone, an azide and a benzophenone moiety.[60] Later, Chin and co-workers also reported a way to incorporate multiple unnatural amino acids by using quadruplet codons.[61] To date many noncanonical amino acids with a plethora of new functional groups have been introduced via this approach.[62][63][64] In section 2.2 this approach will be discussed in more detail.

2.1.3.2 Genetic code reprogramming

In genetic code *reprogramming* the translation system is reprogrammed so that codons regularly coding for canonical amino acids can be used for noncanonical amino acids. This is in contrast to the genetic code *expansion* methodology where only nonsense or quadruplet codons, which are not involved in the translation of canonical amino acids, are used for introduction of noncanonical amino acids. Genetic code *reprogramming* can only be carried out in a cell free system. The molecular machinery required for protein production is typically obtained from cell lysates and used in a buffered solution. By using only purified components, Ueda and co-workers were able to attain greater reproducibility and improved RNA stability.[65][66] Additionally, this approach allows for custom protein production in a cell free system.

Suga and co-workers developed a cell free translation system termed 'Flexible *in vitro* Translation' (FIT). Key in the FIT system are the flexizymes (flexible tRNA acylation ribozymes), these are crucial for the amino acylation of tRNA. In 2001 a prototype flexizyme, r24, consisting of 130 nucleotides (nt) capable of aminoacylating a tRNA with phenylalanine cyanomethyl ester (Phe-CME), was identified.[67] The first flexizyme, Fx3 (45 nt in length) was adapted from this prototype. The Fx3 requires an aromatic moiety in the side chain for recognition (figure 2.11 a, blue) and a cyanomethyl ester leaving group (red). Later generation flexizymes were developed with greater activity and improved substrate specificity, where the aromatic moiety for recognition is included in the leaving group (figure 2.11 b-c). These improved flexizymes offer more structural freedom for amino acids that can be aminoacylated by flexizymes. In theory, any amino acid with a suitable leaving group can be aminoacylated to the tRNA using flexizymes.[68][69]

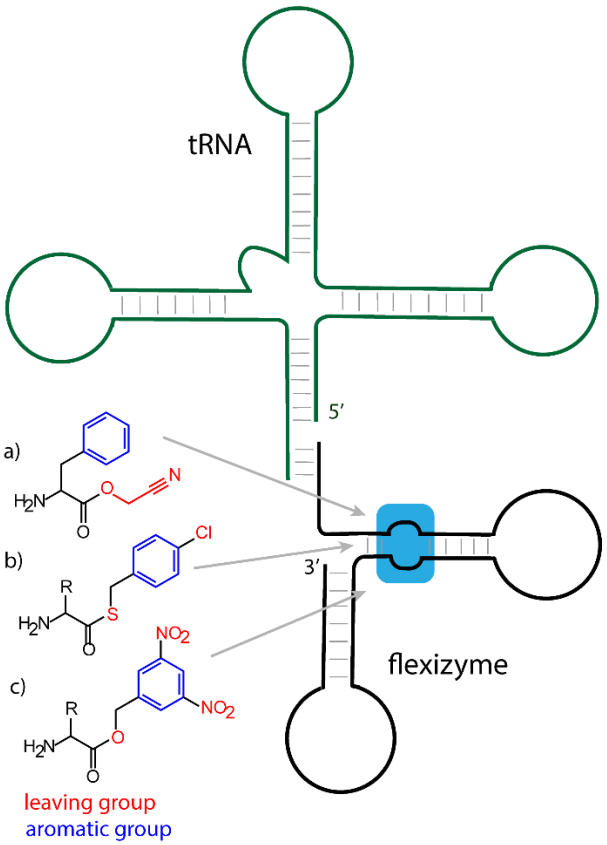


Figure 2.11 Schematic representation of a flexizyme hybridized with a tRNA in the process of aminoacylating a noncanonical amino acid to the 3' end of the tRNA. The blue box indicated on the flexizyme indicates the aromatic recognition site. a) Fx3 is capable of aminoacylating an amino acid substrate with an aromatic moiety in the side chain and a cyanomethyl ester leaving group. b-c) later generation flexizymes are capable of introducing amino acid substrates with any side chain since the aromatic moiety is included in the leaving group.

The use of these flexizymes which allow aminoacylation of tRNA's with noncanonical amino acids in combination with cell free protein expression allow for a very powerful *in vitro* translation method. In section 2.2 this topic will be discussed in more detail.

2.1.4 Chemical synthesis of proteins

In all the methods described above the POI is synthesized by the power of nature's enzymes. In sections 2.1.1 and 2.1.2, the POI was a pre-engineered protein which had a recognition sequence for either enzymes or chemical modification agents. These POI's were expressed in a host organism and consist solely of a combination of the 20 canonical amino acids. In section 2.1.3 the POI was still synthesized by natural enzymes but with more control over the translational system (especially in part 2.1.3.1). These POI's can bear multiple noncanonical amino acids besides the canonical residues. In the current section, the POI is entirely made by chemical means thus offering total control over all incorporated functionalities.[70] In 1994, Dawson *et al.* reported on a simple method for the generation of a native peptide bond between two unprotected synthetic peptides.[71] This method was termed native chemical ligation (NCL) and

relies on the reversible chemo-selective formation of a thioester followed by an irreversible intramolecular *S* to *N* acyl shift leading to the amide bond (figure 2.12).

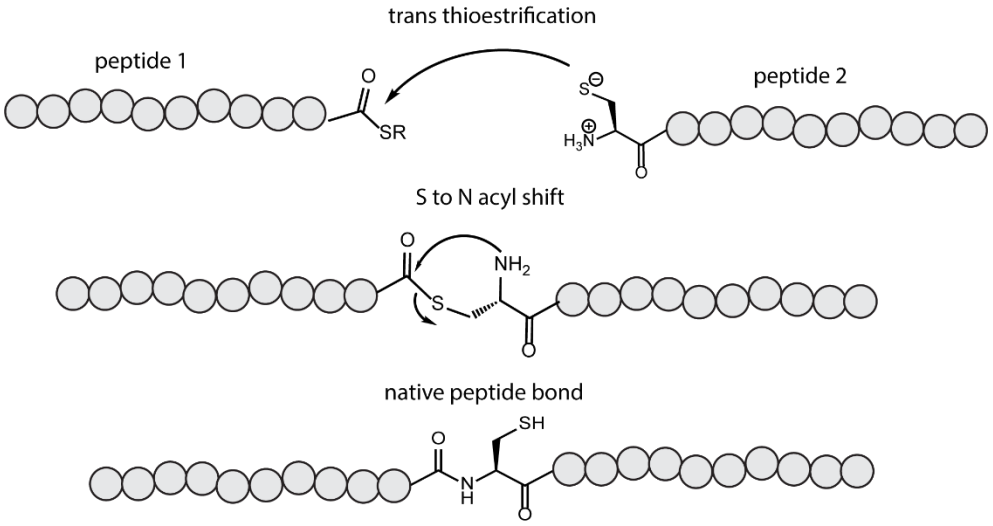


Figure 2.12 Schematic representation of the native chemical ligation. The first step involves a reversible trans-thioesterification followed by an intramolecular *S* to *N* acyl shift leading to a native peptide bond.

The cysteine dependence of this method was remediated by a metal catalysed desulfurization of the cysteine to an alanine residue by Dawson and co-workers in 2001.[72] Additional refinements include the use of a latent thioester for the ligation of glycopeptides and free radical desulfurization by Danishefsky and co-workers.[73][74]

Several other thiol-independent chemical ligation strategies have been developed including a method based on the Staudinger ligation between a thioester and an azide,[75] a method making use of serine or threonine residues in combination with peptide salicylaldehyde ester[76] and another method in which α-ketoacid-hydroxylamine (KAHA) peptides are used for ligation resulting in a homoserine residue at the ligation site.[77]

2.1.5 Native protein modification

In native protein modification, all the functional groups of a protein compete for a modification reagent. New reactions for the modification of native proteins are typically reported with a study of the reactivity on several levels: amino acids/small molecules, peptides and proteins. The analysis of the modification reaction on these levels requires a different approach. For the amino acid/small molecule level simple LC-MS or NMR experiments can be used for analysis. On the other hand, for proteins a combination of MS and MS/MS experiments allows to carry out an in depth study of the modified protein. Since the reactivity of the amino acid side chains is influenced by their environment, it is especially important to investigate the amino acid selectivity of the modification reagent. Peptides or proteins specifically not containing the targeted residue should be used as a negative control reaction, similar peptides and proteins encompassing the targeted residue can then be used as a positive control to confirm the selectivity of the modification reaction. Some methods are more site-selective than others, however side reactions should always be considered in the context of the specifically envisaged application.

Cysteine labelling

As a result of its low abundance and high nucleophilicity, cysteine has become one of the prime targets for amino acid selective protein modification. The most widely used method for cysteine modification is based on maleimide chemistry and maleimides were already used in the 1950's for reaction with cysteines (see Figure 2.13, (a)).[78][79] This robust reaction proceeds via a Michael addition and the product is a thiosuccinimide bond. There are some disadvantages associated with the thiol maleimide reaction like the possibility for retro Michael reaction and the high tendency towards hydrolysis.[80][81][82] Despite these drawbacks maleimides are often used for the generation of antibody drug conjugates.[83] Many refinements to address the shortcomings of the classical maleimides have been made: Bernardes and co-workers reported carbonylacrylic reagents for irreversible cysteine conjugation(see Figure 2.13, (b)),[84] Zhou and co-workers reported the use of 5-methylene pyrrolones for specific and traceless bioconjugation (see Figure 2.13, (c)).[85] De Geyter *et al.* recently reported a method for efficient single site multi-functionalisation using 5-hydroxy-pyrrolone structures (see Figure 2.13, (d)).[86] Kalia *et al.* reported the use of exocyclic olefinic maleimides for more stable conjugates (see Figure 2.13, (e)).[87]

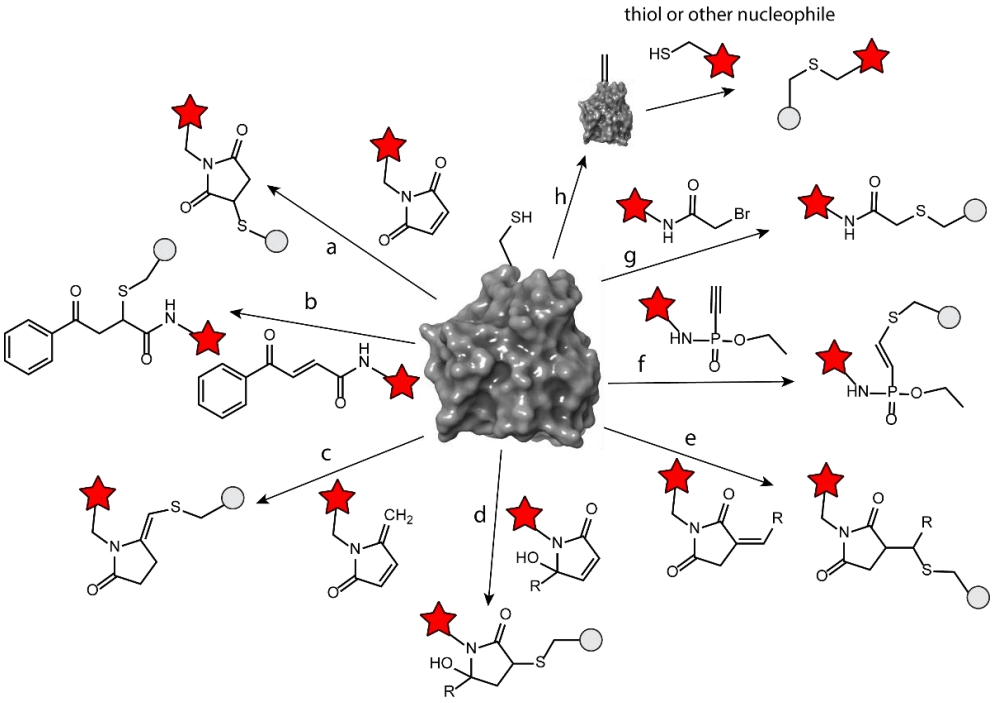


Figure 2.13 Schematic representation of cysteine modification methodologies.

Several other approaches for cysteine modification have been developed. Hackenberger and co-workers reported on a phosphonamidite reagent based on the thiol-yne reaction (see Figure 2.13, (f)).[88] Another popular strategy relies on the use of bromo- or iodoacetamides for S_N2 substitution reaction with thiols (see Figure 2.13, (g)).[89] However the drawbacks of these methods are associated with other cross reactivity of nucleophiles (lysine and histidine).

The Davis group developed an interesting and complementary approach based on cysteine modification. Dehydroalanine (Dha) is a natural amino acid that can occur in proteins, as a consequence of the post translational modification of certain serine residues and to lesser extent cysteine residues.[49] Davis and co-workers reported an

efficient method to convert cysteine to dehydroalanine (see Figure 2.13, (h)).[90] This unique electrophilic group can be targeted with nucleophiles and especially thiol groups in a second stage.

Lysine labelling

As a nucleophilic residue, lysine has attracted considerable attention for protein modification. As opposed to cysteine however, lysine is a high frequency residue with an abundance of 5.8 %[48] (cfr. cysteine abundance 1.3 %). Moreover, lysine residues can be positively charged and are often displayed on the surface of the protein. These combined properties make lysine an interesting but omnipresent target. Additionally there can be competition of the N-terminal amine with the N^{ε}-amine of lysine residues. The most used method for lysine modification is the reaction with N-hydroxysuccinimide (NHS) esters (see Figure 2.14, (a)). NHS esters are also used for generation of ADC's, and since several lysines are present on the protein surface the antibody will be decorated with a number of drug molecules.[91][92] Other approaches for lysine modifications involve the use of sulfonyl chlorides (see Figure 2.14, (b)) and iso(thio)cyanates to form ureas or thioureas (see Figure 2.14, (c)).[8] Additionally, a reductive amination reaction for modification of lysines was developed by McFarland and Francis (see figure 2.14, (d)), this method is Iridium catalysed and takes place at pH 7.4.[93]

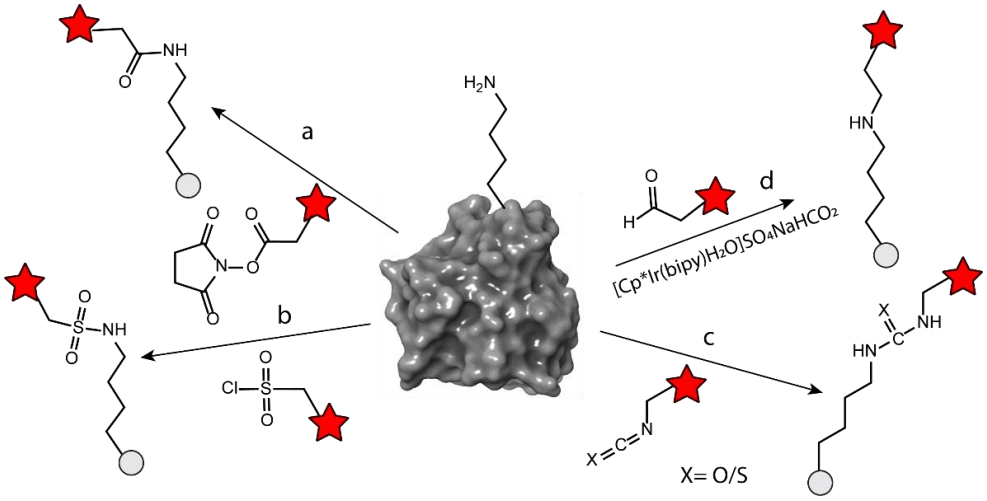


Figure 2.14 Schematic representation of lysine modification strategies.

Tryptophan labelling

As the least abundant amino acid (just over 1%)[48], tryptophan is regarded as a highly interesting target for site selective modification since labelling. Tryptophan is chemically not the most tractable amino acid; nevertheless, several approaches have been developed for protein modification. Francis and co-workers showed rhodium carbenoid-based Trp labeling at mild pH (see Figure 2.15, (a))[94] An organoradical Trp conjugation was demonstrated on peptides and proteins by Kanai and co-workers (see Figure 2.15, (b)).[95] Very recently, a novel biomimetic approach for the selective conjugation of tryptophan in peptides and proteins was developed by Tower *et al*, this method employs UV irradiation and N-carbamoylpyridinium salts (see Figure 2.15, (c)).[96]

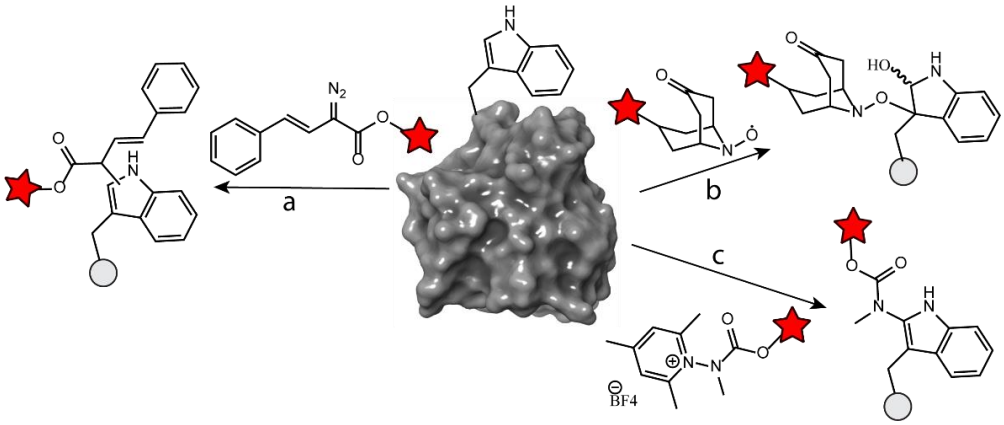


Figure 2.15 Schematic representation of tryptophan modification strategies.

Tyrosine labelling

Tyrosine, like tryptophan is an aromatic amino acid with a moderate to low presence on protein surfaces making it an interesting target for protein modifications. An elegant three component Mannich reaction was reported by Francis and co-workers (Figure 2.16, (a)).[97][98] This approach involves the protein tyrosine, an aldehyde and an electron rich aniline. Another method for oxidative crosslinking between two proteins via tyrosines was reported by Kodadek et al.[99] This method is transition metal mediated and results in a dityrosine crosslink(Figure 2.16, (b)). In an adaptation of this work a similar reaction was used for the site selective modification of tyrosine residues (Figure 2.16, (c)).[100] Also ene like reactions were reported with triazolinedione (TAD) modifications as the most prominent example (Figure 2.16, (d)).[101][102][103] In this PhD work, we have observed an important off target reaction taking place at the tryptophan residues*, an effect which was overlooked for more than a decade by many researchers. In chapter 6.3 of this thesis, the tyrosine-ene reactions, especially with TAD compounds will be discussed in more detail.

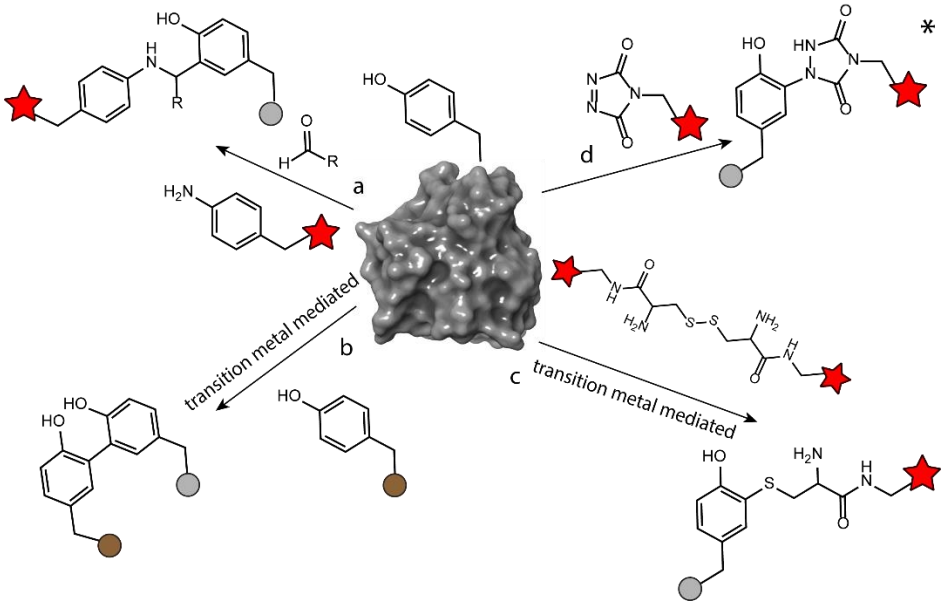


Figure 2.16 Schematic representation of tyrosine modification strategies.

Histidine labelling

With a p*K*a of 6 the histidine side chain has a p*K*a value which is, among all amino acids, most close to neutral.[104] This implicates that histidine can effortlessly function as acid-base catalyst which is the reason why histidine is the residue most often found in the catalytic site of enzymes. Besides its acid-base properties, histidine can also act as a nucleophile. Strategies for histidine modification exploit this nucleophilic property. While other nucleophilic residues are tough competitors in these strategies, several groups have recently reported histidine selective labelling strategies. In 2019, Chang and co-workers presented a bioinspired thiophosphorodichloridate approach (Figure 2.17, (a)).[105] This elegant strategy is inspired by natural histidine phosphorylation and was used for labeling proteins with several azide and alkyne tags for second stage functionalization. An alternative two-step approach was reported by Rai and co-workers in the same year.[106] They harnessed 2-cyclohexenone for selective reaction with histidine followed by oxime ligation to the newly introduced ketone moiety (Figure 2.17, (b)).

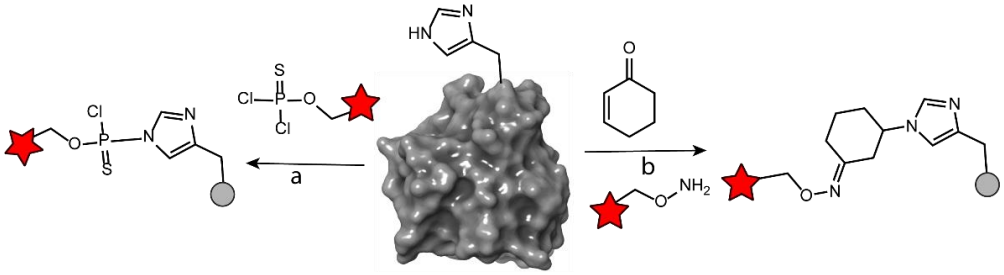


Figure 2.17 Schematic representation of histidine modification strategies.

Carboxylic acid labelling

The side chains of aspartic and glutamic acid as well as the C-terminus contain carboxylic acid functionalities. Both aspartic and glutamic acid are abundant residues and often present on protein surfaces. Together with the C-terminus, this abundance makes the carboxylic acid a very challenging target. Nonetheless, MacMillan and co-workers reported a photoredox method for selective C-terminal modification (figure 2.18).[107] This approach relies on the difference in oxidation potential between the aspartic and glutamic acid side chains on one side and the C-terminus on the other hand.[108] Since the C-terminus is more readily oxidized compared to the side chain carboxylic acids an oxidation approach allows to develop a chemoselective method. Moreover, since on most protein structures there is only 1 C-terminus, this method offers potential for single site modification on native proteins.

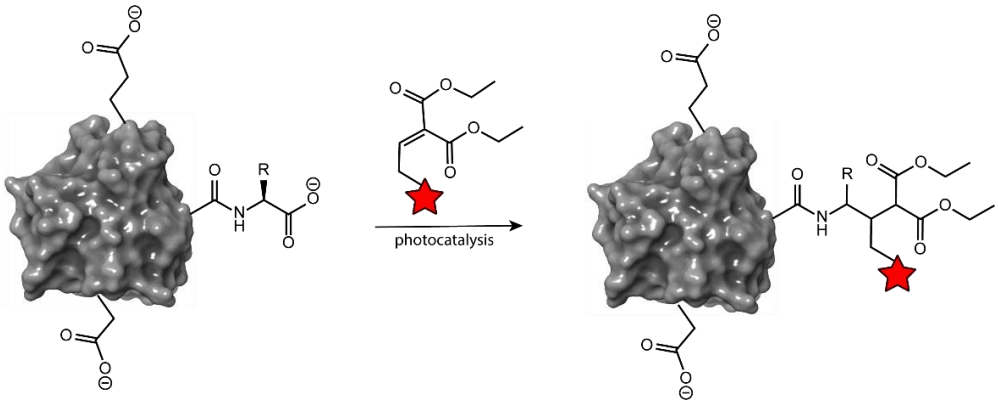


Figure 2.18 Schematic representation of a C-terminal selective labelling strategy.

Other methods for general carboxylic acid labelling include reactions with diazo reagents.[109][110]

N-terminus labelling

Similar to the C-terminus, the N-terminus of a protein is often exposed, a study demonstrating that a protein terminus is exposed in 87.1% of the cases.[111] Several methods were developed for the selective modification of the N-terminus in presence of lysine residues which are also a source of protein amines. The p*K*a value of N-terminal α-amines is situated between 6 and 8 and thus substantially lower than the p*K*a of ε-amines which resides around 10.5. This effect is caused by the nearby carbonyl group in the case of N-terminal α-amines. Note that in eukaryotic cells the N-terminal amine can be blocked by acylation via N-terminal acyltransferases.

By using modification reactions in a pH controlled environment, the N-terminus can be targeted in the presence of lysine residues. Similar as for the labelling of lysine amines, NHS modification and reductive amination reactions were used to aim for N-terminal selective modification by controlling the pH at 4-5.[112] In another approach, azides could be introduced on all amines by reacting at pH 11-12 and additionally a single azide could be introduced at the N-terminus when the reaction was done at pH 8.5 (figure 2.19, top).[113] Alkynes on the other hand can be introduced via a ketene reaction at pH 6.3 (figure 2.19, bottom).[114] Both methods lead to a site selectively installed click chemistry handle for subsequent labelling reactions.

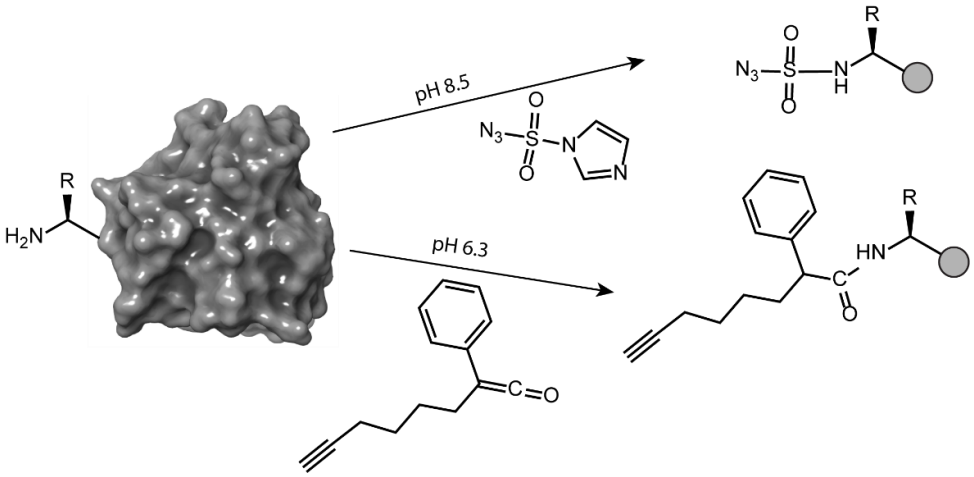


Figure 2.19 Schematic representation of N-terminal modification strategies.

Besides reactions involving pH control, efforts were also undertaken which involve the side chain of the N-terminal residue. The most well-known example of such an approach has been discussed in section 2.1.4 on native chemical ligation where a N-terminal cysteine residue plays a pivotal role. Since natural protein translation starts with a methionine residue in response to the start codon, N-terminal residues different from methionine are relatively rare. This causes the methods below to be typically used with engineered proteins. I chose not to present these methods in the section of engineered proteins since there is no real recognition sequence required but rather one specific amino acid at the N-terminus. Gois and co-workers have shown addition of benzaldehydes and reversible addition of boronic acid benzaldehydes to N-terminal cysteine residues.[115] Another method for modification of N-terminal cysteine residues uses modified 2-cyanobenzothiazole compounds for selective modification.[116] For N-terminal tryptophan residues a method involving a Pictet-Spengler reaction at the modification site was developed by Tam and co-workers.[117]

Ligand directed modifications

Ligand directed modifications exploit the selective interaction of a protein with a ligand to modify the protein in a site-selective fashion. These strategies use a ligand connected with a linker to a label moiety; upon binding of the ligand, a proximate nucleophile on the protein can react with the linker to form a covalent bond with the label moiety (see Figure 2.20). Hamachi and co-workers have developed a series of linker chemistries for ligand directed modifications.[118]

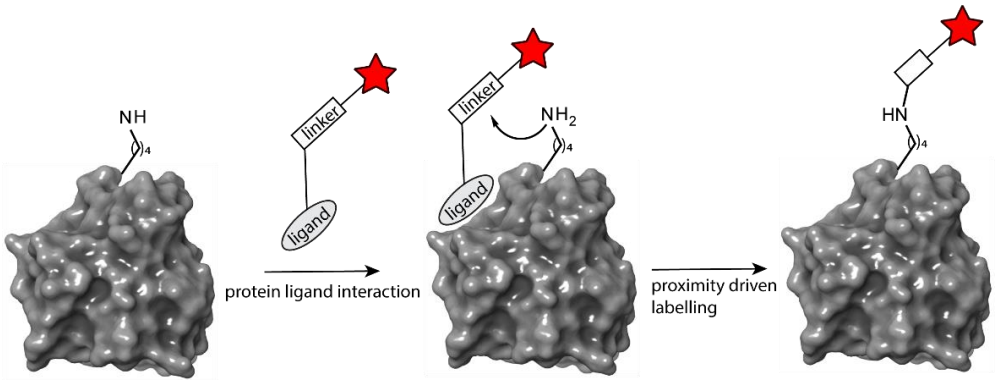


Figure 2.20 Schematic representation of ligand directed methods.

Although such ligand directed methodologies constitute a very elegant approach for site-selective native protein labelling, they are based on a protein ligand interaction which is not available for all POI's.

2.1.6 Protein modification flowchart

In figure 2.21 a protein modification flowchart from Stephanopoulos *et al*. is presented.[119] Several important distinctions are made in this schematic overview. A first important aspect is the environment in which the POI is situated. The options for the bioconjugation of a POI in cell lysate versus a pure POI solution in buffer are very different. A second crucial question is: "is site specificity required?". Indeed, for almost all cases control over the site of the modification is beneficial. Also, the possibility to introduce a new (non)canonical amino acid, peptide sequence or even an entire protein domain is highly important. Especially in the case where the POI is situated in a complex mixture, almost all strategies for bioconjugation depend on the control of the POI sequence by introduction of a noncanonical amino acid, peptide or protein domain.

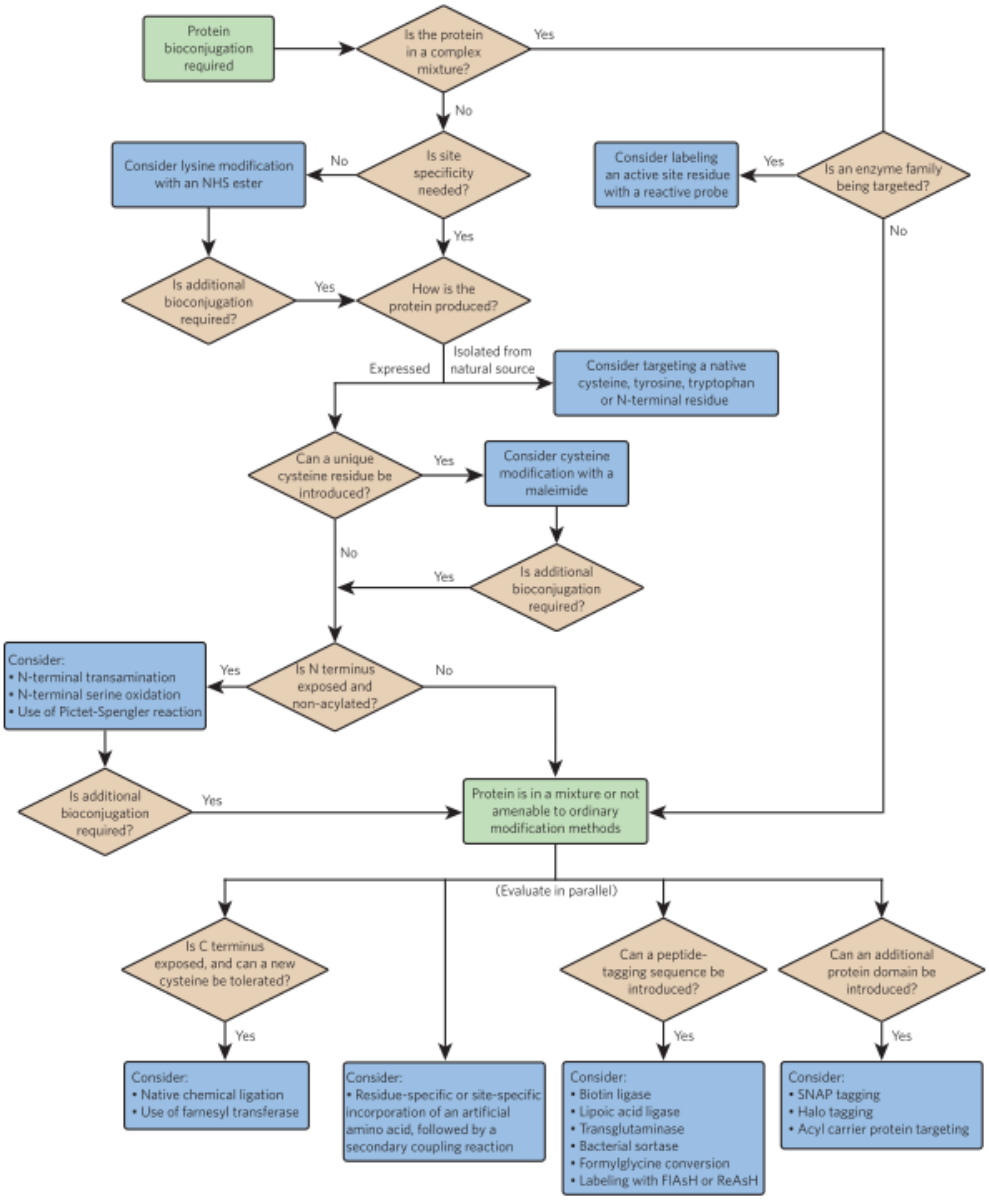


Figure 2.21 Protein bioconjugation flowchart from Stephanopoulos *et al.*[119]

2.2 Furan chemistry platform

A novel chemical tool for biomolecule labeling and crosslinking was developed within the Organic and Biomimetic Chemistry Research (OBCR) group under the guidance of prof. dr. Annemieke Madder. The chemical mechanism at the core of these methods is based on the natural metabolism of furan in the body. The harmful effects from furan exposure are a result of the oxidation of furan by cytochrome P450 (figure 2.22, a).[120] Upon oxidation, a reactive α,β-unsaturated dicarbonyl electrophile is generated. The toxicity of furan and furan containing compounds is triggered by the reaction of natural nucleophiles in the cell with this α,β-unsaturated dicarbonyl species.[121][122]

Stevens *et al.* incorporated a furan modified non-nucleoside building block in a DNA strand, this modified DNA strand was able to form an inter-strand crosslink with the unmodified complementary DNA strand (figure 2.22, b).[123][124] After mixing the modified and unmodified DNA strands to allow hybridization, chemical oxidation of the furan moiety is achieved using *N*-bromosuccinimide (NBS). It has been demonstrated that a covalent crosslink is selectively formed with a complementary adenine and cytosine. This methodology for DNA interstrand crosslinking was further refined by using a uridine modified with a furan moiety to achieve highly selective crosslinking towards complementary cytidine.[125] An important advance to the use of the furan chemistry was the use of visible light to trigger furan oxidation.[126] In this photo-oxidation approach visible light is used to excite a photosensitizer, which generates singlet oxygen which in turn can oxidize the furan moiety.

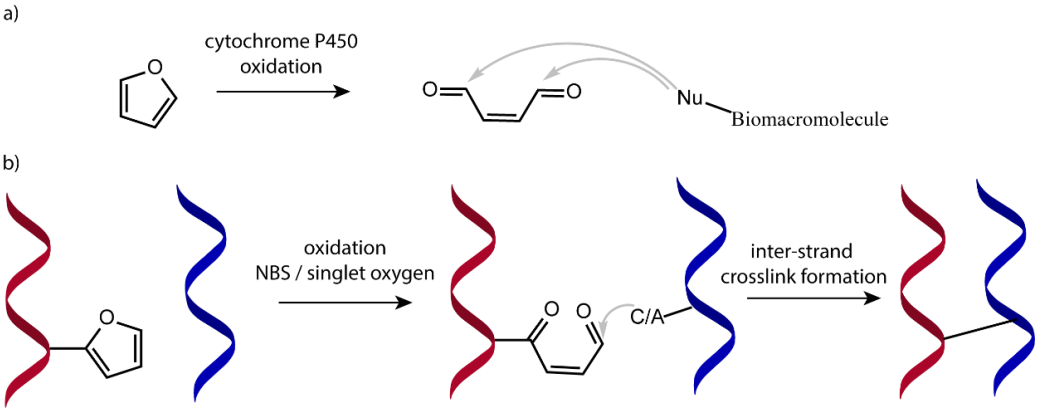


Figure 2.22 a) Schematic representation of cytochrome P450 mediated furan oxidation followed by nucleophilic attack. B) Schematic representation of the use of furan chemistry for the generation of selective inter-strand crosslinks. The oxidation can be effected using *N*-bromosuccinimide (NBS) or via generation of singlet oxygen.

Carrette *et al.* reported on furan modified DNA for probing DNA-peptide interactions. In this approach a DNA binding peptide is crosslinked to a furan containing DNA duplex.[127] In that specific case the amine of a proximate lysine residue is the reacting nucleophile. The furan oxidation approach has also been demonstrated for crosslink formation in RNA duplexes and in triplex crosslinking.[128][129] Additionally also applications using furan modified peptide nucleic acid (PNA) structures are reported for crosslinking to single and double stranded DNA.[130]

The versatility of furan chemistry is further demonstrated by the fact that its use is not limited to the oligonucleotide field. Indeed, also multiple peptide and protein applications were reported.[131] The furan containing amino acid furylalanine was incorporated into a peptide via SPPS. It was shown that introduction of aromatic residues in the vicinity of the furan moiety can protect the furan during the acidic cleavage conditions used in SPPS (figure 2.23, a).[132] Subsequently, the furan can be oxidized followed by a labelling reaction with a strong external nucleophile. In this regard, the furan moiety can be considered a caged aldehyde which becomes available after an oxidation step. The labeling of resin bound peptides was reported by Deceuninck *et al.,* the furan containing peptide was oxidized via NBS followed by a labelling reaction with an amine nucleophile.[133] Interestingly when the resulting imine was reduced to the amine with $NaBH_3CN$ a pyrrole was formed via reaction of the amine with the α,β-unsaturated ketone, subsequent loss of water and re-aromatization yields the pyrrole.

Later also singlet oxygen was used to oxidize furan containing peptides allowing for peptide derivatisation.[134] In these studies strong external nucleophiles like hydrazides and semicarbazides are used to derivatise furan containing peptides (figure 2.23, b). The selective furan oxidation is achieved via irradiation with visible light in presence of a photosensitizer to generate singlet oxygen. Additionally by using a furan containing peptide and another peptide bearing an hydrazine or hydrazide group peptides could be chemo selectively ligated in a one pot manner.[135] In the peptide labeling approaches, typically strong (unnatural) nucleophiles are required as opposed to the DNA crosslinking case where natural amines can act as nucleophiles. The crucial difference lies in the proximity between the reacting species. In the DNA crosslinking example the two DNA strands (one furan modified strand and one natural strand) are complementary and form a stable duplex. This means that the oxidized furan and nucleophile are in close proximity. This pre-paid entropy cost allows for an efficient crosslinking reaction with amine nucleophiles. In the case of the peptide labelling, there is no pre-existing proximity between the oxidized peptide and the label molecule, therefore for an efficient reaction a strong nucleophile is required.

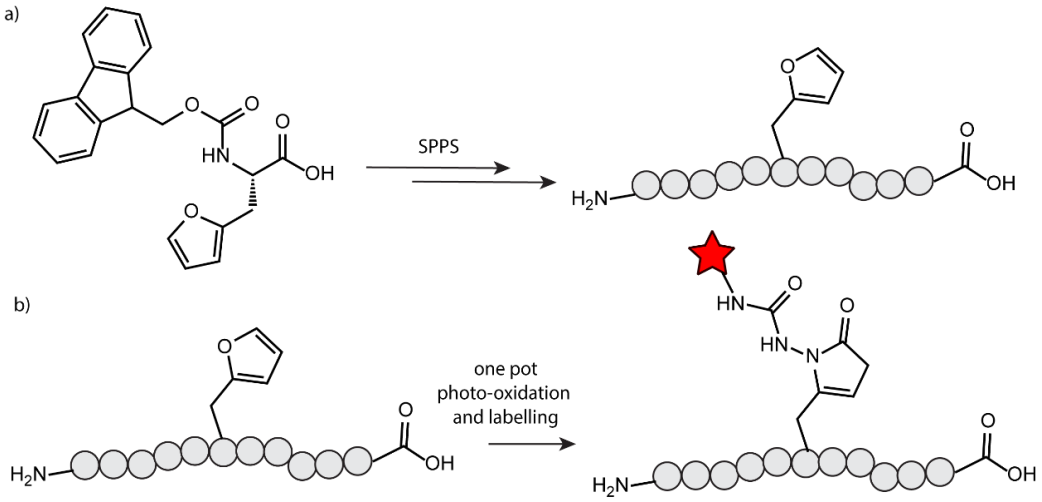


Figure 2.23 a) Structural representation of Fmoc-furylalanine for incorporation in a peptide using SPPS. b) One pot photo-oxidation and labeling reaction of a furan containing peptide using a hydrazide label molecule.

In a subsequent more complex application, the proximity between a peptide ligand and a receptor protein was leveraged for crosslink formation. Vannecke *et al.* reported a furan modified Kisspeptin-10 peptide for *in vivo* crosslinking to a cell surface receptor (figure 2.24).[136] Also in this application, the induced proximity between receptor and ligand upon binding allows for crosslink formation with a natural nucleophile. Remarkably, the furan peptide was able to crosslink to its receptor without the need for oxidation via NBS or photo-oxidation. It was demonstrated that endogenous reactive oxygen species (ROS) are capable of furan oxidation. The ROS species are generated by membrane proteins NADPH oxidases (NOX) among others.[137] This endogenous oxidation opens up the possibility to use furan modified peptides for *in situ* crosslinking to receptors without external intervention.[138]

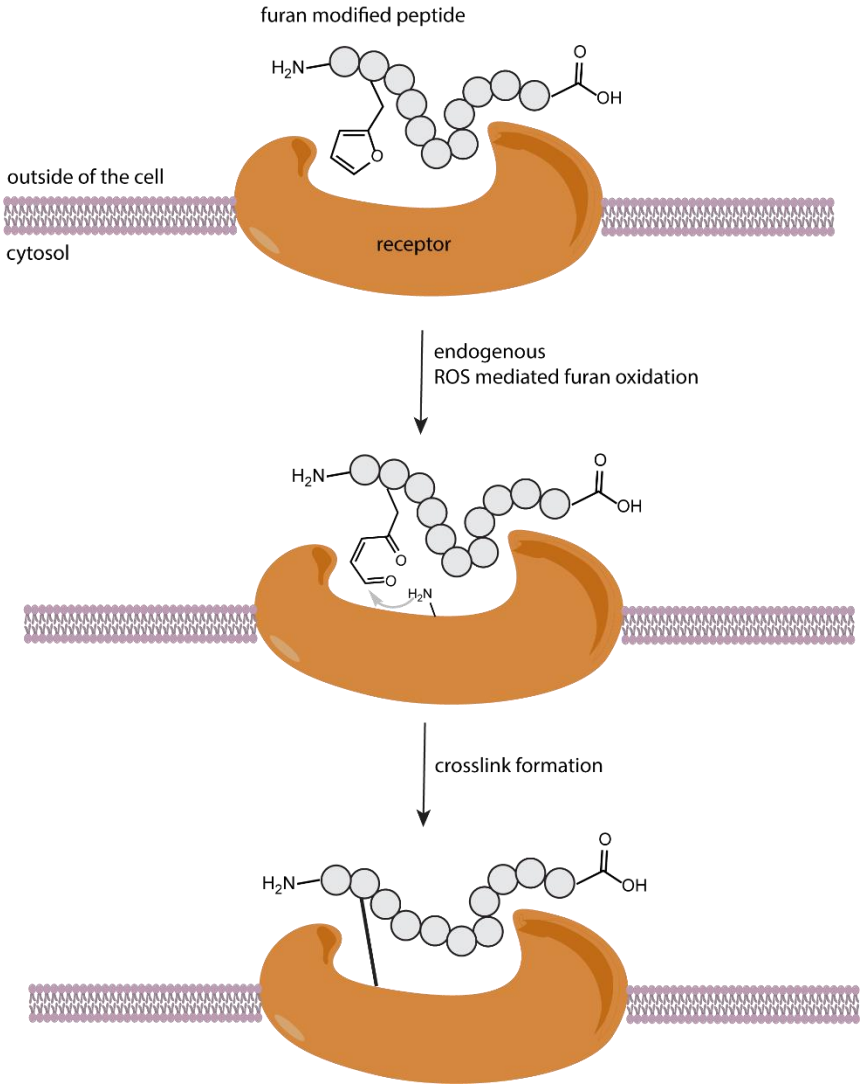


Figure 2.24 Schematic representation of the endogenous oxidation of a furan modified peptide bound to its cell surface receptor. In a second step a proximate nucleophile (presumably lysine) of the receptor can form a crosslink between the peptide and the receptor.

In this peptide-protein crosslinking application and also in protein-protein crosslinking applications (v*ide infra* chapter 4 aims and objectives) the side chain of a canonical amino acid will react with the oxidized furan moiety. Currently, It is not yet certain which residue is responsible for this crosslink formation. From earlier work on oligonucleotide-peptide[127] and oligonucleotide-protein[139] crosslinking, it is likely that lysine (amine functionality) is involved. On the other hand, also cysteine (thiol) could react with the oxidized furan moiety, as can be derived from the mechanism of natural furan metabolization.[121] Although based on previous reports lysine and cysteine seem most likely to react with the oxidized furan moiety, other natural nucleophilic residues cannot be excluded.

2.2.1 Furan incorporation through genetic code expansion

State-of-the-art for introduction of non-canonical amino acids (ncAA) via genetic code expansion

In section 2.1.3.1 genetic code expansion was introduced, a method that is based on the natural incorporation of pyrrolysine (Pyl) in certain *archaea* species. Indeed, nature has found ways to incorporate two additional amino acids beyond the 20 standard proteinogenic amino acids during protein translation. Pyl is the 22nd genetically encoded amino acid next to with selenocysteine (Sec) being the 21st. Both of these amino acids are incorporated in response to a stopcodon.

To allow Sec incorporation in response to the UGA codon rather than protein termination, a sequence element in the 3' untranslated region of the mRNA is required, this sequence was termed the SECIS element.[140] Selenocysteine is found in all three domains of life (*archaea, bacteria and eukarya)* however, the absence of the SECIS element in genomic analyses confirmed that it is not present in many species.[141][142][143] Remarkably in *Mammalia* there are more than 20 proteins that contain a Sec amino acid.[144] The incorporation mechanism is slightly different compared to other amino acids. To date there is no tRNA synthetase discovered to aminoacylate Sec to its tRNA^{Sec}. What is actually taking place in eukaryotes and archaea is that the serine tRNA synthetase (SerRS) aminoacylates serine to tRNA^{Sec} resulting in Ser-tRNA^{Sec}.[145][146] Afterwards a specific kinase enzyme phosphorylates the Ser-tRNA^{Sec} and in a second step the phosphate is enzymatically replaced by a selenium. Sec is incorporated in response to the UGA stop codon, also called the opal codon.

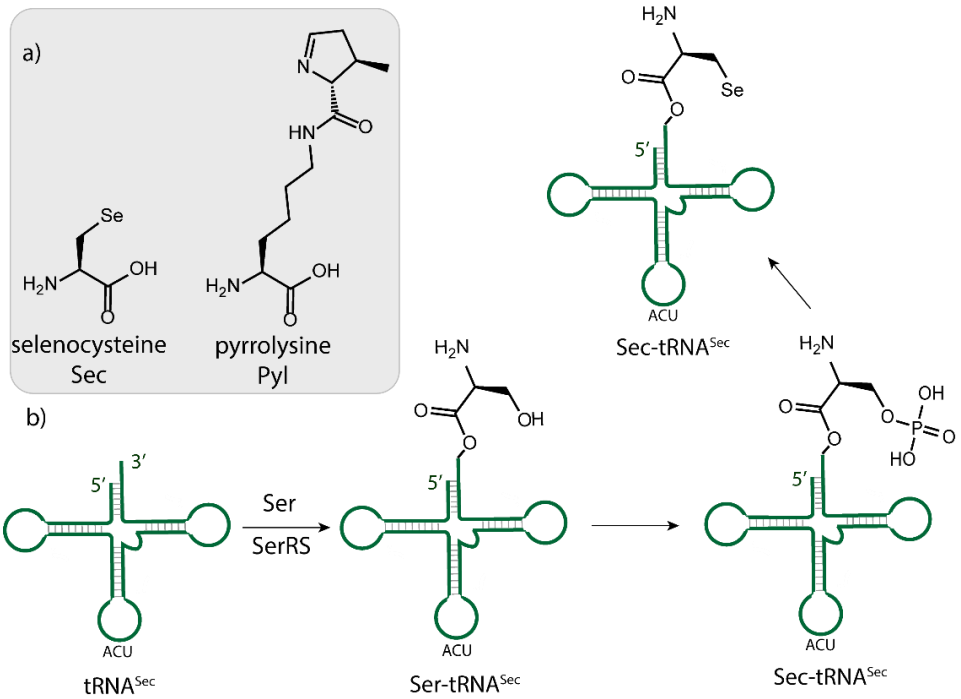


Figure 2.25 a) Chemical structures of selenocysteine (Sec) and pyrrolysine (Pyl), respectively the 21st and 22nd proteinogenic amino acids. B) Schematic representation of the $tRNA^{Sec}$ aminoacylation with Sec.

Pyrrolysine (Pyl) is introduced in the active site of three enzymes involved in the catabolism of methylamines. Pyl is found in a limited number of organisms including *archaea* from the *methanosarcina* family and some *eubacteria*.[147] Pyl is incorporated into proteins by means of an orthogonal tRNA/tRNA synthetase pair ($tRNA^{Pyl}$ / PylRS). Unlike Sec, Pyl is directly charged onto $tRNA^{Pyl}$ via pyrrolysine tRNA synthetase (PylRS).[148] Pyl is incorporated in response to an in frame UAG amber codon. Using three subsequent enzymatic reactions, catalysed by three different enzymes (PylB, PylC and PylD respectively), Pyl is biosynthesized from two lysine amino acids (Figure 2.26).[149][150] In a first step, PylB uses a radical mechanism and *S*-adenosyl methionine for the carbon skeleton rearrangement from lysine to a methyl-ornithine analogue. Subsequently, PylC ligates the intermediate to the ε-amine group of another lysine to generate an amide bond. Finally PylD catalyses the cyclization step using NAD^+ as oxidizing agent resulting in pyrrolysine.[147]

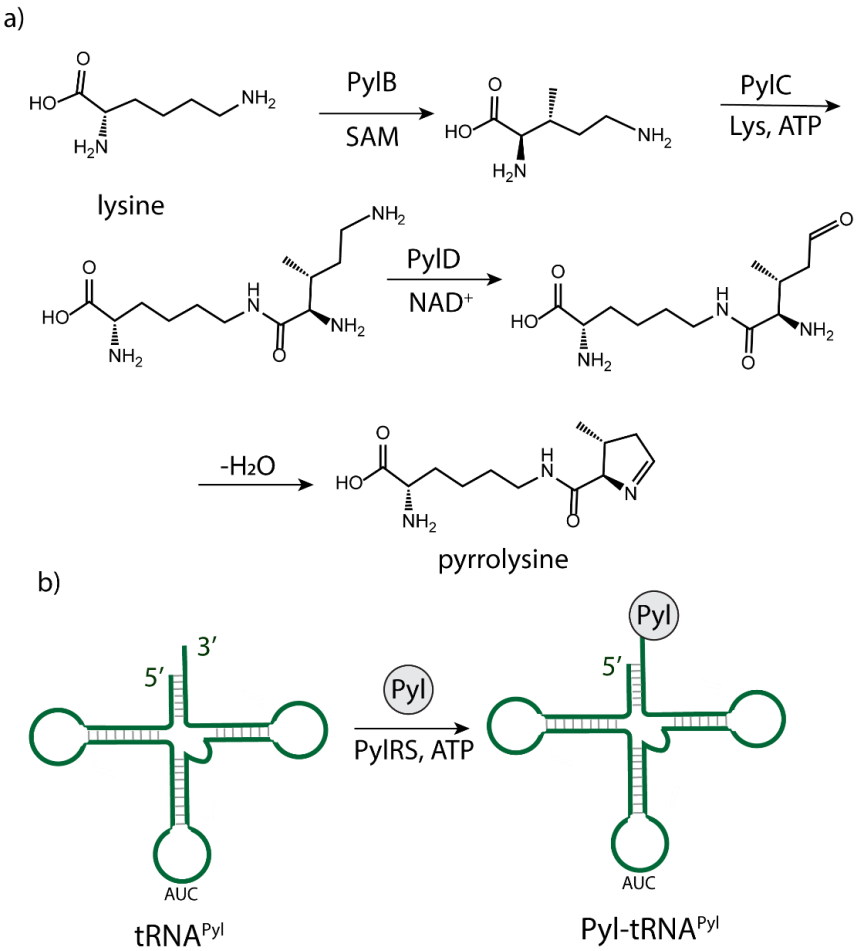


Figure 2.26 a) Biosynthesis scheme of pyrrolysine involving three sequential enzyme mediated reactions catalyzed by PylB, PylC and PylD respectively. b) Schematic representation of direct $tRNA^{Pyl}$ aminoacylation with Pyl, mediated by PylRS.

Scientists have modified the Pyl incorporation system for the introduction of an ncAA into a POI. In this approach a ncAA is site specifically introduced into a growing protein in response to the amber stop codon. Since the amber codon is in essence a stop codon the incorporation of an ncAA is in competition with the termination of the growing protein. The termination at a UAG (amber) codon in protein translation is mediated by release factor 1 (RF1) in *E. Coli* and, the result is a truncated protein. The competition between ncAA incorporation and protein termination is affecting the yield of the envisaged POI, therefore RF1 knock-out bacterial strains were developed. These allow for better POI yields and for multiple site incorporation of the same ncAA.[151]

The POI is expressed in a host organism with the ncAA on a predefined position. Of special importance is the need for an evolved orthogonal tRNA/tRNA synthetase pair that specifically aminoacylates the ncAA to its cognate tRNA (figure 2.27). As a starting point for the evolution of a new tRNA and amino acid tRNA synthetase (aaRS) pair typically $tRNA^{CUA}$ /PylRS is used.[152] A library of aaRS variants with random mutations in the binding pocket is used to identify a mutant that selectively accepts the ncAA of interest and aminoacylate $tRNA^{CUA}$.[153] Crucially the $tRNA^{CUA}$ /PylRS pair should not cross react with other synthetases and tRNA's in the host.[154] To select for a potent PylRS mutant and to avoid cross reaction positive and negative selection rounds are performed.[155][156] The positive selection consists of

an antibiotic resistance gene containing an amber codon. If the host (in presence of the ncAA) is able to survive in medium containing the corresponding antibiotic it means that the mutant PylRS was able to aminoacylate the $tRNA^{CUA}$. If the PylRS variant is not capable of aminoacylating its target tRNA it means the antibiotic resistance protein will be terminated at the amber codon, causing the death of the host. The negative selection is performed with a toxic gene containing an amber codon. If the host is able to survive in absence of the ncAA it indicates that the PylRS variant did not aminoacylate the $tRNA^{CUA}$ with a natural amino acid. If the mutant PylRS is not selective and accepts a natural amino acid for aminoacylation of $tRNA^{CUA}$ the toxic protein will be completed (not terminated at the amber stop codon) and the host will die. A PylRS mutant selected after positive and negative selection will uniquely accept the ncAA for $tRNA^{CUA}$ aminoacylation.[64]

The $tRNA^{CUA}$ / PylRS mutant pair and the gene coding for the POI with amber codon are added to the host organism by means of a vector, typically a plasmid. A plasmid is a circular piece of double stranded DNA that typically contains: an origin of replication for multiplication, an antibiotic resistance gene -to allow selection of a host that has taken up the plasmid- and a multiple cloning site to introduce a gene of interest. The process of adding a plasmid to a bacterial host is called transformation, this can be done either by a heat shock or via electroshock.[157] Additionally also the ncAA is added to the host. Upon translation of the POI during the ribosomal protein synthesis, the ncAA is transferred from the ncAA-$tRNA^{CUA}$ to the growing protein chain in response to the amber codon.

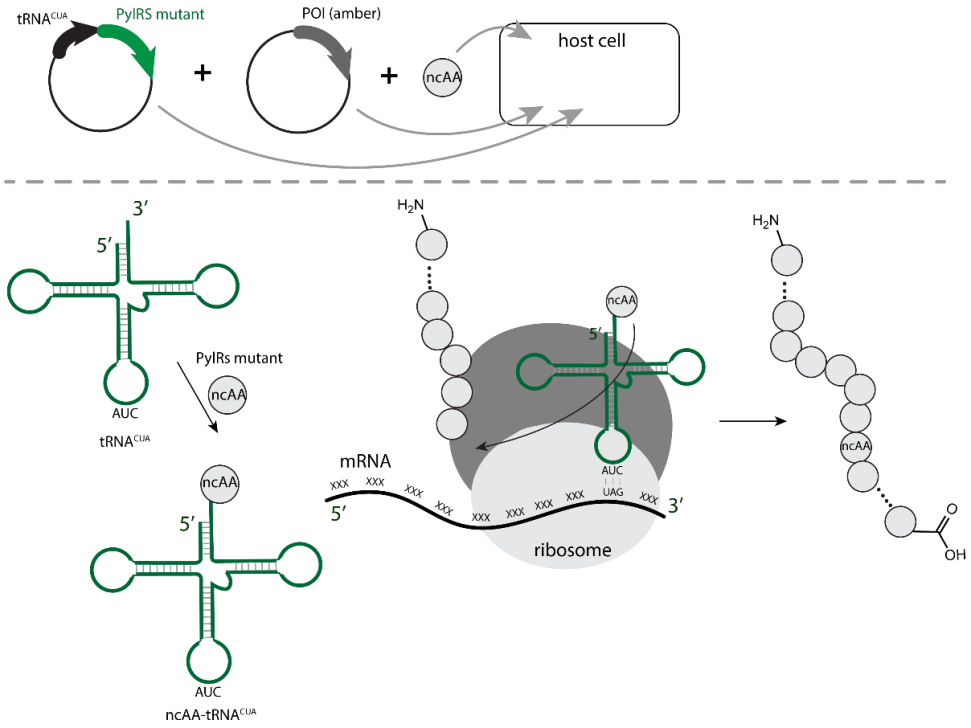


Figure 2.27 Top: Schematic representation of the addition of two plasmids and the ncAA to the host cells. Bottom: Schematic representation of $tRNA^{CUA}$ aminoacylation with a ncAA via the PylRS mutant. Subsequently during the ribosomal protein synthesis the ncAA is transferred from the ncAA-$tRNA^{CUA}$ to the growing protein chain in response to the amber codon.

Often this method is referred to as amber suppression since the regular protein termination of the nonsense amber codon is suppressed and it used for ncAA incorporation instead. To date a vast amount of ncAA's have been introduced into proteins via this approach.[158][159]

To allow uptake of the ncAA's into the cells the concentrations used in the medium for production are high (often around 1-5 mM). However even in these concentrations, some ncAA's are not efficiently taken up by the cells. These are typically permanently charged ncAA's or amino acids that greatly differ in structure from the canonical amino acids and are not recognized by the promiscuous amino acid transporters. Making an ester of the carboxyl group is a strategy that can be used to enhance uptake of an ncAA.[160]

A series of the ncAA's introduced contain a chemical handle allowing for protein modification in a second step (Figure 2.28). Many ncAA variants were constructed containing an azide or alkyne moiety for copper (I) catalysed Azide Alkyne Click chemistry (CuAAC).[161][162] The properties of click chemistry as defined by Sharpless and co-workers include: high efficiency, simplicity, selectivity and preferentially insensitive to water and air.[163] The CuAAC chemistry has become so popular that is sometimes becoming a synonym for click chemistry. Later, copper(I) was found to be cytotoxic and the methodology was refined by using strained alkynes to circumvent the need for copper (I) catalysis.[164][165][166].

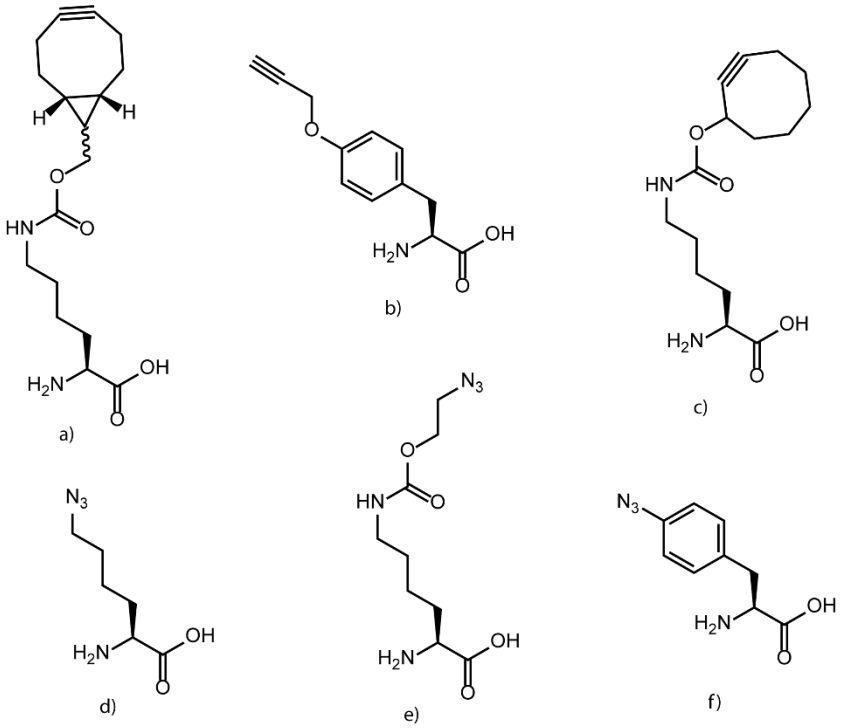


Figure 2.28 A collection of alkyne (a-c) and azide (d-f) containing ncAA's that have been used for CuAAC chemistry after introduction in a POI via amber suppression.

Another class of ncAA's is constructed for use in strain promoted inverse electron demand Diels-Alder cycloaddition click chemistry. This approach was developed by Fox and co-workers and uses strained alkenes for reaction with functionalized tetrazines.[167][168][169] The alkene-tetrazine ligations are known for their very fast

reaction rates.[170] Also strained alkenes in the form of norbornene amino acids were developed for reaction with tetrazine dyes (Figure 2.29).[171][172]

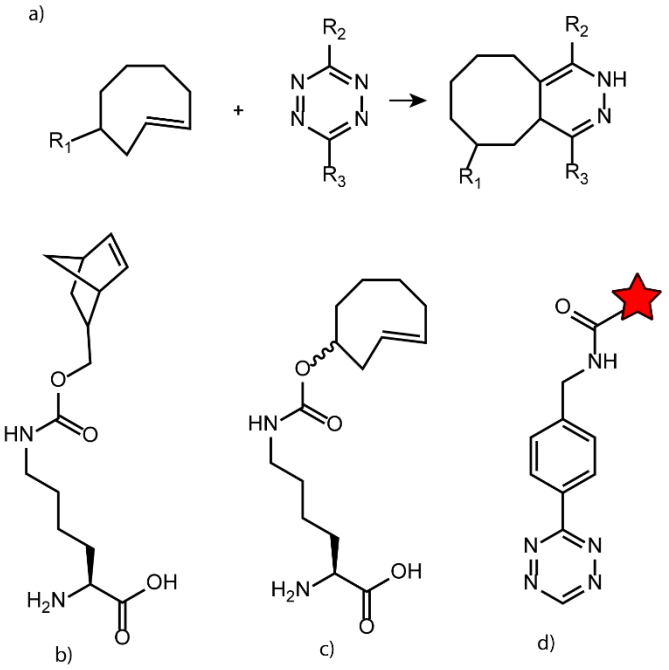


Figure 2.29 a) Schematic representation of the strain promoted inverse electron demand Diels-Alder cycloaddition. b-c) Strained alkene ncAA's, incorporated in PO's using amber codon suppression, used for tetrazine ligations. d) Schematic representation of a tetrazine dye.

Besides the CuAAC and tetrazine ligations also aldehyde or ketone functionalities can be introduced in proteins via genetic code expansion.[173] In 2014 Luo *et al.* even reported a fluorescent ncAA can be introduced via genetic code expansion.[174] Besides labelling of proteins scientists are also interested in trapping protein-protein interactions. An elegant approach towards this goal consists in the incorporation of a ncAA with a photocaged functionality into a POI. Subsequently, when the caged amino acid is activated by UV irradiation, a covalent crosslink is formed between the POI and the interacting protein partner. Typically three types of functionalities are used for photo-crosslinking: benzophenones, diazirines and aryl azides (Figure 2.30).[175][176][177] Amino acids with these functionalities were synthesized and incorporated using genetic code expansion.[178][179][180] When these amino acids are activated by UV irradiation they become very reactive and they can immediately form a crosslink to the interacting protein partner. The downside of the use of UV irradiation is that it is harmful for proteins, DNA and cells in general. Additionally, the generated reactive species are extremely reactive and can form unspecific crosslinks.

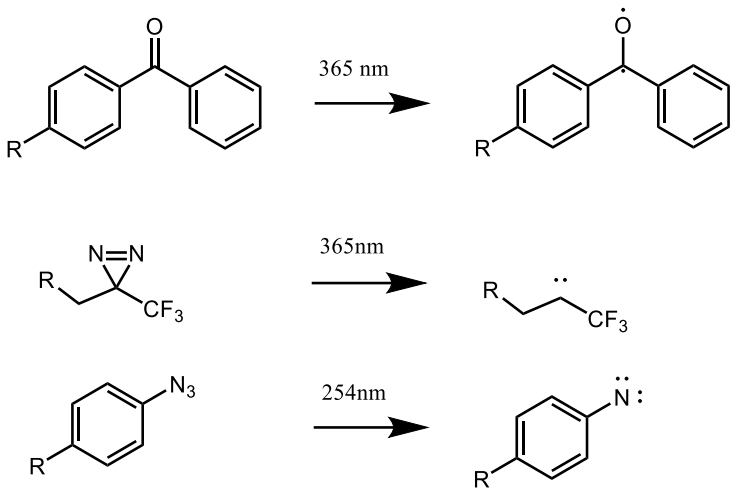


Figure 2.30 Photo-crosslinking functionalities typically used in combination with genetic code expansion to covalently trap protein-protein interactions.

Introduction of furan amino acids by genetic code expansion

In the beginning of section 2.2 of this work the furan chemical platform was introduced and was illustrated to be applicable for peptide labelling and peptide-protein crosslinking purposes. Inspired by our work, Summerer and co-workers were the first to introduce a furan-modified lysine amino acid (FurLys) into a RNA binding protein (Figure 2.31, left). This furan modified protein was subsequently crosslinked to RNA using red light irradiation in presence of a photosensitizer to generate singlet oxygen.[181][182] The possibility for activation of the furan amino acids via red light irradiation is an important advantage over conventional photo-crosslinkers since no UV-irradiation is required. Schultz and co-workers reported the incorporation of the furylalanine (FurAla) ncAA among a series of non-canonical histidine analogues.[183] However it has to be noted that the molecular weight of the furylalanine ncAA is almost identical to histidine (difference of 0.01 Da). This means that it is not possible to confirm the incorporation of the ncAA furylalanine via mass characterisation. The authors provide evidence of the selective FurAla incorporation via a negative control experiment. In this negative control experiment a toxic gene with an amber codon is expressed in absence of the FurAla ncAA. If the host bacterial cell survives it indicates that the toxic gene was not completely expressed since no canonical amino acid was incorporated in response to the amber codon.

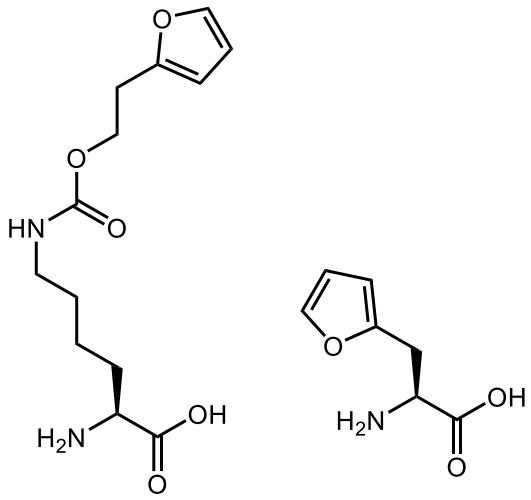


Figure 2.31 Structural representation of both furan containing ncAA's incorporated into proteins via genetic code expansion to date. Furan-modified lysine (FurLys), left, furylalanine (FurAla), right.

In what follows, we will refer to the furan-modified lysine amino acid as FurLys, and to the 2-Furylalanine amino acid as FurAla or Fua.

2.2.2 Furan incorporation via genetic code reprogramming

State-of-the-art for introduction of non-canonical amino acids (ncAA) via genetic code reprogramming

As the name suggests, in genetic code reprogramming arbitrary codons are used for the incorporation of ncAA's, as opposed to the amber suppression methodology which expands the genetic code by employing the nonsense amber codon. A key aspect of genetic code reprogramming is that it is not performed *in-cellulo* but *in-vitro,* the required molecular machinery for protein translation (ribosome, tRNA, amino acids, …) is added in a test tube together with a DNA template. The reconstituted expression system is fully pre-defined and allows for customisation of translation conditions.[184] By not including specific amino acids and their corresponding tRNA's multiple codons can be vacated. The vacated codons can then be used for incorporation of ncAA's. One of the most elegant approaches in genetic code expansion is the flexible *in*-vitro translation (FIT) methodology.

The FIT system was pioneered and developed by the laboratory of Hiroaki Suga. Key in the FIT system are the flexizymes (flexible tRNA acylation ribozymes), which are crucial for the amino acylation of tRNA (*vide supra*). In 2001, a prototype flexizyme, r24, consisting of 130 nucleotides (nt) was identified capable of aminoacylating a tRNA with phenylalanine cyanomethyl ester (Phe-CME).[67] The first flexizyme, Fx3 (45 nt in length) was adapted from this prototype after elaborate studies on r24 revealing the catalytic motif.[68] Later generation flexizymes, termed dFx, eFx and aFx with greater activity and altered substrate specificity, were found in selection campaigns using diverse substrates.[185] The flexizymes bind via base pairing to the 3' end of the acceptor-tRNA. The reaction time for the aminoacylation depends in part on the base pairing and can change for different tRNA's but any canonical tRNA can be aminoacylated by flexizymes. Advantages of the FIT system over amber suppression are the fact that the

noncanonical amino acid to be incorporated can be much more structurally diverse in the FIT system (see figure 2.32 for a few examples), in theory any ncAA can be incorporated via this method.[186][187][188][189]

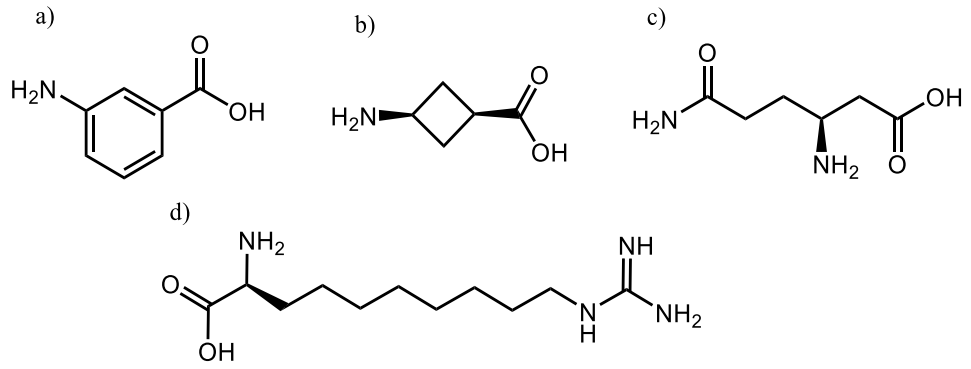


Figure 2.32 Structural representation of several exotic ncAA's incorporated via flexible *in-vitro* translation (FIT): a) 3-aminobenzoic acid, b) 3- aminocyclobutane carboxylic acid, c) β-homoglutamine, d) C8-arginine.

Another advantage of the FIT methodology is that it is straightforward to incorporate multiple noncanonical amino acids into one peptide while in the amber suppression methodology it is already very challenging to incorporate two noncanonical amino acids due to the need to define two orthogonal codons. However, due to the use of purified ribosome's, cofactors and other constituents of the FIT expression system the expression scale is generally very low and more suitable for screening purposes (*vide infra*) than production.

Since it is facile to incorporate multiple noncanonical amino acids, FIT is a powerful technique to produce peptides with a variety of "exotic" amino acids. Moreover, FIT has been used to develop an interesting approach to generate macrocyclic peptides.[184] The most used cyclisation chemistry in this context involves the spontaneous formation of a thioether linkage through reaction of a cysteine (thiol) with a chloro-acetyl moiety.

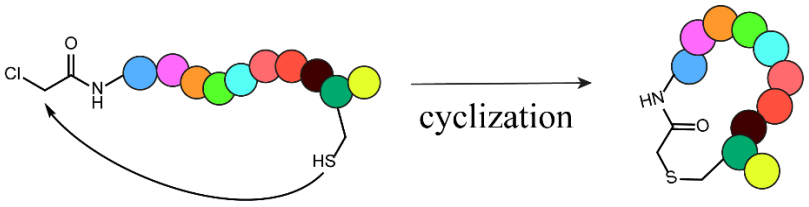


Figure 2.33 Schematic representation of peptide macrocyclization using a thiol (cysteine residue) and chloro-acetyl moiety (ncAA).

FIT is often integrated in the Random non-standard Peptide Integrated Discovery (RaPID) system, where it is combined with mRNA display to efficiently screen large libraries of extremely diverse macrocyclic peptides containing ncAA's with the aim of finding new ligands for various protein targets. The general principle of the RaPID system is depicted in the figure below. Generally RaPID is a combination of mRNA display[190] with FIT.

In a first step, a semi randomised library of mRNA fused to puromycin is generated. The importance of the puromycin is that after translation of the mRNA encoding for the nonstandard peptide, the mRNA is covalently connected to the nonstandard

peptide. When the ribosome reaches the 3' end of the mRNA template it includes the puromycin which is an analogue of an aminoacylated 3' end of a tRNA. This library is then translated via FIT to obtain RNA-peptide fusion structures. Subsequently, the mRNA can be reverse transcribed followed by an affinity selection step with an immobilized target. High affinity peptides will bind to the immobilized target and the non-binding peptides can be washed away. Via the polymerase chain reaction (PCR) the DNA of the binding peptides can be recovered. By performing multiple selection rounds, high affinity macrocyclic peptide binders can be obtained.[184] Since translation is achieved through FIT, the chemical space that can be screened is incredibly more diverse compared to conventional mRNA display.

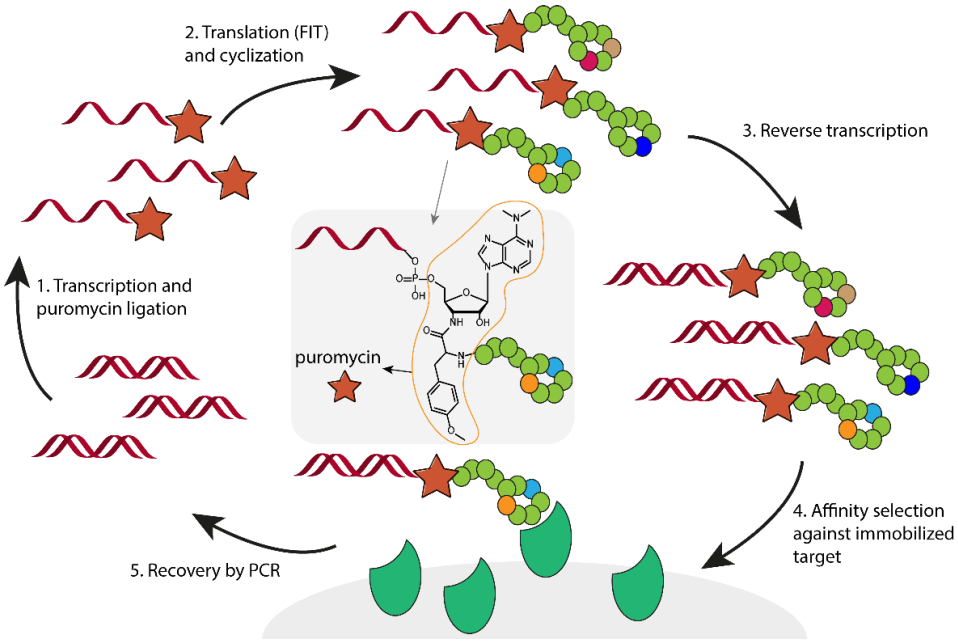


Figure 2.34 RaPID scheme of the screening for macrocyclic peptides. 1. The process starts with the transcription of a semi-randomised library and ligation of puromycin to the 3' end. 2. The library is translated via FIT and the spontaneously cyclized peptides (green, noncanonical amino acids are shown in a different colour) are covalently attached to the encoding mRNA. 3. Reverse transcription of the mRNA yields the cDNA. 4. The pooled mixture is panned against an immobilized target of interest to identify high affinity binders. 5. The DNA sequences attached to high affinity peptides are recovered via PCR. The process is repeated until only peptides with high affinity remain.

Incorporation of a furan amino acid via genetic code reprogramming

In this work we will use genetic code reprogramming to incorporate a furan containing amino acid into peptides. The tRNA was aminoacylated with furan cyanomethylester using the eFx flexizyme (figure 2.35)

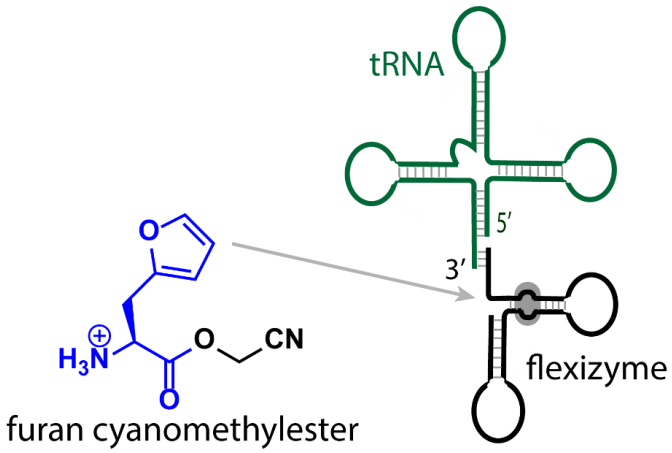


Figure 2.35 schematic representation of the aminoacylation of tRNA with the furan cyanomethylester amino acid.

2.3 Triazolinedione (TAD) chemistry platform

In the 1960s Cookson *et al.*[191] reported the synthesis of 4-phenyl-triazolinedione (Cookson's reagent). This work sparked the interest in the reactivity of these remarkable (di)enophiles. Before that triazolinediones were already observed in the late 19th century[192] but troublesome synthesis and purification hampered their use until the work of Cookson. TAD reagents are structurally similar to maleimides and also their reactivities have some similarities.[193] On the other hand TAD reagents can react with a much broader array of substrates and much faster compared to maleimides. There are four major reaction types for TAD reagents: a) Diels-Alder reactions, b) Alder-ene, c) electrophilic aromatic substitution (EAS) and d) 2+2 cycloaddition (figure 2.36).

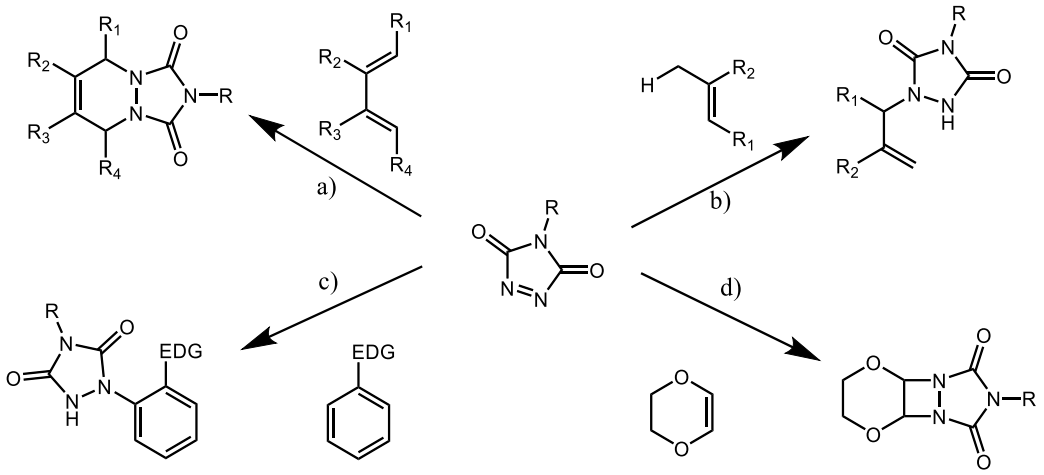


Figure 2.36 Major reaction types for TAD reagents. a) Diels-Alder reaction, b) Alder-ene reaction, c) electrophilic aromatic substitution (EAS) with an electron rich aromatic system, EDG: electron donating group, d) 2+2 cycloaddition.

TAD chemistry has found a large number of applications in polymer chemistry. An important application encompasses the modification of polydienes with TAD which was studied in detail by Butler *et al.* in the late 1970's.[194] In this application the Alder-ene reaction was tested at ambient temperatures between several TAD reagents and multiple polydiene polymers. In a more recent application in polymer field, Du Prez, Winne and co-workers demonstrated the use of TAD's as enablers of ultrafast and reversible click agents for dynamic polymer systems.[195] They exploit the thermal reversibility of the Alder-ene reaction of TAD reagents with indole substrates in combination with the ultrafast Diels-Alder reaction with conjugated dienes as depicted in figure 2.37.

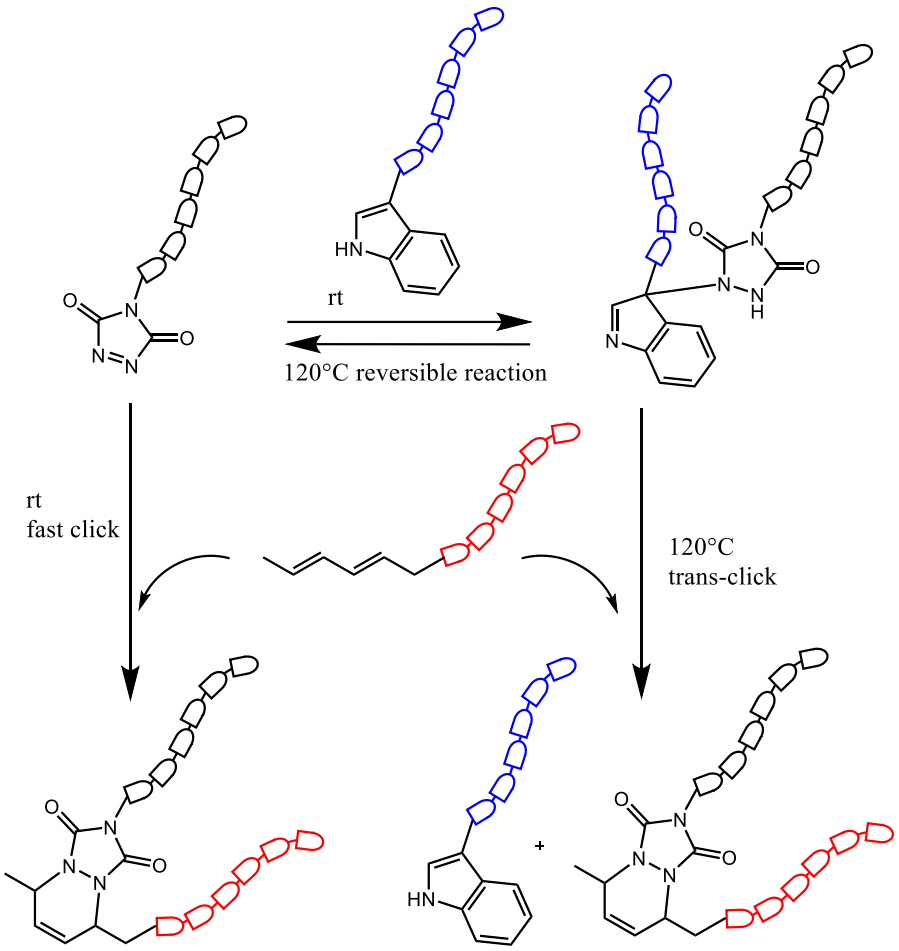


Figure 2.37 General overview scheme for applied TAD chemistry for click and transclick reactions. A TAD polymer can be used for a fast click with a diene polymer in a fast click reaction. Additionally a TAD polymer can be used for a thermoreversible reaction with indole polymers in an Alder-ene reaction. Subsequently upon addition of a diene polymer and heating to 120°C a transclick reaction product is obtained.

Another application area of TAD reagents outside the polymer field is the detection of low abundant (diene containing) lipid metabolites. In the early 1990's Yamada and co-workers reported the synthesis of a fluorescent TAD reagent: 4-[2-(6,7-dimethoxy-4-methyl-3-oxo-3,4-dihydroquinoxalinyl)ethyl]-1,2,4-triazoline-3,5-dione (DMEQ-TAD). This fluorescent compound was used for detection of vitamin D metabolites in biological samples (figure 2.38.[196][197]

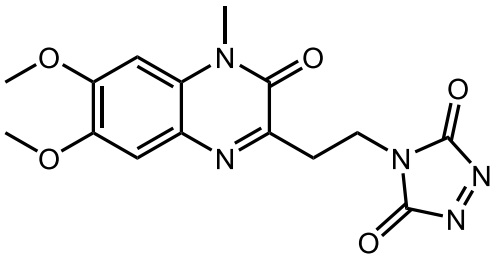


Figure 2.38 Chemical structure of 4-[2-(6,7-dimethoxy-4-methyl-3-oxo-3,4-dihydroquinoxalinyl)ethyl]-1,2,4-triazoline-3,5-dione (DMEQ-TAD).

Another example of the use of DMEQ-TAD was reported by Hoogewijs *et al.,*[198] where DMEQ-TAD was used for the on resin fluorescent labelling of furan containing peptides. In this application, both the furan and TAD chemical platform are combined. Remarkably, the reaction product of the furan-TAD reaction was not the anticipated Diels-Alder product but rather the aromatic substitution product. Unfortunately, TAD-labelling of free furan peptides in solution proved not possible.

A further important application resides in the use of TAD reagents for labeling of tyrosine amino acids as first reported by Barbas and co-workers[199] in 2010. This methodology was reported as a new bioconjugation for selective tyrosine modification. The phenol side chain is partly deprotonated in aqueous medium at physiological pH and this allows reaction with TAD via the EAS mechanism. Many applications and refinements of this method termed the TAD-tyrosine click reaction followed. In section 6.3 the application of TAD reagents for native peptide and protein modification is discussed in more detail.

3. Peptides and proteins as model systems

To start our investigations towards the objectives which are presented in chapter 4 of this work, we require model substrates to test our hypotheses. The objectives in this thesis are all situated in the field of protein modification. Nonetheless, we selected peptides as a first class of model systems. These peptides can be easily synthesised and analysed as opposed to proteins, where both the production and analysis are more tedious. After initial experiments on peptide substrates the hypotheses will subsequently be tested on model proteins.

3.1 Peptides

Peptides consist of a linear combination of amino acids linked together by amide bonds. While the general rule is that a peptide consists of at least two amino acids, the upper limit, where a peptide becomes a protein is somewhat diffuse. As a rule of thumb the limit can be placed at 50 amino acids. Already in 1901 the first unprotected peptide glycine-glycine was synthesized by Emile Fisher.[200] Up until the 1950's, peptide synthesis was very laborious and time demanding. Peptides of around a dozen amino acids were considered to be near the limit of what was achievable at that time.

In 1963 a new concept for peptide synthesis came to light when Merrifield[201] and Letsinger-Kornet[202] simultaneously and independently presented the concept of solid phase peptide synthesis (SPPS). Both methods were published in the same journal (Journal of the American Chemical Society) in the same year (1963) using the same type of polymer support (styrene-divinylbenzene). In this approach a first amino acid is linked to an insoluble polymer support and the second amino acid is attached to the first amino acid and so on. By attaching the growing peptide to a polymeric support the purification process is greatly improved since excess reagents can simply be washed away. However, there was an important conceptual advantage in the work of Merrifield. In his method the first amino acid was attached to the solid support via its carboxylic acid group and each new amino acid was linked to the amine group of the last amino acid. In the approach by Letsinger-Kornet, the first amino acid is attached to the polymer support via its amine group which leads to a huge racemization issue at every step during *on resin* activation of the carboxylic acid for further extension of the peptide. As a result, only the work of Merrifield gained wide acceptance and he received the 1984 Nobel Prize in Chemistry for the development of SPPS.

Even today, SPPS is the method of choice for almost all peptide syntheses. In SPPS, there is need for an orthogonal protecting group strategy, both the N-α amine and the side chain (often) need to be protected. Besides protecting groups, also coupling reagents are required for the amide bond formation. If no coupling reagents are used, only a simple acid-base reaction will occur between the amine and the carboxylic acid instead of amide bond formation. The SPPS procedure starts by attaching the first amino acid to the polymeric support, this is followed by series of N-α amine deprotections, amino acid activation and coupling steps until the desired length of the peptide is reached. Finally, the peptide is cleaved from the solid support (figure 3.1).

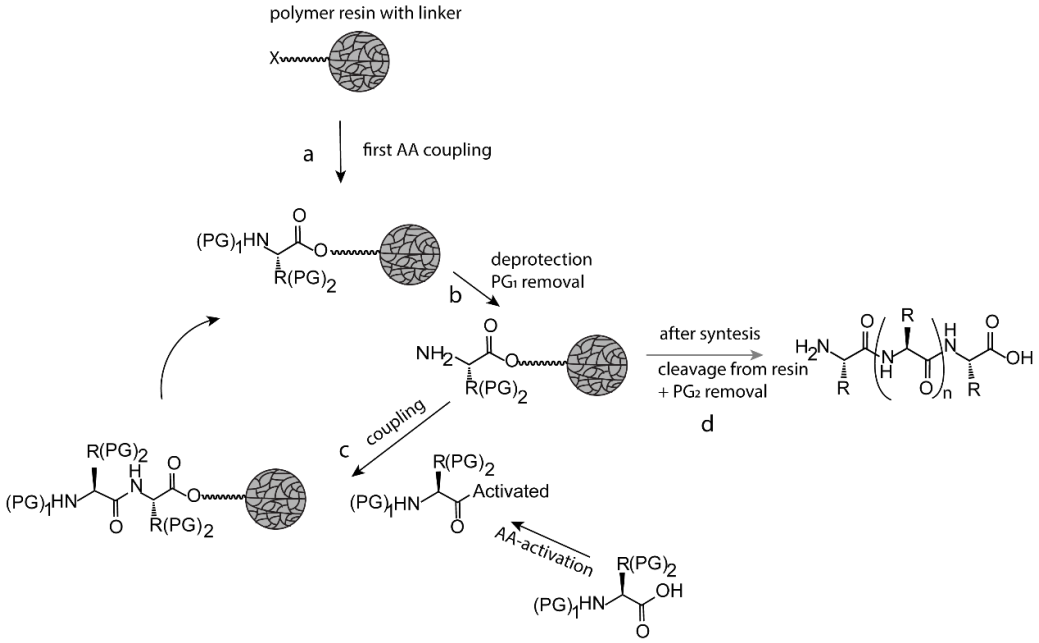


Figure 3.1 General SPPS principle. The first amino acid (AA) is coupled to the polymer through a linker moiety a). The N-α-amine is deprotected and an activated amino acid is coupled to the attached amino acid (b). Subsequently the N-α-amine of the newly added AA is deprotected (PG_1 removal) followed by a new coupling (c). This cycle is repeated until the completion of the peptide, afterwards the peptide is cleaved from the polymer resin and the side chain protecting groups (PG_2) are removed simultaneously (d). The repetitive nature of this cyclic reaction scheme implicates that SPPS lends itself very well for automatization and indeed to date almost all peptides are synthesized in an automated fashion.

As the construction of a peptide is entirely based on chemical steps, the choice of amino acids is broad and natural as well as unnatural amino acids can be easily incorporated at any desired position throughout the sequence. As *N*-Fmoc-furylalanine is a commercially available protected amino acid building block, decoration of peptides with furan moieties can be achieved in a straightforward way. In addition, also the necessary building blocks for exploration of the TAD strategy can be easily incorporated in model peptides via SPPS.

3.2 Proteins

Proteins are larger and more complex structures, also composed of amino acids. The ultimate goal of this thesis is to use furan and triazolinedione chemistries for site-selective protein modification. Once experiments have been conducted on the peptide level we want to test our hypothesis on proteins. This seems a trivial step since proteins can be regarded as very long peptides; however, chemistry on proteins is far more challenging as compared to chemistry on peptides, especially when it comes to introducing the correct handles through incorporation of unnatural amino acids. In this thesis, we use proteins produced via recombinant expression. This approach exploits the biosynthetic power of a host organism to produce a specific protein. It is a very elegant and cost effective way of protein synthesis but the control of the chemical functionalities in a protein much more very limited when compared to peptides.

Additionally, for recombinant expression the DNA coding for the protein is required and changing the amino acid sequence can be expensive and laborious. Besides the limited control over the sequence and functionality, the size of proteins is another important factor to consider. As mentioned before proteins come in all sizes and large proteins are more expensive to produce and they can also have a more complex folding pattern. Additionally, larger proteins are typically more difficult to analyse, rendering the chemical modification reactions on large protein structures more difficult.

In this thesis, so-called nanobodies are one important class of model proteins that we use. In 1993 Hamers-Casteman *et al*. reported a new kind of antibody which was found in species of the *Camelidae* family.[203] These antibodies were much smaller (90 kDa) compared to conventional antibodies (150 kDa) found in humans for example. Remarkably, these smaller antibodies called: heavy chain antibodies (HCAb) were able to bind their antigens with the same affinity as conventional antibodies. HCAbs lack the light chain which is present in conventional antibodies (Figure 3.2). HCAbs were used as a basis for the generation of nanobodies which are even smaller and generally encompass only the variable part of the heavy chain from heavy chain only antibodies (VHH).[204][205]

Since the discovery of heavy chain only antibodies in 1993, nanobodies which are derived from these heavy chain only antibodies have become polyvalent tools for a wide range of applications from therapeutics to imaging tools. Indeed nanobodies are being developed as therapeutics and going through clinical studies.[206][207] In 2019 the first nanobody therapeutic was FDA approved, Caplacizumab was developed by Ablynx (member of the Sanofi group).[208] Besides therapeutic use because of their small size, high conformational stability, high binding affinity and ease of production nanobodies are ideally suited for diagnostics as *in vivo* imaging tools for SPECT[209][210] and PET[211] scans.

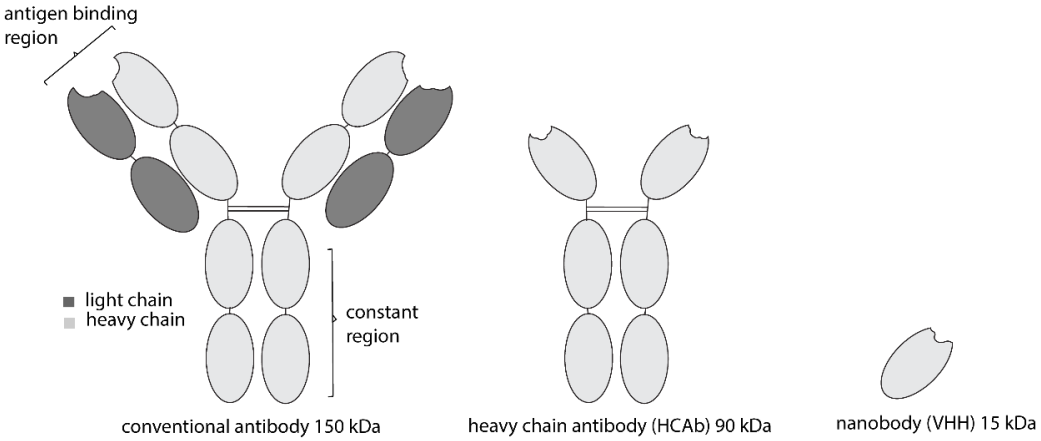


Figure 3.2 Schematic representation of conventional antibody (left) heavy chain antibody, a heavy chain antibodies or HCAb (middle) and a VHH nanobody (right).

Due to their small size, nanobodies have particular advantages as antigen binding proteins compared to conventional antibodies. They are cheaper and easier to produce, they are thermally more stable and have good solubility and they are more stable towards detergents. Because of these interesting properties nanobodies are currently developed for a series of therapeutic and diagnostic purposes.[212][213][214][215][216][217] Next to the clinical applications in therapeutics and diagnostics, nanobodies can be used in cell biology for protein visualisation. Upon generation of nanobodies for a selected POI a labelled version of such nanobody will allow the visualisation of that POI.[213] Since nanobodies are much smaller in comparison with antibodies, the label will be much closer to the POI resulting in smaller so-called linkage error (Figure 3.3). Another advantage of nanobodies over conventional antibodies is that they are capable of targeting cryptic and concave epitopes that are not accessible for antibodies.[218][219] As advances are made in the field of microscopy especially with the development of superresolution microscopy methods the need for a small linkage error is increasing.[220][221] The objective for accurate visualization is to bring a label moiety as close to the POI as possible. In the past, typically a labelled primary antibody or even a labelled secondary antibody were used for protein visualization.[222] Nanobodies fused with fluorescent proteins are an interesting approach and the linkage error is smaller compared to labelled antibodies but the GFP tag is still quite large in size.[223] Therefore a directly labelled nanobody is an excellent candidate to reduce the linkage error for super resolution microscopy purposes.[224]

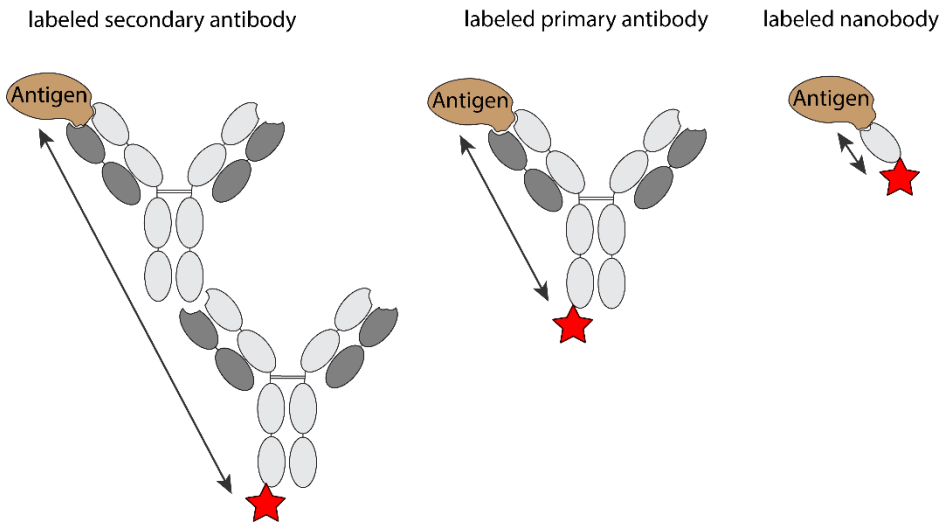


Figure 3.3 Schematic representation of the linkage error. Left: usage of a primary antibody and a labeled secondary antibody can lead to a very large linkage error of up to 30 nm. Middle: a labelled primary antibody reduces the linkage error up to 12-15 nm. Right: a labelled nanobody leads to a very small linkage error of several nm.

In our approach, we aim to incorporate a furan amino acid into a nanobody and to use the furan oxidation methodology for selective oxidation and labeling with a hydrazide or semicarbazide dye. Using this selective furan chemistry allows to avoid the use of the classically applied lysine functionalization which leads to heterogeneously functionalized nanobodies with potential loss of function if binding site lysines are modified.

To test our approach we collaborated with the nanobody lab headed by prof. Jan Gettemans (UGent). The nanobody lab specializes among other things in investigation of nanobodies directed against invadosome constituents e.g. fascin and cortactin.[217][225] These invadosomes are dynamic protrusions of the plasma membrane. These actin based structures can cause degradation of the extracellular matrix and are involved in cell invasion and cancer metastasis.[226][227] Both a cortactin as well as a fascin nanobody construct were prepared containing an amber stop codon suited for furan incorporation. The amber codons are located at the C-terminus of the amino acid sequence of both nanobodies since X-ray structures have demonstrated that the C-terminus is solvent accessible and unlikely to interfere with antigen binding (unpublished data). Specifically fascin nanobody 5 (FasNb5) and cortactin nanobody 2 (CortNb2) are selected because fascin nanobody 5 blocks the actin bundling activity of fascin and cortactin nanobody 2 blocks the interaction of the cortactin SH3 functional domain with WASp interacting protein.

Besides nanobodies which are based on naturally occurring antigen binding proteins, we also use alphabodies as model proteins. These alphabodies are not derived from natural proteins but have specifically been designed *in silico* to bind certain antigen targets.[228] Alphabodies exist of a triple helix coiled coil structure (Figure 3.4), which ensures a very robustly folded protein. Similar to nanobodies, alphabodies are around 10-15 kDa in size and they are easily recombinantly accessible. In this thesis, we use several alphabodies as model proteins for our experiments.

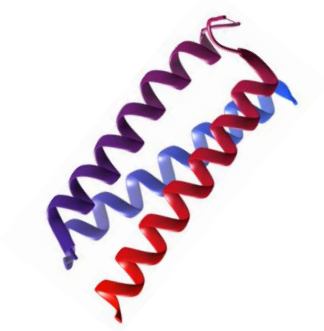


Figure 3.4 Schematic representation of the triple helical coiled coil alphabody structure.

Besides nanobodies and alphabodies we also use several other proteins in this work. The extracellular part of the epidermal growth factor receptor (sEGFR) is used in the part on protein crosslinking in combination with furan modified nanobody variants. Galectin-7 was selected as a model protein for TAD conjugation experiments. H-Ras, which is an important protein involved in cellular signal transduction, is used in this thesis as a substrate for enzymatic farnesylation mediated by farnesyltransferase. Subsequently the farnesylated protein is used as substrate in the experiments on the TAD-farnesyl reaction.

4. Aim and objectives

As described in chapter 2 of this thesis, a large number of methods are available for the site selective modification of proteins. However, there remains a continued interest from academia and industry to further expand the toolbox with additional modification strategies and better and more selective approaches. More specifically, the single-site selective labelling of native proteins remains particularly challenging. In this thesis we set out to develop the furan chemical platform and triazolinedione (TAD) chemical platform, each with their own unique characteristics and advantages, towards applications in site-selective protein modification.

General aim I

In this first part of the work we aim to further develop the furan chemical platform for applications in peptide and protein modifications.

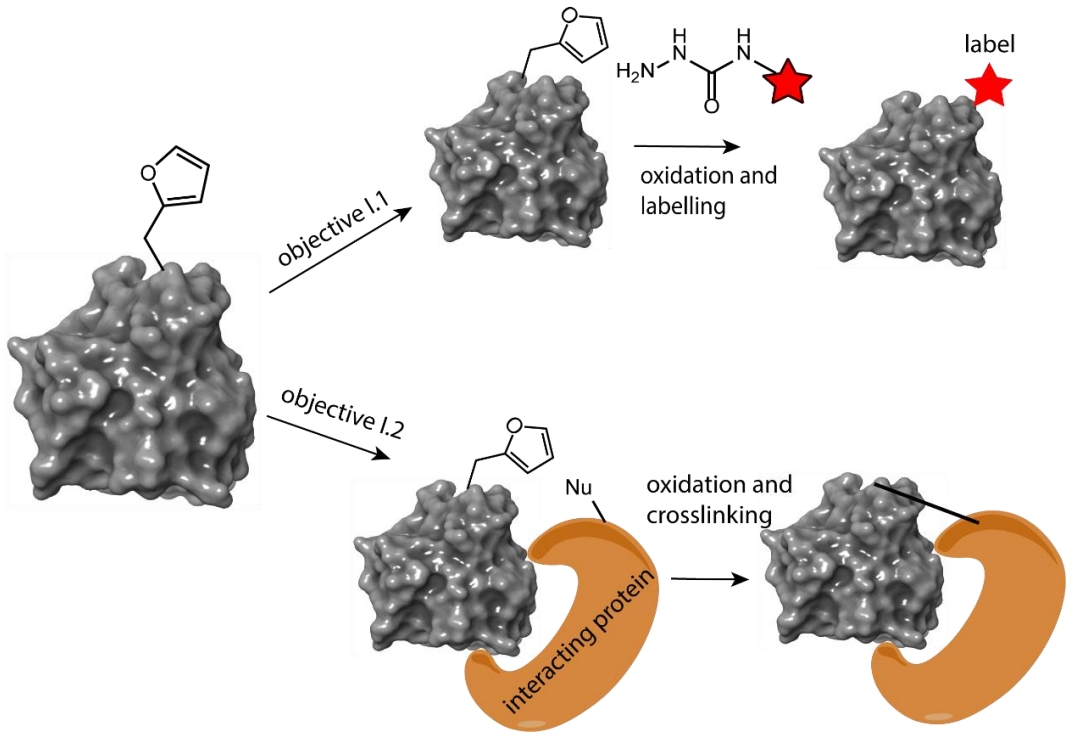


Figure 4.1 Schematic representation of objective I.1 (top) and objective I.2.

Objective I.1

The furan oxidation methodology was originally developed for nucleic acid modification[125][126] and subsequently applied for peptide labelling with hydrazide and semicarbazide functionalized labels[134] (*vide supra)*. In this work, we investigated if a similar labelling strategy could be effective and useful for protein modification / labelling (Figure 4.1, top). To test this hypothesis the first problem we need to overcome is the introduction of a furan moiety in a protein. This will be done using a genetic code expansion approach where a noncanonical amino acid bearing a furan moiety will be incorporated in a protein. As test proteins we selected several nanobody proteins, belonging to class of small (15 kDa) antigen binding proteins derived from antibodies from species of the *Camelidae* family. Subsequently we will use the furan oxidation procedure developed for peptide labelling and apply this on the scale of proteins. Labelled proteins will be analysed via LC-MS and potentially fluorescence measurements.

Objective I.2

Besides labelling applications the furan chemistry was effectively used for crosslinking of a variety of biomacromolecules ranging from DNA-DNA,[126] DNA-protein[229] to peptide-protein crosslinking,[136] which were previously investigated. In this thesis, we aim to further explore the crosslinking capabilities of furan chemistry by applying it in a protein-protein setting (Figure 3.1, bottom). In general, the genetic code expansion approach will be used to generate furan containing nanobodies for crosslinking purposes. Two test systems were selected, the first is the gelsolin nanobody which is known to protect a mutant gelsolin fragment from proteolysis and thus preventing the formation of amyloidogenic peptides. These amyloidogenic peptides are the cause of plaque formation in several tissues of patients suffering from familial amyloidosis Finnish type (FAF). The objective is to generate a crosslinking variant of the gelsolin nanobody that can bind covalently to its target and provide improved protection from proteolysis. However, before a crosslinking furan containing nanobody protein can be generated, the binding epitope on the mutant gelsolin fragments needs to be investigated in more detail. As a second test system, we selected the EgA1 nanobody which binds the extracellular part of the human epidermal growth factor receptor (EGFR). The crystal structure of this binding interaction is known and the objective is to use the crystal structure to generate a small set of nanobodies each containing a furan amino acid on a predefined position. These modified nanobodies will then be tested for crosslink formation with EGFR in *in vitro* experiments.

Objective I.3

As stated in section 2.2, previous reports seem to indicate that lysine (amine functionality) and cysteine (thiol functionality) are the natural amino acids involved in reaction with the oxidized furan moiety. Using the flexible *in* vitro translation methodology for the synthesis of modified peptides, we aim to confirm and examine which other nucleophilic residues are prone to react with an oxidized furan moiety and induce crosslinking. In order to achieve this, different nucleophiles will be tested in an intramolecular reaction setup. More specifically, peptides where both the furan moiety and a nucleophile are located on the same peptide will be oxidized and analysed for macrocyclization.

General aim II

In the second part of this work we aim to further develop the triazolinedione (TAD) chemical platform for applications in peptide and protein modifications.

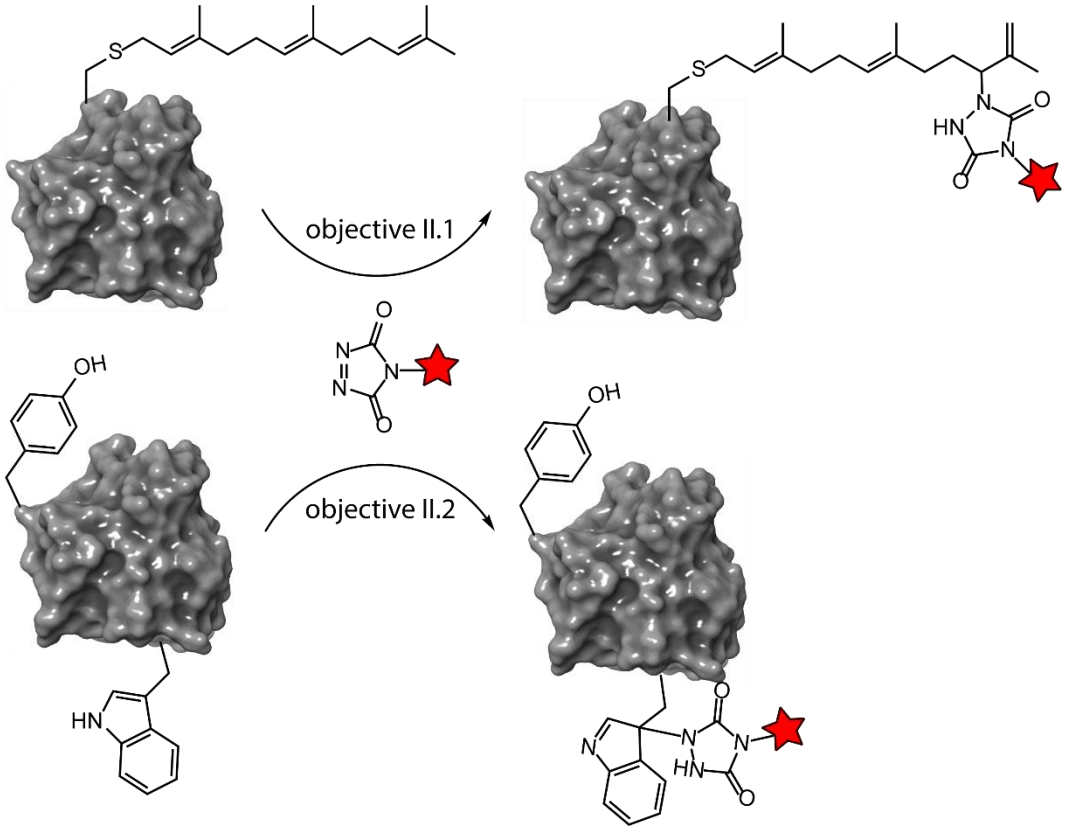


Figure 4.2 Schematic representation of objective II.1 (top) and objective II.2.

Objective II.1

TAD chemistry was previously used for several applications in the polymer chemistry field, including for the development of a self-healing class of polymers called vitrimers.[230] Besides applications in the polymer field, TAD chemistry was developed for putative tyrosine selective modification.[199] Here we aim to find conditions for selective labelling of the farnesyl post translational modification (PTM) on peptides and proteins (Figure 4.2, top). The farnesyl PTM occurs on cysteine residues in a specific CaaX sequence at the C-terminus. From a chemical perspective, the farnesyl group consists of three isoprene units connected in a linear fashion. This hydrophobic group associates with the lipid in the membrane to effect membrane association of farnesylated proteins. The electron rich double bonds in the isoprene units can react with triazolinediones.

Objective II.2

In the process of the experiments carried out in the context of reaching objective II.1, we discovered an important off target effect of the TAD reaction on native peptides and proteins. We found that, in contrast to the current believe, tryptophan amino acids can react with TAD molecules in a pH independent manner while targeting tyrosine residues with the TAD-tyrosine click reaction. This off target reaction was overlooked by many researchers for over a decade. Since there is an inherent difference in the reaction mechanism and pH dependence between the TAD-tyrosine and TAD-tryptophan reactions, a new objective is set to develop a new amino acid-selective protein modification method for tryptophan (Figure 4.2, bottom).

5. Further development of the furan chemistry platform

The nanobody expression and purification experiments in section 5.1.2 were performed in collaboration with Dr. Tim Hebbrecht and performed in the nanobody lab under guidance of Prof. J. Gettemans.

Nanobody expressions in section 5.2.1 were carried out in collaboration with the nanobody lab of Prof. J. Gettemans. Nanobody expression and purification were performed in collaboration with Dr. Tim Hebbrecht and Olivier Zwaenepoel.

Nanobody expression in section 5.2.2 was performed by Laure Tack and Olivier Zwaenepoel in the nanobody lab under guidance of Prof. J. Gettemans, according to the modelling performed by Klaas Decoene. The initial crosslinking experiments were carried out by Laure Tack and Dr. Laia Miret.

The *in-vitro* translation experiments in section 5.3.2 were carried out in the laboratory of Prof. Hiroaki Suga at the University of Tokyo and performed under guidance of Dr. Toby Passioura. Synthesis of the furylalanine cyanomethyl ester was performed by Dr. Willem Vannceke.

5.1 Towards protein labeling

5.1.1 C-terminal furan peptide

To optimize the furan oxidation and subsequent labelling conditions we decided to first test these on peptide level and use SPPS to synthesize the C-terminal part of the finally envisaged nanobody incorporating a furylalanine residue at the C-terminus. Working on a peptide substrate for the initial optimisation offers considerable advantages such as the availability of a large amount of stable test substrate, pure starting material and straightforward analysis of the reaction mixtures. However, a downside that needs to be taken into account is that in all previous studies we had worked with furylalanine modified peptides, while the furan-containing residue that we had incorporated in proteins through genetic code expansion was the above mentioned lysine derivative. It was earlier demonstrated (unpublished results of our group) that whereas furylalanine can be incorporated through classical SPPS, the carbamate linker in the furan-modified lysine derivative was proven not to be stable enough to survive the acidic conditions required for the peptide deprotection and cleavage from the resin (standard Fmoc SPPS chemistry; 95% TFA, full SPPS synthesis protocol was similar as described in section 9.3.3.1).

On the other hand, while, next to the incorporation of the furan-modified lysine derivative, also the incorporation of furylalanine using genetic code expansion was reported in 2014 and we requested the plasmid containing the required tRNA/tRNA synthetase pair, we received it much later. Unfortunately, this means that we had to work with different furan amino acids in our peptide test system when compared to the subsequent application on the furan-modified nanobody. The C-terminal sequence of the fascin and cortactin nanobodies (FasNb5 amber and CortNb2 amber) were adapted with a GGGSGGGX sequence with X representing the amber position. We synthesized peptide **5.1** and the N-acylated version peptide **5.2** by regular SPPS with incorporation of furylalanine at the X-position. Note that the N-terminally acylated variant **5.2** is included to examine the difference of a free/protected N-terminus, which is a potential nucleophile for reaction with the oxidized furan moiety.

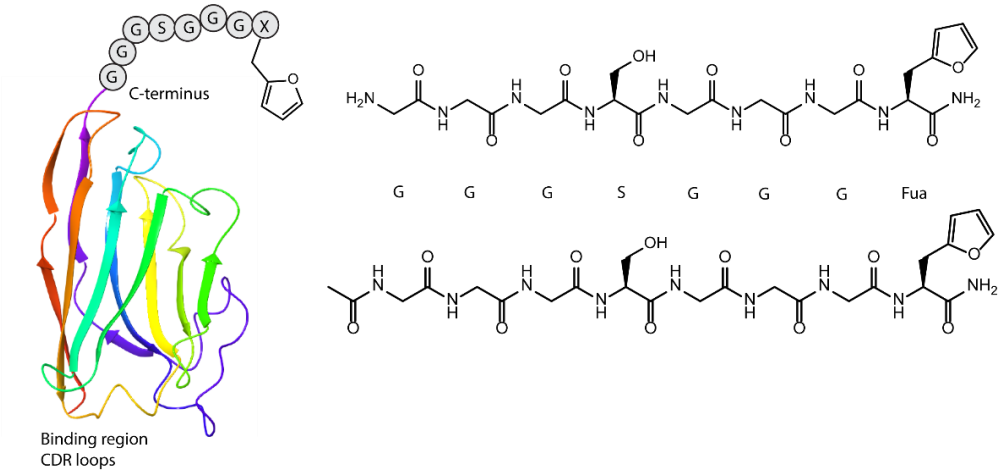


Figure 5.1 Schematic representation of a nanobody with the C-terminus pointing away from the binding region complementary determining region (CDR) (left). Structural representation of peptides **5.1** (top) and **5.2** (bottom).

The C-terminal peptide mimics were first activated by photo-oxidation via the protocol developed in Antonatou *et al.*[134] followed by labelling with methylhydrazine or Lucifer Yellow carbohydrazide (LYCH, Figure 5.2). The mild oxidation takes place in water and involves the use of a photosensitizer, Rose Bengal (RB) in this case. RB, upon irradiation with visible light, is responsible for the *in situ* generation of singlet oxygen, which in turn acts as the oxidant in the furan oxidation mechanism (Figure 5.3). RB is being used in different formulations in clinical trials for treatment of several cancers[231] and is further applied in photodynamic therapies.[232] The LYCH dye is a water soluble fluorescent dye used to stain nerve cells.[233]

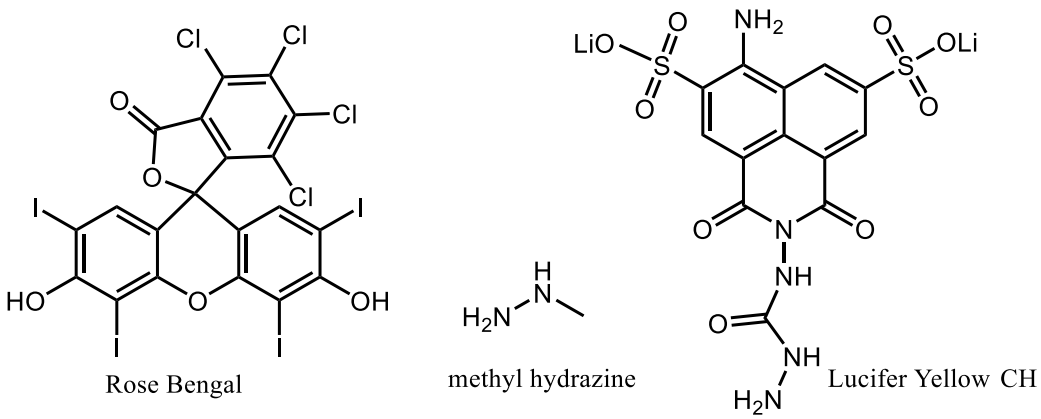


Figure 5.2 Structural representation of the photosensitizer Rose Bengal (RB) used in the furan oxidation process and of the labeling molecules methyl hydrazine and Lucifer Yellow carbohydrazide (LYCH).

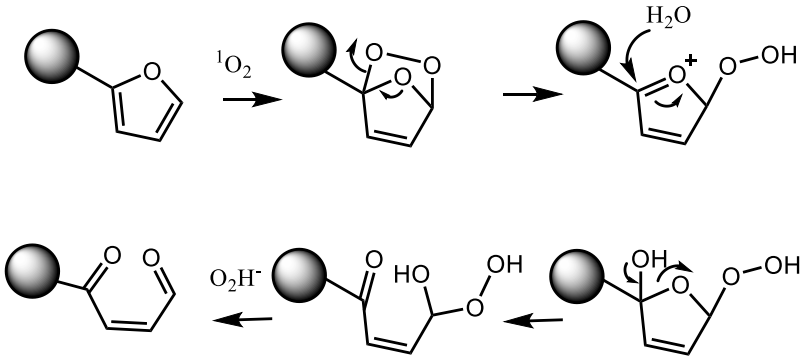


Figure 5.3 Schematic representation of a singlet oxygen ($^1O_2$) mediated furan oxidation reaction mechanism.

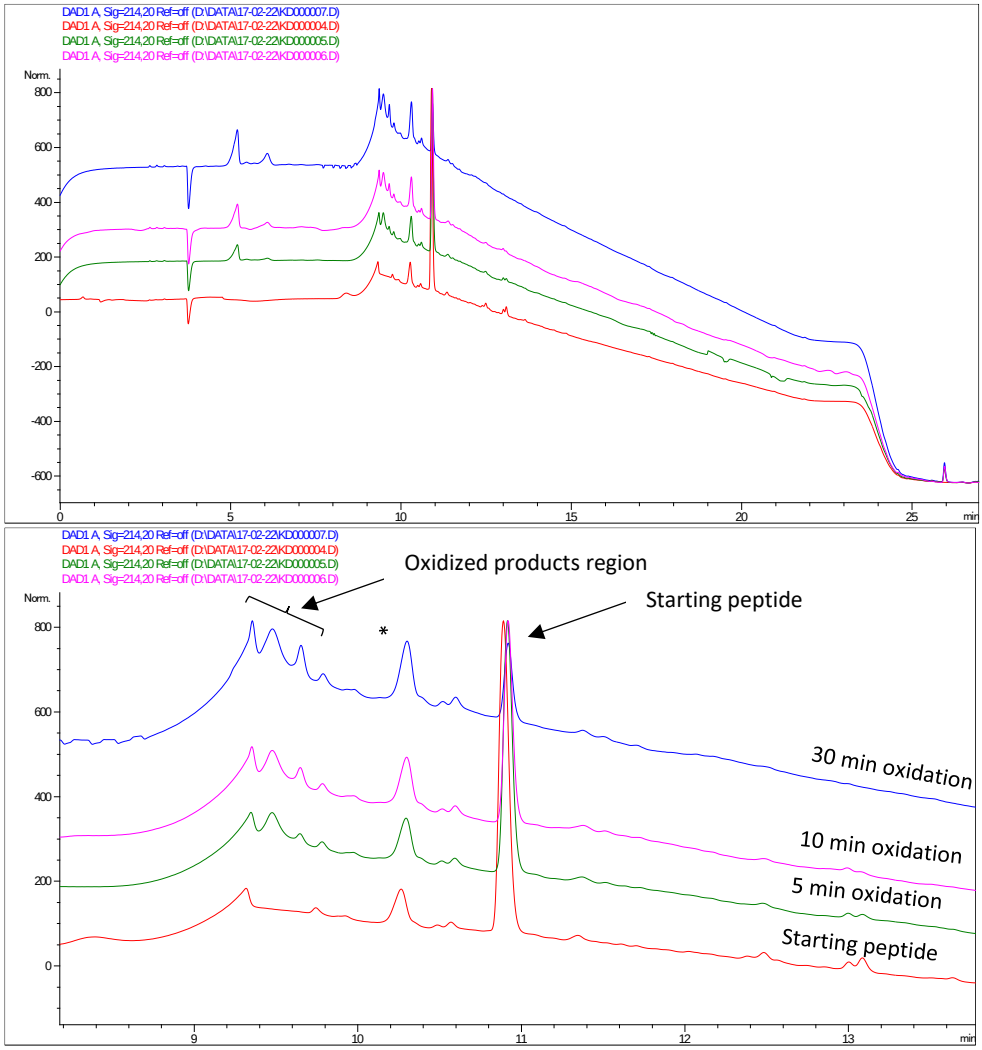


Figure 5.4 top: Overlay of complete HPLC chromatograms of the oxidation of **5.1** using white light irradiation in presence of RB, chromatograms at 214 nm at different irradiation times. Peptide concentration: 1 mM, RB concentration: 20 µM. Bottom: zoom in of the HPLC overlay between 8 and 14 min. Starting material: red; 5 min ox: green; 10 min ox: pink; 30min ox: blue. Peak indicated with "*" is an impurity in the peptide starting material.

Figure 5.4 shows the HPLC analysis of the oxidation of **5.1** with 0.02 eq RB at 0, 5, 10 or 30 min irradiation with white light at room temperature. The peptide peak is the main peak at 10.9 min; with increasing irradiation time we observe a less intense peptide peak. This result is in line with the expectation since upon oxidation, new products are formed visible to the left of the original peptide peak in the chromatogram, indicating the formation of (a) product(s) with a more hydrophilic character. Another observation is that even after 30 min irradiation the oxidation of **5.1** is not yet complete. Note that we observe not one product but several new peaks can be detected after oxidation.

In a subsequent experiment, the oxidation reaction was repeated with lower peptide and RB concentrations and at 10°C instead of at room temperature. In figure 5.5 an overlay of HPLC chromatograms at 214 nm is presented of a photo-oxidation experiment at several time points (blue: peptide starting material; red: 10 min irradiation; green: 20 min irradiation; pink: 30 min irradiation). Even after 10 minutes the starting peptide is fully consumed and several new small peaks appear. Note that

there is a significant difference in starting material consumption between the experiments in figures 5.4 and 5.5. Besides the fact that the sample figure 5.5 is more diluted and the oxidation is performed at 10 °C instead of at RT, we believe suboptimal positioning of the lamp and lower light intensity cause only partial consumption of starting peptide in the initial experiment represented in figure 5.4.

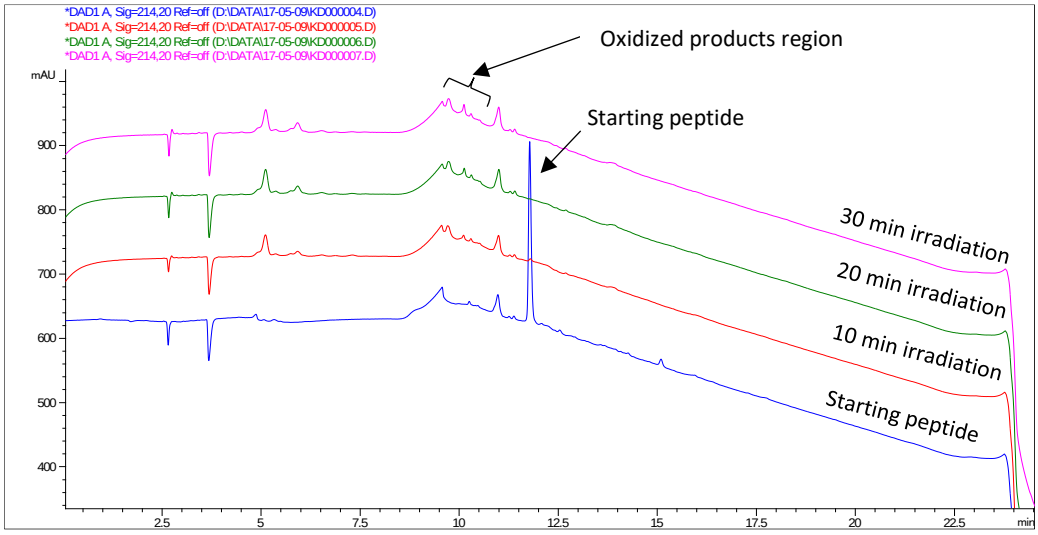


Figure 5.5 Overlay of HPLC chromatograms of the oxidation of **5.1** using white light irradiation in presence of RB, chromatograms at 214 nm of different irradiation times. Peptide concentration: 0.5 mM, RB concentration: 10 µM. Blue: peptide starting material; red: 10 min irradiation; green: 20 min irradiation; pink: 30 min irradiation.

In a following experiment guided by Eirini Antonatou, we repeated the oxidation experiment with the same peptide and RB concentration (0.5 mM and 10 µM respectively), the oxidation reaction was performed at 10°C and for 30 minutes. Subsequently 2 eq. of LYCH were added and after 3 hours the sample was analysed by HPLC.

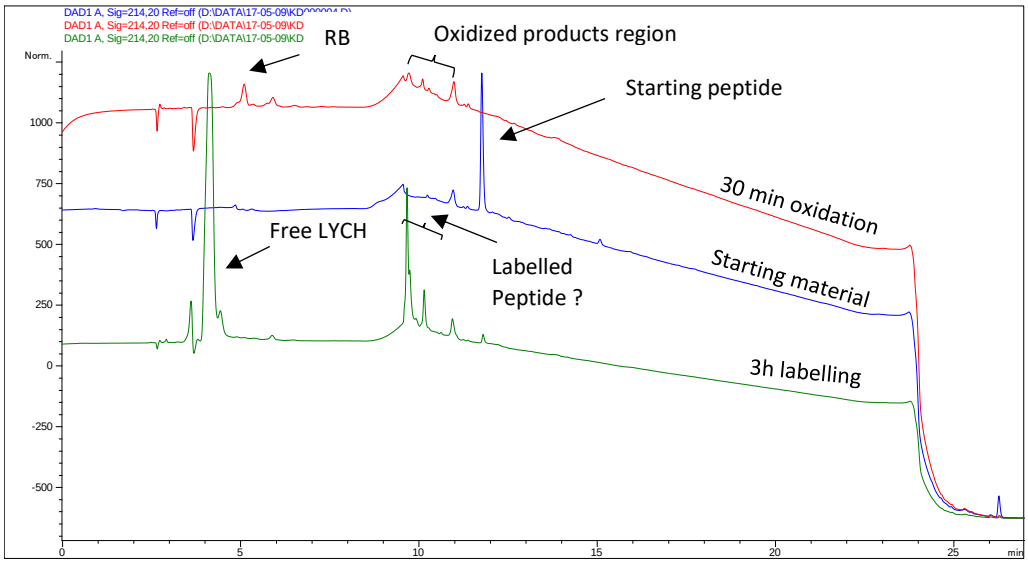


Figure 5.6 Overlay of the HPLC chromatograms at 214 nm of three samples. Blue: starting peptide, red: after 30 min oxidation, green: after 3h labelling with 2 eq. LYCH.

The oxidation at 10 °C is complete after 30 min since the starting peptide signal is completely consumed after oxidation (red chromatogram). After the LYCH labelling one intense and one smaller peak appear shortly after the injection peak (at around t_r: 9.5 min and 10.2 min), likely one of these peaks corresponds to the labelled peptide product. To confirm the identity of the labelled product the crude mixture after labelling was analysed on a matrix assisted laser desorption ionisation time of flight (MALDI-TOF) mass-spectrometer (Figure 5.7). Note that the mass deviation between the calculated mass (1025.9 Da) and the observed mass (1023.0 Da) is due to measurement on an old MALDI-TOF equipment. The oxidation/labelling reaction was repeated and confirmed with LYCH and methylhydrazine and analysed via LC-MS. Both products were formed as expected, LC-MS analysis in section 9.2 of this thesis.

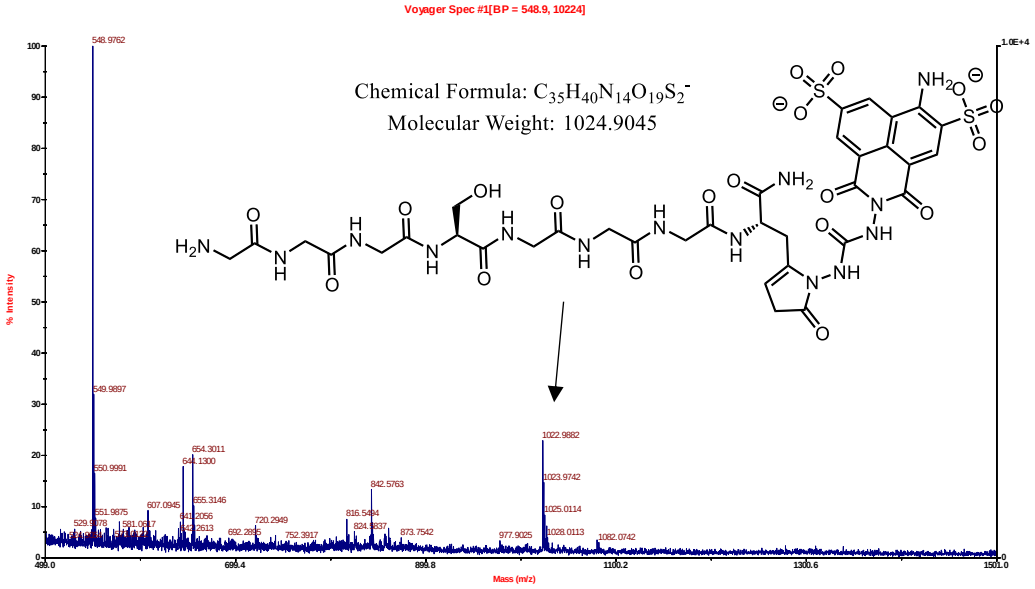


Figure 5.7 MALDI-TOF MS spectrum (negative mode) of the crude sample of oxidized and labelled peptide **5.1** with LYCH. Structural representation of peptide **5.1** labelled with LYCH. Calculated mass: 1025.9 Da, observed mass: 1023.0 Da. Note that the intense peak at 549 Da is not product related and likely originates from matrix components which are typically found at low molecular weight.

In the nanobody expression protocol, the use of DTT is required in an on column cleavage step (*vide infra*). DTT is a reductant and could potentially interfere with the furan oxidation and or labelling we aim to apply to the furan modified nanobodies. We tested to see if furan oxidation and labelling was still possible on peptides **5.1** and **5.2** in presence of different amounts of DTT. As reported in Hoogewijs *et al*.[234], where DTT was used as a scavenger in the peptide cleavage, several side products can be formed. As a starting point we used a DTT concentration of 50 mM since that is the concentration used in the nanobody cleavage buffer. Via HPLC analysis of a sample after different irradiation times we were able to demonstrate that the starting peptide is still consumed upon singlet oxygen oxidation (figure 5.8).

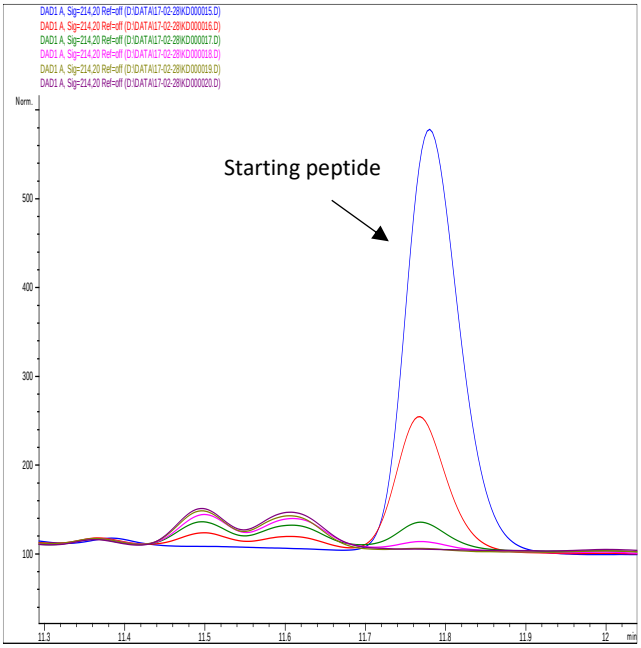


Figure 5.8 Zoom from an HPLC chromatogram overlay at 214 nm from a photo-oxidation experiment of peptide **5.1**, HPLC samples are analysed after different irradiation times. Blue: starting peptide (0 min); red: 5 min; green 10 min; pink: 20 min; brown: 30 min; purple: 60 min.

It can be noticed that **5.1** is quickly consumed in the presence of DTT but in the region before the starting peptide peak two new peaks appear in this experiment. These new peaks likely correspond to products resulting from side reactions caused by DTT and oxidized peptides. In subsequent experiments, the same oxidation reaction was carried out followed by labelling with methylhydrazine. Successful labelling could be confirmed via LC-MS analysis, although the labelled compound was only observed in small amount.

NBS oxidation is an alternative approach for furan oxidation, the final product of the oxidation is also believed to be the α,β-unsaturated ketone (figure 5.9). In theory, the labelling reaction with methylhydrazine and LYCH should also be possible.

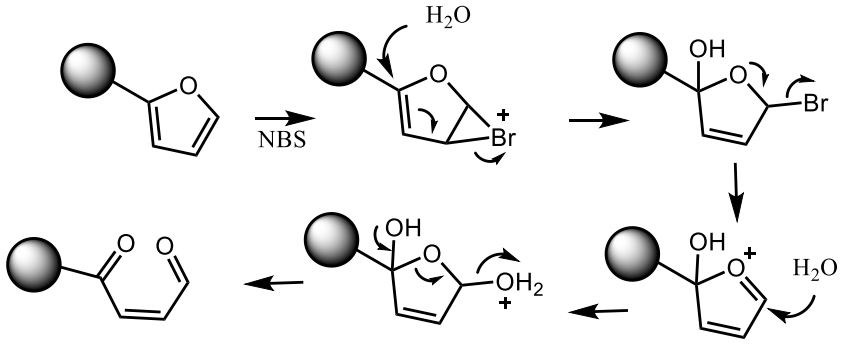


Figure 5.9 Mechanism of furan oxidation using NBS as an oxidant.

The oxidation reaction was performed with 3 eq. NBS compared to the peptide as in Vannecke *et al.*[136] also different amounts of DTT were present during the oxidation to examine the influence of DTT on the oxidation. At this time, we realised that the DTT concentration in the final nanobody sample could be reduced using small spin filters (BioRad Micro Bio-Spin®). These filters allow separation of large molecules over 6 kDa from smaller compounds. The separation mechanism is based on gel

permeation, larger molecules (over 6 kDa) are unable to migrate in and through the gel beads packed in the column. Smaller compounds (like DTT) are able to permeate in the gel beads and therefore have a longer path length to elute from the column. This allows their separation from large molecules. Additionally, these columns can also be used to change the buffer of a protein sample. Loading the desired buffer repeatedly on the column (at least 4 times) followed by the protein sample allows for the exchange of the buffer.

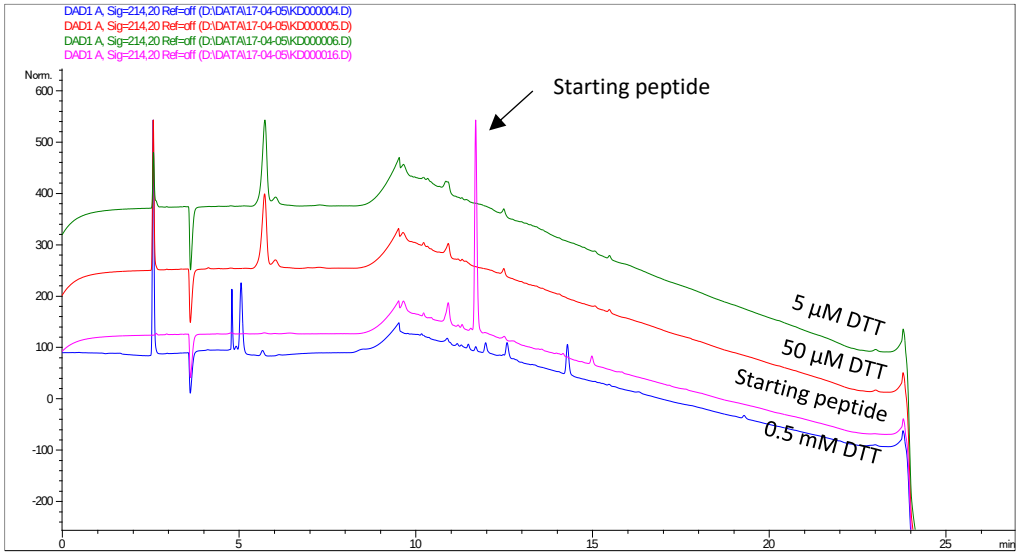


Figure 5.10 HPLC overlay of 214 nm chromatograms from an oxidation experiment with 3 eq. NBS of peptide **5.1**, HPLC samples with different DTT concentrations are analysed. Pink: starting peptide; blue: 0.5 mM DTT; red 50 µM DTT; green: 5 µM DTT.

The NBS treatment allows to completely oxidize the starting peptide independent of the level of DTT concentration in this experiment (blue: 0.5 mM DTT; red: 50 µM DTT; green: 5 µM DTT). Although the starting peptide is completely consumed, it can be noticed that in the blue, red and green chromatograms in figure 5.10 there are no intense new peaks appearing in the oxidized samples. Since it is always difficult to analyse the formed oxidation products at the oxidation stage, the subsequent LYCH labelling was performed.

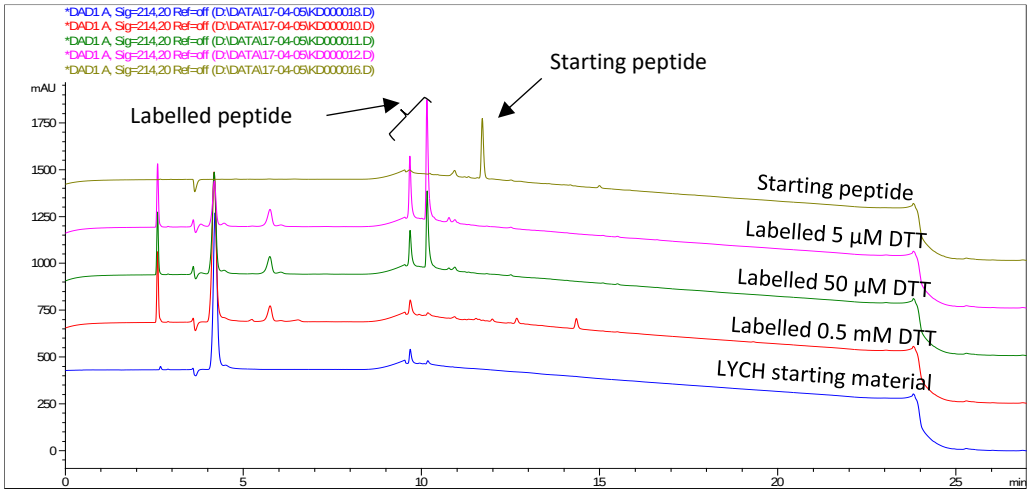


Figure 5.11 HPLC overlay of 214 nm chromatograms from oxidation with 3 eq. NBS and labelling with 2 eq. of LYCH of peptide **5.1**. Yellow: peptide starting material; blue: LYCH starting material; red: **5.1** oxidized and labelled 0.5 mM DTT; green: **5.1** oxidized and labelled 50 µM DTT; pink: **5.1** oxidized and labelled 5 µM DTT.

In the green and pink chromatograms in figure 5.11 after oxidation and labelling of peptide **5.1** we observe 2 intense peaks (eluting around t_r = 9.7 and 10.3 min) that are not present in the starting sample of **5.1**. These peaks appear similar to the peaks we have observed after labelling of the photo-oxidation product. Additionally, the intensity of both new peaks is dependent on the amount of DTT in the sample during oxidation/labelling. In the red sample with the highest DTT concentration, there is no sign of the new peaks, in the green sample with lower DTT the new peaks are present but in the pink sample with the lowest DTT concentration, the newly formed peaks likely corresponding with labelled peptide **5.1** are the most intense. Surprisingly the mass of the labelled peptide was not detected in MALDI-TOF or LC-MS analysis. This prompted us to look more in detail to the HPLC chromatograms of the labelling after photo-oxidation versus NBS oxidation. In figure 5.12 a HPLC overlay (zoom) of these samples is shown (red: NBS oxidation; blue: photo-oxidation). Besides a small shift in retention time we observe that two large peaks in the red chromatogram near the injection peak. Additionally, we observe similar peaks for the blue chromatogram which also has an extra peak as a prominent shoulder in the first large peak.

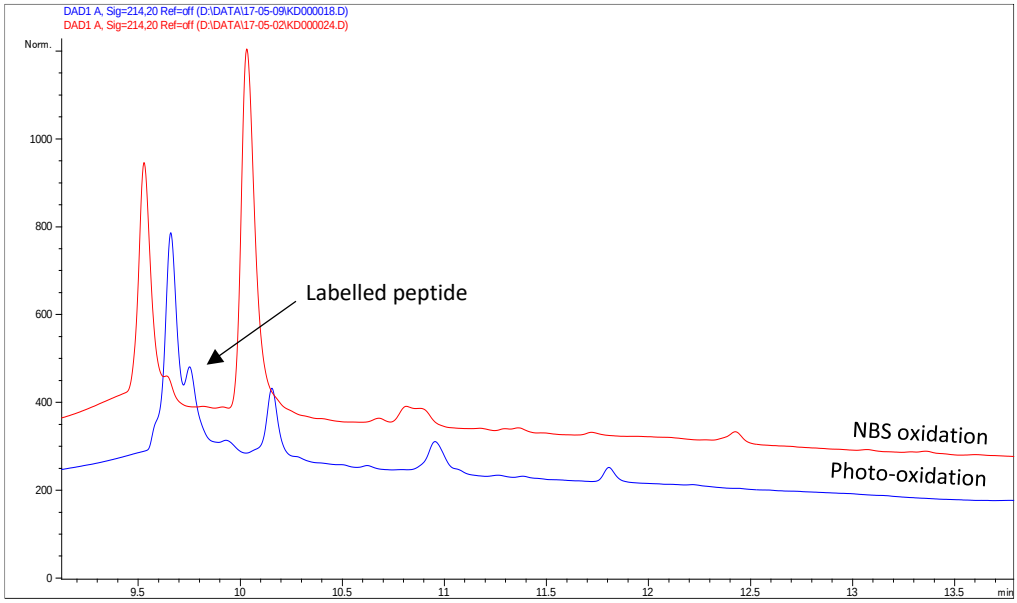


Figure 5.12 Zoom of overlay HPLC-chromatograms at 214 nm of the labelled samples after NBS oxidation (red) and photo-oxidation (blue). A small shift in retention time between both sample runs is present.

An overlay HPLC-chromatogram from samples taken at different time points in a LYCH labelling reaction after photo-oxidation was constructed and is shown in figure 5.13. For reference, also a sample of LYCH in water is included as a control (blue chromatogram).

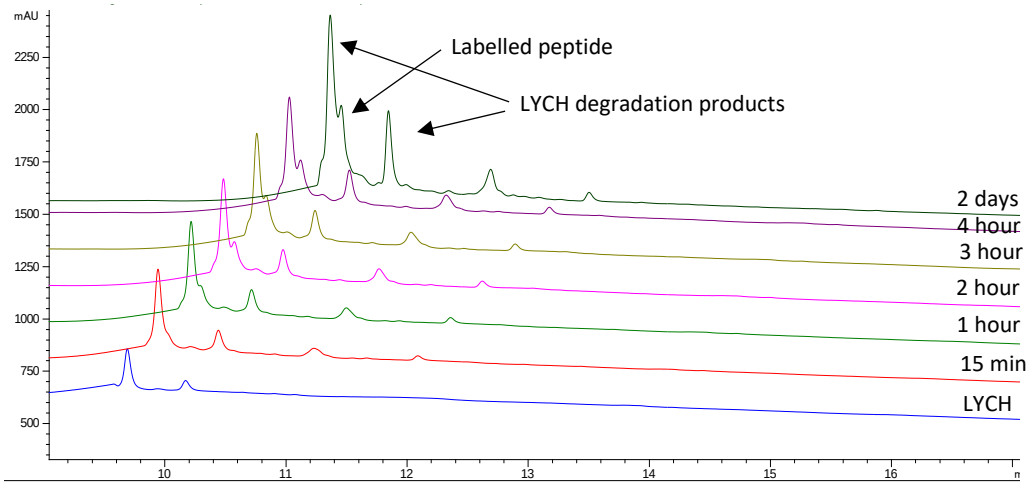


Figure 5.13 Zoom of overlay HPLC-chromatograms at 214 nm of a LYCH labelling reaction of peptide **5.1** following a photo-oxidation. A sample for HPLC analysis was collected at different time points in the labelling reaction. For reference also a sample of LYCH in water is included as a control. To avoid overlap of the peaks the chromatograms are artificially shifted.

In figure 5.13 we can see that the two peaks initially believed to correspond to the labelled peptide product are in fact already present in the LYCH control sample (blue). We hypothesise that these two peaks are LYCH degradation products. The shoulder peak corresponding to the actual labelled peptide is absent in the control sample. It seems that in the case of NBS oxidation the labelling reaction is not possible.

5.1.2 Nanobody expression

The nanobodies were produced as a N-terminal fusion construct with an intein tag, this allows for purification of the protein according to the IMPACT™ protocol.[2] The intein tag contains a chitin binding domain (CBD), which allows for purification over chitin beads (Figure 5.14). After expression of the POI fusion construct in the host, the host cells are lysed after which the POI needs to be purified from the cell lysate mixture containing many other proteins. The intein tag that is N-terminally fused to the nanobody protein contains a binding domain for the polysaccharide chitin.

During purification in the IMPACT™ protocol the purification takes place over a solid phase made of chitin beads. The crude protein lysate is loaded on the chitin beads column; the chitin binding domain in the intein tag of the POI fusion constructs binds to the chitin beads. In subsequent washing steps all other proteins are washed from the column, only the POI fusion construct remains. By adding a dithiothreitol (DTT) containing buffer for overnight on column cleavage the POI is cleaved from the chitin resin. In a following step, the POI is eluted from the column and collected for analysis and further use. Afterwards the N-terminal part of the fusion construct, the intein tag, and some remaining complete fusion proteins are still bound to the chitin beads. The chitin beads can be regenerated for reuse by treatment with 1% sodium dodecyl sulphate (SDS), these denaturing conditions allow for the release of the intein tag and leftover fusion protein.

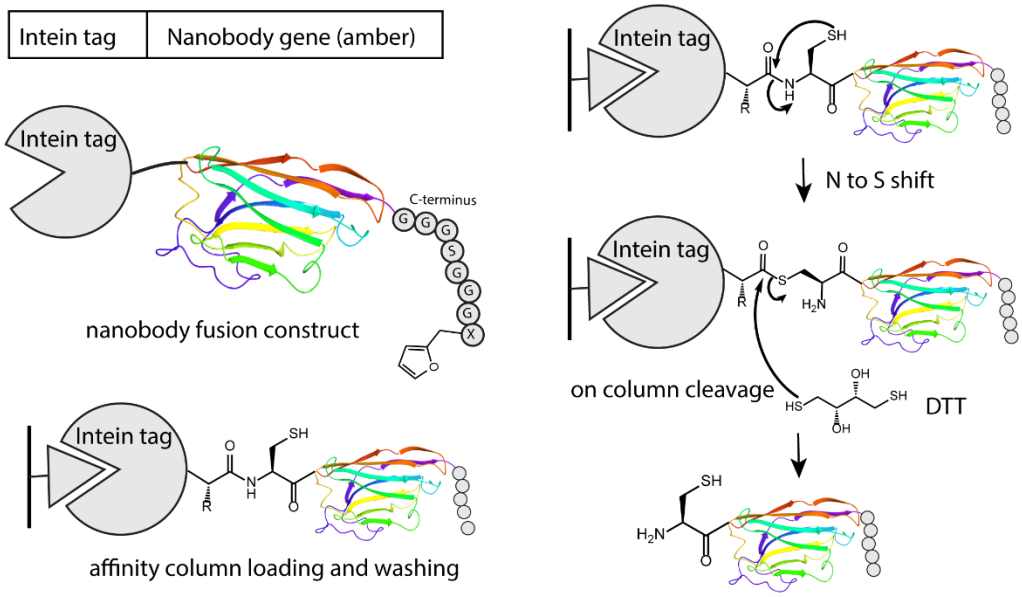


Figure 5.14 Schematic representation of the N-terminal intein-tag nanobody fusion construct for IMPACT™ purification. The fusion construct is loaded on a chitin beads column. After washing steps to remove other proteins the nanobody is cleaved on column using DTT.

The first nanobody expressed for ncAA incorporation in this work was fascin nanobody 5 (FasNb5) and, two constructs containing an intein-tag, nanobody sequence, a linker and one or two amber codons were considered (figure 5.15). Since the construct with two amber codons was readily available in the nanobody lab it was considered to gain

[2] https://international.neb.com/products/e6901-impact-kit#Product%20Information

a better understanding of the incorporation efficiency of the ncAA. The plasmid with the tRNA and tRNA synthetase for incorporation of furan-modified lysine derivative FurLys was received from Summerer.[229] After expression and purification the nanobody samples were analysed via LC-MS.

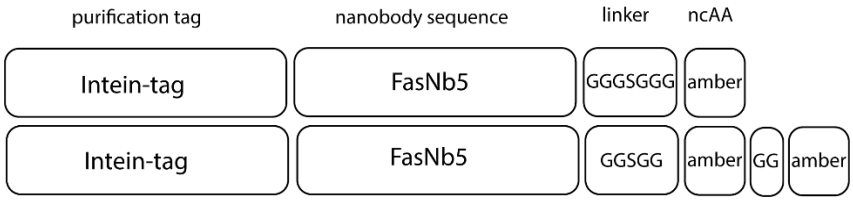


Figure 5.15 Schematic representation of the two fascin nanobody (FasNb5) constructs used in this work. In the first construct one amber codon is located at very C-terminus of the protein after a GGGSGGG linker. The second construct is almost identical to the first but has shorter linker GGSGG and an additional amber codon after the first amber codon and a small GG spacer.

The first expression attempt to incorporate FurLys in FasNb5 was carried out according to the procedure but the LC-MS analysis of the final nanobody samples yielded a surprising result (figure 5.16). The expected masses for the nanobodies with incorporated furan lysine were not detected (see Table 5.1). For the slightly larger FasNb5 amber GG amber nanobody a lower mass was detected compared to the FasNb5 amber nanobody.

The explanation for these unexpected masses is that the incorporation of FurLys did not take place and instead the protein is terminated at the amber codon by the release factor (RF 1). We realised that the incorporation was not successful because we did not induce the tRNA synthetase. This means that there was no tRNA synthetase present in the cell to charge the tRNA with the FurLys amino acid. The calculated masses for both truncated constructs (in the case of FasNb5 amber GG amber the termination occurs at the first amber codon) are in agreement with the detected mass (see Table 5.1). Although the incorporation had failed, with the LC-MS analysis of the nanobody products we were able to characterise the formed truncation products within a few Daltons. Typically nanobodies are characterised via gel electrophoresis for an estimation of the molecular weight. We could clearly show here that using LC-MS analysis, more detailed information can be obtained about the structure of the obtained constructs thus allowing us to analyse and remediate in case of unexpected outcomes.

Table 5.1 Expected and observed masses for the expressed nanobody constructs FasNb5 amber and FasNb5 amber GG amber. The expected masses are given for both successful incorporation of FurLys and protein termination at the amber codon (failed incorporation).

	Successful incorporation Expected mass (Da)	Observed mass (Da)	Failed incorporation Expected mass (Da)
FasNb5 amber	14590.18	14321	14324.01
FasNb5 amber GG amber	14856.36	14207	14209.91

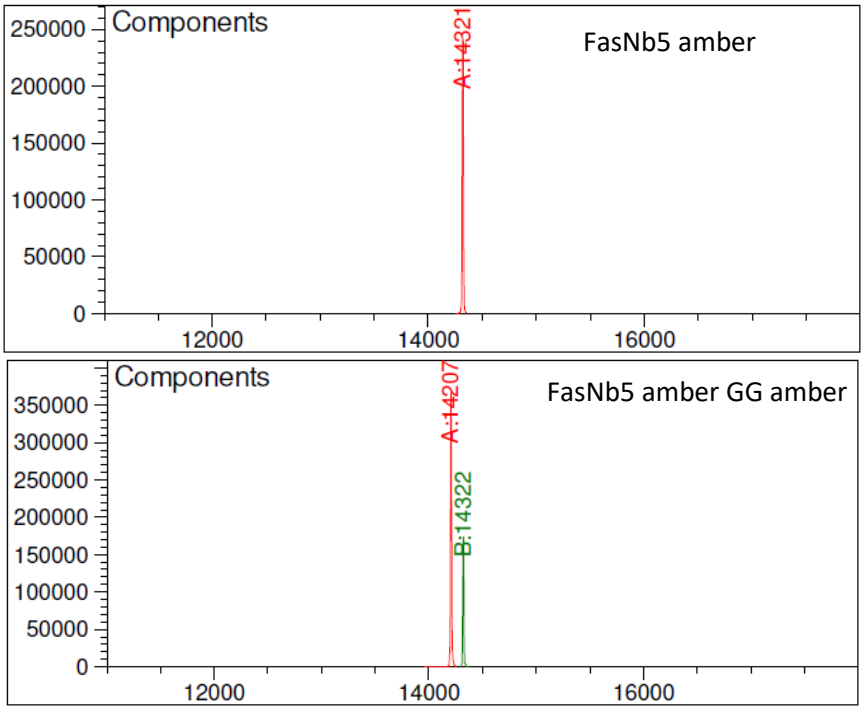


Figure 5.16 Deconvoluted MS spectra for the FasNb5 amber (top) and FasNb5 amber GG amber nanobody productions. The expected mass for FasNb5 amber with an incorporated FurLys is 14590.18 Da (found 14321 Da) and for FasNb5 amber GG amber with 2 FurLys amino acids incorporated 14856.36 Da (found 14207 Da). Additionally there is a second peak (14322 Da) present in the bottom deconvoluted MS spectrum. This mass originates from the FasNb5 amber nanobody with failed incorporation, which was analysed directly before this nanobody and was present as a column contamination without extra washing of the column.

In the second series of FasNb5 expression experiments the tRNA synthetase was induced as required. LC-MS analysis of these nanobody samples demonstrated successful incorporation of FurLys in response to the amber codon (figure 5.17). For the FasNb5 amber sample two signals are observed in the deconvoluted mass spectrum 14322 Da and 14588 Da, corresponding with the truncated and FurLys containing product. Indeed, as expected the incorporation of FurLys is in competition with termination of translation. This competition results in a mixture of two nanobody components. In the FasNb5 amber GG amber three nanobody components are observed 14207 Da; 14588 Da and 14855 Da which correspond with a termination product at the first amber codon, FurLys incorporation at the first amber codon but termination at the second amber codon and double incorporation of FurLys respectively.

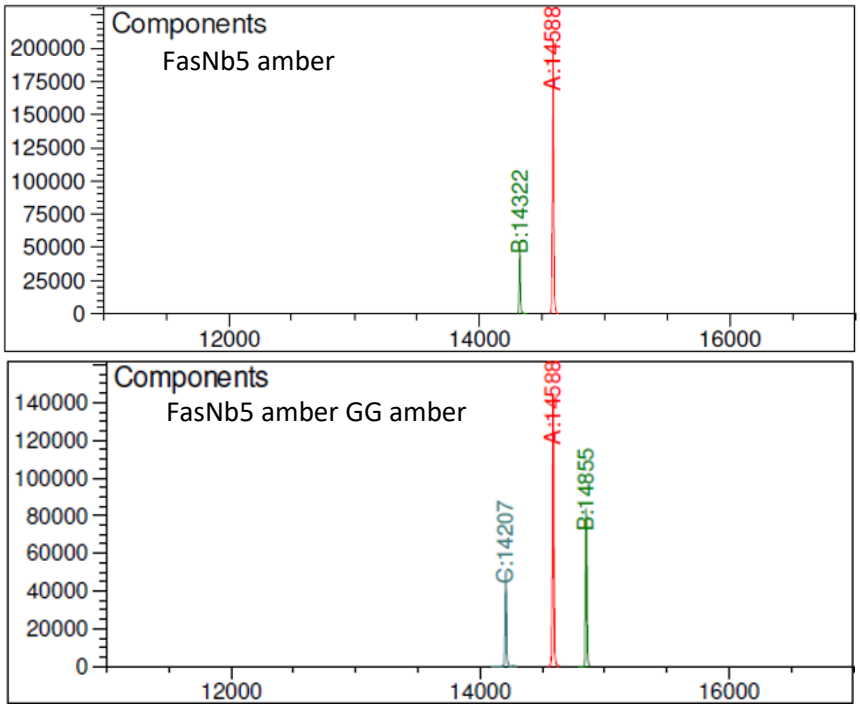


Figure 5.17 Deconvoluted MS spectra for the FasNb5 amber (top) and FasNb5 amber GG amber nanobody productions. The expected mass for FasNb5 amber with an incorporated FurLys is 14590.18 Da (found 14588 Da) and for FasNb5 amber GG amber with two incorporated FurLys amino acids 14856.36 Da (found 14855 Da). Besides the expected masses of the nanobodies with incorporated FurLys residues, also the respective truncation products are observed in the deconvoluted MS spectrum.

The ratio of the signal intensity for the nanobody species in the deconvoluted MS spectrum can be interpreted in a quasi-quantitative way since ionisation is expected to be very similar for these nanobody species. In the top panel of figure 5.17, the incorporation rate can be estimated at around 80%. By changing several parameters such as ncAA concentration and time interval between tRNA synthetase induction and nanobody induction the incorporation rate could be increased to around 90%.

The presence of the truncated nanobody species in the samples was initially regarded as a disadvantage since the truncated species is not the desired nanobody and it is not easy to remove from the final sample. However, we soon realized that the presence of the truncated and non-furan containing nanobody actually offers the benefit of serving as a very good internal control. The truncated nanobody has indeed almost exactly the same amino acid sequence as the furan-modified one except for the fact that it lacks the C-terminal FurLys ncAA. It was further used to monitor collateral oxidative damage during later labeling and crosslinking experiments.

Also the cortactin nanobody CortNb2 amber was successfully expressed using the same fusion construct. The incorporation rate of FurLys was lower in this expression, around 50 % (see figure 5.18).

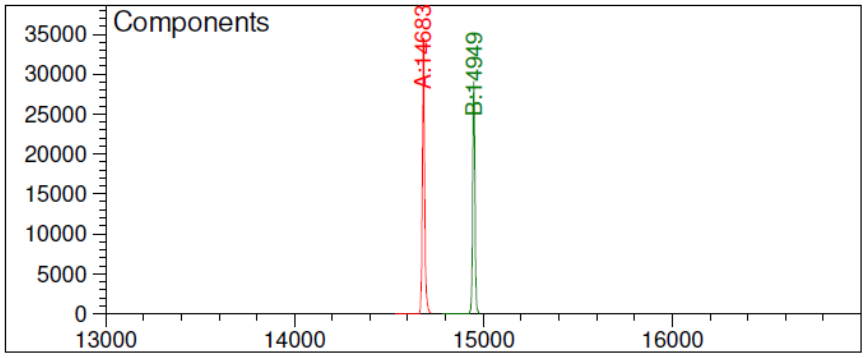


Figure 5.18 LC-MS analysis of CortNb2 amber nanobody expressed for incorporation of FurLys in response to the amber codon. Expected mass for the nanobody with incorporated FurLys: 14952.59 Da, observed mass: 14949 Da. Additionally a second mass is detected in the deconvoluted MS spectrum corresponding to the mass of the nanobody where translation is terminated at the amber stop codon 14683 Da.

5.1.3 Nanobody labelling

When considering furan moieties for labelling purposes, it is important to realise that in order to exploit the introduced furan, activation through oxidation by either NBS or singlet oxygen is required. It is however known that several amino acids are particularly sensitive to oxidative damage; these are methionine, tryptophan, histidine, tyrosine and cysteine. Such structural alterations on the amino acid level can potentially have an effect on the protein function or binding, for example if a residue important for ligand binding is oxidized. On the other hand, a protein with oxidation damage could also remain fully functional if no important residues are affected.

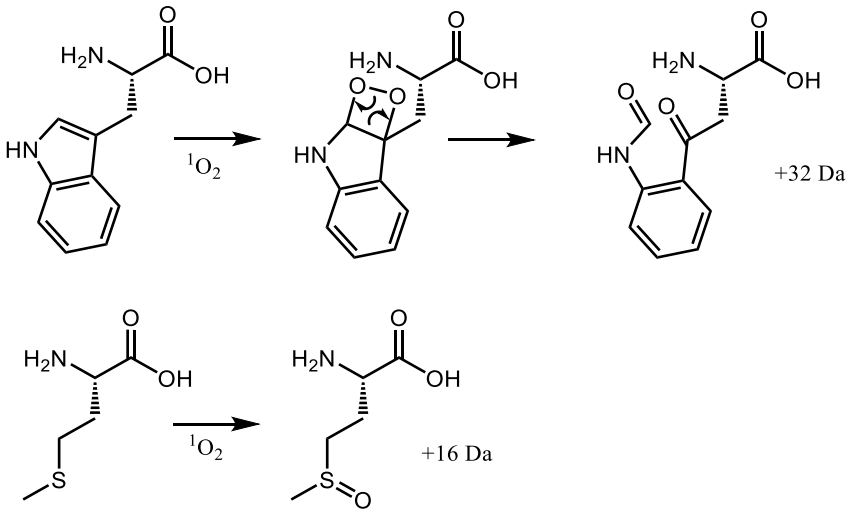


Figure 5.19 Structural representation of a proposed singlet oxygen oxidation mechanism for tryptophan (top) and the oxidation product for methionine (bottom).

To test if CortNb2 is still binding to its antigen cortactin a new type of experimental setup was developed based on a sandwich ELISA experiment (figure 5.20). The idea is to coat CortNb2 that was irradiated in presence of RB for different amounts of time

and a control that was not irradiated and had no RB present. CortNb2 samples were coated in protein binding 96 well plates. Afterwards the remaining binding sites were blocked with BSA. Cortactin is then added and allowed to bind with the coated CortNb2. Unbound cortactin is washed away and a mouse anti cortactin antibody is added and allowed to bind to cortactin. Again, unbound mouse anti cortactin antibody is washed away and a secondary anti-mouse antibody linked to horse radish peroxidase (HRP) is added and allowed to bind to the mouse antibody. The HRP is capable of oxidizing a substrate, 3,3',5,5'-tetramethylbenzidine (TMB) with hydrogen peroxide as the oxidant (figure 5.21). TMB is a chromogenic substrate that becomes blue upon oxidation. The oxidation reaction can be terminated by the addition of sulfuric acid and this causes a colour shift from blue to yellow. Finally, the yellow colour can be quantified by measuring absorbance at 450 nm by spectrophotometric methods.

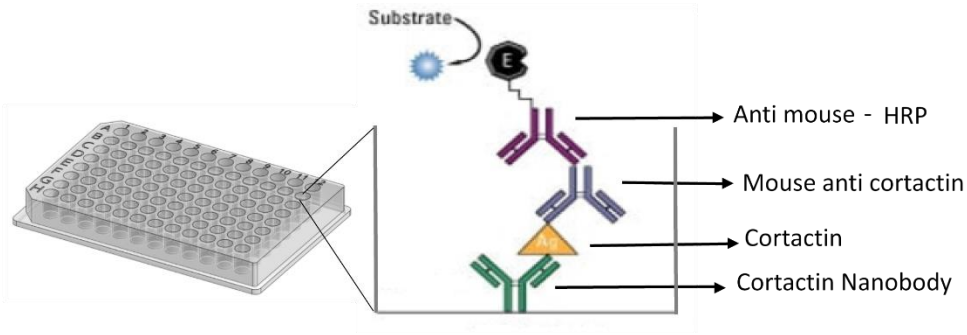


Figure 5.20 Schematic representation of the ELISA setup used for control of oxidative damage.

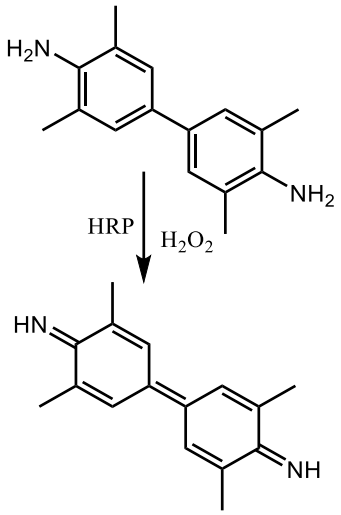


Figure 5.21 Reaction scheme for the oxidation of the TMB substrate catalysed by the HRP enzyme, hydrogen peroxide is used as an oxidant.

In a first experiment CortNb2 samples were irradiated for 0 min, 2 min, 5 min, 10 min and 20 min irradiated with white light in presence of 3 µM RB. 7.5 equivalents of the cortactin antigen were used and four replicates were made for each sample.

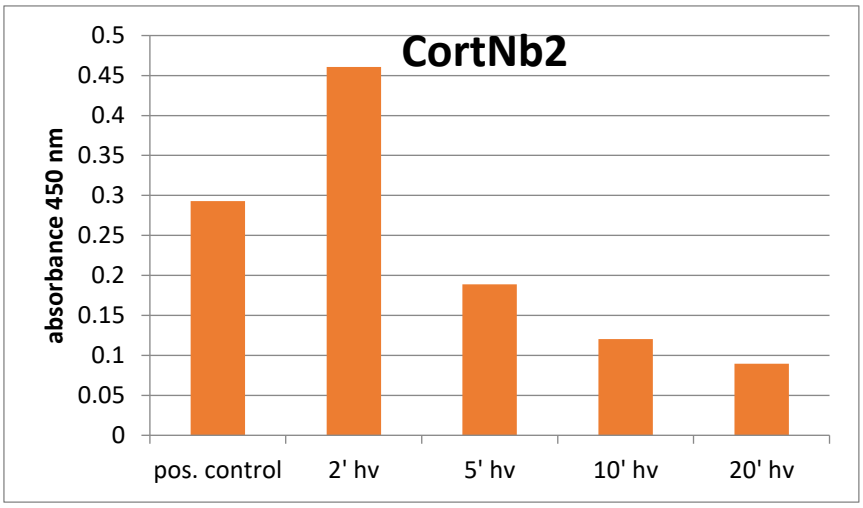


Figure 5.22 Bar graph for the oxidative damage ELISA response at 450 nm for CortNb2. Positive control: not irradiated CortNb2; 2' hv: 2 min irradiation; 5' hv: 5 min irradiation; 10' hv: 10 min irradiation; 20' hv: 20 min irradiation.

The results from the first ELISA experiments to check for oxidative damage were a bit surprising since the response from the CortNb2 sample that was not irradiated is not the most intense. The signal from CortNb2 that was irradiated for 2 minutes was substantially higher. The samples with longer irradiation times 5 min, 10 min and 20 min result in a gradually lower response with increasing irradiation times. A lower signal for longer irradiation times was indeed expected and can potentially be attributed to oxidative damage to the CortNb2 which is expected to hamper the antigen binding. The increase in ELISA response for CortNb2 sample that was irradiated during 2 min, was not expected.

A similar experiment was performed with different amounts of cortactin and a similar trend was observed with a higher signal for the CortNb2 sample that was irradiated for 2 min when compared to the positive control. In this experiment the signal decreased with increasing irradiation time as expected but all signals for irradiated samples remained more intense compared to the positive control (experimental data figure 9.16). This was in contrast to the data presented in figure 5.22. In additional experiments with 8 different cortactin concentrations we saw an increase in ELISA response for 10 min irradiation and an even higher response for prolonged irradiation for 60 min (experimental figure 9.17).

In a following oxidative damage ELISA experiment both the CortNb2 and the CortNb2 amber with incorporated FurLys were tested. The experimental setup was identical to the first oxidative damage experiment.

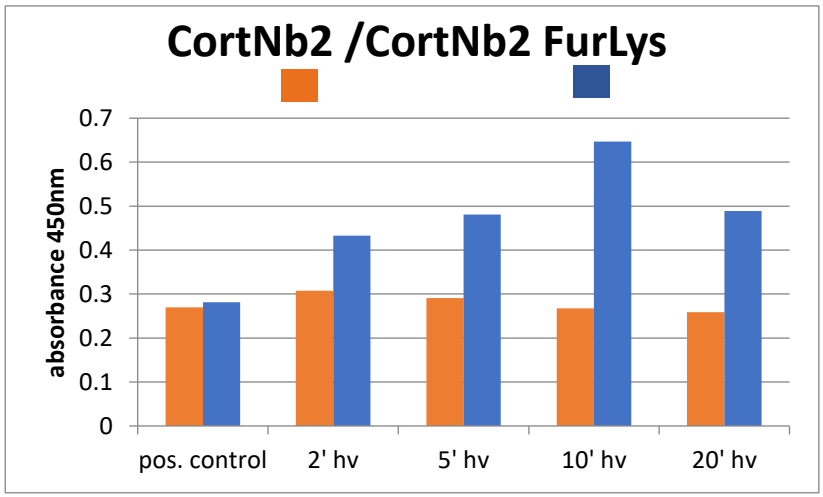


Figure 5.23 Bar graph for the oxidative damage ELISA response at 450 nm for CortNb2 (orange) and CortNb2 amber FurLys (blue). Positive control: not irradiated CortNb2; 2' hv: 2 min irradiation; 5' hv: 5 min irradiation; 10' hv: 10 min irradiation; 20' hv: 20 min irradiation.

In the ELISA results in figure 5.23 the same effect for the CortNb2 (orange) is observed but to a smaller extent, namely a slightly higher signal for the sample that was irradiated for 2 min compared to the positive control (non-irradiated sample). A further increase in irradiation time resulted in slightly lower ELISA responses for the CortNb2 samples. For the furan modified nanobody samples (blue) a different trend is observed. For the non-irradiated sample a similar ELISA response is detected compared to the CortNb2 sample.

On the other hand, the intensity of the ELISA response is substantially higher when compared to the respective CortNb2 samples for all irradiated samples. The samples irradiated for 10 min resulted in the highest response, which was more than double the value for the CortNb2 sample. These results were repeated with different equivalents of the cortactin antigen resulting in a similar outcome. Although the results for the oxidative damage ELISA were interesting, especially the substantial increase in signal for furan-modified nanobodies, the results proved to be not consistent over consecutive experiments as evidenced by the data on CortNb2.

Potential explanations for these unexpected ELISA results:

- The nanobody samples are purified after irradiation over a small spin column (separation based on size exclusion) to remove small molecule compounds such as the photosensitizer. The oxidation of nanobody samples could have an impact on the retention of these proteins on such a spin column leading to a potential concentration difference in comparison with a non-oxidized sample, ultimately leading to altered ELISA signal.
- The nanobody samples are coated on a 96-well plate after irradiation and size exclusion purification. Potentially, oxidation of the nanobody could affect the adsorption to the well-plate leading to a different amount of coated nanobody. This difference in coating could in turn lead to a different ELISA signal.
- Additionally, oxidative damage could potentially lead to an enhanced binding of the nanobody to its target protein. Methionine residues can be oxidized to sulfoxides, this could lead to an additional binding contact strengthening the binding affinity.

While the nanobody is ionising well which allows for the generation of a strong signal in MS analysis, HPLC analysis is hampered by poor absorption in the UV region. Therefore, we typically rely on LC-MS for analysis of the experiments with the furan-modified nanobodies. We demonstrated on the peptide level that photo-oxidation is required for subsequent peptide labelling and NBS oxidation does not allow for successful labelling (*vide supra*). Proteins are large and intricate structures that are more sensitive to oxidation compared to peptides.[137][235] The general photo-oxidation was tested on CortNb2 (no FurLys). The deconvoluted mass spectrum in the LC-MS analysis of a CortNb2 nanobody after 10 minutes irradiation at 10 °C in presence of RB (figure 5.24) showed in total five +16 Da oxidation signals. These results indicate indeed that several oxidation sensitive residues in the nanobody are oxidized (cfr. Figure 5.19) resulting in multiple additions of 16 Da.

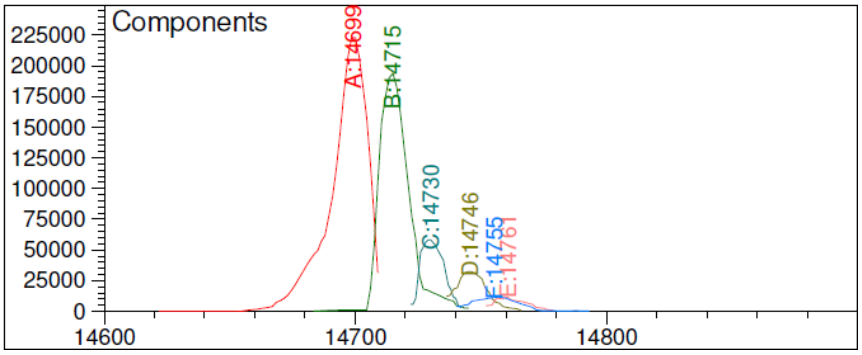


Figure 5.24 Deconvoluted MS spectrum of CortNb2 after 10 minutes irradiation with white light at 10 °C in the presence of a photosensitizer (RB). The expected mass for the CortNb2 nanobody is 14683 Da, In the MS spectrum the expected mass is absent and in total 5 +16 oxidation peaks are observed (1 ox: 14699 Da; 2 ox: 14715 Da; 3 ox: 14730 Da; 4 ox: 14746 Da; 5 ox: 14761 Da).

The FurLys containing CortNb2 amber was subjected to the same photo-oxidation procedure followed by LC-MS analysis. The deconvoluted mass spectrum is depicted in figure 5.25 and a similar result was obtained with respect to the non FurLys containing truncated nanobody species (which is present in view of the non-complete incorporation of the FurLys ncAA, *vide supra*). Indeed, also in this case several +16 oxidation peaks were observed for the truncated nanobody (1 ox: 14699 Da; 2 ox: 14715 Da; 3 ox: 14730 Da; 4 ox: 14746 Da). However the FurLys nanobody of interest or any +16 oxidized variants were not detected at all in the deconvoluted MS spectrum in figure 5.33. The observed mass for the CortNb2 amber FurLys nanobody was 14949 Da but in the MS spectrum no peaks are detected with a molecular weight higher than 14873 Da. The FurLys nanobody appears to be completely consumed and unexpected peaks appear starting from 14828 Da. This could be an addition of a small compound to the truncated CortNb2 nanobody but no label molecule was added in this experiment and moreover, this would not explain the absence of the CortNb2 amber FurLys nanobody.

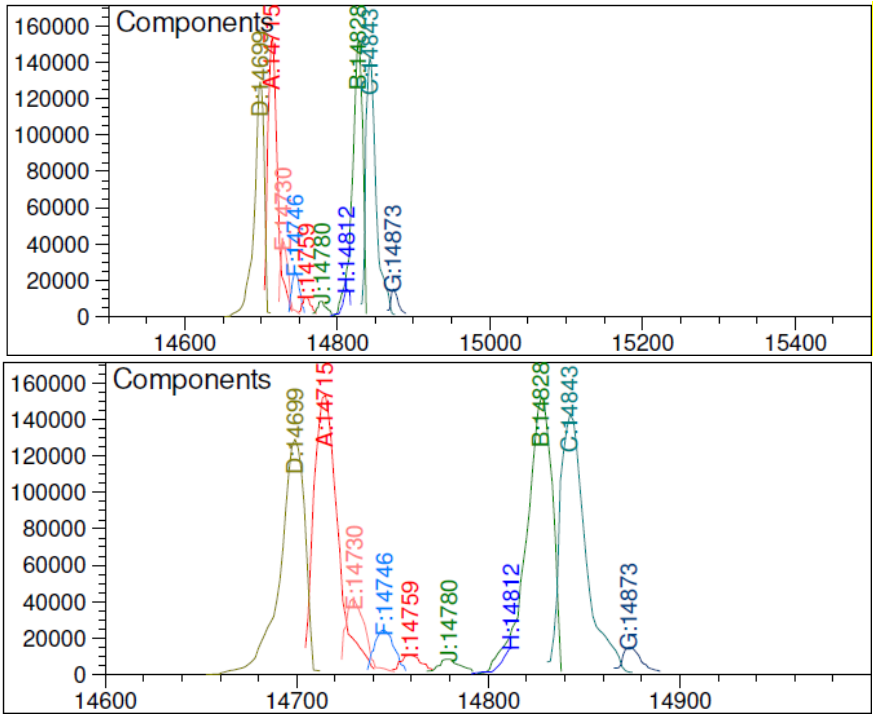


Figure 5.25 Deconvoluted MS spectrum of FurLys containing CortNb2 amber after 10 minutes irradiation with white light at 10 °C in the presence of a photosensitizer (RB), 14500-15500 Da (top panel) and 14600-15000 Da (bottom panel). The expected mass for the CortNb2 nanobody is 14683 Da, In the MS spectrum the expected mass is absent and in total four +16 oxidation peaks are observed for the nanobody species without FurLys (1 ox: 14699 Da; 2 ox: 14715 Da; 3 ox: 14730 Da; 4 ox: 14746Da). Additionally other prominent peaks are observed: 14828 Da, 14843 Da in combination with several other small peaks.

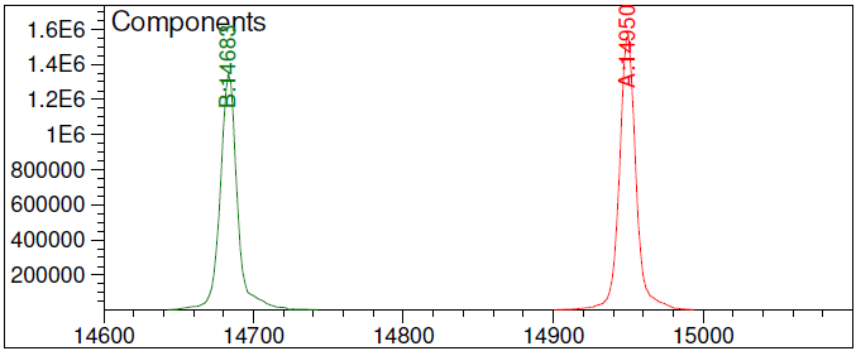


Figure 5.26 Deconvoluted MS spectrum of CortNb2 amber FurLys after 10 minutes irradiation with white light at 10 °C (in the absence of a photosensitizer). The expected mass for the CortNb2 nanobody is 14686 Da, observed 14683 Da. The expected mass for CortNb2 amber FurLys is 14952 Da, observed 14950 Da.

As a control experiment, the same irradiation was performed in the absence of a photosensitizer and the sample was analysed via LC-MS. The resulting deconvoluted mass spectrum is presented in figure 5.26. Both the CortNb2 at 14683 Da and the FurLys containing CortNb2-amber at 14950 Da are present without any oxidation products. The peaks at 14828 and higher are also completely absent in this sample. These results indicate that, as expected, the oxidation process is responsible for the

degradation of the furan modified nanobody and the generation of an extra nanobody species with a mass in between that of both starting nanobody species.

We were puzzled with this unexpected result for a while until we hypothesized that the FurLys amino acid was in fact breaking down during the oxidation process. The carbamate bond between the furan moiety and the protein backbone is cleaved during the oxidation process and the furan part is lost. The amino acid remaining is a regular lysine residue after degradation (figure 5.27). The mass loss upon degradation is 138 Da. The mass difference between 14950 Da (FurLys-modified CortNb2 amber) and 14828 Da (most intense degradation peak) is 122 Da. The discrepancy of 16 Da between the expected mass difference and the found mass difference is caused by an oxidation (+16 Da) of the degraded protein. It is not clear if the degradation occurs only in a labelling setup where there is no proximate nucleophile present or if the degradation also takes place in a crosslinking setup with a nucleophile in close proximity.

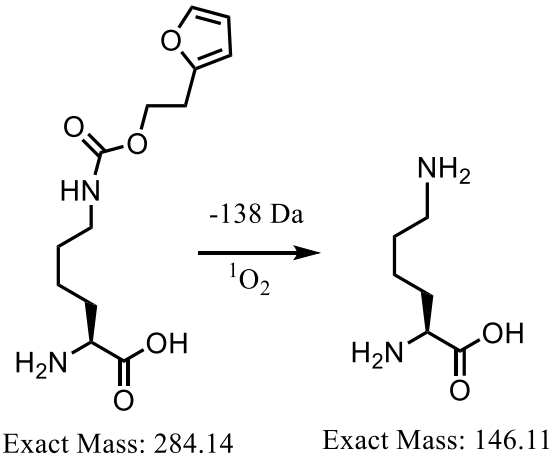


Figure 5.27 Structural representation of the oxidative degradation of the FurLys ncAA to lysine. The corresponding mass loss is 138 Da.

Since the furan oxidation step, which is needed to activate the furan for labelling, proved to cause degradation of the FurLys ncAA we attempted to label via another approach. Previously our group reported the use of TAD reagents for the labelling of furan-modified peptides on solid support.[236] Inspired by this, we tried to label FurLys-modified FasNb5 amber with TAD. Although TAD based strategies for protein modifications are in detail described in chapter 6, in this specific experiment a combination of furan and TAD chemistry is explored. FurLys-modified FasNb5 amber was labelled with 20 equivalents of phenyl-TAD (PTAD). Remarkably, an identical degradation of the FurLys amino acid was observed upon LC-MS analysis (figure 5.28).

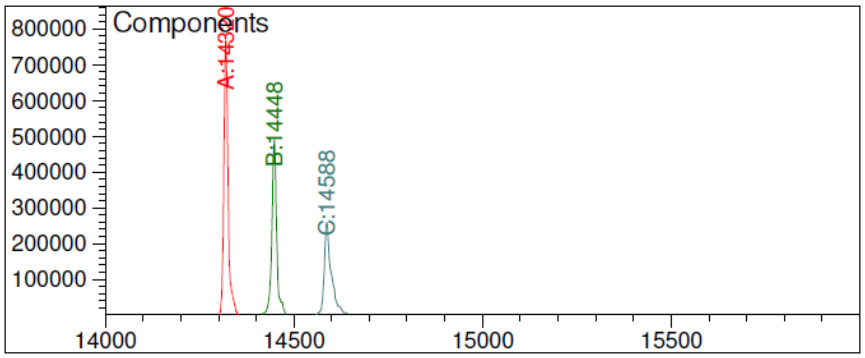


Figure 5.28 LC-MS analysis of FurLys-modified FasNb5 amber after conjugation with 20 equivalents of PTAD. Both the FasNb5 with FurLys and the truncated nanobody are observed as expected (14588 Da and 14320 Da respectively). However no species is detected corresponding with the desired conjugated nanobody with a higher mass than the FurLys-modified FasNb5 amber nanobody. Additionally a third compound is detected with a mass lower than the furan lysine nanobody (14448) corresponding to the loss of the furan moiety in FurLys as described in figure 5.35. .

It seems that also the TAD labelling of the furan modified nanobody results in degradation of FurLys since no species with a higher MW is detected while a peak corresponding to ncAA degradation appears.

Although the tRNA/tRNA-synthetase pair for genetic code expansion with FurAla was received towards the end of this thesis we were not able to prove incorporation of this ncAA. The molecular weight of FurAla is nearly identical to histidine (0.01 Da difference), this small difference renders the confirmation of FurAla incorporation via LC-MS impossible. In the initial paper reporting the incorporation of FurAla via GCE, selective incorporation was demonstrated via a negative control (*vide supra*). The difficulty in the LC-MS analysis in combination with the limited time led to the fact that GCE with FurAla was not fully investigated.

5.1.4 Conclusions: towards protein labelling

The C-terminal amino acid sequence of the nanobody, containing a furan amino acid was successfully synthesized as a first model compound via SPPS. The FurAla was used for testing the peptide test system since it had been shown earlier that the FurLys side chain is not stable during the acidic conditions required for peptide cleavage. On the other hand, as for the protein model system, we were only able to prove incorporation of the FurLys amino acid in proteins using genetic code expansion (GCE). Although the tRNA/tRNA-synthetase for GCE with furylalanine was received towards the end of this thesis we were not able to prove incorporation for this ncAA (*vide supra*). As a result, this unfortunately means that our peptide model system of the C-terminus contains a different furan ncAA compared to the nanobody model system. We were able to successfully use the previously developed furan oxidation methodology for the labelling of the furan-modified model peptide. This labelling was also shown possible in the presence of DTT.

However, when the furan-oxidation approach was explored for labelling at the protein level, two issues were revealed.

- The first issue concerns the oxidative damage that is observed, with several +16Da peaks showing in the LC-MS analysis after analysis of a non-furan containing nanobody that was exposed to the oxidative irradiation conditions. These extra oxidation peaks correspond to oxidation of amino acid side chains as a result of the photo-oxidation process required for the furan oxidation. Such oxidative damage to amino acids can have an impact on the protein binding to its target. This was investigated using an ELISA setup where the binding of furan modified nanobodies and wildtype nanobodies was tested after irradiation. From these oxidative damage ELISA results it seems that for wildtype nanobodies the binding increases after a very brief two minute irradiation, while longer irradiation substantially lowers binding affinity. On the other hand, for the furan containing nanobodies the results indicate that the binding affinity substantially improves after longer irradiation. This increase in ELISA response was unexpected and several repetition experiments were performed. The oxidative ELISA experiments proved not very reproducible casting doubt on the results. Additionally secondary effects could be at play resulting in artificially higher ELISA response signal. A potential secondary effect is possible positive influence of photo-oxidation on the absorption of the nanobodies to the 96-well plate. If more nanobody is absorbed to the 96-well plate or the nanobody is better oriented this could lead to an artificial higher response. The oxidative damage can likely be remediated to some extent by using a different photosensitizer and reduced irradiation time. The oxidative damage was not optimized in this instance since a second unexpected issue came into play.
- The second issue we observed when applying the furan oxidation method on furan-containing nanobodies came rather unexpected. Upon LC-MS analysis of furan-modified nanobody samples a peak was observed after oxidation with a mass value corresponding to a nanobody with a lysine residue instead of a FurLys residue. Additionally the peak corresponding to the FurLys-containing nanobody was either reduced or completely absent. This indicates that, upon oxidation, the carbamate linker in the FurLys ncAA is cleaved resulting in a lysine residue on the amber position. Remarkably also an attempt to label the furan modified nanobody with PTAD resulted in similar FurLys degradation.

In light of the side effects we observed with the use of the FurLys residue, we decided that labelling of furan modified proteins might be possible with an appropriate furan containing ncAA, potentially furylalanine, after careful optimisation of the oxidative damage. However the added value over existing click chemistry applications would be rather limited and we therefore decided to focus our attention on applications in the protein crosslinking area.

5.2 Towards protein crosslinking

In the previous section of this chapter on the further development of the furan chemical platform the objective was to use furan modified nanobodies for labelling purposes with external nucleophilic label molecules. In such a labelling approach there is no pre-existing proximity between the label molecule and the furan-modified protein, as a result a label molecule with a strong nucleophilic properties is required for reaction with the oxidized furan moiety as was demonstrated for furan-mediated peptide labelling (*vide supra*).[134] In this section we aim to further develop furan chemistry for protein crosslinking. In this crosslinking approach model nanobodies which bind to a certain antigen are required.

We aim to use GCE to generate furan-modified nanobodies and upon incubation of these nanobodies, binding of the nanobody with its antigen results in a proximity between the furan moiety and a nucleophilic residue on the antigen. When the furan moiety which is present on the nanobody is oxidized it becomes reactive towards the proximate nucleophile on the antigen, potentially resulting in a covalent crosslink.

5.2.1 Gelsolin crosslinking nanobody

The Gettemans lab has been interested in gelsolin since the late 1990's because of its pivotal role in actin network dynamics.[237] Gelsolin is a calcium regulated actin-binding protein with a molecular weight of around 80 kDa part of the gelsolin family of actin associated proteins. Gelsolin can act as a powerful protein "knife" to disrupt the non-covalent interactions between actin monomers in an actin filament.[238] This breaking up of actin filaments can allow the cell to remodel the actin filament structure. Gelsolin null cells (gelsolin is not present) exhibit many features of reduced motility and actin dynamics. Consisting of six homologous subdomains gelsolin is divided in two segments of three subdomains with a long linker in between the two segments. Each of these subdomains binds calcium. There are two variants of gelsolin: plasma gelsolin and cytoplasmic gelsolin resulting from alternative splicing. The cytosolic variant is involved in actin skeleton remodelling (*vide supra*) and the secreted variant (plasma gelsolin) acts as an actin scavenger in the bloodstream.[239]

Familial amyloidosis-Finnish type (FAF) or gelsolin amyloidosis is a rare disorder resulting from a point mutation in gelsolin.[240] In Finland 400-600 people are affected by this disease and there are 15 cases elsewhere. Amyloid diseases result from aggregation of peptides/proteins in tissue, the depositions of these amyloid fibrils in tissue are termed plaques. Although there are many different amyloid related diseases caused by many different aggregating peptides/proteins, the amyloid fibrils have a common core structure.[241] Typically, the structure of these amyloid fibrils consists of a helical array of parallel β-sheets in the longitudinal direction of the fibre with the β-sheet strands perpendicular to the long axis.

The mutation causing FAF is located in the second subdomain where an aspartic acid is replaced by an asparagine or tyrosine (D187N/Y). This mutation jeopardizes calcium binding for this second subdomain (figure 5.29). The absence of calcium in turn leads to a defective folding of the second subdomain and because of the incorrect folding, a protease cleavage site is exposed (gain of function mutation). This initial cleavage by furin in the trans-Golgi network results in a 68 kDa fragment C68, which is subsequently cleaved by membrane type 1-matrix metalloproteinase (MT1-MMP). The final products of this proteolytic cascade are 5 and 8 kDa amyloidogenic peptides which are the cause of pathogenic plaque formation in multiple tissues of affected patients. The main implications of FAF are amyloid deposits in the cornea, bilateral facial paralysis and "sagging" skin. The symptoms typically begin in an individual's 20s or 30s, the symptoms worsen with age although the progression is typically slow.[3]

[3] https://rarediseases.info.nih.gov/diseases/2339/familial-amyloidosis-finnish-type#:~:text=This%20condition%20is%20inherited%20in,facial%20paralysis%20and%20cutis%20laxa.

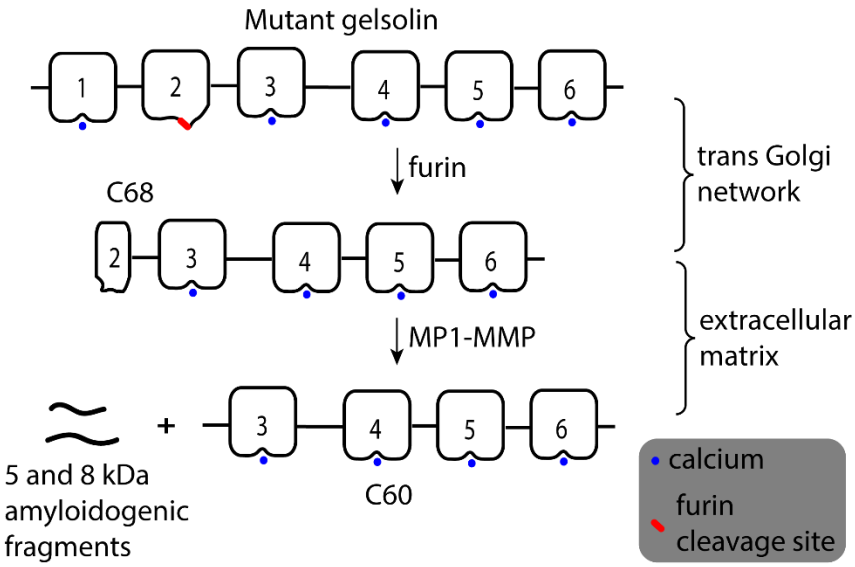


Figure 5.29 Schematic representation of the mutant gelsolin proteolytic cascade, furin cleavage in the trans Golgi network followed by MP1-MMP cleavage in the extracellular matrix. The proteolytic cascade results in amyloidogenic peptide fragments of 5 and 8 kDa.

Currently treatment for this rare disease focuses on specific signs and symptoms; there is currently no treatment available to prevent the formation of these amyloid plaques. In the nanobody lab of professor Gettemans nanobodies were developed against the 8 kDa amyloidogenic peptide.[239] It was demonstrated that these nanobodies bind the C68 fragment and reduce MT1-MMP degradation *in vitro* (figure 5.30). Additionally it was shown that the same effect takes place *in vivo* in transgenic mice expressing D187N human gelsolin. However the *in vivo* reduction of plaque formation was 15-30%.[239]

The binding of the FAF nanobody (FAFNb1) to C68 is transient and the nanobody is in an equilibrium between the bound and dissociated stage (Kd ≈ 0.6 µM). This means that when the nanobody is dissociated the C68 fragment is vulnerable for MT1-MMP degradation. To avoid this bound/dissociated issue, we envisaged to use a furan containing crosslinking nanobody which is able to crosslink to the C68. The hypothesis is that such covalent connection should provide the C68 with permanent protection from proteolysis resulting in substantially reduced plaque formation.

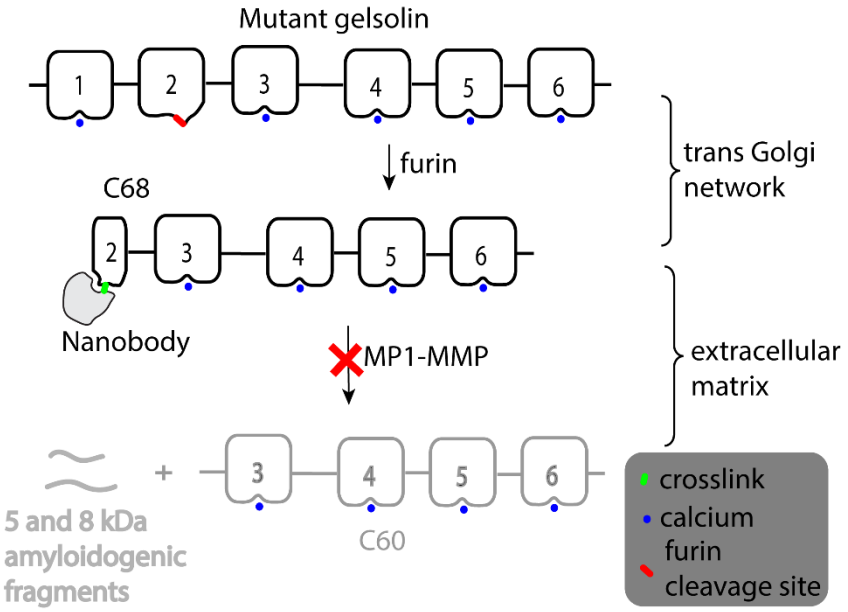


Figure 5.30 Schematic representation of the mutant gelsolin proteolytic cascade protected by a crosslinked nanobody. The furin cleavage in the trans Golgi network takes place however the following MP1-MMP cleavage in the extracellular matrix is blocked by the crosslinked nanobody.

For the furan-modified crosslinking nanobody to be effective, the exact location of the furan ncAA is of crucial importance. The furan moiety can be incorporated in the nanobody on a predefined position using genetic code expansion using amber suppression. The furan moiety should be located in close proximity to an opposing lysine residue to allow for crosslink formation, so that a suitable nucleophile is available for formation of a covalent link with the activated furan moiety. Unfortunately, no X-ray structure of the bound complex is available and the structure of the interaction interface between the nanobody and the C68 is not well known. The anti-gelsolin nanobody was developed against the 8 kDa peptide fragment (figure 5.31) and we decided to start from that peptide to try and get an idea of the exact nature of the interaction site. To have a better idea of where exactly the epitope is located, this fragment was divided in four overlapping peptides spanning the entire sequence. The four peptides were tested for affinity for the nanobody using ELISA, and it was demonstrated that, the C terminal peptide demonstrated binding (underlined sequence, unpublished data).

ATEVPVSWESFNNGNCFILDLGNNIHQWCGSNSNRYERLKATQVSKGIRDNERSGRARVHV SEEGTEPEAM

Figure 5.31 Amino acid sequence of the 8 kDa peptide fragment. The underlined peptide sequence was demonstrated to show affinity for the nanobody indicating that the binding epitope is located within the underlined peptide. Lysine residues (indicated in red) close to the epitope are potential targets to form the crosslink with a furan-modified nanobody.

To further narrow down the location of the epitope in this underlined peptide fragment (figure 5.31), a series of new peptides were synthesized. We decided to make peptide variants containing respectively three and six amino acids less from the N-terminus (figure 5.32). This will allow to make an estimation how close the epitope is located to the lysine residues which are located towards the N-terminus.

WT: RDNERSGRARVHVSEEGTEPEAM

WT-3: ERSGRARVHVSEEGTEPEAM

WT-6: GRARVHVSEEGTEPEAM

Figure 5.32 Sequences of synthesized peptides: WT, WT-3 and WT-6

These peptides were synthesized via SPPS and biotinylated (experimental section 9.2.2) and, the affinity of these peptides for the anti- gelsolin nanobody was tested using ELISA. Remarkably all three peptides still demonstrated binding to the FAFNb1 nanobody. In figure 5.33, the ELISA response is shown for eight tenfold dilution series of the WT-6 peptide. Both a positive (8 kDa wildtype gelsolin fragment) and negative control (gelsolin fragment with no FAFNb1 affinity) is provided. Since we still observe binding of the WT-6 peptide to FAFNb1, this indicates that the epitope is located towards the C-terminal part of the gelsolin-derived underlined sequence in figure 5.31. Unfortunately, these results also indicate that the epitope is likely located rather far away from the nucleophilic lysine amino acids.

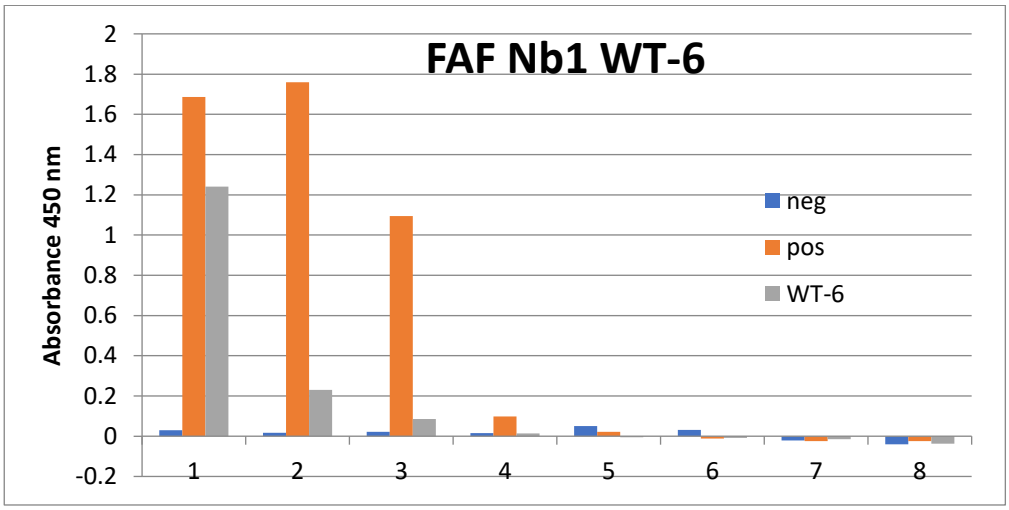


Figure 5.33 Bar graph of the ELISA results for peptide **WT-6** and FAFNb1. ELISA absorbance at 450 nm is plotted against 8 tenfold dilutions of the FAF nanobody. The negative control (blue) consists of a gelsolin fragment with no affinity for FAF nanobody whereas the positive control (orange) consists of the wildtype gelsolin fragment.

During the experiments, it was observed that the peptide concentration in the solution is diminishing over time. Since the peptides are known to be amyloidogenic we decided to use dynamic light scattering (DLS) to determine if particles were formed that cause the solution peptide concentration to decrease. DLS measurements indeed showed particle formation indicating aggregation (experimental data section 9.2.2.3). Moreover, we demonstrated that sonication of the peptide samples followed by sequential DLS measurements causes the particles to grow in size (see figure 5.34).

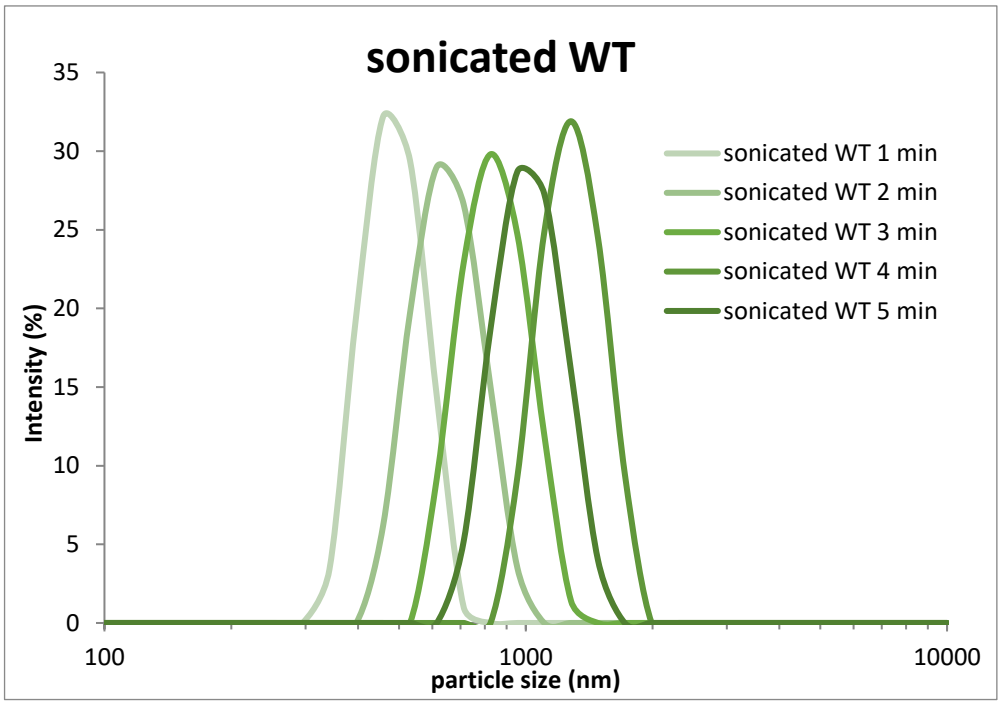


Figure 5.34 Dynamic light scattering (DLS) graph of **WT** peptide. The peptide sample was sonicated for 15 minutes and afterwards the sample was measured at 1, 2, 3, 4 and 5 minutes after sonication.

Due to this peptide aggregation phenomenon in combination with the fact that the epitope on the 8 kDa peptide fragment is seemingly not located in close proximity to a suitable nucleophilic lysine residue, we decided to discontinue the investigation towards a furan-modified crosslinking variant of FAFNb1.

5.2.2 EgA1 crosslinking nanobody

Another model nanobody-target interaction we selected for the development of a furan crosslinking nanobody is the EgA1 nanobody-EGFR interaction. The EGFR is a transmembrane protein acting as a receptor for extracellular epidermal growth factor (EGF) ligands. EGFR is part of a family of receptor tyrosine kinases. Abnormal EGFR activation is involved in an array of human cancers.[242] Not surprisingly, several classes of therapeutic agents that target the EGFR have been / are being developed. It was shown that antibody binding to the EGFR can inhibit receptor activation and tumour growth.[243] However antibodies have some limitations (tumour penetration, cost of production, ...) originating from their size and complexity (*vide supra*).

To remediate these limitations, nanobodies offer an attractive alternative. Several nanobodies have been reported to bind EGFR.[244] EgA1 is a nanobody that binds an epitope of EGFR preventing the conformational rearrangement which is required for receptor activation. In this project, the EgA1-EGFR interaction was specifically selected because a crystal structure of the bound complex is available. Indeed, in the preceding gelsolin nanobody crosslinking project, considerable effort was spent in determining the exact location of the binding epitope due to the absence of crystal structure. The crystal structure of the EgA1 nanobody in complex with the extracellular part / soluble fraction of the EGFR (sEGFR) was studied in detail *in silico* to determine if a suitable

lysine residue is present on the EGFR protein in close proximity of the EgA1 nanobody (figure 5.35, crystal structure PDB: 4KRO).

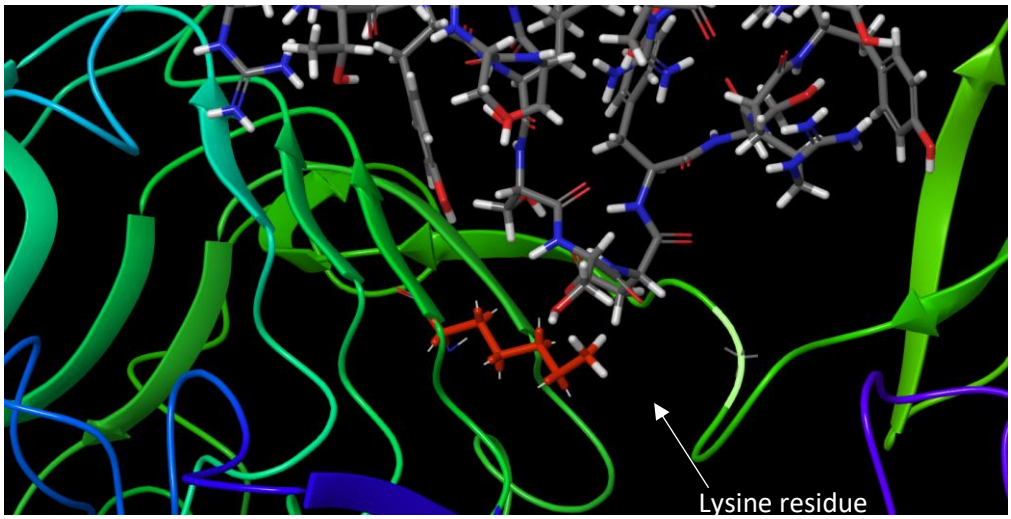


Figure 5.35 Zoom of the crystal structure at the interaction site of the EgA1 EGFR complex. The EGFR is depicted in cartoon representation with the lysine residue in red stick representation. EgA1 is depicted in stick representation.

From the crystal structure it is clear that a strategically located lysine residue is present on the EGFR protein in the interaction site (see figure 5.35). The idea is that a furan-containing EgA1 nanobody will be generated so that this lysine can act as a nucleophile and form a covalent crosslink with the furan-containing EgA1 nanobody upon activation. The objective of this project is to first establish crosslinking *in* vitro between the EgA1 nanobody and sEGFR. In this *in vitro* setup, the oxidation of the furan amino acid will be performed via the photo-oxidation approach as described above. Subsequently we aim to crosslink the EgA1 nanobody to EGFR on the surface of live cells. Previously, or group reported the crosslinking of furan-modified Kisspeptin peptide to the Kiss receptor on live cells.[136] Crucially, it was shown that the furan activation is effected by reactive oxygen species produced by the cancer cells themselves. This means that, in contrast to earlier applications of the furan methodology, where a separate photo-oxidation or NBS oxidation step (*vide supra*) is required. In this *in vivo* setting, the oxidation takes place *in situ* on the surface of cancer cells.

As we know from earlier reports on furan crosslinking in both the DNA and protein realm (*vide supra*), the proximity between the oxidized furan and nucleophile is a highly important parameter. Therefore, also in this crosslinking nanobody application, the location of the furan amino acid is of crucial importance to ensure effective crosslink formation. The crystal structure of the EgA1-EGFR complex was investigated with a focus on those residues within the EgA1 nanobody could be replaced by a furan ncAA. Two furan-containing amino acids can be used for incorporation via GCE (the furan-modified lysine derivative (FurLys) on the one hand and furylalanine (FurAla) on the other hand).

The main difference between both furan ncAA's, is the length of the spacer between the peptide backbone and the furan moiety. To accommodate both ncAA's, several positions were selected for furan incorporation; some very close to the opposing lysine residue, others slightly further away. In total four amber mutants were selected: Y100

amber, S102 amber, S103 amber and R107 amber along with the EgA1 wildtype nanobody as a control. The five nanobody genes were ordered from Eurofins Genomics and cloned in an expression vector (pHen6). The nanobody gene is under control of a Plac operator and gene expression can be induced via addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). Additionally, the nanobody contains a N-terminal PelB sequence that directs the nanobody to the periplasm after expression. The PelB sequence is removed upon arrival in the periplasm. Additionally, the nanobody also has a C-terminal His tag for purification and visualisation purposes. The complete nanobody amino acid sequence and plasmid sheet for pHen6 are included in the experimental section 9.2.1.4.

Table 5.2 Overview of the ordered nanobody genes and the amount (µg) of plasmid DNA that was received.

Nanobody	**received DNA (µg)**
wildtype	7.5
Y 100 amber	3.5
S 102 amber	1.0
S 103 amber	3.6
R 107 amber	6.2

The DNA was amplified in TOP10 cells and each construct was stored in a glycerol stock so that the ordered constructs cannot be lost.

In a first expression experiment, the EgA1 nanobody amber mutants were expressed in heat-shock competent WK6 cells. The nanobody production experiment was performed in collaboration with Laure Tack and Olivier Zwaenepoel. Five double transformations are performed with one of the five nanobody plasmids and the plasmid containing the tRNA/ and tRNA synthetase pair for incorporation of the FurLys ncAA. The transformation is done via a heat shock and the cells are grown in LB medium. The FurLys amino acid is added to the medium and the nanobody expression is induced with IPTG after an optical density at 600 nm (OD_{600}) has reached the desired value. After protein expression the cells are pelleted down and lysed. The nanobody mutants are purified via His tag affinity purification followed by gel filtration. The EgA1 nanobody mutants were analysed via SDS-PAGE and LC-MS analysis figure 5.36 and 5.37, for EgA1 Y100 amber furan lysine. SDS-PAGE images and LC-MS analysis data for the other nanobody mutants is included in the experimental data (section 9.2.2.4). For the EgA1 S102-amber FurLys nanobody not enough protein material was retrieved after purification for LC-MS analysis.

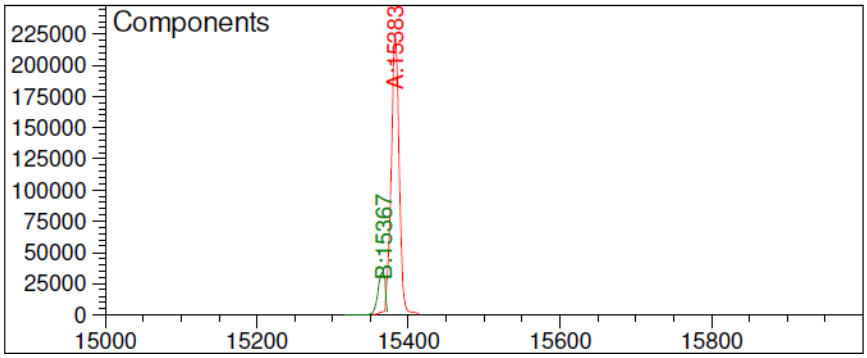


Figure 5.36 Deconvoluted MS spectrum of EgA1 Y100 amber containing FurLys. The calculated mass is 15388 Da, observed 15383 Da.

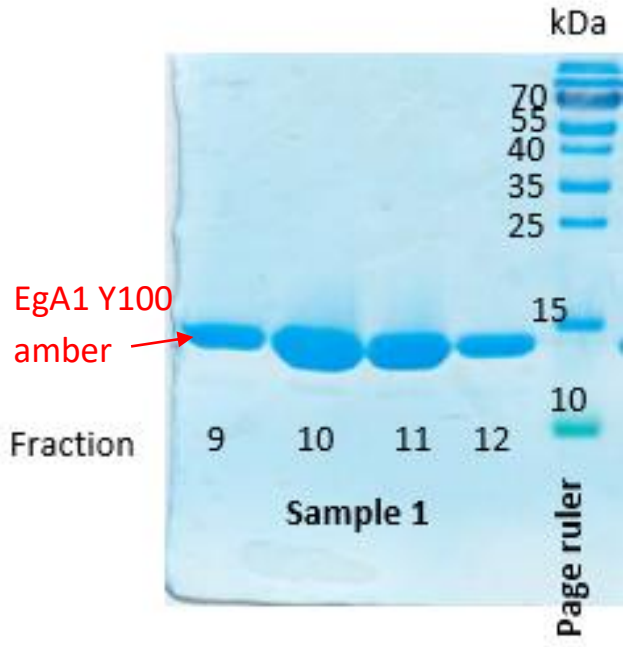


Figure 5.37 Image of the SDS-PAGE analysis of fractions 9-12 from the gel filtration of the EgA1 Y100 amber FurLys sample.

These EgA1 nanobody mutants were used for a first *in vitro* crosslinking experiment with the target sEGFR protein. These initial crosslinking experiments were carried out by Laure Tack and Dr. Laia Miret. Equimolar concentrations of nanobody and sEGFR were used and 5µM Rose Bengal was selected as a photosensitizer and irradiated for 4 minutes. The samples were subsequently analysed on SDS-PAGE (figure 5.38). In another similar experiment a positive control with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and sulfo-NHS crosslinking. Additionally a lane with only sEGFR is included in the analysis (figure 5.39).

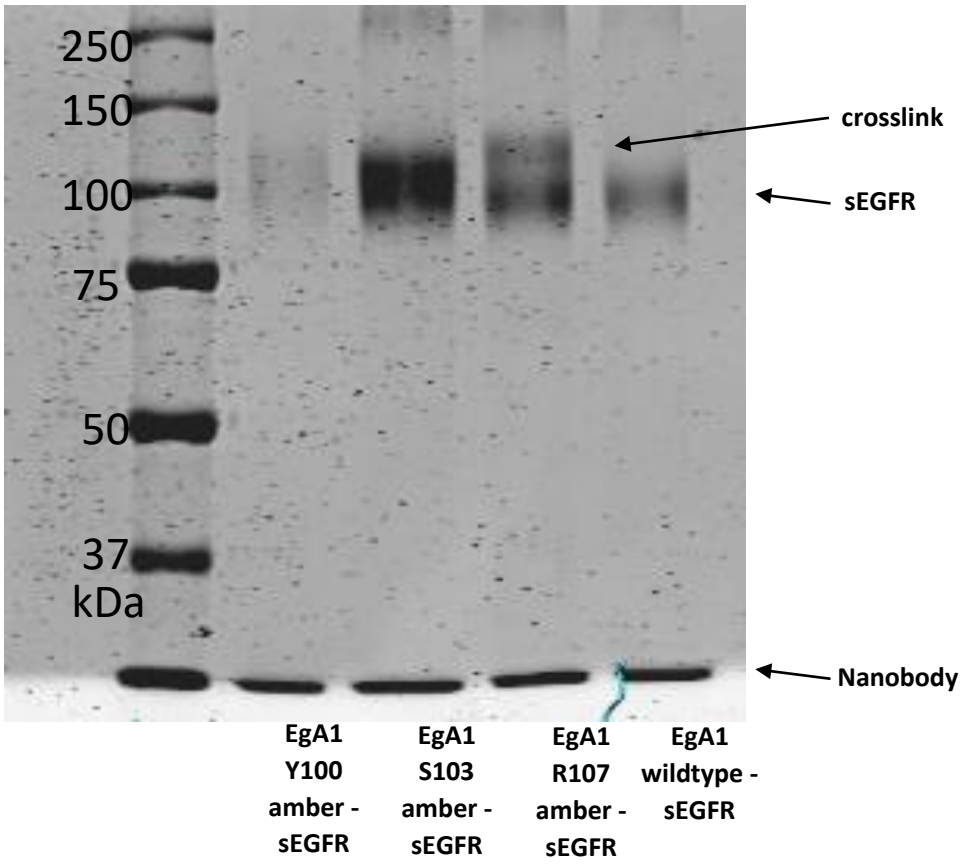


Figure 5.38 Image of the SDS-PAGE gel of *in vitro* eqiumolar crosslinking experiments between nanobody variants and sEGFR. Photo-oxidation is achieved via irradiation in presence of Rose Bengal.

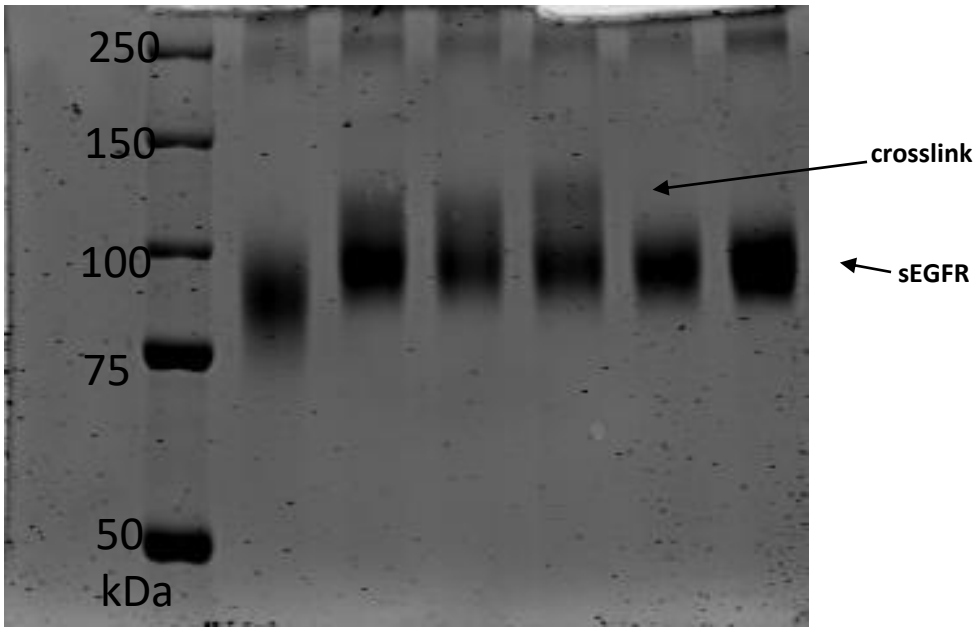


Figure 5.39 Image of the SDS-PAGE gel of *in vitro* equimolar crosslinking experiments between nanobody variants and sEGFR. Photo-oxidation is achieved via irradiation in presence of Rose Bengal. The EDC/Sulfo-NHS lane is a positive control where a combination of EDC and Sulfo-NHS crosslinkers were used.

From the SDS-PAGE analysis in figure 5.38, it can be observed that there seems to be a crosslink band visible for the R107 amber nanobody sample. This crosslinked product is present directly above the sEGFR band. Also for the S103 amber sample there is potentially crosslinked product present since the sEGFR band appears to be broad. Likely something went wrong with the Y100 amber sample since almost nothing is detected in that lane. The lane with the wildtype nanobody shows no signal right above the sEGFR band as expected since no furan moiety is present on the wildtype nanobody.

In the second SDS-PAGE crosslinking experiment these results were confirmed for the furan-containing nanobodies. Indeed, for the three furan-containing nanobodies crosslink signal is visible above the sEGFR band. For the R107 amber sample the separation between the crosslink band and the sEGFR band is again most clear. Y100 amber and S103 amber on the other hand seem to show a more diffuse crosslink band rendering the visual distinction between the sEGFR and a potentially crosslinked complex more cumbersome. There is clearly a large difference between the wildtype nanobody sample and the furan containing nanobody samples. In the wildtype nanobody sample, no signal is visible above the sEGFR band. As expected also in the lane where only sEGFR was present, also only the sEGFR band is present.

In the EDC/sulfo-NHS positive control lane a broad signal was observed but the corresponding molecular weight was lower than expected. It appears that the crosslinked product has a lower mass compared to the sEGFR itself. It is known that, as electrophoresis is based by separation on hydrodynamic radius of the compounds,

this does not always correlate directly with molecular weight. This means that it is possible for a densely crosslinked product to migrate faster than expected for its molecular weight. This could form a potential explanation for the location of the band in the positive control sample (containing EDC/SulfoNHS). These very promising results form a solid basis for further *in vivo* experiments where the furan modified nanobodies will be exploited and further investigated for a covalent crosslink formation to the extracellular part of the EGFR on live cells.

5.2.3 Conclusions: towards protein crosslinking

Furan chemistry was used in two different nanobody-target interaction model systems, the gelsolin nanobody model and the EgA1 nanobody model. Both of these model systems were selected in view of their therapeutically relevant properties.

The gelsolin nanobody (FAFNb1) interaction with the 8 kDa fragment was investigated as a first test case. The objective was to develop a furan crosslinking variant of the FAFNb1 nanobody able to crosslink to the 8 kDa gelsolin fragment and thus prevent the further proteolysis of that fragment. The first obstacle in this project was the absence of a crystal structure of the nanobody target interaction. We tried to locate the binding epitope on the 8 kDa fragment using ELISA of synthetic peptide variants. We were able to further narrow down the location of the epitope more towards the C-terminus of the 8 kDa peptide. This unfortunately indicates that the lysine on the 8 kDa fragment is likely further away from the binding furan modified FAFNb1. Additionally we observed peptide aggregation in the purified peptide samples. Although we did not anticipate this aggregation it is not entirely surprising since proteolysis of the 8 kDa fragment leads to pathogenic aggregation, ultimately resulting in plaque formation. The combination of these results and insights lead to the decision that or attention is better focused on a model system where a crystal structure is available and where a suitably located lysine residue is located on the nanobody target protein.

An example of a nanobody-target interaction with a known crystal structure and a proximate lysine residue located on the target protein is the EgA1-EGFR interaction. This interaction was identified after a thorough literature study for suitable nanobody protein interactions on the protein databank. The EgA1 nanobody was ultimately selected due to its well-studied and therapeutically relevant target. Five EgA1 nanobodies were constructed, in four of which an amber codon was inserted on a different predefined position. The location of the amber codon is important for the final position of the furan amino acid and these positions were selected based on the crystal structure. The nanobody variants were successfully expressed with a furan-modified lysine derivative incorporated on the amber position. Subsequent *in vitro* crosslinking experiments in presence of the sEGFR target protein demonstrated promising crosslinking results as derived from the SDS-PAGE analysis. This work forms a solid proof of concept for future *in vivo* experiments with these furan crosslinking nanobodies.

5.3 Furan incorporation into peptides via genetic code reprogramming and their one-pot pyrrole mediated cyclisation

The contents of this section were published in Klaas W. Decoene, Willem Vannecke, Toby Passioura, Hiroaki Suga and Annemieke Madder *Biomedicines* **2018**, 6, 99.

In this chapter, we aim to demonstrate that a furan amino acid can be incorporated via FIT and, following peptide translation, selective furan oxidation within the translated peptides can be achieved using N-bromosuccinimide (NBS). We aim to examine the reactivity of natural nucleophiles towards the oxidized keto-enal moiety via analysis of potential peptide macrocycles.

From a pharmaceutical perspective, peptides occupy a transition zone between small molecule chemicals and larger biologics.[245] Compounds in this intermediate size range can theoretically combine the advantages of biologics (e.g., the specificity of monoclonal antibodies and their ability to disrupt protein-protein interactions) with the desirable aspects of small molecule drugs (e.g., oral availability and biostability). In particular, relatively small peptides (up to ~15 amino acids) constrained through cyclizing linkages are capable of interfering with protein–protein interactions, exhibit high target specificity, and can be highly resistant to serum proteases, and yet are small enough to be orally available and avoid immune responses.[245][246] For these reasons, several constrained peptides are used as therapeutics (e.g., octreotide, cyclosporine A, nisin) and more are in late-stage clinical trials. Moreover, new approaches for constraining the inherently flexible conformation of short peptides are highly sought after and form a subject of continuous investigation. From an organic synthetic perspective, constraining peptide structure can be achieved in numerous ways.

Four conceptually different approaches can be adopted to obtain cyclic peptides, depending on the functional groups present in the side chains of the envisaged sequence: head-to-tail, head-to-side-chain, side-chain-to-tail, and side-chain-to-side-chain connection. In this way, not only head-to-tail cyclic peptide construction but also loop stabilization and stabilization of alpha helices through so-called peptide stapling can be achieved by creating covalent connections between two of the side chains of a specific peptide sequence.

Methods can also be conceptually subdivided into either two component systems, in which a bireactive external component is added to react with two functional groups in the peptide (commonly side-chain moieties), or one-component systems, in which two moieties in the peptide react to form a cyclizing link. When considering the specific nature of the side-chain functionalities of the relevant amino acids, further distinctions can be drawn based on whether two natural amino acid residues, one natural and one unnatural amino acid residue, or two unnatural amino acids are involved in the reaction.

A number of chemistries used in two-component systems are based on the use of two unnatural amino acids that react with double “click” linkers.[247] Alternatively, reactions using two canonical amino acids include double thioether[248] linkages via reaction with cysteine residues. Arylation[249] and alkylation[250,251] chemistry on both lysine and cysteine side chains have also been reported. In view of the higher complexity of such two-component systems and the possibility for formation of regio-isomeric products when applying unsymmetrical linker components, alternative, one-component systems have

been developed. One-component approaches where two natural amino acids are involved in the peptide cyclisation include lactam formation between lysine and Glu/Asp[252] and formation of a disulphide bridge between two Cys residues[253] in addition to enzymatic cyclisation strategies, for example, between Lys and Trp[254]. However, one-component systems employing two natural amino acids inherently lack selectivity due to the possible presence of other natural residues in the sequence of interest.

Therefore, chemistries using two unnatural amino acids have been developed, including ring-closing metathesis (RCM) hydrocarbon linkages,[255] copper-catalysed azide alkyne click (CuAAC)[256] reactions, alkene-tetrazole[257] cycloaddition, Glaser coupling,[258] and oxime formation.[259] It is, however, not always straightforward to incorporate two unnatural amino acids in a peptide sequence. Therefore, strategies using one unnatural and one naturally occurring amino acid functionality (thereby combining a degree of selectivity with ease of synthesis) have gained increasing attention, such as the formation of a thioether[260] between a specifically located Cys residue and a second residue featuring an alkylating moiety, or peptide macrocyclization via reaction of a noncanonical aldehyde moiety and the peptide N-terminus.[261] In addition to the chemical peptide synthesis approaches discussed above, biotechnological techniques for the generation of constrained peptides have recently been developed. In particular, genetic code reprogramming approaches, such as flexible in vitro translation (FIT)[185] (vide supra) allow for the incorporation of diverse unnatural amino acids into peptides synthesized ribosomally, thereby enabling diverse one-component cyclizing chemistries in translated peptides.[262–267]

In a FIT reaction, one or more unnatural amino acids are aminoacylated onto specific tRNAs, which are then introduced into a fully reconstituted in vitro translation reaction deficient for one or more natural amino acids. In this way, specific unnatural amino acids of interest can be translated in place of their natural analogues. The aminoacylation reaction itself is achieved through the use of flexizymes (flexible tRNA acylation ribozymes), which catalyse the aminoacylation of tRNA 3'-terminal hydroxyls using an amino acid substrate activated via an ester (or thioester) leaving group. This highly versatile approach has been used for the synthesis of peptides containing diverse unnatural amino acids including multiple cyclizing chemistries. In the context of previous work in our group on furan-based peptide labeling[133,134] and peptide/protein crosslinking,[136] we became interested in exploiting FIT for the exploration of the potential utility of furan side chains in a peptide-constraining context. To identify which of the canonical nucleophilic amino acids is prone to react with an oxidized furan moiety, we made use of FIT to produce a set of peptides encompassing both a furan amino acid and a nucleophilic amino acid that could potentially react to form a side-chain-to-side-chain linked structure (Figure 5.40).

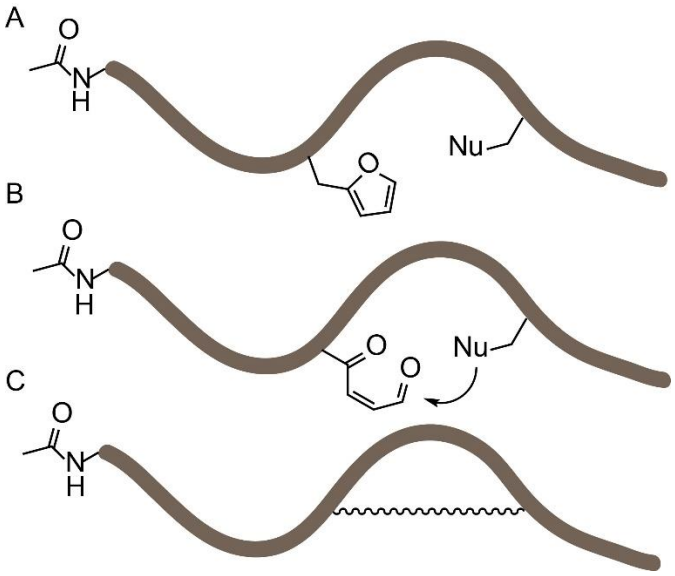


Figure 5.40 Schematic overview of the objective of the flexible in vitro translation screening experiment. A set of peptides was produced by flexible in vitro translation (FIT) (A) all bearing a furan amino acid and a nucleophilic residue. In a second stage, the furan moiety was oxidized (B) and nucleophiles can potentially react with the keto-enal moiety. The peptides were analysed via mass spectrometry to identify any formed constrained peptides (C).

5.3.1 FIT-Based Synthesis of Furan-Modified Peptides

Taking advantage of the utility of FIT for the facile synthesis of diverse peptides at small scale, a screening experiment was designed to identify amino acids able to react with an oxidized (activated) furan amino acid. Commercially available furylalanine (Fua), which is isosteric with His and isoelectronic with respect to Tyr, was selected as the furan-containing amino acid. For this screening experiment, simple template peptides were designed containing no other reactive functional groups aside from those functionalities to be examined for their capacity to intramolecularly react with the (oxidized) furan moiety. Since the N-terminus is also a potential nucleophile and we were interested in reactions of the oxidized furan moiety and the side chains of natural amino acids (peptide-constraining) rather than head-to-side-chain macrocyclization, each peptide was translated with an acetylated N-terminus through the use of N-acetyl Phe (AcF).

To favour eventual cyclisation events by enhancing the proximity between the two reactive groups, a Pro residue was positioned in between. The final template peptide sequence used was AcPheAlaGlyAlaFuaGlyProGly**X**AlaGlyAla with **X** being a natural residue featuring a nucleophilic side chain cysteine histidine, lysine, arginine, serine, or tyrosine (figure 5.41). FIT-mediated translation of peptides initiated with N-acetyl Phe has been described previously,[268] however, to the best of our knowledge, use of this approach for translation of Fua has not been reported. To achieve this, cyanomethylester (CME)-activated Fua was synthesized starting from boc-furylalanine, through esterification with chloro-acetonitrile, followed by boc deprotection. This activated Fua was then charged onto a tRNA for decoding of the Met AUG elongator (i.e., not initiator) codon in a flexizyme (in this case eFx)-catalysed reaction (Figure 5.41B). A similar process was used for N-acetyl Phe with the exception that a tRNA specific for peptide initiation was used.

Due to the fact that the acetyl-Phe residue is located at the N-terminus, both unnatural residues are incorporated in response to the same codon (AUG), a consequence of the difference between translation initiation and elongation. Addition of these two unnatural amino-acid-charged tRNAs to an in vitro translation reaction deficient for Met thus produced a genetic code in which N-acetyl Phe replaced the initiating Met, with Fua replacing any downstream Met residues and all other codons encoding their cognate natural amino acid

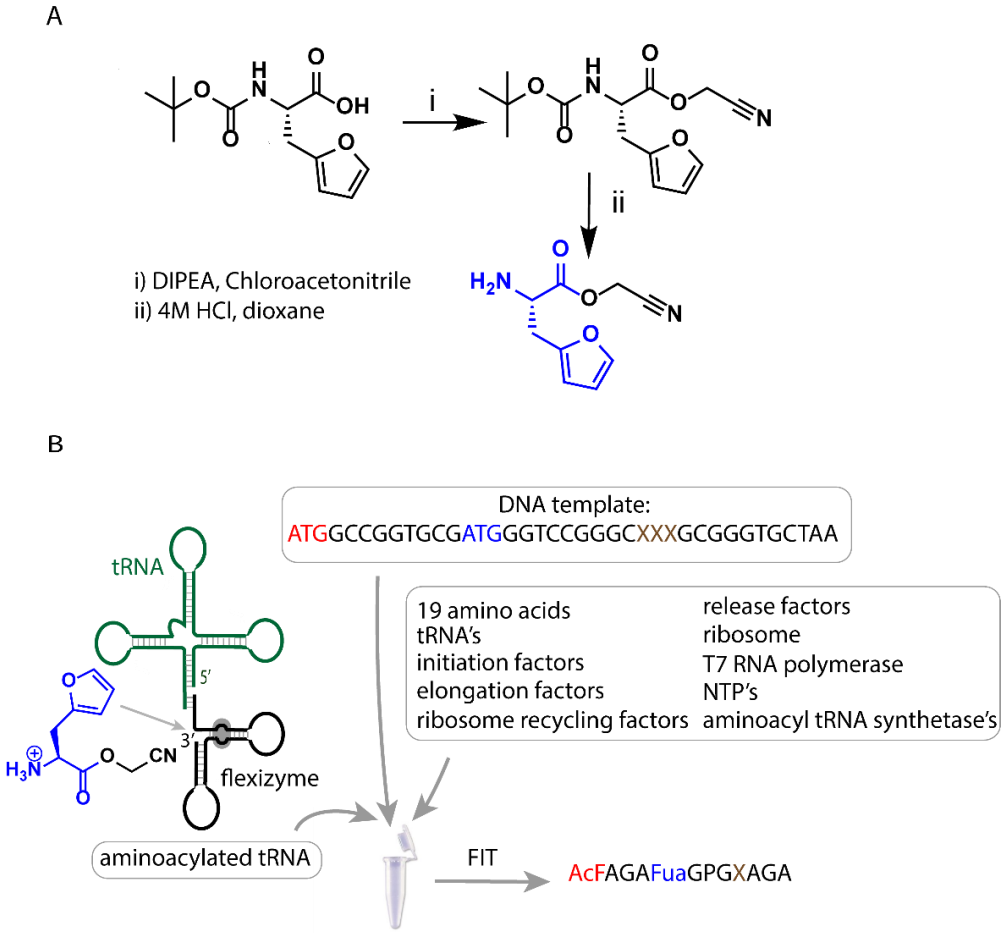


Figure 5.41 Schematic representation of the flexible in vitro translation system used for expression of furylalanine-containing peptides. On top (A), the structure and synthesis of the CME-activated furylalanine is presented. On the bottom (B), a simplified structure of the flexizyme eFx and a tRNA molecule (green) is shown. The translated peptides contained an N-acetylated Phe (red) at the N-terminus and a Fua (blue), as well as a range of nucleophilic residues (brown). For all amino acid residues, one-letter codes are used in this table except for furylalanine, abbreviated as Fua.

To confirm translation of the test templates under the genetic code used (both described above), small-scale translation reactions (2.5 µL each) were assessed by MALDI-TOF MS. Table 5.3 lists the calculated expected mass for each peptide, followed by the detected ions (original MALDI-TOF mass spectra are included in the experimental data section 9.2.3.4). These experiments demonstrate successful translation of all peptides, with the $M+K^+$ and $M+Na^+$ peaks detected in all cases, and the $M+H^+$ peaks also found for the more easily protonated species (i.e., those including

His, Lys, and Arg). The use of the FIT system allowed us to have rapid access to the peptide library we envisaged for this study on a small scale. Once the CME-activated amino acids are synthesized, the tRNA aminoacylation, peptide translation, oxidation, purification, and MALDI-TOF analysis can all be done in a single day.

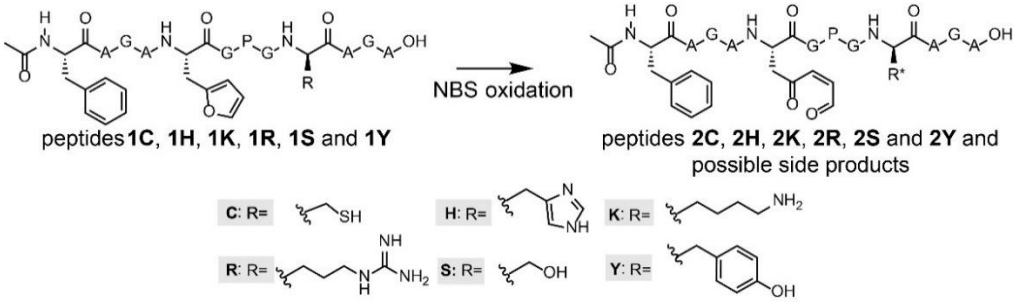
Table 5.3 MALDI-TOF analysis of the translated peptide library. For all amino acid residues, one-letter codes are used in this table except for furylalanine, which is abbreviated as Fua. Colour codes are in accord with those used in Figure 5.52. "/": not observed.

Peptide name and sequence		Calculated Exact mass (Da)	experimental $M+H^+$ (Da)	$M+Na^+$ (Da)	$M+K^+$ (Da)
1C	AcFAGAFuaGPGCAGA	1056.43	/	1079.58	1095.55
1H	AcFAGAFuaGPGHAGA	1090.48	1091.66	1113.67	1129.68
1K	AcFAGAFuaGPGKAGA	1081.52	1082.69	1104.69	1120.66
1R	AcFAGAFuaGPGRAGA	1109.53	1110.75	1132.71	1148.70
1S	AcFAGAFuaGPGSAGA	1040.46	/	1063.65	1079.63
1Y	AcFAGAFuaGPGYAGA	1116.49	/	1139.70	1155.69

5.3.2 NBS Oxidation of Furan-Containing Peptides Obtained through FIT

One of the advantages of the furan methodology relies in the fact that furan as such is not reactive towards nucleophiles present in a biological system, but reactivity can be unveiled by oxidation. Therefore, the translated, furan-containing peptides undergo an oxidation step prior to screening for cyclisation reactions. In this work, N-bromosuccinimide (NBS) was used to selectively oxidize the furan moiety in the template peptides, with oxidation and/or cyclisation being monitored by MALDI-TOF. Since the furan moiety as well as the nucleophilic residue are present in a single peptide, we carefully analysed the spectra for the presence of products resulting from potential intramolecular reactions. In Table 5.4, the results of the MALDI-TOF analysis are shown (original mass spectra are added in the experimental data section 9.2.3.5).

Table 5.4 MALDI-TOF analysis of the oxidized template peptides after C-tip reverse phase purification. Peptides 2 and their observed mass values are depicted in bold. "/": not observed.



Peptides 1 **Peptides 2**	$M+H^+$	$M+Na^+$	$M+K^+$	Other
1C **2C**	/	/	/	1214.97 (cysteinylation) **1230.97** (cysteinylation +16)
1H **2H**	/	/	1130.18 **1146.21**	/
1K **2K**	1082.90	1105.13	1121.01 **1137.10**	**1062.13 (M+H-20 Da)**
1R **2R**	1111.13 **1127.17**	/	1149.17	/
1S **2S**	/	1064.06	1080.01 **1096.22**	**1021.43 (M+H-20 Da)**
1Y **2Y**	/	/	/	1313.91 (dibromination) **1329.87** (dibromination +16)

For the Cys-containing template peptide **1C**, the original protonated mass was not observed (nor were sodium or potassium adducts), but a new cysteinylation product (1214.97 Da) was detected in addition to its furan oxidation product +16 Da (1230.97 Da). While the Cys thiol moiety is the most nucleophilic functional group tested, these results suggest that under the oxidizing conditions used for formation of the keto-enal group, disulphide bond formation between the peptide Cys and free Cys in solution prevents reaction with the oxidized furan. In the case of the His-containing template peptide **1H**, next to the potassium adduct of **1H**, also the furan oxidation product **2H** was identified, but no product resulting from a potential cyclization was observed. The mass spectrum for the Arg-containing template peptide **1R** showed the $M+H^+$ peak as well as the potassium adduct ($M+K^+$), along with a product resulting from oxidation **2R**.

For the Tyr-containing peptide **1Y**, a mass corresponding to dibromination of Tyr in the starting peptide as well as the furan oxidation product thereof were observed, not unexpected due to the use of a huge excess of NBS. For the Lys- as well as Ser-containing template peptides **1K** and **1S**, we observed the sodium and potassium adducts of the starting peptides as well as the potassium adduct of the furan oxidation product. Additionally, in both cases, a particular signal corresponding to M+H−20 Da was detected (highlighted in grey in Table 5.4). This observation of a reduced mass could indicate, for example, an oxidation (resulting in a peptide mass M+16), followed by an intramolecular cyclization reaction accompanied by the loss of water (resulting

in a peptide mass M+16−18), and an additional loss of water during ionisation (M+16−18−18 = M−20).

5.3.3 Peptide Scale up by SPPS and Subsequent NBS Oxidation

Based on the results of the FIT screening, the Lys- and Ser-containing template peptides **1K** and **1S** were resynthesized via solid-phase peptide synthesis (SPPS) in order to generate sufficient material for more detailed characterisation of the reaction following oxidation. In addition, a control peptide with a Gly residue rather than a nucleophilic residue was synthesized. Similar to the FIT screening, the peptides were treated with NBS to oxidize the furan moiety. To render the oxidized furan moiety more electrophilic and to promote imine formation (in the case of peptide 2K), this reaction was performed in the presence of NaOAc buffer of pH 5.2 as per Malins *et al.*.[261] In Figure 5.42, the structure of the three resulting peptides is shown (top) with the LC chromatogram (red) as well as the MS spectrum corresponding to the oxidation peak (bottom) for each of the peptides.

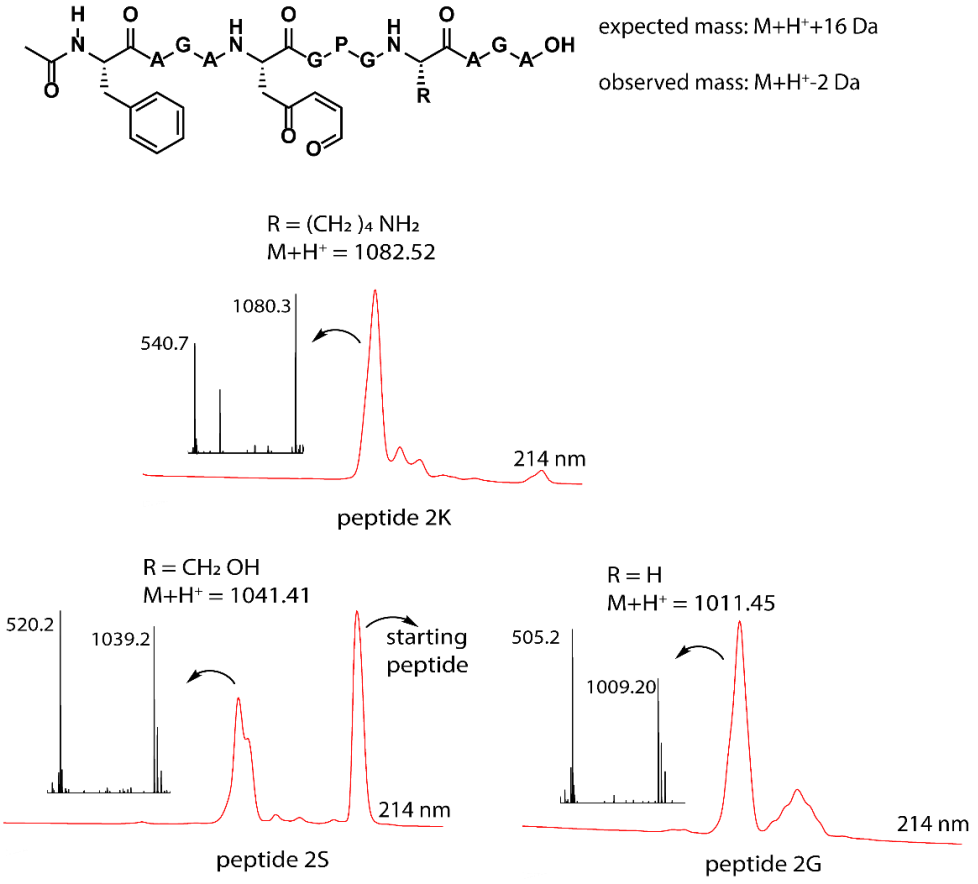


Figure 5.42 Top: general structure of the oxidized peptides 2, the expected mass (in case of oxidation only) and the observed mass. M+H$^+$ mass for the individual peptides (2K, 2S and 2G), LC chromatogram at 214 nm (zoom of the relevant region) red and the MS spectrum of the oxidation peak for the three oxidized peptides.

From the MS spectrum corresponding to the LC signal of the product formed upon oxidation, each time a mass of −2 Da compared to the M+H$^+$ mass can be observed for the three peptides. This is 18 Da lower than the expected +16 Da mass for the oxidation product (Figure 5.42), which can be explained by oxidation of the furan

moiety followed by loss of water (+16 Da−18 Da = −2 Da). The obtained mass data should, however, be interpreted with care. Indeed, the observed loss of water can be the result of an intramolecular reaction, thus indicating formation of a new species resulting from a cyclisation event, but can also occur during the ionisation (the signal thus corresponding to the oxidized product without further cyclisation).

To determine whether the $M+H^+$ − 2 Da peak and the $M+H^+$+16 Da peak originated from one and the same compound in the LC, the chromatograms for both extracted ions (−2 Da and +16 Da) were compared. The results indicated that for peptides **2G**, both ions originated from the same (oxidized but not cyclized) compound in the LC, indicating that the $M+H^+$−2 Da peaks resulted from the ionisation process and were artefacts of mass spectrometry. However, for peptide **2K**, it was clear that the $M+H^+$−2 Da and $M+H^+$+16 Da ions originated from different products, implying that the $M+H^+$−2 Da peak may be due to imine formation (see figure 5.43).

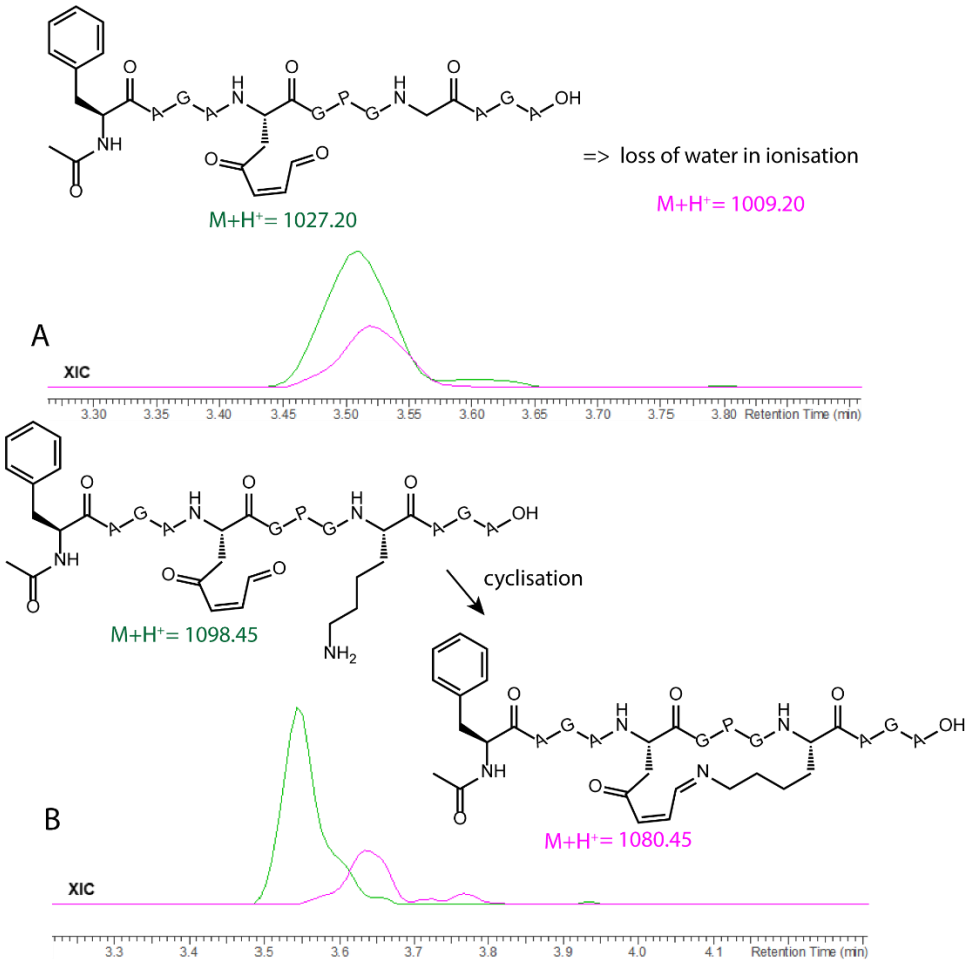


Figure 5.43 A: Zoom of the extracted ion chromatogram for two selected ions for an oxidation reaction with petpide **2G**; (green = $M+H^+$-2 (1009.20); purple = $M+H^+$+16 (1027.20)). The ion with a mass of 1027.20 Da is the oxidized petpide and the 1009.20 Da ion is 2 Da less compared to the $M+H^+$, the presence of this last ion can be attributed to loss of water in the ionisation process, evidenced by the co-elution of the two selected ions. B: Zoom of the extracted ion chromatogram for two selected ions for an oxidation reaction with petpide **2K**; (green = $M+H^+$-2 (1098.45); purple = $M+H^+$+16 (1080.45)). The ion with a mass of 1098.45 Da is the oxidized petpide and the 1080.45 Da ion is 2 Da less compared to the $M+H^+$, the presence of this last

ion can not be explained by loss of water in the ionisation process since both ions originate from products that elute at a different time. The mass reduction of 2 Da can be explained by intramolecular imine formation.

5.3.4 Identification of a Pyrrole Moiety as Cyclisation Motif

To quench the rather labile imine-constrained peptide (Figure 5.44, peptide **2Ka**), we decided to try and reduce it to the corresponding secondary amine with sodium cyanoborohydride. This reduction would theoretically result in a product with the same mass as the original peptide. However, MS analysis of the sample, in which NBS oxidation was followed by $NaCNBH_3$ reduction, revealed the presence of a compound with a mass corresponding to $M+H^+-18$ Da (Figure 5.44). This mass can be explained by an additional subsequent attack of the formed amine (upon reduction) on the remaining keto functionality followed by aromatization through loss of water.

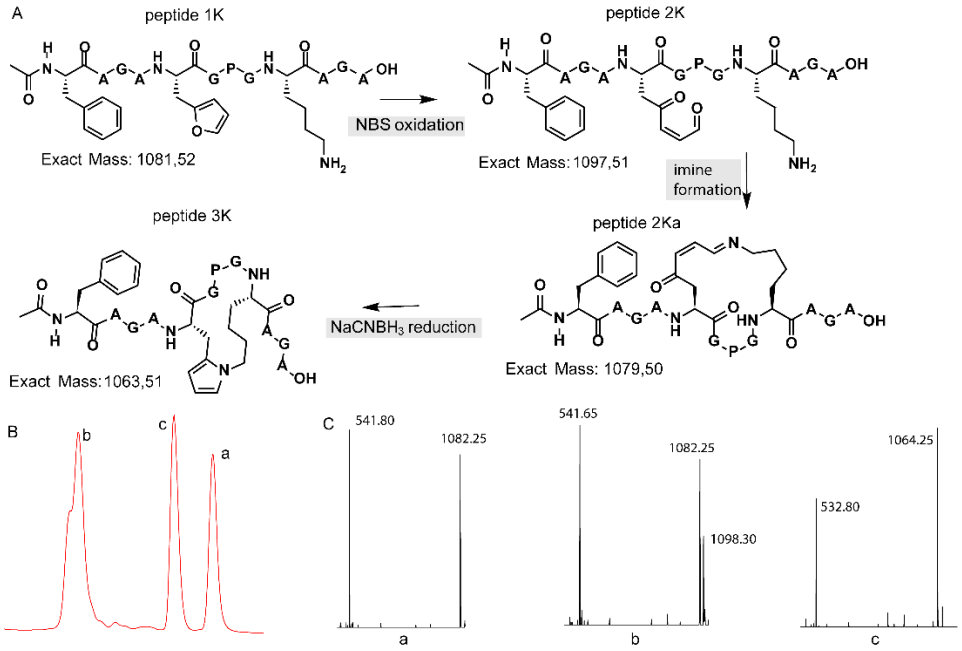


Figure 5.44 (A) Schematic overview with structural representation of the one-pot peptide oxidation and reduction reaction. (B) A zoom-in of the LC chromatogram (214 nm) of the reaction mixture is shown. (C) A zoom-in of the ESI-MS spectra (positive mode) of the three prominent peaks. a: starting peptide, b: oxidation products, c: pyrrole-constrained peptide.

In the chromatogram shown above (Figure 5.44, B), three major peaks can be observed. The peak at the far right (a) corresponds to the starting peptide which was not oxidized. Signal b in part B of Figure 5.44 can be correlated to oxidation products. The MS spectrum of peak b indicates two products, the expected oxidized peptide with a mass of 1098.30 Da and a product with the mass of the starting peptide. This last product can be explained by the fact that oxidation took place and an imine was formed (with loss of water) which was further reduced to a secondary amine upon $NaCNBH_3$ treatment without further reaction to the pyrrole-constrained peptide. Finally, the MS spectrum of peak c shows a mass of 1064.25 Da, which can be correlated with the formation of a pyrrole-constrained peptide. We decided to further explore the scope of this pyrrole-based cyclisation while at the same time generating additional structural evidence for formation of the pyrrole moiety as cyclisation element.

5.3.5 Extending the Scope of Pyrrole-Mediated Cyclisation: Varying the Positioning and Conformation within the Template Peptide

The original template peptide used in this work featured a Pro positioned in between the Fua and the Lys, resulting in a form of preorganisation of the peptide conformation inducing proximity between the reactive moieties. To demonstrate that the pyrrole-constrained peptide formation is not limited to the specific peptide sequence chosen here, two additional variants were synthesized via SPPS. In one of these (peptide **4K**), the distance between the two reacting residues was increased by interchanging both Fua and Lys with a neighbouring Ala. On the other hand, in peptide **5K**, the internal Pro residue was replaced by an Ala (figure 5.45).

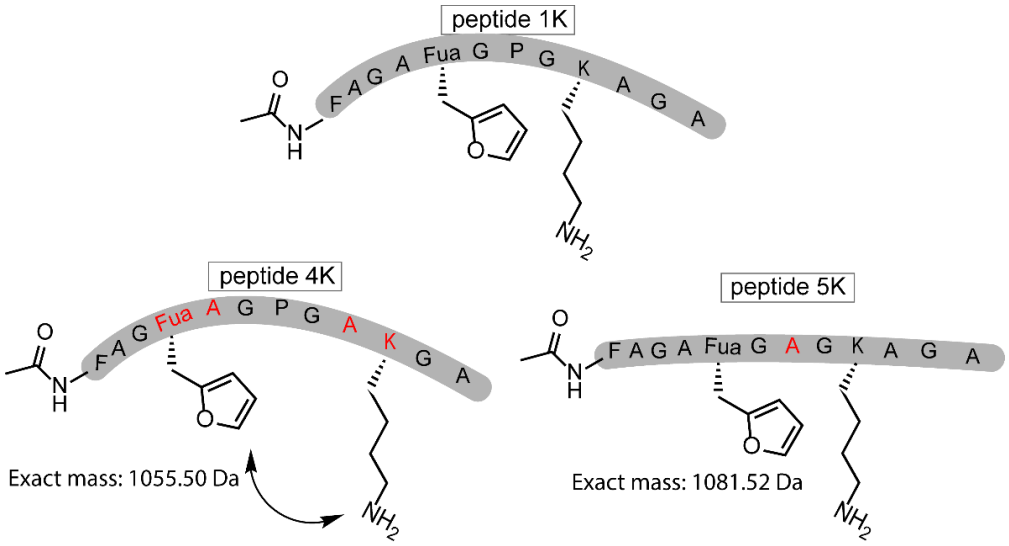


Figure 5.45 Schematic representation of the original template peptide (peptide **1K**) and two variations on the template peptide which were synthesized by SPPS. In peptide **4K**, the distance between the two reacting residues was increased, and in peptide **5K**, the Pro was exchanged for an Ala residue. The altered residues (Fua, A, K) in peptides **4K** and **5K** compared to peptide **1K** are depicted in red.

The pyrrole-mediated, peptide-constraining reaction proved successful in the case of both peptide **4K** and **5K**, and LC-MS data for the oxidation and reduction reactions towards the cyclic peptides **6K** and **7K** are included in the experimental figures 9.71 and 9.73. These one-pot oxidation and reduction reactions were also repeated at larger scale for peptide **4K** to allow structural characterisation of the cyclised product. The optimized reaction for the formation of **6K** was used to determine the conversion via HPLC peak integration. It was shown that near-complete conversion of the starting peptide to the oxidized products was achieved. However, the oxidized products did not convert completely to **6K**. The total conversion of **4K** to **6K** was 32% (figure 5.46).

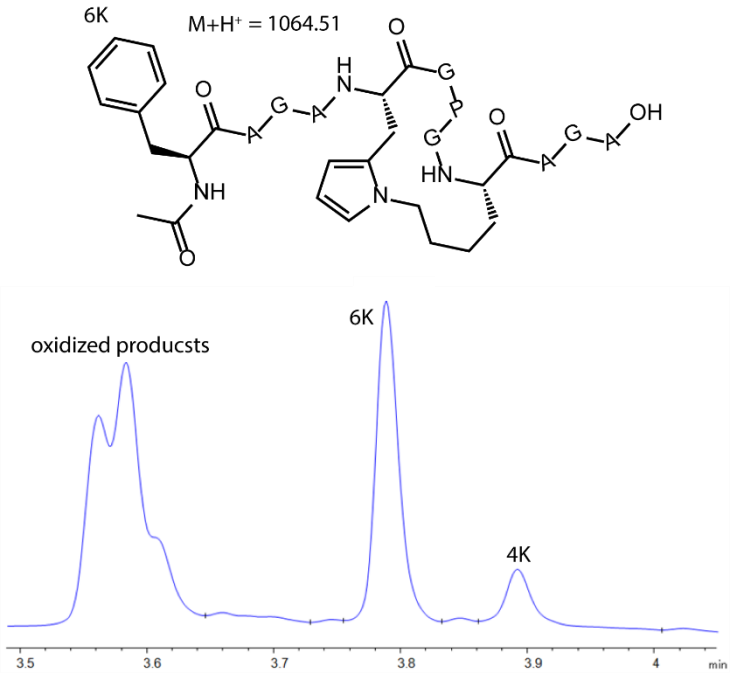


Figure 5.46 Top: Structure of cyclized peptide **6K.** Bottom: Zoom of the LC chromatogram at 214 nm. Starting peptide **4K,** oxidized products and **6K** are indicated.

The pyrrole compound (peptide **6K**) was purified by RP-HPLC and after lyophilization, the product was resolubilized in 90/10 H2O/D2O for NMR analysis on a 700 MHz (Bruker) spectrometer, confirming the structure of the pyrrole cyclized product. A zoom of the 1HNMR overlay spectrum of **4K** and **6K** is provided in Figure 5.47. The three expected signals for the pyrrole unit are present, a triplet for the proton on C4 and a hidden triplet for C3 and C5. For the furan unit in starting material **4K**, the signals of the protons on C3 and C4 are present; the signal for the proton on C5* is not visible in this zoom since its chemical shift is larger than 7 ppm. Complete NMR spectra (1H, TOCSY, ROESY, HSQC) with assignment of the signals for peptide 6K and full 1H spectrum for **4K** are included in the experimental data (figures 9.75, 9.76 and 9.77).

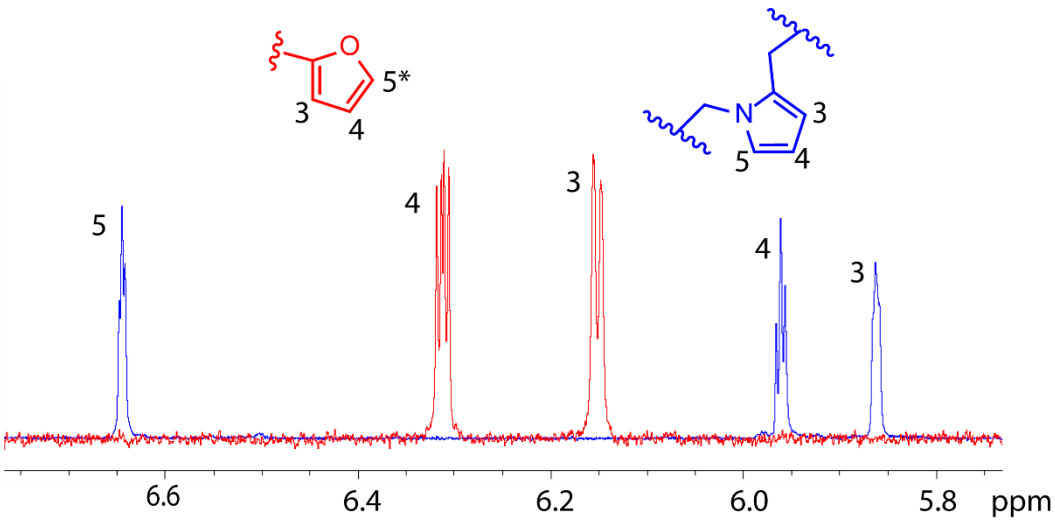


Figure 5.47 Zoom of the 1H NMR spectrum of peptides **4K** (red) and **6K** (blue) between 5.8 and 6.8 ppm combined with the structures of the furan and pyrrole unit (right). The signals corresponding to the protons on C3, C4, and C5 are indicated except for the signal on C5* of the furan moiety which has a chemical shift larger than 7 ppm and is not shown in this figure.

5.3.6 Conclusions

In summary, a new pyrrole-mediated peptide cyclisation methodology was developed using FIT as an initial screening strategy. We here have shown the successful translation of template peptides containing both a furylalanine and a nucleophilic side chain residue using FIT. Oxidation of the furan moiety was achieved with NBS and MALDI-TOF analysis was used to screen for cyclized peptides. SPPS synthesis was used to generate more peptide material for LC-MS and structural analysis. NMR spectroscopy allowed confirming the formation of the pyrrole moiety as cyclisation motif. Further variation of the peptide structure indicates the generality of this novel methodology which thus promises to be useful in a broader context for the efficient cyclisation of peptides containing a furylalanine moiety and a suitably positioned lysine residue.

6. Development of further applications of the triazolinedione (TAD) chemistry platform

Both the synthesis of the urazole precursor of TAD-propanol as well as the development of the urazole oxidation towards TAD using TCCA were performed by Kamil Unal in the lab of Prof. J. Winne.

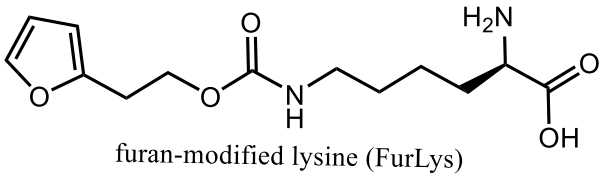
Prof. J. Winne performed the NMR analysis in section 6.3.

ESI-CID/HCD/ETD mass analysis of peptides as well as intact protein analysis in section 6.3 were performed by An Staes in the lab of Prof. K. Gevaert.

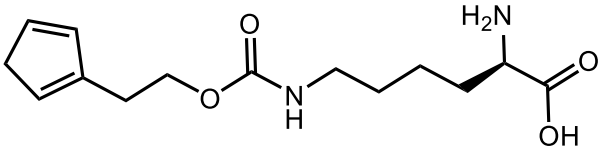
In this chapter various approaches are explored to either equip proteins with a non-canonical TAD-reactive moiety or to target proteins that contain naturally TAD-reactive moieties. The cyclopentadiene modified lysine amino acid CpLys belongs in the first category. Cyclopentadiene is known to be one of the best substrates for reaction with TAD,[269] this implies that a cyclopentadiene containing ncAA could be a very interesting option for protein labelling. On the other hand there are already TAD-reactive moieties present in natural proteins. Prenyl groups, which are post-translationally installed on specific cysteine residues, contain multiple tri-substituted double bonds. As potent (di)enophiles, TAD reagents are known to react with these double bonds. However, besides the post translationally installed prenyl moieties, also tyrosine residues can readily react with TAD in buffered conditions, which led to the development of the so-called tyrosine click chemistry.[101]

6.1 Cyclopentadiene amino acid

In 2019 a new cyclopentadiene (Cp) containing ncAA was reported for incorporation in proteins using genetic code expansion.[270] The Cp lysine can be used for Diels-Alder labelling reactions as well as proximity dimerization between two entities both containing the Cp lysine. The authors expressed a human IgG1 antibody with Cp lysine on one of two positions. They used functionalized maleimides for modification of the Cp lysine if the ncAA was located on a position pointing away from each other in the dimeric antibody structure. On the other hand if the Cp lysine were incorporated on a different position pointing toward each other in the dimeric structure spontaneous dimerization took place. A key concern related to the highly reactive Cp moiety in the ncAA, that was addressed by the authors is the stability of the Cp lysine during the ncAA synthesis and protein expression steps. Structurally the Cp lysine is very similar to the furan-modified lysine derivative furLys that was earlier used in this work (*vide supra*).



furan-modified lysine (FurLys)



cyclopentadiene-modified lysine (CpLys)

Figure 6.1 Chemical structures for the furan-modified lysine (top) and cyclopentadiene (Cp)-modified lysine (bottom) derivatives.

Due to the structural similarity of both ncAA's we came up with the idea to synthesize the CpLys and use it for incorporation into a test protein using the same PylRS mutant we used for the incorporation of the FurLys ncAA. In a second step, we wanted to explore the possibility of using the CpLys containing protein for labelling reactions with functionalized TAD reagents.

6.1.1 Synthesis of the cyclopentadiene-modified lysine derivative CpLys

The CpLys synthesis reported in St. Amant *et al.*[270] was used as depicted in figure 6.2, the first step involves the formation of a cyclopentadiene alcohol in two steps from sodium cyclopentadienide. The alcohol is reacted with 4-nitrophenyl chloroformate followed by generation of the carbamate by reaction with N_α(Fmoc)-lysine in DMF. Remarkably the N_α(Fmoc)-lysine was not soluble in DMF in contrast to what was reported by the authors. Several other solvents and solvent combinations were tried and in the end, a mixture of MeOH and chloroform seemed to work. Finally, the N_α is Fmoc deprotected using piperidine, this affords the CpLys derivative which was stored at -80°C to avoid dimerization. The final product was analysed via LC-MS and showed a major product peak for the expected CpLys product, a small amount of dimer and an unidentified impurity. In a small scale synthesis the final sample was briefly heated up at the rotavapor to remove residual DMF and this caused the formation of large dimer product peaks. At a larger synthesis scale, this heating was avoided resulting in only minor dimer product peaks in the final product (experimental data section 9.2.1).

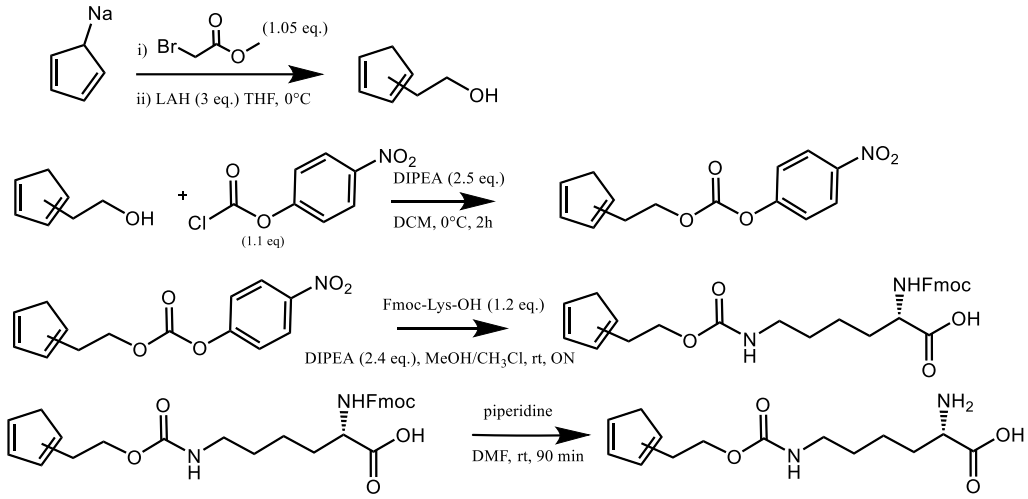


Figure 6.2 Synthesis scheme for cyclopentadiene (Cp)-modified lysine.

Although the synthesis was completed successfully, due lack of time this CpLys could not be used anymore for nanobody incorporation experiments.

6.2 Targeting the native prenylgroup

A prenylgroup consists of 5 carbons with a tri-substituted double bond. TAD reagents are potent (di)enophiles and can react with prenyl moieties via an Alder-ene mechanism (figure 6.3). Naturally specific cysteine residues in certain proteins can be prenylated in a post translational process with three or four of these prenyl groups, respectively termed farnesylation or geranylgeranylation. Besides the natural protein prenylation also chemical prenylation and *in vitro* enzymatic prenylation of peptides and proteins have been developed. Using the substrate promiscuity of the enzymes involved in prenylation also several prenylgroup mimics, called isoprenoids have been introduced into proteins *in vitro* and *in vivo*.

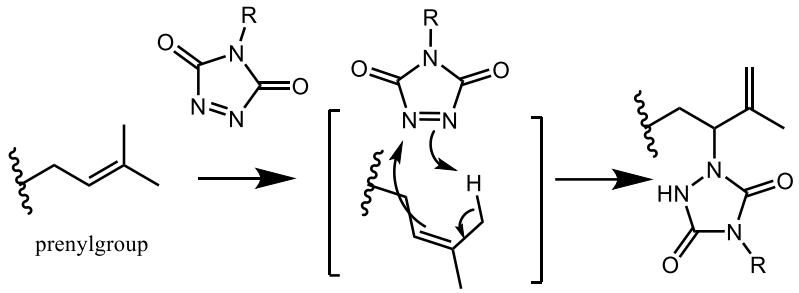


Figure 6.3 Schematic representation of a single prenyl moiety and the Alder-ene reaction mechanism to form the conjugated product with a TAD reagent.

6.2.1 The prenylation post translational modification

Prenylation can be subdivided in farnesylation or geranylgeranylation corresponding to the covalent addition of a 15-carbon or 20-carbon polyisoprene unit respectively. This post translational modification (PTM) occurs at a specific conserved cysteine residue near the C-terminus of proteins (*vide supra*).[26,27] The attachment of isoprenoids to a protein promotes membrane association, this effect can be attributed to hydrophobic interactions between the lipids in the membrane and the prenyl groups. The first report documenting the discovery of a (poly)peptide modified with a prenyl group appeared in 1978 where the structure of rhodoturucine A including the novel amino acid Cys (S-farnesyl) was published.[271]

The discovery of prenylation of mammalian proteins took another decade until research into the cholesterol biosynthesis pathway revealed that the envelope-protein lamin B was farnesylated.[272] Almost simultaneously it was demonstrated that Ras proteins are farnesylated at their C-terminal Cysteine.[273] Ras proteins are guanine triphosphate (GTP) binding proteins involved in cellular signal transduction. Ras proteins are of high significance in cancer research as mutations in Ras genes are one of the most common alterations in cancer.[274][275] It was found that the farnesylation of Ras proteins is a critical step in Ras maturation which is required for oncogenic forms of Ras to transform cells.

This discovery fuelled the interest in the protein farnesylation process and the development of inhibitors of farnesyltransferase (FTase) which is the enzyme responsible for this post translational modification using farnesyl pyrophosphate (FPP) as a substrate.[30][28][29][274] Since there was interest in inhibitors of FTase, scientists also became interested in the analysis of the prenylated proteins or the so-called prenylome. A common C-terminal CaaX motif (see figure 6.4, C: cysteine, a: aliphatic amino acid, X: determines if the protein will be farnesylated or geranylgeranylated)

was identified in the prenylated proteins from lower eukaryotes to mammalian systems. Geranylgeranyltransferase 1 and 2 (GGTase 1 and 2) are involved in geranylgeranylation of proteins. GGTase 1 is responsible for geranylgeranylation of the CaaX motif. GGTase 2 on the other hand is necessary for double geranylgeranylation of a CXC or CCXX motif also located near the protein C-terminus.[276] After the attachment of a farnesylgroup to the C-terminus CaaX motif, a farnesylated protein is the further processed. The first step is the removal of the aaX C-terminal tripeptide by an endoprotease. Finally, the new C-terminal carboxylic acid is converted into a methylester by isoprenylcysteine carboxylmethyltransferase. These additional maturation steps add to the protein hydrophobicity to promote membrane association.[276] Nevertheless, it has been reported that often prenylation alone is not sufficient for stable membrane association. In some instances other lipidations are occurring in combination with prenylation to promote membrane association.[277]

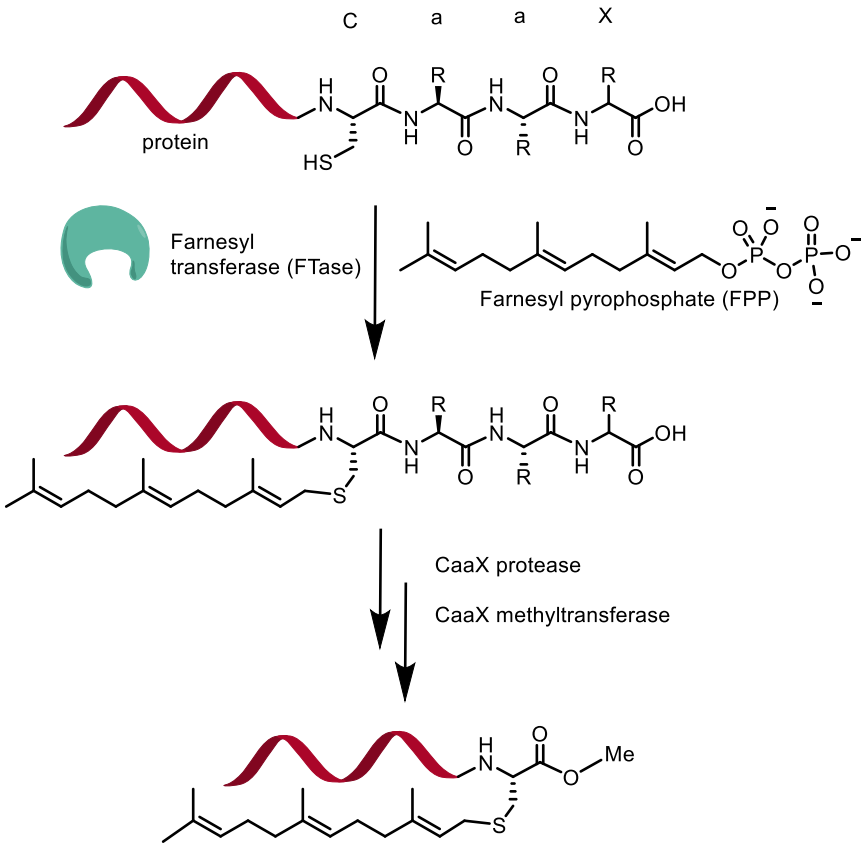


Figure 6.4 Schematic representation of protein farnesylation and downstream processing towards a fully mature farnesylated protein. In a first step mediated by FTase the cysteine residue in a C-terminal CaaX motif is farnesylated using FPP as a substrate. Subsequently the C-terminal aaX tripeptide is removed by an endoprotease. Finally the carboxylic acid is transformed into a methylester by a methyltransferase enzyme.

6.2.2 Detection of farnesylated proteins

Initially, the detection of prenylated proteins was accomplished by the metabolic incorporation of radiolabelled isoprenoid precursors (figure 6.5). In fact there was already compelling evidence reported in 1984 for a PTM involving a product of mevalonic acid in mammalian cells.[278] This study even points to an initial conversion

of mevalonic acid to an isoprenoid compound several years before mammalian prenylation was reported for the first time.

Later, radiolabelled mevalonic acid was used for detection of farnesylated proteins.[279] Additionally, radiolabelled substrates for farnesyl- or geranylgeranyltransferase [^{3}H]FPP or [^{3}H]GGPP were used in mammalian cells for prenylation detection.[280] The main drawback of these methods resides in the long exposure times required for detection which are typically in the range of 1-3 months. Moreover, these methods based on radioactivity offer no means to enrich and identify these proteins.[281] In attempts to overcome the disadvantages of the radiolabelled analogues, chemical probes were developed. These probes are isoprenoid analogues containing a fluorescent/biotin[282][283][284] moiety or a bio-orthogonal group. After addition of these isoprenoid variants to cultured cells, they become metabolically incorporated into proteins (figure 6.6).

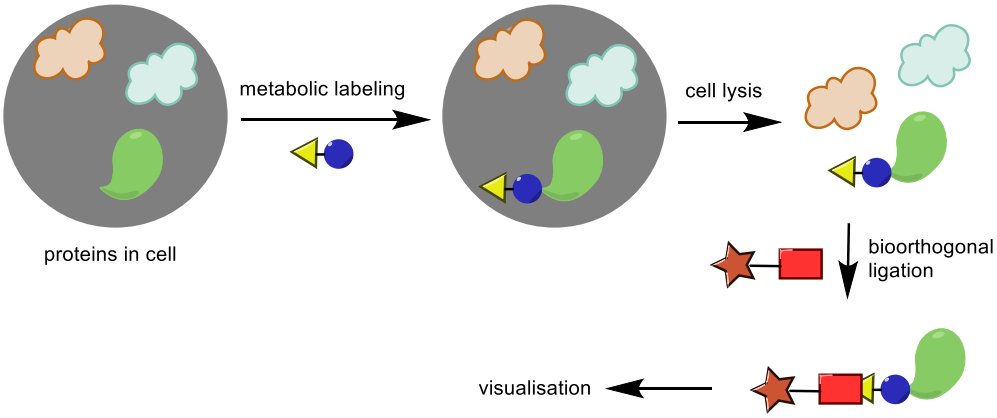


Figure 6.5 Simplified schematic representation of the metabolic incorporation of an isoprenoid structure. Following incorporation in proteins the cells are lysed and in a follow up bio-orthogonal reaction an affinity handle (e.g. biotin) or fluorophore (star) is linked to the prenylated proteins allowing for affinity purification or detection respectively.

In a subsequent step these proteins can be selectively conjugated via Staudinger ligation[31] or (CuAAC) click chemistry[285][286][287][288][289][290][291] and tetrazine ligation.[292] Other chemical handles such as aldehydes for oxime ligations[35] as well as photoactivatable benzophenone,[293] diazotrifluoropropanoyl[294] and diazirine[295] analogues can be introduced into proteins via metabolic incorporation of isoprenoid analogues (figure 6.6). Conjugation reactions are often carried out with reagents that have a biotin moiety for affinity purification, or fluorophores for visualisation purposes.[296] An inherent disadvantage of the metabolic incorporation of functionalized prenyl groups and subsequent orthogonal labeling is that this system requires these modified analogues to be prepared and added to the growth medium of the cells.

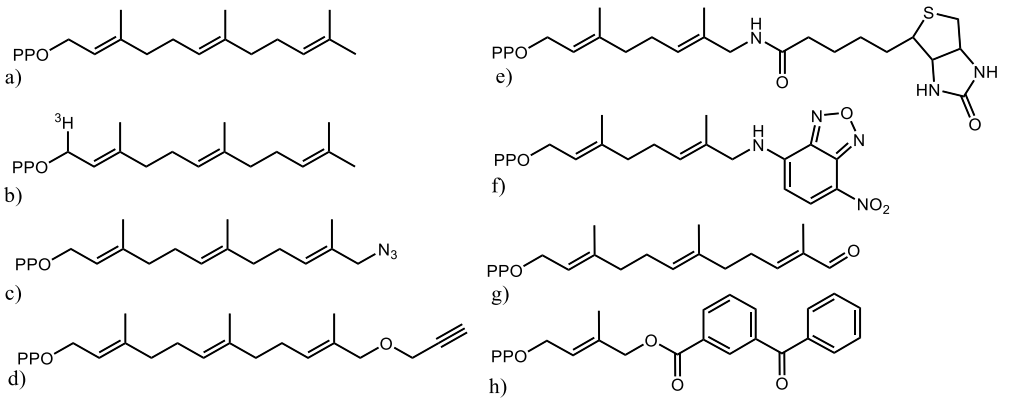


Figure 6.6 Structural representation of a selection of isoprenoids used for metabolic labelling for radiolabelling (b) and bio-orthogonal labelling purposes (b-d, g), affinity purification (e), fluorescence imaging (f) and photo-activatable crosslinking (h).

An inherent disadvantage of the metabolic incorporation of functionalized prenyl groups and subsequent orthogonal labeling is that this system requires these modified analogues to be prepared and added to the growth medium of the cells. In this respect the ability to target the native prenylgroup would offer a substantial advantage. Typically, unsaturated hydrocarbons are quite unreactive and therefore developing selective chemistry is challenging. TAD reagents (*vide supra*) are known to react with double bonds and we decided to take on the challenge and develop a TAD based method for the selective modification of the native prenylgroup.

6.2.3 Farnesyl selective TAD modification

6.2.3.1 Introduction

The first hurdle in the use of TAD reagents for the selective detection of native prenyl groups in proteins is that triazolinedione (TAD) reagents were already reported for protein conjugation. In 2010, Barbas and co-workers reported a click like reaction for tyrosine residues using TAD reagents in aqueous buffers.[199][102] The authors describe an elegant methodology for selective tyrosine modification with multiple TAD reagents on several peptides and proteins. A key objective to use TAD's for selective farnesyl modification in proteins (in presence of all canonical amino acids including tyrosine) will be to find conditions to selectively target the farnesylgroup.

Barbas and co-workers performed a pH study on the TAD conjugation of bovine serum albumin (BSA), which is a well-known model protein of 66 kDa in size. In this pH study, a range of conditions from pH 2 to pH 10 were tested for the conjugation reaction, and conjugation was shown to be efficient for the entire tested pH range. Several years ago, the TAD tyrosine bioconjugation reaction was investigated in mechanistic detail. Engels and co-workers demonstrated via computational studies on cresol and different TAD's, that the deprotonated tyrosine side chain -the phenolate rather than the phenol- is the reactive species in the reaction with TAD's (figure 6.7).[297] Via computational analysis the authors were able to demonstrate that the reaction of phenolate with TAD in buffered aqueous media proceeds according to an electrophilic aromatic substitution (S_EAr) mechanism. This is in

contrast to the initial report of the tyrosine bioconjugation reaction were it was denoted as an aqueous ene-type reaction.

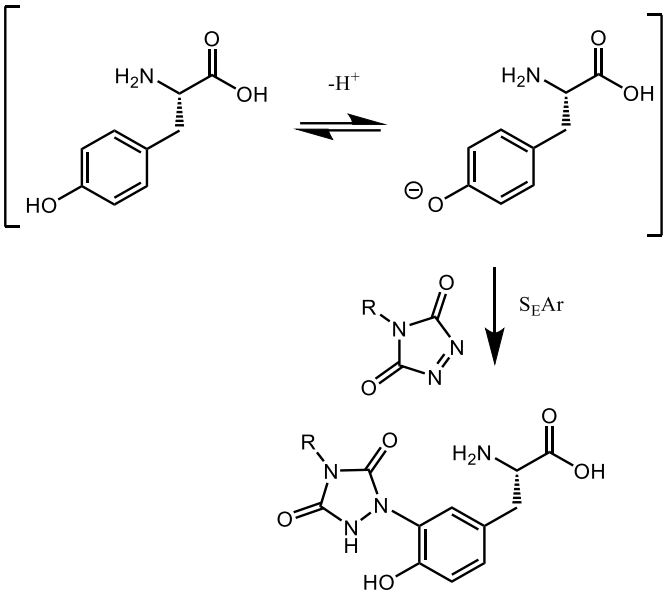


Figure 6.7 Schematic representation of the S_EAr reaction of a substituted TAD reagent with the deprotonated tyrosine side chain.

These computational studies effortlessly explain the excellent experimental results obtained with the TAD tyrosine click reaction in buffered aqueous media at pH 7-10. They also explain why the reaction does not proceed well in pure water or mixed organic solvents and water. However, there is an obvious issue with the reported protein conjugations at lower pH, as Barbas and co-workers report protein conjugation conversions over 50 % in a KCl/HCl buffer pH 2. At pH 2, the concentration of phenolate is 100.000 times lower compared to the value at pH 7.

Since hydrolysis of TAD is known to compete with the conjugation reaction on tyrosine, the reaction speed is of crucial importance for the conjugation conversion. The combination of these findings indicate that something else is happening at low pH that is not taken into account in the current reports on the TAD tyrosine conjugation reaction. During the research carried out in this chapter, we found out that tryptophan is the culprit of conjugation at lower pH and that even at pH 7 there is also competitive tryptophan modification. These findings are further discussed in detail in section 6.3 of this thesis.

6.2.3.2 Peptide experiments

The approach of performing initial peptide experiments and analysis, which are then followed by testing on proteins is a workflow that was used throughout this PhD work and which is also applied in this section on prenylation detection. The first peptides that were synthesised via SPPS were GGGAAY,**6.2a** and GGGCAAS, **6.2b1** (figure 6.8). These peptide sequences were selected since they possess an N-terminal GGG sequence which makes a later stage attachment to a protein via sortagging possible (as described in section 2.1.1).[14] On the other hand, the four C-terminal amino acids (CAAS) correspond to the C-terminal CaaX motif in naturally farnesylated proteins.

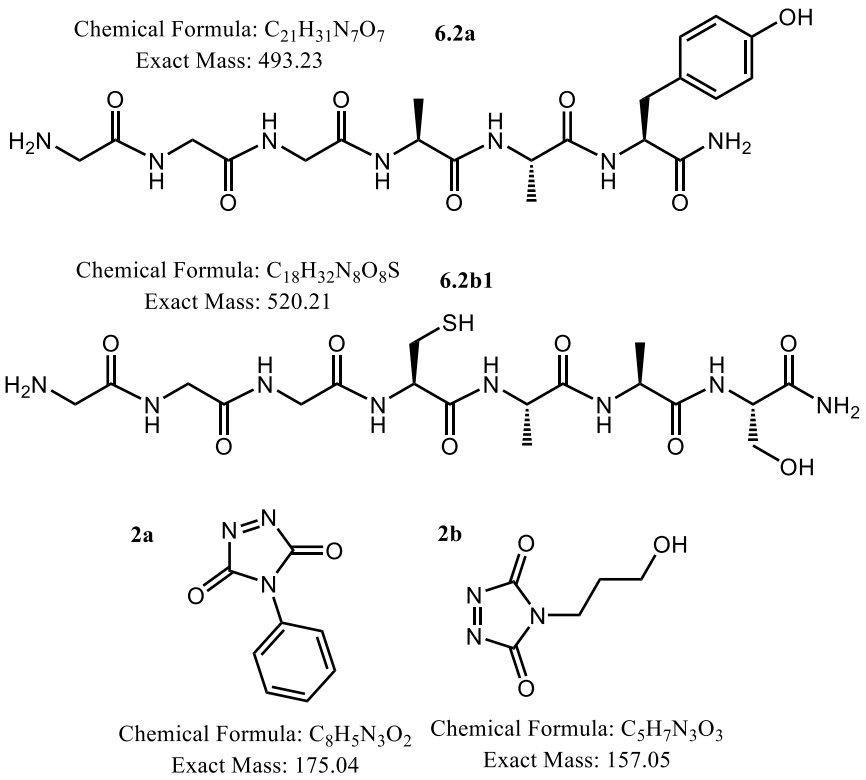


Figure 6.8 Structural representation of peptides GGGAAY,**6.2a** and GGGCAAS, **6.2b1**. Structures of Phenyl-TAD (PTAD), **2a** and TAD-propanol, **2b**.

A series of pH experiments was performed with peptide **6.2a** and PTAD to examine to what extent the pH influences the TAD tyrosine conjugation reaction. A conjugation reaction of **6.2a** with PTAD **2a** was performed in pure water, at pH 5, pH 7 and pH 9 to examine the influence of the buffer presence and the pH. In theory, the conversion should be lower for experiments performed in lower pH due to competitive TAD hydrolysis. The analysis of the conjugation samples was done via LC-MS. For clarity the extracted ion chromatograms (XIC) for the **6.2a** ($[M+H]^+$ 494.2 Da, purple) and for the conjugated product **6.2a2a** ($[M+H]^+$ 670.2 Da, blue) are depicted in figure 6.9.

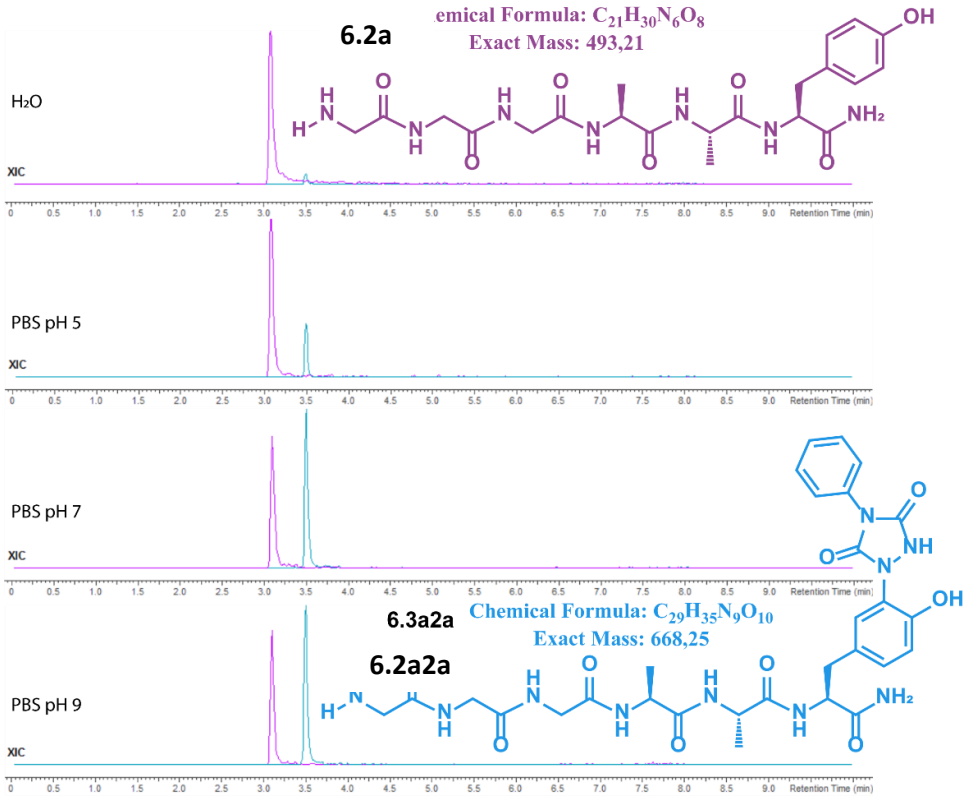


Figure 6.9 Extracted ion chromatograms for a pH screen of the reaction of peptide **6.2a** with PTAD, **2a** (2 eq.) in MilliQ H_2O or PBS buffers (pH 5, pH 7, pH 9). Peptide **6.2a** ($[M+H]^+$ 494.2 Da, purple) peptide **6.2a2a** ($[M+H]^+$ 669.2 Da, blue).

From the LC-MS analysis of these peptide conjugation experiments, we can indeed observe that almost no conjugated peptide product **6.2a2a** is formed when the conjugation is carried out in milliQ water. Although neutral water has a pH around 7, only a very small blue peak corresponding to product **6.2a2a** can be observed. The formed product (**6.2a2a**) is a urazole with a pK_a of around 5, which results in a lowering of the pH upon its formation. In turn, the lowering of pH (increased hydroxonium concentration) will push the phenol-phenolate equilibrium towards more towards the phenol, slowing down further reaction.

Slowing down the conjugation reaction means that the hydrolysis of PTAD which is in competition with the conjugation, takes the upper hand in pure water, almost excluding the formation of any conjugation product. In the conjugation sample at pH 5 the **6.2a2a** ion peak is still much smaller in comparison to the **6.2a** ion peak, even if the **6.2a2a** peak intensity is significantly higher compared to the intensity in the sample in milliQ water. The presence of the PBS buffer maintains the pH value at 5, this ensures that a small but constant fraction of phenolate is available for reaction with PTAD. In the samples at pH 7 and pH 9 the peak intensity for the product ion **6.2a2a** is higher compared to that of the peptide ion **6.2a**. The PBS buffers at pH 7 and pH 9 are able to ensure a signification fraction of phenolate for reaction with PTAD **2a**. Actually, in the pH 9 sample, the fraction of phenolate is 100 times larger compared to the sample at pH 7, which should lead to the formation of more product **6.2a2a**. However, at the same time, the hydrolysis of PTAD is also accelerated at increased pH (higher hydroxide ion concentration) leading to a roughly comparable outcome for the samples at pH 7 and pH 9.

A side product of the reaction with TAD reagents and free amines can be detected in the samples at pH 7 and pH 9 (see experimental data section 9.2.2.1). Free amines are able to react with a TAD rearrangement product that has lost N_2 and CO, leading to the formation of a urea bond (figure 6.10). The unwanted side product formed on the N-terminal amine is denoted with **N***.

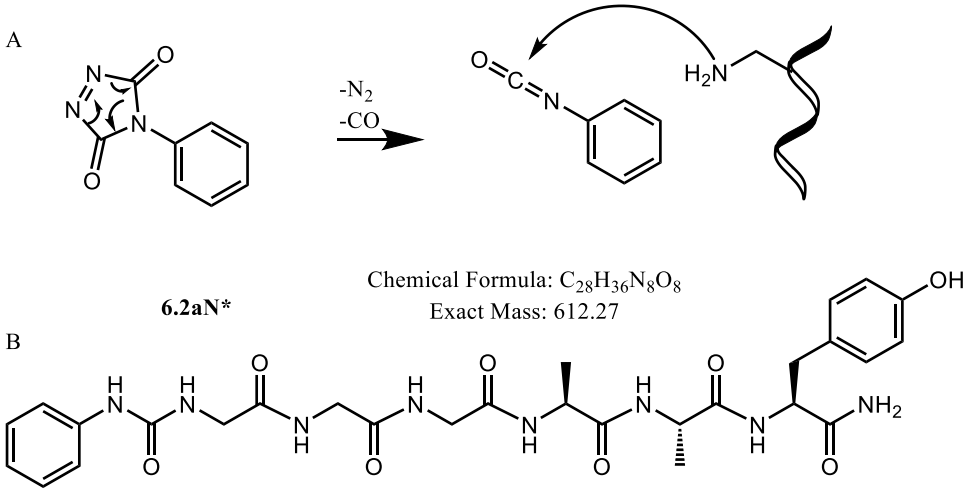


Figure 6.10 A) schematic representation of a known side reaction with amines (N-terminus/lysine): TAD loses N_2 and CO in a rearrangement and the resulting isocyanate is able to react with amines leading to a urea final product. B) Structural representation of a side product, **6.2aN*** formation in the reaction of peptide **6.2a** with **2a** .

As expected, this side reaction is observed at higher pH values when the N-terminal amine is more nucleophilic. The mechanism This side reaction is known in literature to occur on lysine residues and at the N-terminus.[298][299] This unwanted side reaction can be avoided by using trisaminomethane (Tris) buffers, since Tris contains a free amine itself which allows to quench this side reaction. In figure 6.11 the HPLC chromatogram at 214 nm for the reaction of **6.2a** with **2a** in PBS buffer at pH 7 is presented in combination with the extracted ion chromatogram for **6.2a** and **6.2a2a**. Besides the starting peptide and conjugated product, also 2 smaller peaks are observed, these correspond to the unwanted side reaction on the N-terminus amine in the starting peptide, **6.2aN*** and the conjugated product, **6.2a2aN***. Additionally also a peak is observed corresponding with the urazole precursor of **2a** indicated as **2aU.**

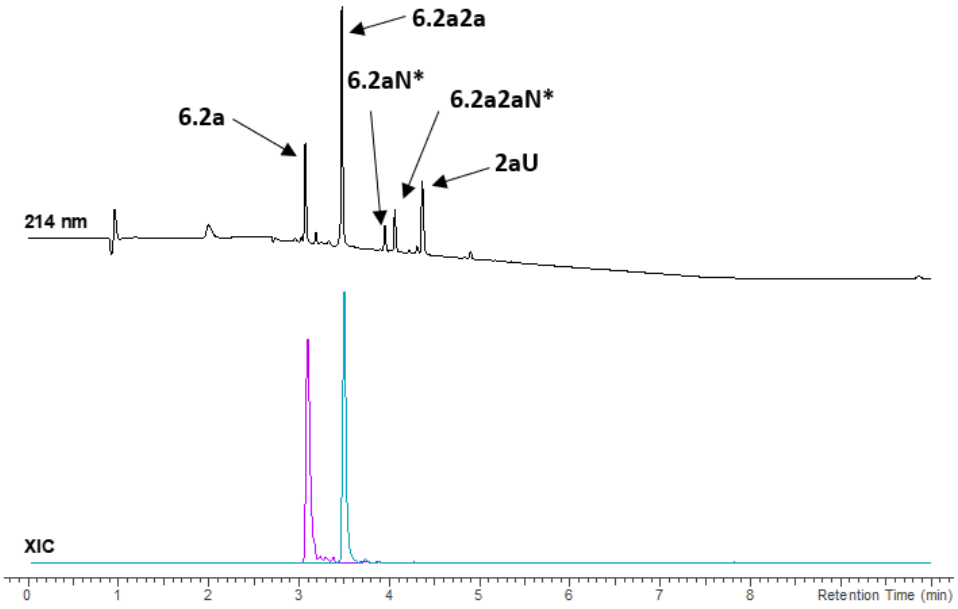


Figure 6.11 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with PTAD **2a** (2eq.) in PBS pH 7. LC chromatogram at 214 nm with annotated peaks including the side products resulting from reaction of the N-terminal amine with a **2a** rearrangement product: **6.2aN*** and **6.2a2aN*** (top), XIC of **6.2a** (purple, 494.2 Da) and **6.2a2a** (blue, 669.2 Da) (bottom).

In the experimental section 9.2.2.1, the complete LC chromatograms of these experiments are included along with the structures for the side products from this reaction. Additionally these pH experiments were repeated with a different TAD reagent: TAD propanol **2b** and the results lead to the same conclusions. With these data, we have confirmation that the TAD-tyrosine reaction is indeed highly dependent on the presence of buffer and the pH value of the buffer. We have observed that in pure water, as expected from computational studies, the TAD-tyrosine reaction does almost not take place. On the other hand, in buffered PBS conditions we have found that there is only a small amount of product formed at pH 5. This means that it should be possible to achieve selective farnesyl modification with TAD in pure water or at a pH value below 5, as these conditions will practically inhibit the TAD reaction with tyrosine.

Peptide GGGCAAS **6.2b1** is synthesized via SPPS, however we need to modify the cysteine residue to obtain the farnesylated peptide (see figure 6.12). In a first step, farnesol (mixture of isomers) is reacted with phosphorous tribromide to yield farnesyl bromide. The obtained farnesylbromide was used without purification in the reaction with peptide **6.2b1**, in the peptide coupling reaction $Zn(OAc)_2$ is added to coordinate the attacking sulphur atom. The reaction is performed in a solvent mixture (DMF/1-butanol/0.1 % aqueous TFA) and left overnight in the dark. The final product **6.2b2** was purified by reversed phase preparative HPLC on a C18 column.

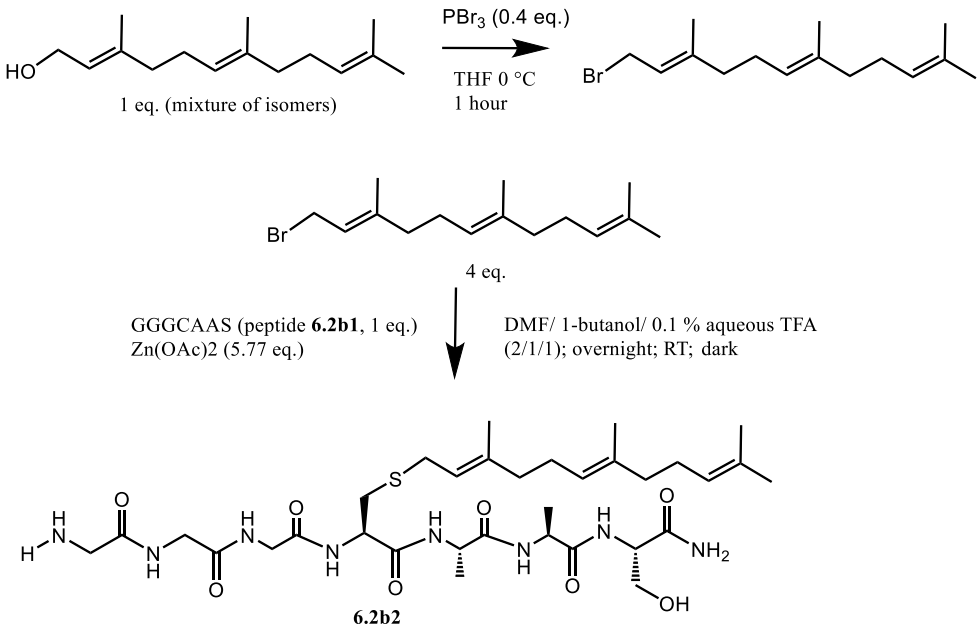


Figure 6.12 Reaction scheme for chemical farnesylation of peptide **6.3b1.**

Although it was possible to purify **6.2b2** via reversed phase HPLC, we observed that the farnesylated peptide was particularly sticky to the column and the product peak in the chromatogram was broad and tailing. This is caused by the hydrophobic farnesylgroup, which is added to the peptide. The type LC-MS analysis experiments that were performed with peptide **6.2a,** will not be possible with peptide **6.2b2.** An alternative analysis method is MALDI-TOF, unfortunately via this way only qualitative information is obtained since ionisation of modified versus non-modified peptides will be different, thus interfering with any quantitative interpretations. In general, direct MALDI-TOF analysis of experiments that are carried out in buffer solution, is more difficult as buffer ions suppress the ionisation of the analytes (peptides).

However, we were able to analyse the reaction of PTAD, **2a** with **6.2b2** in 10 X PBS pH 5 after diluting the sample in milliQ water. In the MALDI-TOF MS spectrum (figure 6.13) we can observe the farnesylated peptide **6.2b2** [$M+K^+$ = 763 Da] as well as three PTAD conjugated products corresponding with PTAD **2a** conjugated to each of the three double bonds in the farnesyl **6.2b2(2a)1, 6.2b2(2a)2, 6.2b2(2a)3**. These data indicate that all three double bonds in the farnesylgroup can react with a PTAD moiety leading to a mass addition of 175 Da for each added PTAD, **2a**. The reaction of TAD with farnesyl takes place via the alder-ene mechanism and is theoretically much less dependent of the presence of a buffer and the pH value of the buffer.

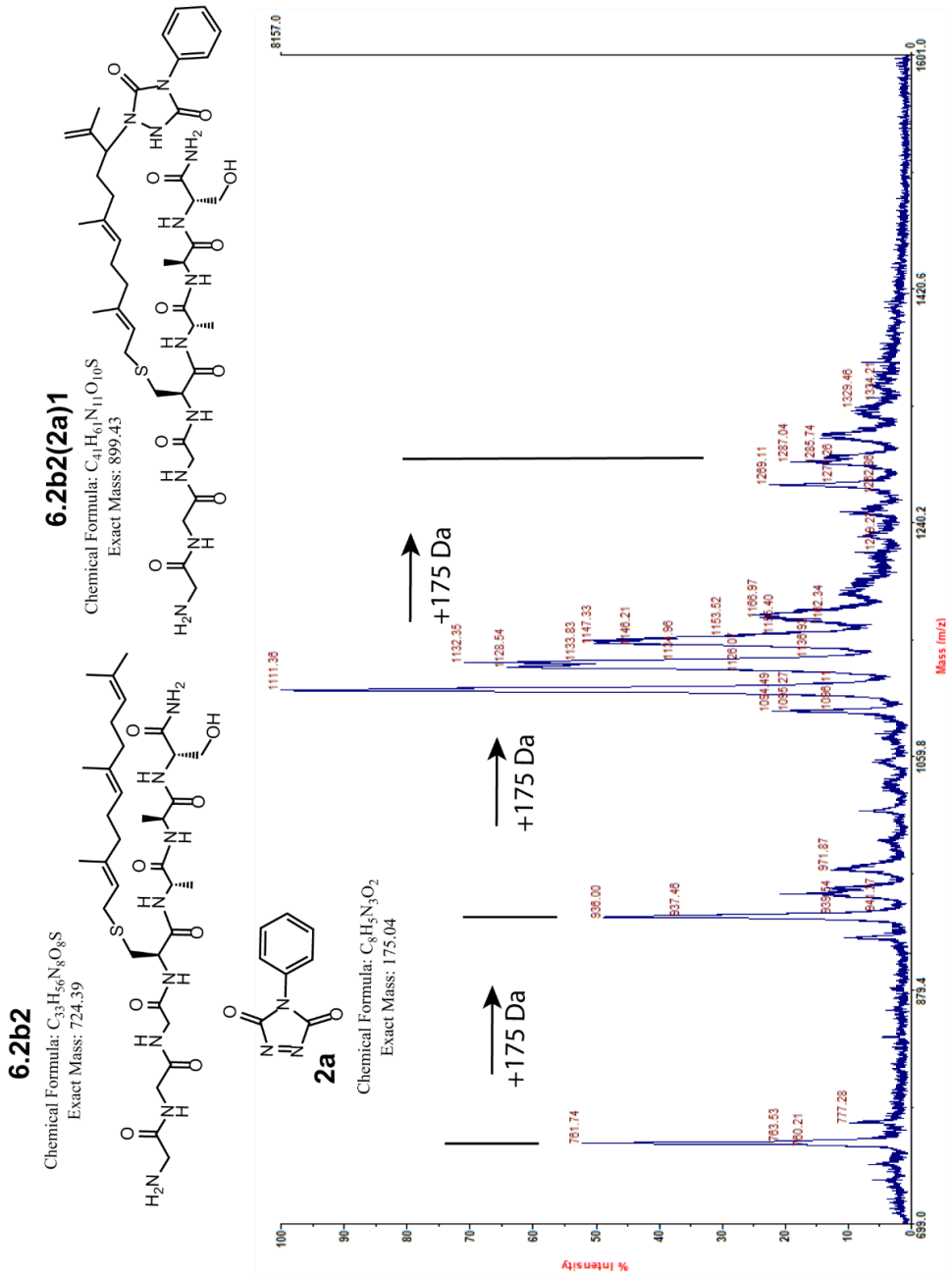


Figure 6.13 MALDI-TOF MS spectrum of the reaction of **6.3b2** with **2a** in PBS buffer at pH 5. Farnesylated peptide **6.2b2** is observed in this mass spectrum, calculated [M+K^+ = 762.4 Da], observed: 761.7 Da. Additionally, three sequential additions of **2a** (+175 Da) can be observed, calculated masses for the potassium adducts of **6.2b2(2a)1, 6.2b2(2a)2** and **6.2b2(2a)3** are: 937.4 Da, 1112.4 Da and 1287.4 Da respectively, the found masses are 938.0 Da, 1111.4 Da and 1287.0 Da respectively.

With standard peptides (e.g. peptides from a tryptic digest) a Zip-Tip procedure is often used to desalt the peptide samples before MALDI-TOF analysis to get rid of the buffer components. However, the Zip-Tip approach is based on a reverse phase column which is not compatible with our hydrophobic peptides. Due to difficulties in the MALDI-TOF MS analysis of experiments in buffer, the following experiments were performed in pure water.

A MALDI-TOF competition experiment between GGGAAY **6.2a** and GGGC(farnesyl)AAS **6.2b2** in pure water was performed. This competition experiment was performed with TAD-propanol **2b** since **2b** is more water soluble compared to **2a**. The competition experiment between tyrosine and farnesyl is performed to examine if the TAD reagent will preferentially react with the farnesylgroup even in presence of tyrosine. An equimolar amount of **6.2a** and **6.2b2** was dissolved in milliQ water and a MALDI-TOF MS spectrum was recorded (figure 6.14 top). Subsequently two samples were prepared with 5 and 10 equivalents of TAD-propanol added to the **6.2a** and **6.2b2** mixture (figure 6.14 bottom).

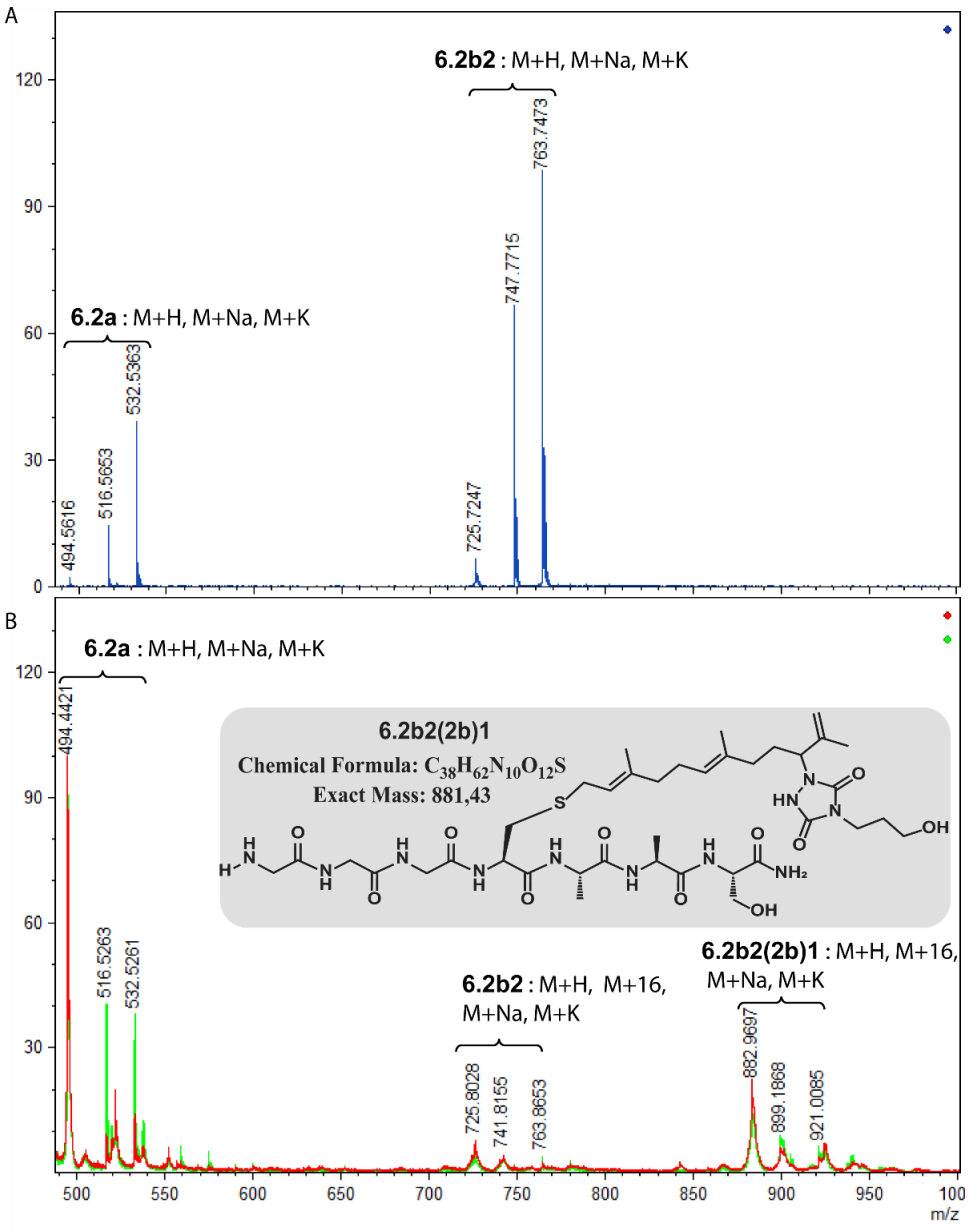


Figure 6.14 A: MALDI-TOF MS spectrum of the starting mixture of peptides **6.2a** ($[M+H]^+$ 494.6 Da, $[M+Na]^+$ 516.6 Da, $[M+K]^+$ 532.5 Da) and **6.2b2** ($[M+H]^+$ 725.7 Da, $[M+Na]^+$ 747.8 Da, $[M+K]^+$ 763.7 Da) B: MALDI-TOF MS spectrum of the reaction mixture in H_2O of **6.2a** ($[M+H]^+$ 494.4 Da, $[M+Na]^+$ 516.5 Da, $[M+K]^+$ 532.5 Da), **6.2b2** ($[M+H]^+$ 725.8 Da, $[M+H+16]^+$ 741.8 Da, $[M+K]^+$ 763.9 Da) + **2b** (green: 5 eq.; red: 10 eq.). The reaction product **6.2b2(2b)1** ($[M+H]^+$ 883.0 Da, $[M+H+16]^+$ 899.2 Da, $[M+K]^+$ 921.0 Da) is observed for both reactions with 5 or 10 eq. of **2b**, product **6.2a2b** is not observed. The chemical structure of **6.2b2(2b)1** with chemical formula and exact mass is depicted in the grey box.

The MALDI-TOF spectra in figure 6.14 B demonstrate the consumption of peptide **6.2b2** and formation of conjugated peptide **6.2b2(2b)1** and further additionally confirm the absence of peptide conjugate product **6.2a2b** which could be expected for TAD-tyrosine conjugation. These results indicate that the TAD-propanol reagent is selective for the farnesylated peptide in pure water in the presence of a peptide containing a tyrosine residue. Oxidation products corresponding to the $[M+H+16]^+$ ion for peptides **6.2b2** and **6.2b2(2b)1** are detected as side products. These

products are the sulfoxide analogues of the respective peptides resulting from oxidation of the farnesyl sulphide. The oxidative properties of TAD compounds have been described in literature.[300][191] Note that this oxidative side product is still a substrate for the reaction with **2b**. These data demonstrate that selectivity of TAD reagents for the farnesyl group over tyrosines can be obtained on peptides in milliQ H_2O. In a next phase in this study, the selectivity for selective reaction with a farnesyl moiety, was examined on farnesylated proteins.

6.2.3.3 Protein experiments

Peptide **6.2b2**, which was used in the previous subchapter, was designed in such a way that it was possible to ligate **6.2b2** to a protein of interest by making use of the sortase ligation procedure[301]. Sortase is a transpeptidase enzyme that can selectively cleave off a certain peptide from the C-terminus of a protein upon recognition of the sortag sequence and ligate another peptide to the remaining (N-terminal) part of the original sequence (*vide supra*, section 2.1.1). Via this method, a farnesylated protein can be obtained. As a first model protein a fascin nanobody FasNb5[302][303], compatible with sortase ligation, was selected (FasNb5 Sort-GG HA/HIS). The sortag sequence that is required for sortase recognition is LPETG, sortase will cleave the peptide bond between the threonine and glycine residues and ligate any peptide starting with three glycine residues, such as peptide **6.2b2** in this instance. A schematic representation of the preparation of the farnesylated nanobody is depicted in figure 6.15.

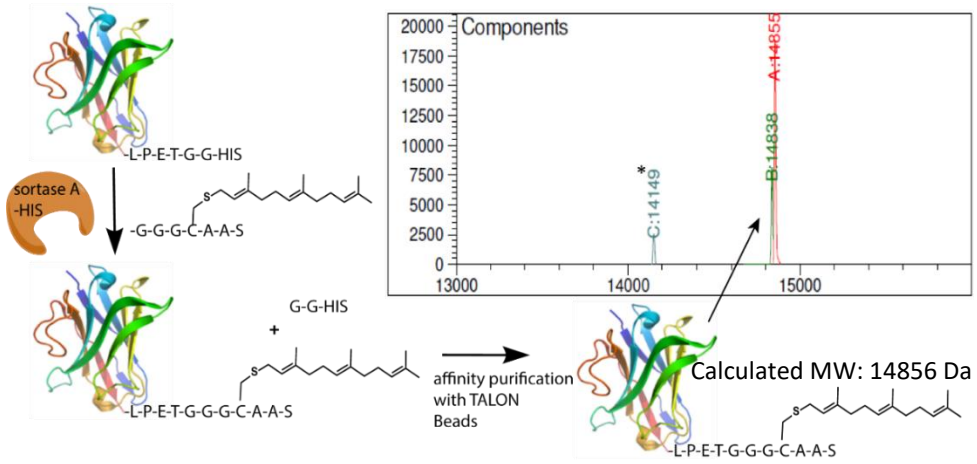


Figure 6.15 Schematic overview of the sortase mediated ligation of peptide **6.2b2** to a fascin nanobody FasNb5 sort-GG-HA/HIS and deconvoluted LC-MS spectrum of the final product (calculated: 14856 Da, observed: 14855 Da). The first step is the ligation of peptide **6.2b2** to the fascin nanobody and cleavage of the C-terminal GG-HIS sequence. In the second step the farnesylated nanobody is purified from the mixture via affinity purification with TALON® beads (Co^{2+}). An amount of nanobody where the GG-HA/HIS sequence was cleaved off but **6.2b2** was not ligated is present in the final mixture*.

The result of this sortase ligation was a mixture of the desired nanobody ligated with peptide **6.2b2** (observed deconvoluted mass: 14855 Da) and a nanobody resulting from cleavage of its C-terminus (as predicted by the sortase mechanism) without peptide **6.2b2** ligated (observed deconvoluted mass: 14149 Da). The expected mass difference for the addition of **6.2b2** to the FasNb5 nanobody is 706 Da, corresponding to the mass of the **6.2b2** peptide reduced with 18 Da for the loss

of water upon amide bond formation with the nanobody. The observed mass difference is 706 Da in the electrospray ionisation (ESI) LC-MS deconvoluted MS spectrum. The same sample from the sortase reaction was also analysed on MALDI-TOF MS where also the mass difference of 706 Da was observed.

The fact that a mixture of farnesylated and unfarnesylated nanobody was obtained, seemed like a disadvantage at first. However, the nonfarnesylated fascin nanobody in the mixture can actually be used as highly relevant internal control. Indeed, the non-farnesylated product is structurally almost identical to the farnesylated nanobody except for the last seven C-terminal amino acids including Cys(farnesyl). This internal control is useful as we aim to selectively target the farnesyl group with TAD reagents in the presence of other amino acids. Exclusive selectivity of TAD for farnesyl on the model peptides was obtained in milliQ H_2O. However it is not trivial to obtain a protein solution in milliQ H_2O and additionally proteins are not always stable in pure water.

On the other hand, a simple buffer exchange can afford a protein in acidic buffer solution which could also offer selectivity for the farnesyl over tyrosine. Indeed, we observed very limited conjugation to tyrosine residues in PBS pH 5 when no other competing substrate for TAD is present. Therefore, we expect that when a competitive substrate for TAD is present, the farnesyl group in our case, the large majority of TAD conjugation will take place on the farnesyl group rather than on the tyrosine. Reaction of PTAD **2a** with the sortase reaction mixture of farnesylated and non-farnesylated nanobody was carried out in PBS pH 5 using 100 eq. of **2a**, (based on an estimation of 50 % farnesylated nanobody in the mixture). The reaction was analysed via LC-MS (figure 6.16).

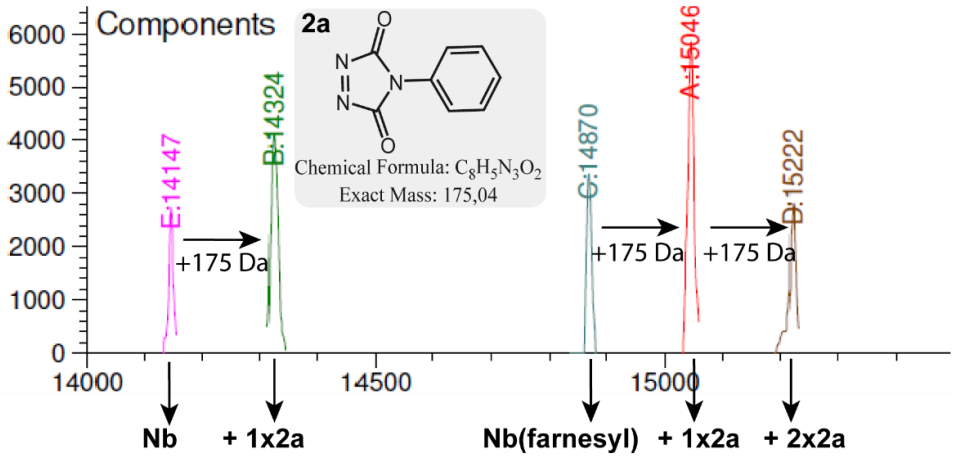


Figure 6.16 LC-MS deconvoluted mass spectrum (14 kDa - 15.5 kDa) of FasNb5 farnesylation reaction mixture after reaction with 100 eq. PTAD **2a** (based on an estimation of 50% farnesylated nanobody in the mixture) in PBS buffer at pH 5. ESI deconvoluted mass peaks are indicated, regular nanobody: 14147 Da, regular nanobody + **2a**: 14324 Da; farnesylated nanobody: 14870 Da; farnesylated nanobody + **2a**: 15046 Da; farnesylated nanobody + 2 x **2a**: 15222 Da. The chemical structure and formula along with exact mass of **2a** is shown in the grey box.

Upon LC-MS analysis, it can be noted that PTAD **2a** modifications are observed on both the farnesylated nanobody as well as the non-farnesylated nanobody. For the

farnesylated nanobody a single and a double PTAD **2a** modification were detected, ESI deconvoluted masses are 15046 Da and 15222 Da respectively. Note that the ESI deconvoluted mass for the farnesylated nanobody is 14870 Da which is 15 Da more compared to the value found after sortase ligation, which can be explained by an oxidation (+16 Da). Small deviations of a few Dalton off the expected value in the LC-MS deconvoluted masses are common for proteins of this size.

The non-farnesylated nanobody is modified one time with **2a**, ESI deconvoluted mass 14324 Da. It is now clear that the control nanobody without farnesyl group demonstrates its importance in this experiment. If only the farnesylated nanobody was present in the sample, a double modification of PTAD **2a** would likely be attributed to two modifications to the farnesyl group, since we have previously observed up to three individual additions of a TAD moiety. However, we observe also a single modification on the nanobody without farnesyl, indicating that something else is happening. This also indicates that in the case of the farnesylated nanobody where double PTAD **2a** modification is observed, likely only one is originating from modification on the farnesyl group.

At pH 5, we did not expect tyrosine to be entirely responsible for this modification as typically at such low pH we have demonstrated earlier (*vide supra*) that the TAD-Tyr almost does not occur. We subsequently considered that in principle also tryptophan could react with TAD (even if this was earlier communicated not to be the case).[199] A competition experiment between farnesylated peptide **6.2b2,** tyrosine peptide **6.2a** and a *N*-Boc protected tryptophan amino acid for TAD-propanol **2b** in water was performed. The sample was analysed via LC-MS (figure 6.17). To our surprise, a large amount of TAD conjugate product formation for the tryptophan amino acid could be observed. Note that there are two separate peaks in the LC chromatogram corresponding with the mass for the TAD propanol-conjugated *N*-Boc-tryptophan (light blue trace in figure 6.17). At this point we did not know the exact structure of both conjugated products. In section 6.3 of this thesis the reaction of TAD with tryptophan was investigated in more detail. For the farnesylated **6.2b2** which was observed to be oxidized to the sulfoxide **6.2b2(ox)**, only minor reaction with **2b** was observed (light green trace), and for tyrosine containing peptide **6.2a** no conjugation with **2b** could be detected (brown trace only contains compound corresponding to the starting peptide).

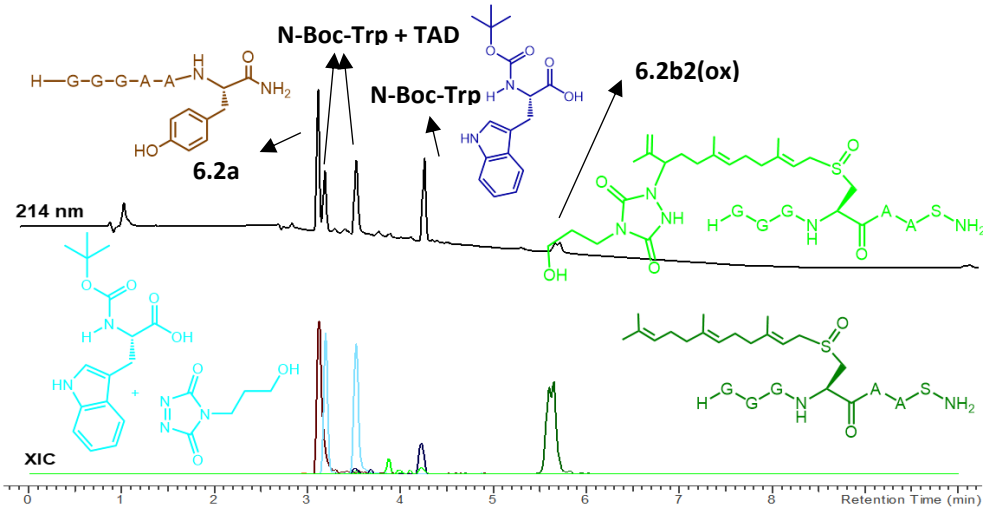


Figure 6.17 LC (214 nm) chromatogram for a competition experiment between farnesylated peptide **6.2b2** and tyrosine peptide **6.2a** and a N-Boc protected tryptophan amino acid for TAD-propanol **2b** in water (top). XIC's for **6.2b2(ox)** (dark green), **6.2b2(2b)ox** (light green), **6.2a** (brown), N-Boc protected tryptophan (dark blue) and the combined mass of N-Boc-tryptophan + **2b** (light blue).

We did not expect reaction of TAD with tryptophan, and this competition experiment was actually designed as a test to confirm this. However, the design of this competition was certainly not ideal, as the tryptophan amino acid was compared to a tyrosine and farnesyl moiety that were incorporated in a larger peptide, which could affect the reaction. Nonetheless, these initial results indeed indicated that tryptophan can and will react with TAD in water even in the presence of other substrates for reaction with TAD. Additionally it led us to start an in depth investigation of the reaction of tryptophan with TAD reagents which is presented in chapter 6.3.

With the knowledge that tryptophan can react with TAD, the observation that a conjugation product is formed from the non-farnesylated FasNb5, could now be attributed to tryptophan conjugation. There are two tryptophan amino acids in the amino acid sequence of the FasNb5 nanobody. The crystal structure of the nanobody is not known indicating that it is not sure whether both tryptophan residues are solvent accessible and available for modification. Detailed sequence information indicates that both tryptophan residues are actually located in framework regions of the nanobody. These framework regions are conserved parts of the sequence, involved in the general folding of the nanobody. in These framework regions are highly conserved in different nanobodies. This indicates that the crystal structure of other nanobodies can provide information on the solvent accessibility of the tryptophan residues in FasNb5. Based on two nanobodies with a resolved crystal structure (gelsolin nanobody and EgA1), we could derive that one of the two framework tryptophan residues is indeed likely solvent accessible, while the other tryptophan is buried in the interior of the crystal structure (figure 6.18 for example of EgA1 nanobody).

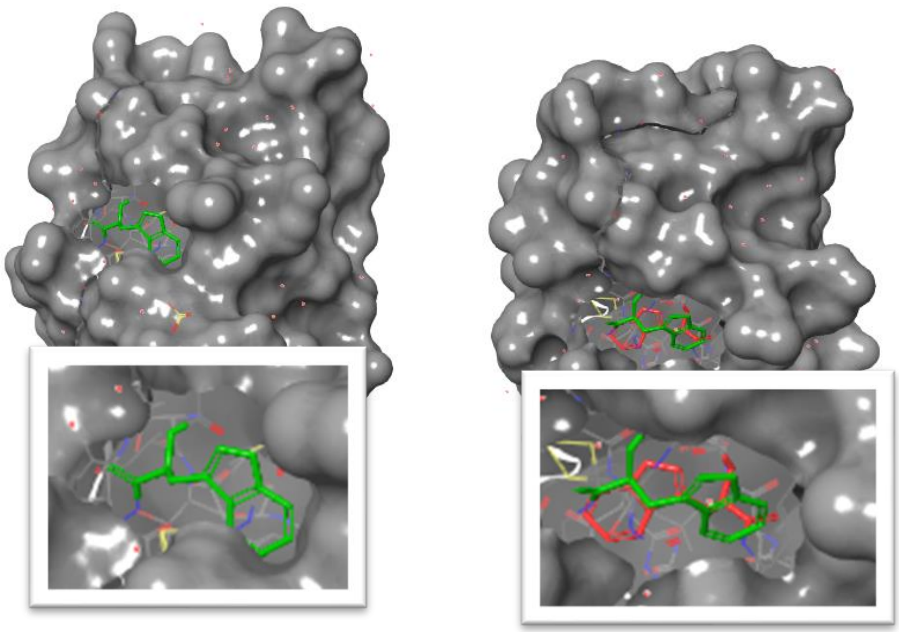


Figure 6.18 Two views of the X-ray structure of the EgA1 nanobody with protein surface and indicated tryptophan residues, with zoomed in panels of the relevant parts. Tryptophan at the beginning of framework region 4: green and tryptophan at the beginning of framework region 2: red. On the left view a tryptophan residue (green) can be seen on the surface of the nanobody. On the right view a buried tryptophan (red) can be observed buried in the internal structure of the protein underneath the exposed tryptophan (green). PDB:4KRO

The fact that there is competition between tryptophan and farnesyl residues for reaction with TAD-reagents is unfortunately bad news for the development of a TAD-based methodology for site-selective targeting of the native farnesyl PTM. On the other hand, TAD-reagents might still be interesting for detection purposes, moreover tryptophan is the least abundant amino acid of the 20 canonical amino acids.[48] Almost 10% of the human proteins do not contain a tryptophan in their primary sequence as was calculated from the data available on Uniprot.[4] Among the proteins that do not contain a tryptophan residue are the members of the Ras family, which are from an oncology perspective the most important farnesylated proteins (*vide* infra).

[4] https://www.uniprot.org/downloads

6.2.3.4 *In vitro* farnesylation

From an oncology point of view, Ras proteins are the most interesting proteins that are farnesylated in humans. Remarkably none of the three clinically most relevant Ras protein family members[304] (H-Ras, K-Ras and N-Ras) contain a tryptophan residue in their primary amino acid sequence (see experimental data section 9.2.2.5). The four C-terminal amino acids of the members in this protein family follow the CaaX motif. This motif is recognized by farnesyltransferase, the enzyme responsible for farnesylation of the Cysteine in this motif. Recombinant human H-Ras (21.3 kDa) was selected as a model protein without tryptophan amino acids in this study. This protein was *in vitro* farnesylated using recombinant farnesyltransferase, subsequently the reaction was analysed with LC-MS.

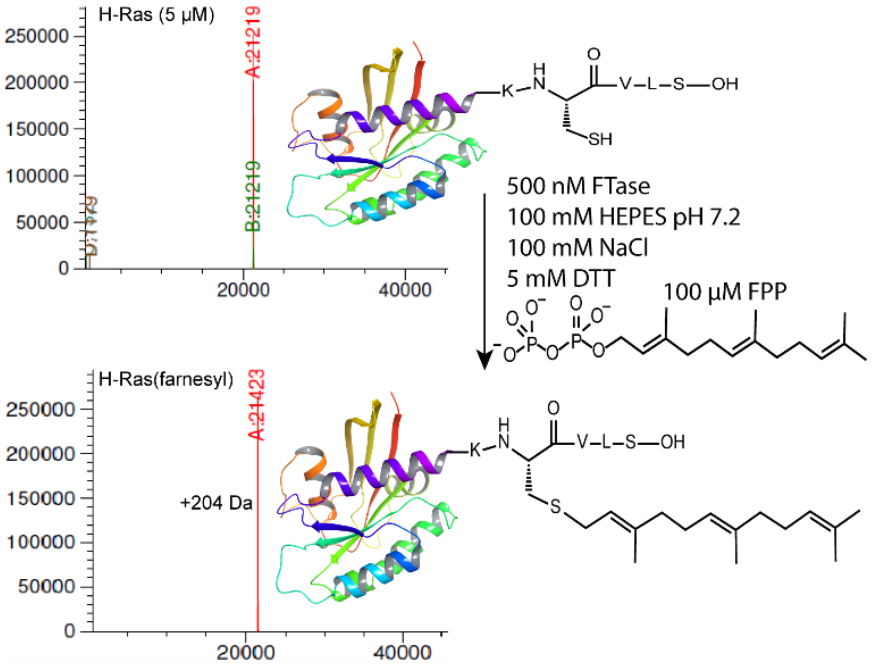


Figure 6.19 LC-MS deconvoluted MS spectra of the H-Ras protein, ESI deconvoluted mass: 21219 Da (top) and H-Ras protein after farnesylation, ESI deconvoluted mass: 21423 Da (bottom). Cartoon representation of H-Ras with the chemical structure of the final five amino acids: starting protein (top right) and after farnesylation (bottom right). Conditions of the enzymatic reaction (middle right).

The enzymatic farnesylation reaction was performed in a 100 mM HEPES buffer at pH 7.2 with 100 mM NaCl and 5 mM DTT. H-Ras (5 µM), FTase (500 nM) and FPP (100 µM) were added and the reaction was shaken for 8h at 37°C. The ESI deconvoluted mass for H-Ras is 21219 Da, after farnesylation the deconvoluted mass of the farnesylated H-Ras is 21423 Da, corresponding with a 204 Da increase as was expected ($+C_{15}H_{26}$ $-H_2$ = +204.2 Da). Farnesylation was perceived to be quantitative since no starting H-Ras protein was observed after the farnesylation reaction. The *in vitro* farnesylation of H-Ras proceeded very efficiently with full conversion in 8h. While in literature only farnesylation at the C-terminus is reported, we proceeded to try whether *in vitro* farnesylation of a peptide with an internal CaaX motif is possible, as this would then allow to develop a two-step farnesylation/TADylation labelling protocol.

In vitro farnesylation by FTase of short peptides, even tetrapeptides, has been demonstrated in the past.[305] Therefore, two new peptides were synthesised via SPPS QYKKGCVLS **6.2c** and QYKKGCVLSGQKS **6.2d**. The CVLS motif was selected as a CaaX motif since it was shown to be farnesylated most efficiently.[305] Peptide **6.2c** with a conventional C-terminal CaaX motif was used as a positive control and **6.2d** was a test peptide with an internal CaaX motif. An *in vitro* farnesylation experiment was performed with **6.2c** and **6.2d** and the reaction mixture was analysed by MALDI-TOF MS after 2h enzymatic reaction (see experimental data figures 9.104 and 9.105). For peptide **6.2c** farnesylation was clearly observed, with a mass addition of 204 Da corresponding to the introduction of the farnesyl group. On the other hand for peptide **6.2d** no mass addition was detected. This result indicates that internal CaaX farnesylation might not be possible. However, we here tested only one internal CaaX test peptide and a much broader study would be required to prove that internal CaaX farnesylation cannot take place.

6.2.3.5 TAD modification of farnesylated H-Ras

Subsequently, the farnesylation mixture was used directly in the same HEPES buffer at pH 7.2 and without further purification for reaction with TAD-propanol **2b** (figure 6.20). TAD-propanol **2b** (200 equivalents) was used for the reaction with farnesylated H-Ras. The reaction was subsequently analysed with LC-MS.

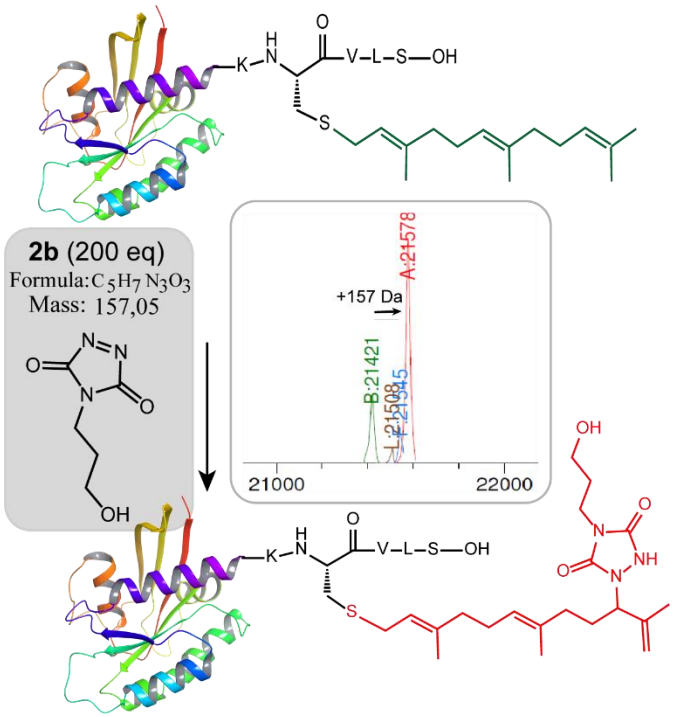


Figure 6.20 LC-MS deconvoluted mass spectrum of the TAD-propanol **2b** reaction with farnesylated H-Ras (insert). Farnesylated H-Ras (21421 Da, green) and farnesylated H-Ras TAD-propanol conjugate (21578 Da, red) are the two most abundant peaks in the spectrum. Cartoon representation of H-Ras with the chemical structure of the final five amino acids: farnesylated H-Ras (top) and farnesylated H-Ras TAD propanol conjugate (bottom). Structure, chemical formula and exact mass of TAD-propanol **2b**. Conjugation reaction was done in HEPES pH 7.2 with 200 equivalents of **2b**.

In the deconvoluted mass spectrum (figure 6.20 insert) the two most prominent peaks correspond with the mass of the farnesylated H-Ras (21421 Da) and

farnesylated H-Ras TAD-propanol conjugate (21421 Da +157 Da = 21578 Da). The mass difference between farnesylated H-Ras and the TAD-propanol conjugate peak corresponds with 157 Da as expected for the addition of one TAD-propanol moiety **2b**. From literature as well as previous experiments, we know that tyrosine residues can react as well in buffer at pH 7.2. H-Ras contains a total of 9 tyrosine residues of which several are exposed to the solvent (experimental figure 9.103).

To confirm whether the farnesyl group or one of the tyrosine residues reacted with **2b**, a series of control experiments were performed. In the first control reaction, TAD-propanol **2b** labelling was performed on H-Ras that had not been subjected to a farnesylation reaction. In a second control reaction, a similar H-Ras farnesylation experiment was done in the absence of FPP and the TAD propanol **2b** labelling was performed subsequently. Both control samples were analysed via LC-MS (experimental figures 9.101 and 9.102). Remarkably only low abundant masses were detected that were not corresponding with the starting H-Ras (21219 Da) or TAD-modified H-Ras (21376 Da) or any multiply TAD-modified H-Ras species. We hypothesize that the excess FPP used for the enzymatic farnesylation acts as scavenger for TAD-propanol **2b** and prevents reaction between tyrosine and TAD. These data seem to indicate that the TAD-propanol reacts selectively with the farnesyl group under these conditions. However, further control experiments and possible protein digestion experiments will be required to allocate the TAD modification to the farnesyl group. Additional peptide competition experiments between tyrosine and farnesyl are currently investigated by Laure Tack to further pursue this investigation.

6.2.3.6 Conclusion

In conclusion we present the use of triazolinedione (TAD) chemistry for selective modification of the native farnesyl group. The TAD-tyrosine reaction was reported in 2010 and has since been widely adopted as a potent tyrosine modification method. One of the main goals for this project was identifying conditions that allow selective reaction of TAD with farnesyl in presence of tyrosine. We were able to use TAD to selectively target a farnesylated peptide in a competition experiment with a tyrosine containing peptide in milliQ water. Based on TAD modification experiments with a tyrosine containing peptide in different buffers, we anticipate that it will be possible to attain the same farnesyl selectivity over tyrosine in buffered conditions below pH 5. Additionally, we have shown that all three double bonds in the farnesyl group can react with TAD leading to a triply modified peptide. On the other hand, we have observed competitive reaction of tryptophan with TAD reagents; this off target effect was not reported before and is further discussed in more detail section 6.3. Because of this off-target tryptophan effect, it will not be possible to attain complete amino acid selectivity. However, we still believe that TAD reagents can be useful for the scientific community as detection tools for prenylated proteins. The interest in labelling of prenylated proteins is further evidenced by a very recent publication where the use of TAD reagents for labelling of prenylated proteins is reported.[306] The authors mention the known TAD-tyrosine reaction but do not report on the selectivity of tyrosine versus farnesyl of their TAD reagent. Moreover the authors do not realize that they in fact also face

competition of tryptophan (*vide infra*). By using buffered conditions at a pH below 5, we can avoid tyrosine competition and improve labelling selectivity towards farnesyl.

6.3 Triazolinedione protein modification: from an overlooked off-target effect to a tryptophan selective bioconjugation strategy

The contents in this chapter were published as a preprint on ChemRxiv: Protein Conjugation with Triazolinediones: Switching from a General Tyrosine-Selective Labeling Method to a Highly Specific Tryptophan Bioconjugation Strategy; Klaas Decoene, Kamil Unal, An Staes, Kris Gevaert, Johan Winne and Annemieke Madder.

6.3.1 Introduction on tryptophan selective bioconjugation

Site-selective protein modification is of utmost importance for many applications from fundamental biology (fluorescent tagging) to therapeutic development (antibody-drug conjugates).[307],[2],[308],[309] While amino acid selectivity can be achieved by exploiting the nucleophilic functionalities of e.g. lysines and cysteines,[310],[49] genuine site selectivity depends on their representation density on the protein surface. In this regard, tryptophan (Trp) is an interesting target for native conjugation strategies, with an abundance of just over 1% in proteins.[48] Despite the indole side chain not being the most chemically tractable target, several groups have reported methodologies for selective modification of tryptophan in peptides and proteins.[311],[312],[313] ,[314] Many of these strategies employ transition metal catalysed reactions and/or conditions limiting downstream biochemical applications. These reactions are typically alkynylations and C-H arylations of the indole.[315],[316],[317],[318],[319] Also, Trp sulfenylation was demonstrated for peptide ligation.[320] While Francis and co-workers showed rhodium carbenoid-based Trp labeling at mild pH figure 6.21 a),[94] this method is dependent on transition metal catalysis and requires long reaction times. An organoradical Trp conjugation was demonstrated on peptides and proteins (figure 6.21 b) [95] and even if the method is devoid of transition metals, it requires acidic conditions, acetic acid in water, and is not compatible with buffers. Very recently, a novel biomimetic approach for the selective conjugation of tryptophan was developed, this method however employs UV irradiation and needs to be performed in absence of oxygen (figure 6.21 c).[96]

In 2010, Barbas and co-workers reported a click like reaction for the more abundant tyrosine (Tyr, 3.3% abundance[48]) using triazolinedione chemistry,[101] and several applications for protein conjugation followed.[321],[322],[103],[298],[323],[324],[299] Interestingly, when exploring these powerful TAD reagents on farnesylated peptides and proteins, we observed a high degree of off-target labeling on Trp residues, even in aqueous buffers. While the swift reaction of indoles with triazolinediones was reported by Baran, Guerrero and Corey in 2003,[325],[326] Barbas and co-workers demonstrated that tyrosine labeling is kinetically favoured in buffers. The selectivity of TAD reagents for tyrosine and tryptophan was investigated in detail on several peptides and proteins.

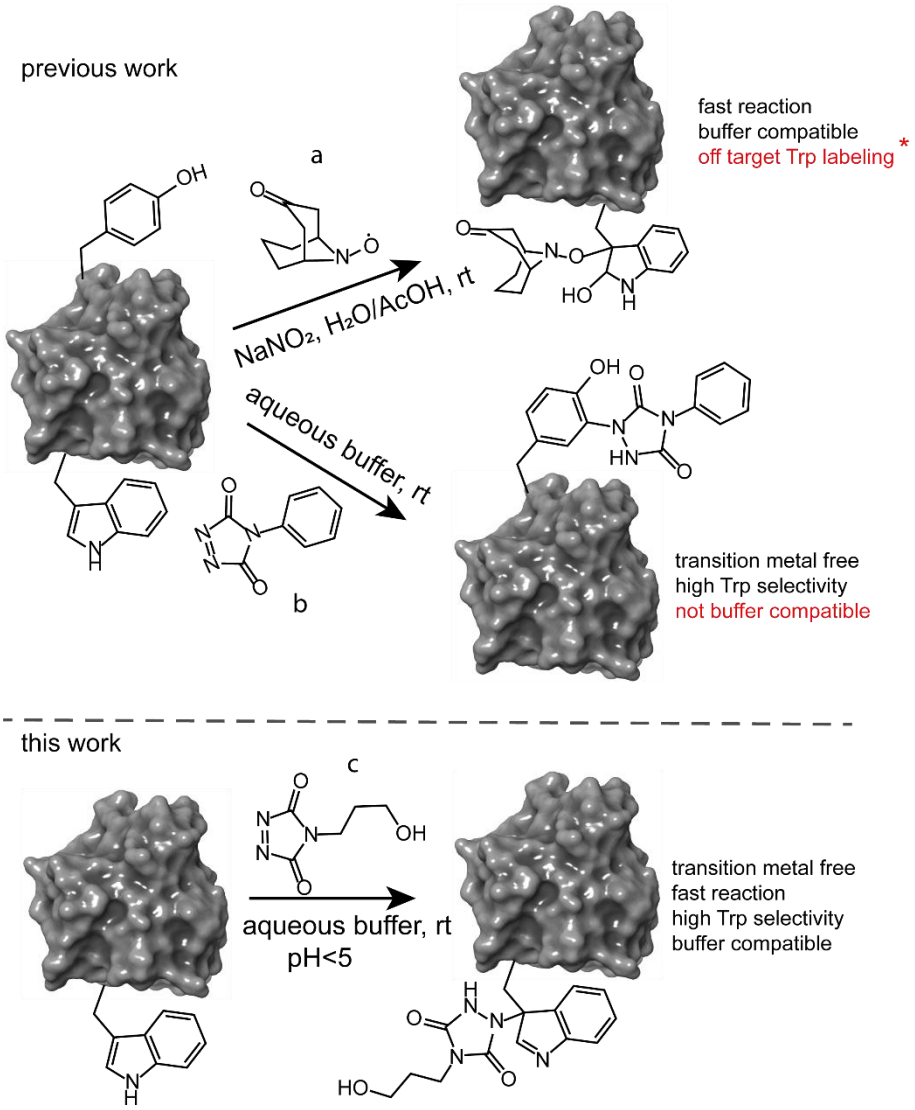


Figure 6.21 Schematic representation of: (a) organoradical tryptophan modification, (b) general triazolinedione tyrosine modification at pH 7, (c) selective tryptophan modification with triazolinediones at pH <5.

6.3.2 Tryptophan versus tyrosine selectivity

6.3.2.1 Intermolecular tyrosine versus tryptophan competition for TAD

We thus decided to more closely examine the competition between Trp and Tyr labeling by TADs in order to probe the potential of TAD reagents for selective Trp-bio-conjugation (scheme 1). For that purpose, tetrapeptides NWAS **6.3a** and NYAS **6.3b** were tested in intermolecular competition experiments with phenyltriazolinedione (PTAD **2a**) in PBS-buffer at two different pH values, allowing for head to head comparison between Tyr and Trp side chains embedded in the exact same chemical environment (figure 6.22). Signals for peptide conjugates **6.3a2a** and **6.3b2a** overlap on the HPLC UV chromatogram, therefore extracted ion chromatograms (XIC's) were used for the analysis.

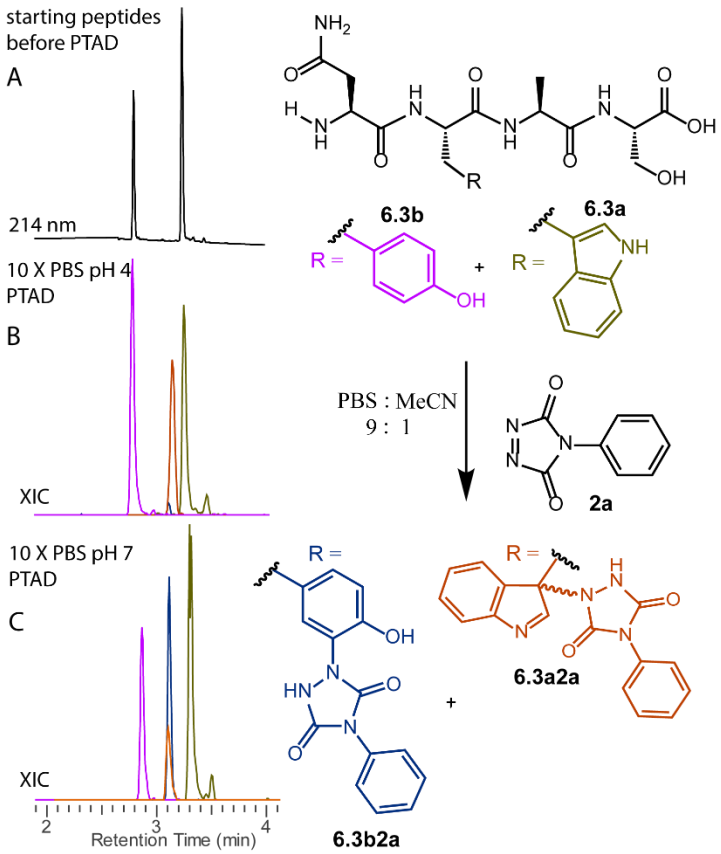
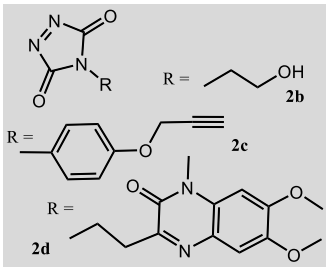


Figure 6.22 Competition experiment between NWAS (**6.3a**, 0.3 mM) and NYAS (**6.3b**, 0.3 mM) peptides for PTAD (**2a**, 0.4 mM) in 10x PBS buffers at pH 4 and pH 7. HPLC chromatogram before reaction (A). Extracted ion chromatograms (XIC) of both starting peptide ions **6.3a** (green) and **6.3b** (pink) and conjugated peptide ions **6.3a2a** (orange) and **6.3b2a** (blue) for reaction in 10x PBS at pH 4 (B) or pH 7 (C).

When analysing the XIC's of the starting peptide-ions NWAS **6.3a** (green) and NYAS **6.3b** (pink) and conjugated peptide-ions NWAS-PTAD **6.3a2a** (orange) and NYAS-PTAD **6.3b2a** (blue), a pronounced difference can be observed between the reaction at pH 4 and pH 7. Indeed, at pH 4 Trp conjugate **6.3a2a** was detected nearly exclusively while at pH 7 a mixture of conjugates was obtained. This observed pH-dependent reactivity of TADs with Tyr is in accord with previous mechanistic studies of the tyrosine-TAD click reaction, which indicate the phenolate species as the prevalent nucleophile.[297] Lowering the pH will effectively decrease the amount of tyrosine-phenolate form and thus decrease the extent of reaction of Tyr with TAD. This was further confirmed using additional peptides (**6.3a**-**6.3h**, table 1) and TAD-propanol **2b**, PTAD-alkyne **2c** and fluorescent DMEQ-TAD **2d** (experimental data section 9.2.3.4).

Table 6.1 Peptide sequences used in this study, structures of TAD reagents **2b**, **2c** and **2d**.

Entry	Sequence
6.3a	Asn-Trp-Ala-Ser-OH
6.3b	Asn-Tyr-Ala-Ser-OH
6.3c	Asn-Ser-Ala-Trp-OH
6.3d	Asn-Ser-Ala-Tyr-OH
6.3e	Trp-Ser-Ala-Asn-OH
6.3f	Tyr-Ser-Ala-Asn-OH
6.3g	Lys-Lys-Ser-Tyr-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH
6.3h	Lys-Lys-Ser-Trp-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH
6.3i	Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH
6.3j	Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH
6.3k	Val-Trp-Ser-Gln-Lys-Arg-His-Phe-Gly-Tyr-OH
6.3l	Lys-Asp-Tyr-Trp-Glu-Cys-Ala-OH



It was also observed that, even without competing Trp-peptide present, lowering of pH causes a significant reduction in Tyr-conjugate formation (figure 6.23). pH experiments were performed to investigate the pH dependency of TAD with tyrosine. Several conjugation reactions of TAD-propanol **2b** with peptide **6.3b** were done in a series of 10x PBS buffers at pH 4, 5, 6 and 7. The reaction between TAD and tyrosine is finished in several seconds, the background hydrolysis of TAD in aqueous solutions is competing with the tyrosine conjugation reaction. The bright purple color of TAD-propanol **2b** and PTAD **2a** swiftly disappears upon addition to the peptide solution. The reaction mixture is analyzed by HPLC (214 nm).

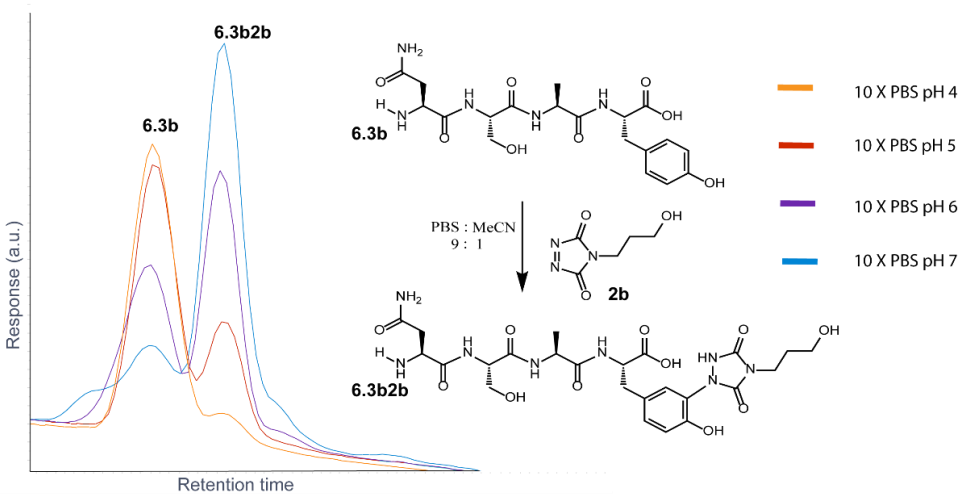


Figure 6.23 Zoom in of an HPLC chromatogram overlay (214 nm) of the reaction of a peptide **6.3b** with TAD-propanol **2b** in several 10 X PBS buffers with different pH ranging from pH 4 to pH 7.

These results point to the same conclusion as the one presented in section 6.2. The reactivity of tyrosine towards TAD is highly dependent on the pH. The reaction proceeds smoothly at pH 7, with almost all starting peptide **6.3b** material consumed and intense peak for product **6.3b2b** (figure 6.23, blue). On the other hand, at pH 4 there is hardly any **6.3b2b** product formation (figure 6.23, orange). The major shift happens between pH 5 where the major peak is the starting peptide **6.3b** and pH 6 where the major peak is the product **6.3b2b**.

6.3.2.2 Peptide conjugate stability

For the application of the TAD-tryptophan strategy as a new protein modification method the modification should be stable in solution. Most applications would require the modification to be stable in physiological conditions. To examine the stability of the TAD-tryptophan connection in solution the stability of the TAD-peptide conjugate was investigated by HPLC analysis of conjugate **6.3e2b**. The conjugate was left for 2 weeks at room temperature in 10 X PBS buffer at pH 7, at several time points HPLC samples were recorded. Analysis of the HPLC chromatograms at 214 nm from these samples demonstrated only a minor reduction of the **6.3e2b** conjugate peak after 2 weeks (figure 6.24).

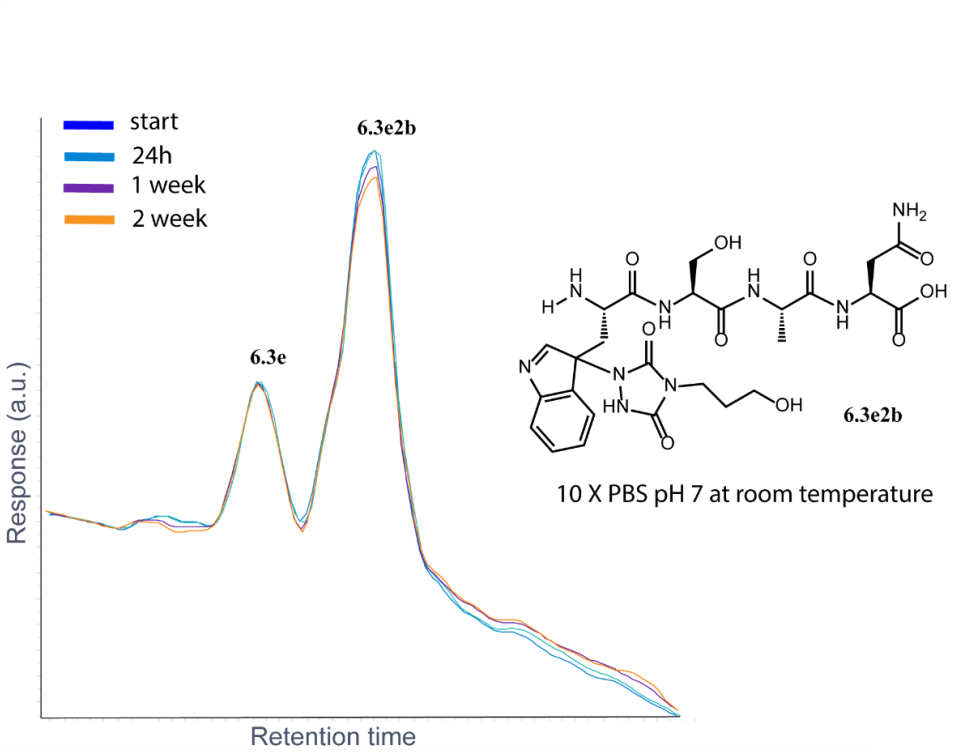


Figure 6.24 Zoom of an HPLC chromatogram overlay (214 nm) of the Trp-peptide conjugate **6.3e2b** in 10 X PBS buffer at pH 7. The conjugate was left at room temperature and at several time points a sample was taken and a chromatogram recorded.

6.3.2.3 TAD protein modification in literature.

These findings at peptide level prompted us to look in more detail to earlier reports on the tyrosine click protein modification. Indeed, off-target Trp-labeling was observed earlier at protein level. In the initial study of Ban *et al*. modification on tryptophan was observed upon myoglobin labeling albeit in a very low amount.[101],[298] Furthermore, careful reinterpretation of the MALDI-TOF MS spectra (kindly provided by the authors) of Vandewalle *et al*.,[322] who labeled BSA with butyl-TAD, showed that Trp-modification was indeed noticeable (figure 6.25).

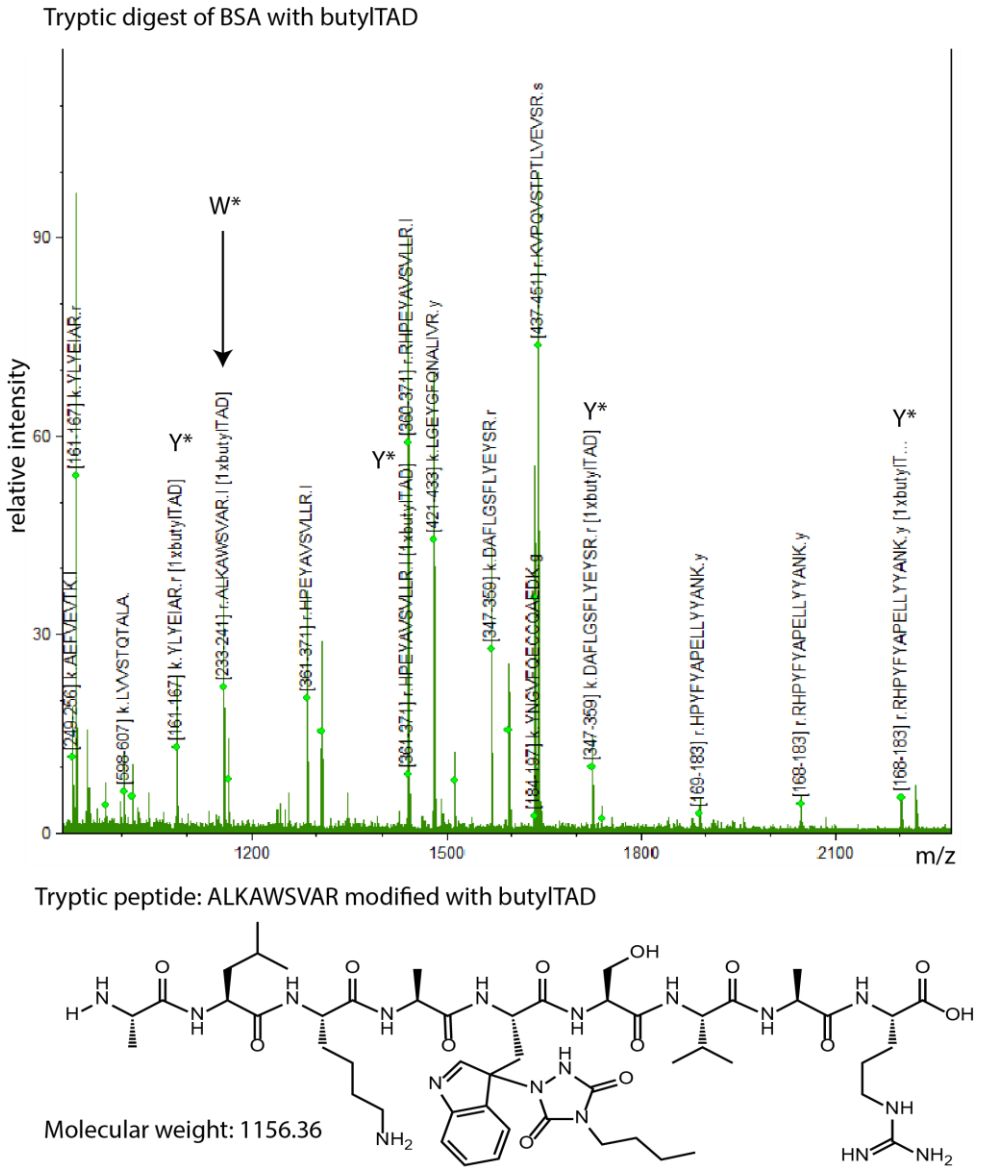


Figure 6.25 MALDI-TOF MS spectrum from a tryptic digest of BSA after reaction with butyl-TAD in PBS/MeCN (0.28/0.72) (top). Chemical structure of a tryptic peptide modified with butyl-TAD (bottom). These data originate from Vandewalle *et Al*. and were reinterpreted. A peptide fragment indicated with a Y* or W* marks a fragment with a TAD modification. As described in the aforementioned article several tryptic peptides modified with butyl-TAD were detected (Y*) , all these modified peptides contain at least one tyrosine residue and no tryptophan. However, after careful reanalysis also a butyl-TAD modified tryptic fragment containing one tryptophan and no tyrosine residue was observed (W*) corresponding with the peptide fragment ALKAWSVAR.

These findings demonstrate that researchers can incorrectly assume that tryptophan will not react with TAD-reagents in protein conjugation reactions, possibly leading to flawed interpretation of data.

6.3.2.4 Influence of the relative Trp position

Triggered by these findings, indicating that a completely Trp-selective modification can be possible, we examined TAD-Trp conjugation in peptides by investigating the influence of the relative amino acid positioning on the outcome of the reaction. Competition experiments between tetrapeptides NWAS **6.3a** and NSAW **6.3c** and TAD-propanol **2b** showed a remarkable difference in reactivity (figure 6.26).

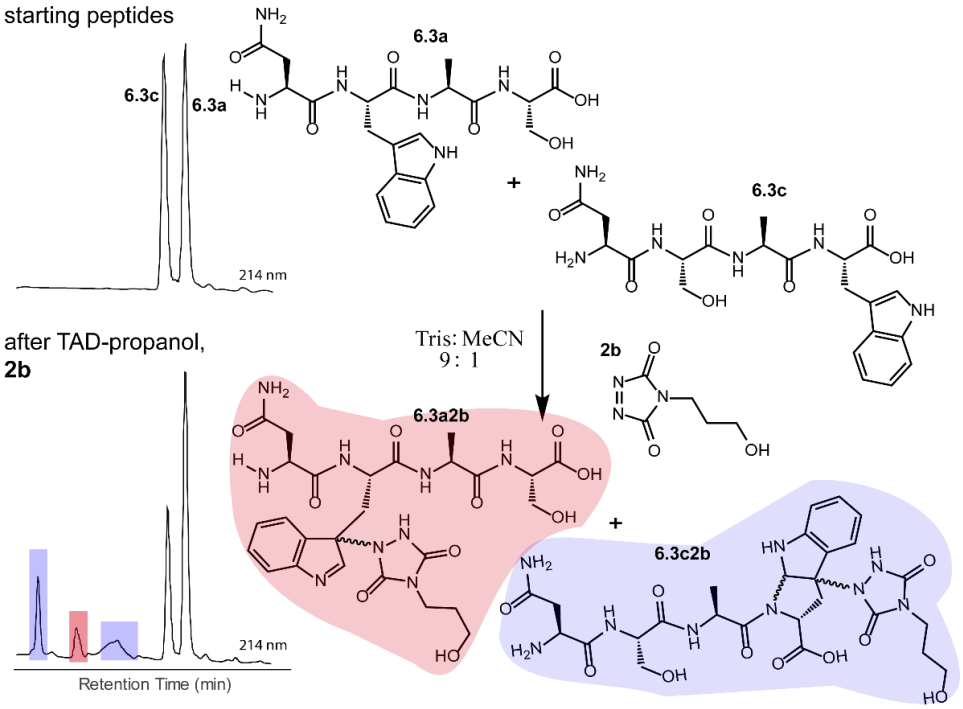


Figure 6.26 Competition experiment of NWAS (**6.3a**, 0.3 mM) and NSAW (**6.3c**, 0.3 mM) peptides with TAD-propanol (**2b**, 0.5 mM) in Tris buffer pH 5.8. Zoom of the HPLC chromatogram at 214 nm of the starting mixture (top left) and after reaction with TAD-propanol **2b** (bottom left).

Intermolecular competition between **6.3a** and **6.3c** clearly demonstrates the position-sensitivity of the Trp-TAD reaction: the C-terminal tryptophan in **6.3c** is labeled to a three times higher extent, as calculated via HPLC peak integration at 214 nm, compared to its internal tryptophan **6.3a** counterpart. This reactivity difference can be attributed to the more exposed reactive centre as well as to the presence of the carboxylic acid which can transiently donate a proton to the TAD moiety rendering it even more electrophilic. A second striking difference resides in the nature of the formed adducts. For the C-terminal tryptophan, two peaks for the labeled product **6.3c2b** are observed, indicating the formation of isomers. Indeed, we found this adduct had undergone an additional annulation caused by the reaction of the lone pair on the backbone nitrogen with the indole C2 after reaction of TAD with the indole C3. These findings were confirmed via NMR analysis of Boc-Trp-OH and N-Ac-Trp-OMe adducts with TAD-propanol **2b** (experimental data

section 9.2.3.13) and are in agreement with the results reported by Baran *et al*.[325] on non-peptide related TAD-indole reactions.

6.3.2.5 Intramolecular Tyr vs Trp competition for TAD

In a subsequent series of experiments, we investigated if the observed intermolecular selectivity, translates into intramolecular Trp versus Tyr selectivity. To this end, competition experiments were performed with peptides containing both tyrosine and tryptophan (**6.3i**-**6.3l**, table 6.1). MS/MS analyses were carried out to determine the modification site. We found that the modification on tryptophan is unstable in all tested MS/MS conditions except for ESI in combination with electron transfer dissociation (ETD). ESI-higher energy collision induced dissociation (HCD), ESI-collision induced dissociation (CID) as well as MALDI-TOF/TOF all largely lead to the loss of the TAD modification on tryptophan (experimental data sections 9.2.3.5, 9.2.3.8 and 9.2.3.9). The TAD modification on tyrosine was found to be stable in all tested conditions. These findings are in agreement with earlier work on the thermoreversibility of indole-TAD reactions.[230]

Peptide VWSQKRHFGY **6.3k** was labeled using TAD-propanol **2b** at pH 4 to selectively target the tryptophan residue in the peptide, or at pH 7 to mainly target the tyrosine. The conjugated peptide will have the same mass regardless of the site of the modification (M+: 1464.71), the triply charged precursor (M+++: 488.91) was selected for electron transfer dissociation (ETD) MS/MS analysis. At pH 4 modification only takes place on tryptophan and Trp-modified fragment ions are observed (figure 6.27, top). If the modification reaction is performed at pH 7 the major product is the tyrosine modified peptide (figure 6.27, bottom). However upon detailed analysis of the full ion chromatograms (experimental figures 9.147 and 9.148) it can be seen that in contrast to Trp-labeling which occurs cleanly at pH 4, the Tyr-labeling of this peptide at pH 7 is accompanied by formation of Trp-modified side products. Further confirmation for the off-target labeling at pH 7 is provided through MS based screening for doubly modified peptides (experimental figure 9.149) showing double TAD modified peptides only at pH 7, implying additional off-target Trp labeling next to the intended Tyr labeling.

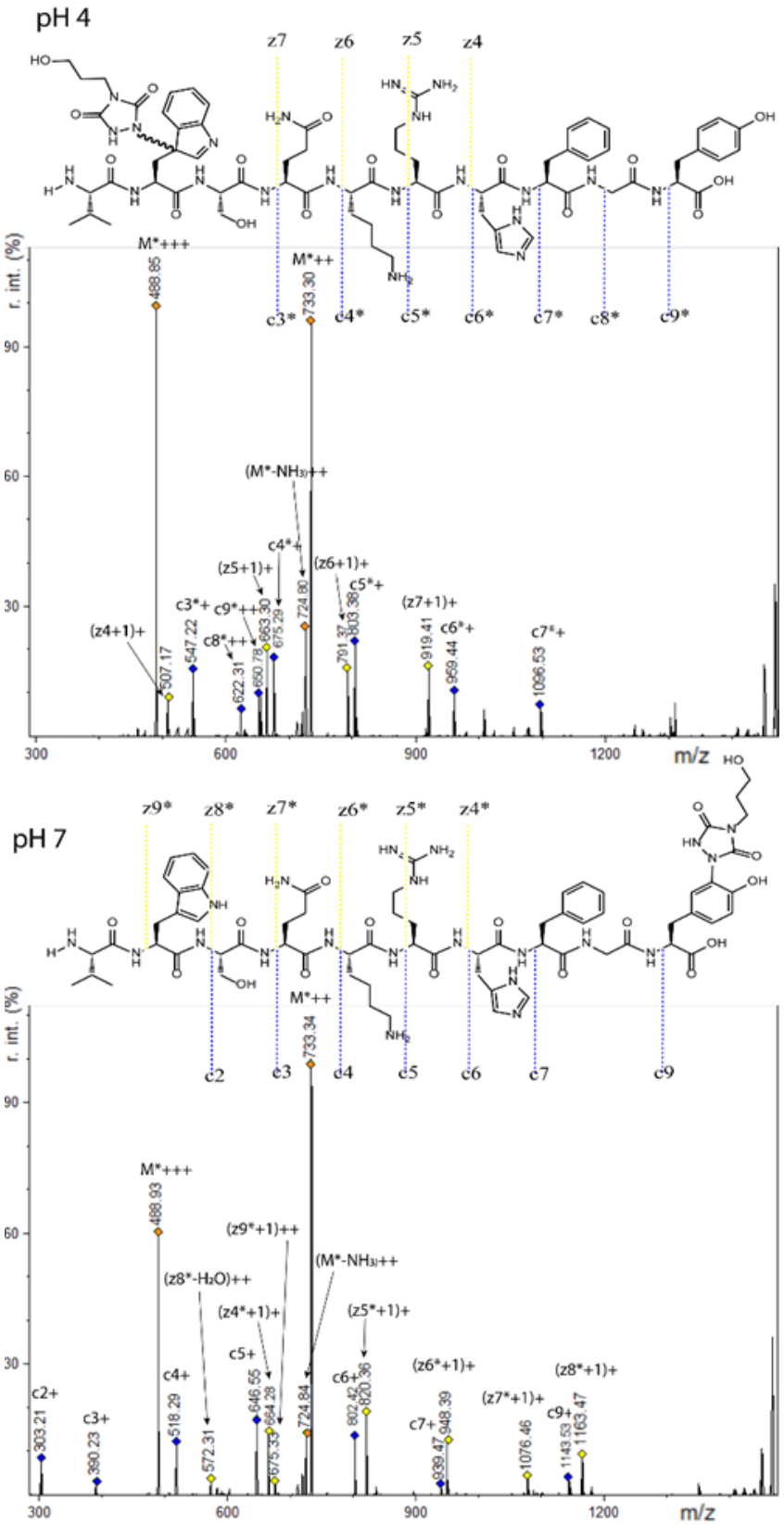


Figure 6.27 ESI ETD MS/MS of VWSQKRHFGY **6.3k** after modification with TAD-propanol **2b** in 10 X PBS pH 4 (top) and pH 7 (bottom). In both MS/MS spectra the chemical structure is depicted with the observed fragment ions. Fragment ions corresponding to TAD-modified peptide fragments are indicated with "*".

6.3.2.6 TAD modification of recombinant proteins

We next explored TAD-click reactions for Trp-based protein conjugation. Alphabodies have a triple helical coiled coil structure and are developed for intracellular protein interaction targets by Complix N.V.[228] Two recombinant alphabodies were used; the valentine alphabody containing no tryptophans and three tyrosines, and alphabody 586D containing one tryptophan residue next to three tyrosines. A competition experiment between the valentine alphabody and tryptophan containing peptide **6.3c** (experimental figure 9.160) with **2b** at pH4, resulted in 80% conversion of **6.3c** while alphabody conjugation was absent. Protein conjugation with 586D at pH 4 was carried out using fluorescent DMEQ-TAD, **2d**. Alphabody 586D contains six methionine's, upon reaction with DMEQ-TAD, **2d** we observed several oxidation (+16 Da) products caused by methionine oxidation. The 586D **2d** conjugation reaction was repeated with a buffer solution that was previously bubbled with Argon to remove as much oxygen as possible. In the intact protein analysis of the sample where buffer pre-treated with Argon was used only one oxidation was observed (figure 6.28).

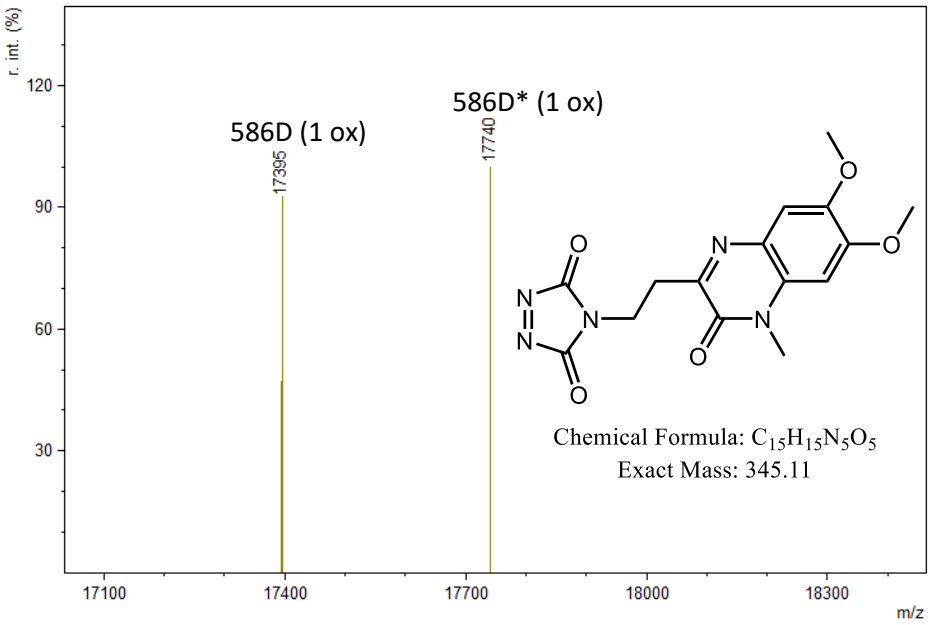


Figure 6.28 Deconvoluted mass spectrum of the alphabody 586D protein after addition of DMEQ-TAD **2d** (30 eq) in 10 X PBS pH 4. Note that the buffer solution was saturated with argon prior to use. We observe the starting alphabody 586D protein with 1 oxidation (+16 Da) 17395 Da. Additionally we observe the DMEQ-TAD modified alphabody 586D also with 1 oxidation 17740 Da.

The Intact protein analysis of the 586D **2d** conjugate (30 eq., pH 4) shows as single modification with almost 60 % conjugation (based on peak intensity). MS/MS analysis of the resulting protein conjugate digest (figure 6.29) confirms the localization of the TAD modification on tryptophan.

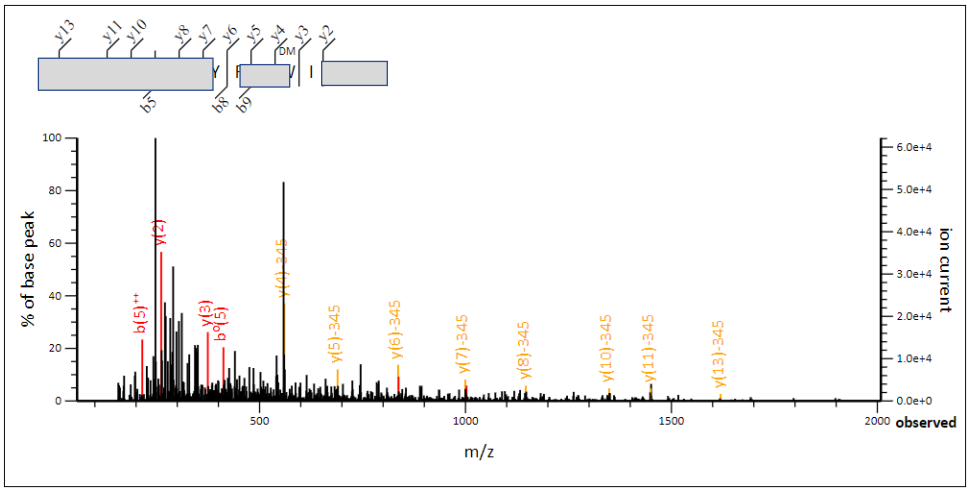


Figure 6.29 HCD MS/MS spectrum of the DMEQ-TAD modified tryptic peptide (XXXXXXXYXXW*XXX). The modification reaction was performed in 10 X PBS pH 4. The "DM" label above the tryptophan residue indicates the location of the DMEQ-TAD. Note that in this figure a fragment ion is marked with "^{0}" this indicates -H_2O. The annotated peaks in yellow contain -345 in their label, this indicates the loss of DMEQ-TAD modification on tryptophan as is expected in HCD MS/MS.

Additionally, the HCD MS/MS analysis demonstrates that the DMEQ-TAD, **2d** modification is lost upon HCD fragmentation. Together, these results indicate that protein TAD modification can be targeted selectively to a Trp side chain, in the presence of tyrosines. As a negative control the valentine alphabody was reacted with 48 eq. of **2b** at pH 4 and no conjugation was observed (experimental figure 9.164). Additionally the alphabody DMEQ-TAD, **2d** conjugate stability was demonstrated by HPLC analysis at 370 nm (DMEQ-TAD absorption, experimental section 9.2.3.11). HPLC peak integration of samples did not show any sign of reduction after 24 hours at room temperature in 10 X PBS pH 7.

Finally, human galectin-7,[327] containing one tryptophan and one tyrosine residue was treated with **2b, 2c** and **2d** at pH 4. The conjugated proteins were observed for all TAD reagents (experimental data section 9.2.3.12). Intact protein analysis of the galectin-7 **2d** conjugate (10 eq. pH 4, figure 6.30) shows over 50 % conjugation.

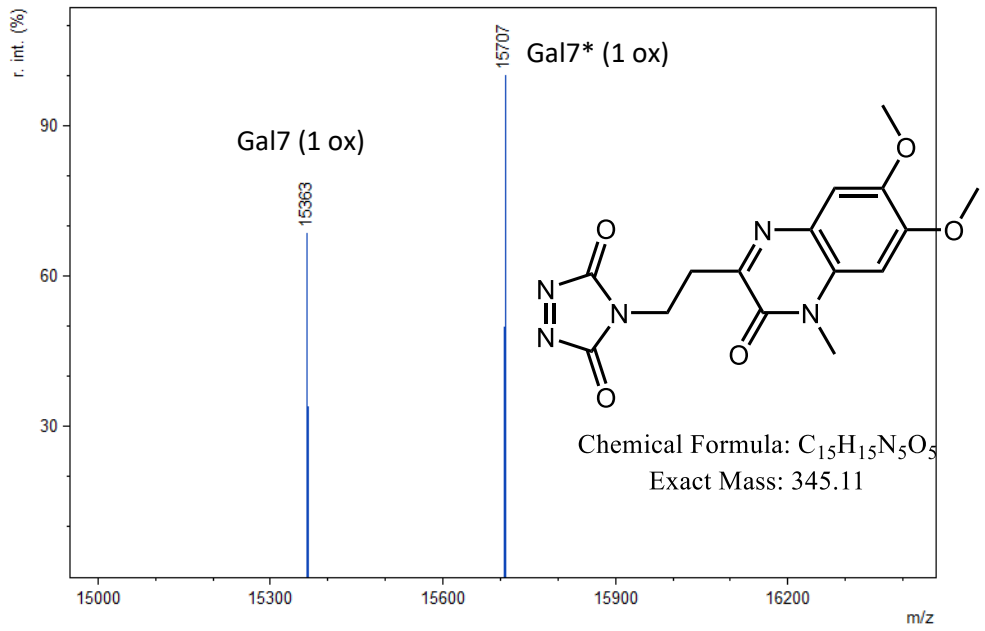


Figure 6.30 Deconvoluted mass spectrum of the galectin-7 protein after addition of DMEQ-TAD **2d** (10 eq.) in 10 X PBS pH 4. We observe the starting galectin protein with an oxidation (+16 Da, *vide supra* methionine presence) at 15363 Da. Additionally, we observe the DMEQ-TAD modified galectin-7 with an oxidation at 15707 Da.

Analyses of the galectin-7 TAD propanol conjugate digests confirm the localization of the TAD moiety on tryptophan. Furthermore, the MS/MS analyses of conjugation experiments with **2b** (20 eq.; pH 4 and 7) demonstrate that at pH 4 the tryptophan has almost exclusively reacted with TAD while at pH 7 both the tyrosine and the tryptophan had reacted (experimental figures: 9.180, 9.181 and 9.182). Additionally, in accord with the findings on the peptide level we found the TAD modification on tryptophan to be labile under the HCD MS/MS conditions used in these experiments.

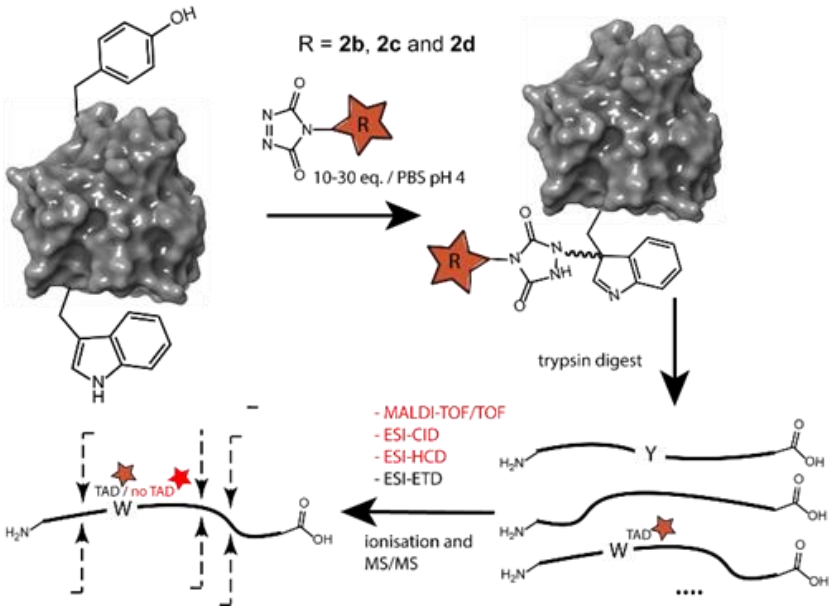


Figure 6.31 Schematic representation of protein conjugation reaction of alphabody 586D and galectin-7 with **2b**, **2c** and **2d**, subsequent trypsin digest followed by different ionization and MS/MS fragmentation of the modified tryptic fragment. Depending on the applied MS/MS technique the TAD modification on W can be observed or not.

6.3.2.7 Conclusion

We report in this section that competitive tryptophan labeling is liable to have so far been systematically over-looked in the current use of triazolinedione (TAD) chemistry for putative tyrosine-selective protein conjugation, a technique that is growing in popularity. The reversibility of the TAD-tryptophan in MS/MS analysis, in combination with the low abundance and low accessibility of tryptophan side chains likely caused this off-target effect to have remained under the radar. We have found that an exposed tryptophan is in fact kinetically favoured over tyrosine in certain conditions. Lowering the buffer pH further enhanced the selectivity resulting in a transition metal free, buffer-compatible selective labeling method for tryptophan. Thus, in addition to a better understanding of the factors that govern the click-like TAD-based protein conjugation, its scope has been expanded for peptide and protein substrates that can withstand buffered conditions at pH 4.

7. General conclusions

In this work we investigated the furan and triazolinedione (TAD) chemical platforms for development of new site-selective protein modification methods. Site-selective protein modification is required for a wide array of applications with the objective to study, adapt and control protein function. Over the years, a substantial toolbox for the site-selective modification of proteins has been developed; however, the drive to come up with better and more selective solutions remains present. The solid foundations upon which this work was built, are provided by the furan and TAD chemical platforms, which were already studied in several different settings in previous work. An important difference in how both methodologies were applied, is situated in the location of the furan and TAD moiety. While the furan moiety is located on the protein, the TAD is found on the modifier side. This difference can be traced back to their respective chemical reactivities. Furan can be regarded as a caged electrophile: meaning that furan is stable in a biological context but becomes very reactive upon oxidation. This offers the possibility to incorporate a stable moiety into a protein and allows for the creation of a highly reactive hotspot upon oxidation. The modifying reagent can be anything from a small fluorescent label to an entire other protein. On the other hand, TAD are reactive compounds which are typically short lived in biological media, but able to react fast with several chemical groups which are naturally present on proteins. This renders TAD moieties less favourable to be incorporated in proteins, since they would likely be degrading or causing side reactions before being useful for protein modification. However, TAD-based reagents are well suited as potent protein modifying compounds.

7.1 Furan chemical platform

7.1.1 Furan incorporation through genetic code expansion

7.1.1.1 Towards protein labeling

Objective I.1 as defined in chapter three of this thesis, is focussed on the development of a furan based protein labeling approach. Furan labeling was previously reported as a potent method for labeling of peptides with hydrazide functionalized dyes. Our aim was to extend to scope of this methodology to a new and more challenging class of substrates: proteins. Introduction of a furan-modified lysine (FurLys) ncAA via genetic code expansion approach proceeded efficiently as was reported in literature. However in our experiments we found that the carbamate linker in FurLys was not stable during the oxidation step leading to degradation of FurLys to lysine (figure 7.1 left). Additionally, when the oxidation procedure was applied to non furan-modified proteins substantial oxidative damage was observed.

Nevertheless, the FurLys degradation can be avoided by using the more stable FurAla ncAA, and optimised oxidation procedures could be explored to circumvent oxidative damage during furan oxidation.

Future perspective:

We believe the added value of furan oxidation-based labelling to the already available toolbox for protein labelling, would be limited since various ncAA's are available that contain bio-orthogonal reactive moieties (e.g. the CuAAC chemistry that was used in section 8.2 of this work, figure 7.1 right). We therefore believe that the future of furan chemistry in the protein context is rather situated in crosslinking applications.

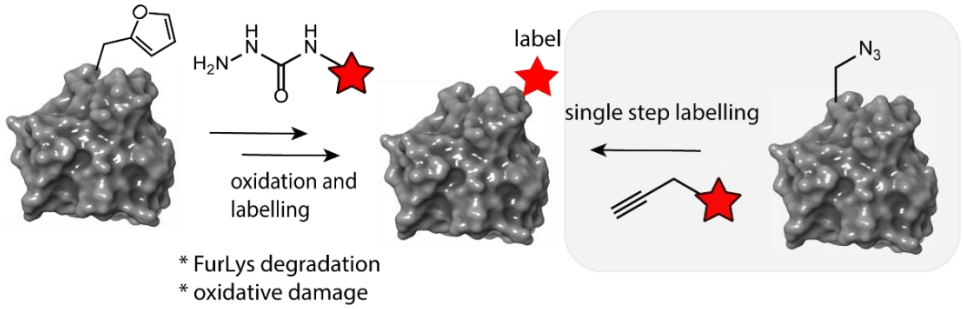


Figure 7.1 Schematic representation of the two-step furan labeling labelling approach (left). Example of a typical single step bio-orthogonal reaction, CuAAC chemistry (right).

7.1.1.2 Towards protein crosslinking

The caged reactivity of furan lends itself particularly well for crosslinking applications. Indeed, a proximate nucleophile, located on the interaction partner of a furan modified POI, is able to react quickly when the furan moiety is oxidized. This allows for the formation of a covalent crosslink. In this work furan, chemistry was used in two different nanobody-target interaction model systems, the gelsolin nanobody model and the EgA1 nanobody model as described in objective I.2. Both of these model systems were selected in view of their therapeutically relevant properties.

The interaction between the gelsolin nanobody (FAFNb1) and an 8 kDa gelsolin fragment was investigated as a first test case. The objective was to develop a furan crosslinking variant of the FAFNb1 nanobody able to crosslink to the 8 kDa gelsolin fragment and thus prevent the further proteolysis of that fragment (which is believed to be at the origin of the disease-related effects). The first and most challenging obstacle in this project was the absence of a crystal structure of the nanobody target interaction. Using ELISA experiments with synthetic peptide fragments we were able to further narrow down the location of the nanobody binding epitope more towards the C-terminus of the 8 kDa peptide (figure 7.2).

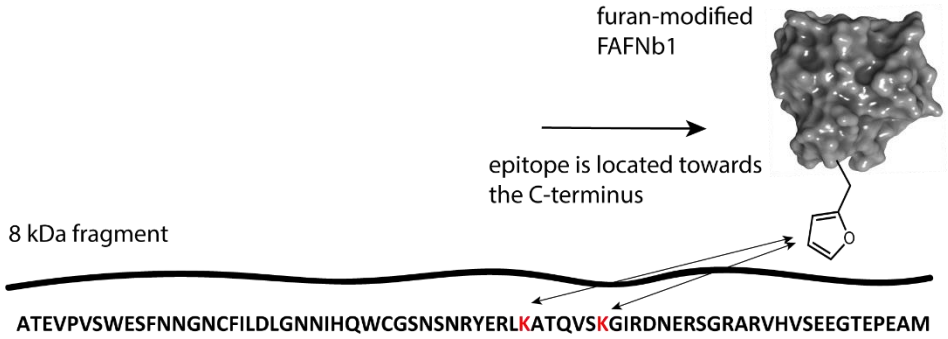


Figure 7.2 Schematic representation of the location of the furan-modified FAFNb1 with respect to the 8kDa fragment.

However, the confirmation of this C-terminal location of the epitope unfortunately also indicates that the lysines that can be targeted for crosslinking on the 8 kDa fragment are most likely too far away from the furan modified FAFNb1. Additionally, we observed peptide aggregation in the purified peptide samples. Although we did not anticipate this aggregation, it is not entirely surprising since proteolysis of the 8 kDa fragment leads to pathogenic aggregation, ultimately resulting in plaque formation. The combination of these phenomena and insights led to the conclusion that it is important to start with a model system where a crystal structure is available and where a lysine residue is located, at a suitable (=crosslinking) distance of the furan moiety, on the nanobody target protein.

An example of such a nanobody-target interaction with a known crystal structure and a proximate lysine residue located on the target protein is the EgA1-EGFR interaction. This interaction was identified after a thorough literature study for suitable nanobody protein interactions on the protein databank. Five EgA1 nanobodies were constructed, in four of which an amber codon was inserted on a different predefined position. The location of the amber codon is important for the final position of the furan amino acid and these positions were selected based on the crystal structure.

The nanobody variants were expressed with a furan-modified lysine derivative incorporated on the amber position. Subsequent *in vitro* crosslinking experiments in presence of the sEGFR target protein demonstrated promising crosslinking results as derived from the SDS-PAGE analysis (figure 7.3).

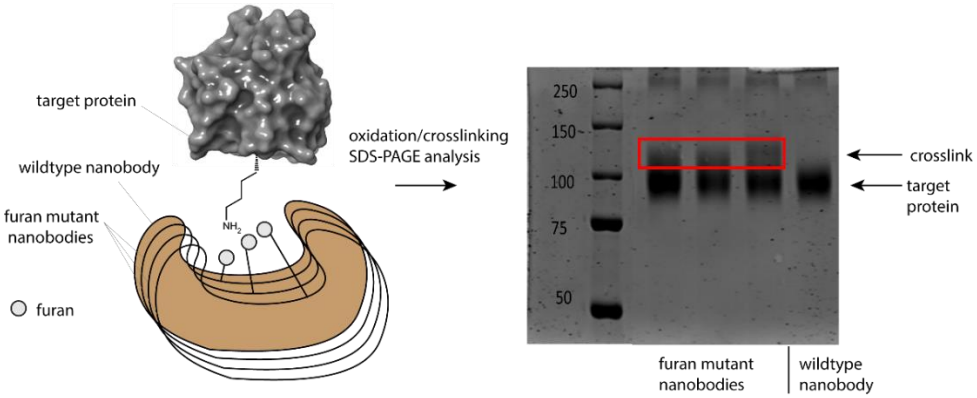


Figure 7.3 Scheme of a nanobody crosslinking experiment. The furan containing nanobody mutants in combination with a target protein containing a lysine (left). *In vitro* oxidation leads to crosslink formation which is clearly visible upon SDS-PAGE analysis as a band slightly higher than the target protein. In the case of the wildtype nanobody (not containing a furan) this extra crosslink band is absent. *These final proof-of-concept crosslink experiments were performed by Laure Tack*

In the SDS-PAGE analyses, an additional band, which is located slightly higher than the EGFR band, is observed for the samples containing furan-modified nanobodies. On the other hand, the sample where the wildtype EgA1 nanobody was incubated with EGFR does not show this additional crosslink band. Although the FurLys ncAA was found to be unstable in the oxidation step in the labelling experiments, it appears that crosslink formation with a proximate nucleophile can occur rather than degradation of FurLys.

Future perspective:

In combination with previous results obtained by our group on the endogenous oxidation of furan containing peptides followed by crosslinking to their cell surface receptor,[328] these initial results form a solid proof of concept for future *in cellulo* experiments with such furan crosslinking nanobodies. This study is currently carried out by Laure Tack in the context of the PhD research project she started in November 2020.

7.1.2 Furan incorporation via genetic code reprogramming

In order to investigate the reactivity of natural nucleophiles towards the oxidized furan moiety, as described in objective I.3, a small library of peptides was designed. The peptides were produced using genetic code reprogramming via flexile *in vitro* translation (FIT). To mimic the proximity found in two binding proteins, each peptide contains a furan moiety and a natural nucleophile. The aim is to oxidize the furan moiety and analyse the sample for potential macrocycle formation (figure 7.4).

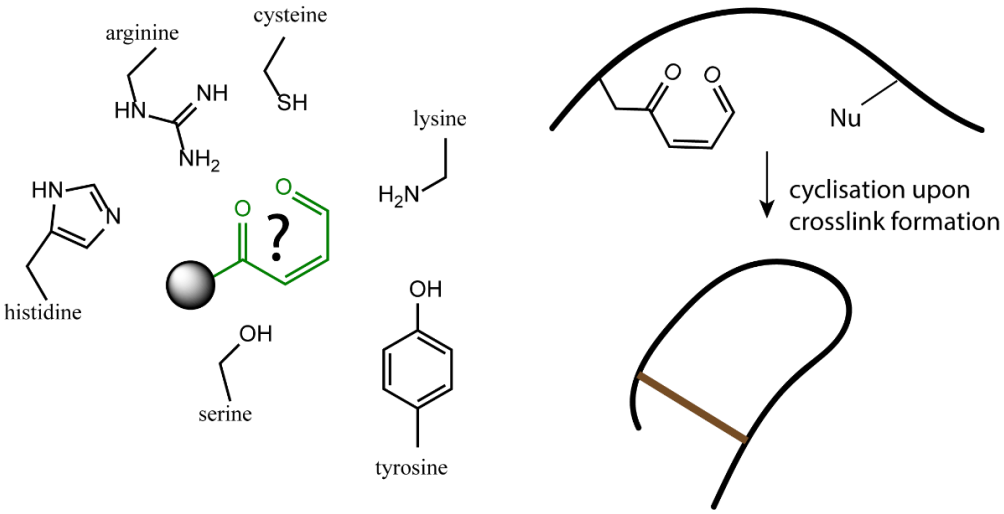


Figure 7.4 Overview of nucleophiles that were tested for their ability to react with the oxidized furan moiety (left). Schematic of the intramolecular crosslink formation between an oxidized furan moiety and a nucleophile that leads to a cyclic product (right).

Based on literature reports lysine was anticipated to be a suitable nucleophile. With the results of these experiments we confirm that lysine is able to form a macrocyclic product, indicating that lysine is indeed a suitable nucleophile in the furan crosslinking approach. Although FIT is very powerful to generate enormously diverse libraries of ncAA containing peptides, the amount of each individual peptide is typically very low. Therefore SPPS synthesis was used to generate more peptide material for LC-MS and structural analysis. Upon reduction of the formed imine, we found that a pyrrole moiety is formed as the connecting motif. Besides the information that lysine is a suitable nucleophile in the furan crosslinking approach, we reported on this novel side-chain to side-chain peptide cyclisation method.

The results in this section are in line with our data on the EgA1-EGFR crosslinking experiments, where nanobody variants were designed *in silico* to locate the furan amino acid in close proximity to a proximate lysine amino acid.

Future perspective:

We hypothesize that the method could also be used in a head to side-chain fashion since the N-terminus is typically more nucleophilic in comparison to the amine of a lysine. Although no cyclic product was observed with other nucleophiles, more experiments are required to confirm that other nucleophiles are not able to react with the oxidized furan.

7.2 TAD chemical platform

7.2.1 Targeting the native prenylgroup

Objective II.1 concerns the development of TAD reagents for selective detection of the native prenylgroup. The TAD-tyrosine click-like reaction was reported in 2010 and has since been widely adopted as a potent tyrosine modification method. Therefore, one of the main goals for this project was identifying conditions that allow selective reaction of TAD with farnesyl in presence of tyrosine. Based on TAD modification experiments with a tyrosine containing peptide in different buffers, we anticipated that it would be possible to attain the same farnesyl selectivity over tyrosine in buffered conditions below pH 5. We were able to use TAD to selectively target a farnesylated peptide in a competition experiment with a tyrosine containing peptide in pure water. Additionally, we have shown that all three double bonds in the farnesyl group can react with TAD leading to a triply modified peptide. The ability to attach multiple label moieties onto a single prenylgroup could increase the sensitivity of a TAD-based detection method. Although the TAD-tryptophan reaction was reported to be sluggish in buffer, we surprisingly observed competitive reaction of tryptophan with TAD reagents (figure 7.5).

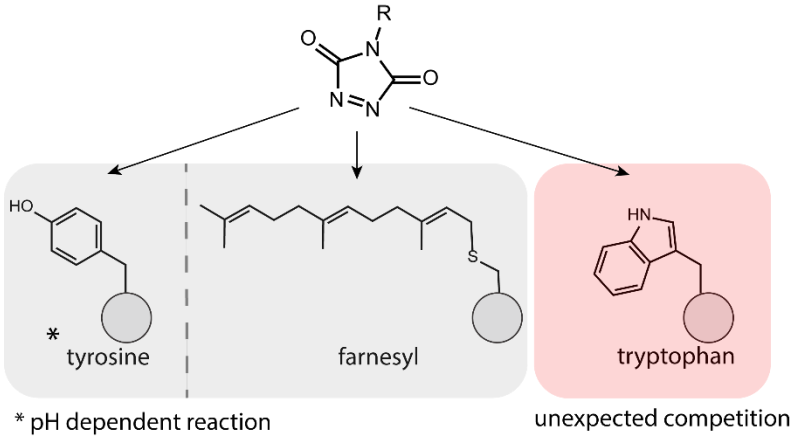


Figure 7.5 Schematic representation of the natural substrates, present on proteins that compete for reaction with TAD-reagents.

Because of this off-target tryptophan effect, it was not be possible to attain complete selectivity of TAD reagents over all 20 canonical amino acids. However, we believe that TAD reagents can be useful for the scientific community as detection tools for prenylated proteins. Moreover, sequence analysis of proteins from the Ras family, which are among the most relevant prenylated proteins from an oncology perspective, revealed that none of these proteins have a tryptophan residue in their AA sequence.

The interest in labelling of prenylated proteins is further evidenced by a very recent publication where the use of TAD reagents for labelling of prenylated proteins is reported.[306] Remarkably, the authors are aware of the known TAD-tyrosine reaction but do not report on the selectivity of tyrosine versus farnesyl of their TAD reagent. Moreover the authors do not realize that they in fact also face competition of tryptophan (*vide supra*).

Future perspective:

We believe that TAD reagents can still be useful for detection purposes of native farnesyl groups. By using buffered conditions at a pH below 5, we can avoid tyrosine competition and substantially improve labelling selectivity towards farnesyl.

7.2.2 Triazolinedione protein modification: from an overlooked off-target effect to a tryptophan selective bioconjugation strategy

Triazolinediones (TAD's) have been used for protein modification since the report of the TAD-tyrosine click like reaction in 2010 by Barbas and co-workers.[199] The authors report that TAD's react selectively with tyrosine in an ene-type reaction. Later the reaction mechanism for the TAD-tyrosine was studied via computational methods leading to the conclusion that the reaction actually proceeds along an electrophilic aromatic substation pathway.[329] Moreover, the authors demonstrate that the phenolate rather than the phenol is the reactive species for reaction with TAD. This improved understanding of the TAD-tyrosine mechanism is in agreement with the experimental observations that the reaction proceeds very efficiently in buffered conditions at pH 7. The same conclusion can be drawn from the pH experiments we performed with a tyrosine containing peptide and TAD reagents in section 6.2.3 of this work. However, in a simple competition experiment between tyrosine, tryptophan and farnesyl for TAD in water, we found that the TAD-tryptophan reaction was in fact anything but sluggish. At this point two important research questions remained. The first question is: how could this important off-target effect have been overlooked for over a decade by so many scientists. And the second question is: if the tyrosine-TAD reaction is inhibited at low pH, can a genuine amino acid selective method be developed for TAD-tryptophan modification.

7.2.2.1 How was the off-target effect of TAD-reagents on tryptophan overlooked for so long

In their report on the TAD-tyrosine click like reaction, Barbas and co-workers conducted a very poor amino acid selectivity study. The authors actually report on formation of a tryptophan TAD conjugate in a separate LC-MS experiment with TAD and N-acyltryptophan methylamide. However, it was subsequently determined and claimed that the TAD-tryptophan reaction did not interfere with the TAD-tyrosine reaction in a competitive experiment only analysed via NMR. After the initial report, many applications and refinements of the TAD-tyrosine reaction followed. Since the original paper described the TAD-Tyr approach as an amino acid selective method for tyrosine labelling, the amino acid selectivity was not interrogated in detail in later works. There are two additional factors that contributed to the fact that this off-target effect could remain unnoticed. The first is that the TAD modification on tryptophan is labile under most MS/MS conditions. Indeed, in our experiments both on peptides as well as on proteins we found that MS/MS analysis via MALDI-TOF/TOF, ESI-CID and ESI-HCD all lead to the loss of the modification on tryptophan during fragmentation (figure 8.8). We could however demonstrate that by using the softer electron transfer dissociation (ETD) fragmentation the modification on tryptophan is preserved, thus confirming unambiguously the off-target reaction of TAD-reagents with Trp residues.

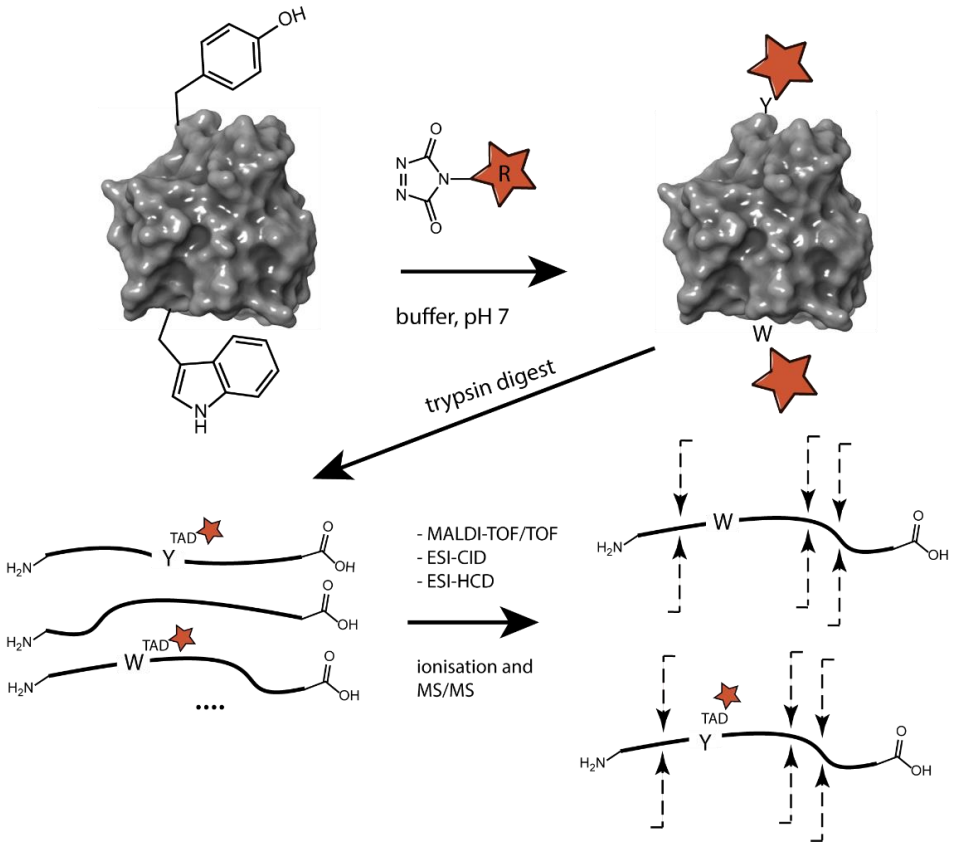


Figure 7.6 Schematic overview of a TAD protein modification reaction at pH 7. Both tryptophan and tyrosine residues are modified. The protein is digested using trypsin to determine the exact modification site. Subsequently, the peptides are analysed using MS/MS and while the TAD-tyrosine modification is stable during MS/MS fragmentation, the TAD tryptophan modification is not, leading to a perceived amino acid selectivity for tyrosine.

The second factor that contributed to the fact that the off-target effect remained under the radar is the low abundance of tryptophan. With an abundance of just over 1 % it is the most rare of all 20 canonical amino acids.[48]

Only during MS-analysis of the resulting conjugates, the TAD moiety is removed, but the original protein does indeed bear that modification. In this way, the off-target effect can lead to false interpretations and conclusions whenever the TAD-tyrosine modification is used in applications ranging from the generation of protein conjugates for construction of ADCs[321] and vaccine research[298] all the way through to fundamental work on mapping the structure of proteins.[324][330]

Nevertheless, the MS/MS lability of the conjugate and the low abundance of tryptophan may have helped in obscuring this off-target effect. In reality, the fact is that the TAD-tyrosine click reaction is not amino acid selective and the tryptophan modification should have been observed immediately. In our opinion, more attention should be focussed on rigorous investigation of the amino acid selectivity when a novel method is reported for protein modification.

7.2.2.2 Development of a genuine TAD-based tryptophan-selective modification method

By tuning the pH value below 5, we were able to use TAD reagents to selectively modify the tryptophan amino acid in presence of tyrosine. Different TAD reagents were used in this study. This new tryptophan selective modification strategy was demonstrated on several peptides and recombinant proteins (figure 8.9).

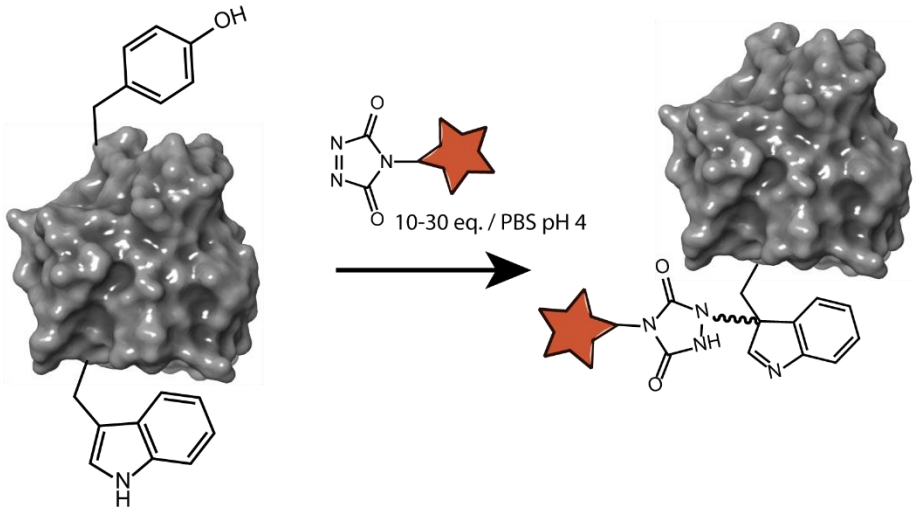


Figure 7.7 Schematic overview of selective tryptophan modification with TAD chemistry in PBS buffer at pH 4.

This novel approach for tryptophan modification compares favourably with current methods for selective tryptophan modification. Francis and co-workers reported a method for selective tryptophan modification in peptides and proteins.[94] The method proceeds at pH 6-7 and generally yields 40-60 % using 100 eq. of a label molecule. The disadvantage of this approach is that a transition metal is required (Rhodium) and that the procedure requires long reaction times. Kanai and co-workers reported a transition metal free tryptophan conjugation method using organoradicals to selectively conjugate to tryptophan residues in peptides and proteins.[95] However, the reaction proceeds under acidic conditions (acetic acid in water, pH 1-2) and the reaction is terminated by addition of buffer. The authors have shown protein modifications with functionalized compounds with a yield of around 50 %. Although the reaction occurs in a transition metals-free fashion, it needs to be run in water at pH 1-2 and is not buffer compatible. Recently, Taylor and co-workers reported a photochemical tryptophan modification process. The authors show modifications of several peptides and lysozyme. The downside of this approach is that it requires the use of UV light, the addition of glutathione and needs to be performed in absence of oxygen. Additionally, the authors show modification of short peptides with functionalized (biotin, alkyne) N-carbamoylpyridinium salts. However, crucially the modification of longer peptides and a protein (lysozyme) was only reported with a non-functional N-carbamoylpyridinium salt. Our TAD-based approach for selective tryptophan modification can selectively modify tryptophan in proteins under buffered conditions at pH 4. We have shown over 50 % modification using 20-30 equivalents

of fluorescent TAD reagent. Our approach does not require the use of transition metals or other additives; it does not require absence of oxygen or UV irradiation. The approach we present takes place in a matter of seconds in buffered medium. Additionally, the fluorescently modified TAD (DMEQ-TAD) is readily commercially available.

Future perspective:

With these data, we intend to set the record of TAD based protein modification straight. Additionally we hope that this work can help to prevent similar mistakes in the future by placing more weight on the rigorous characterisation of amino acid selectivity.

We fully realize that our new site-selective approach for tryptophan modification will not be applicable to all proteins, due to the fact that mildly acid conditions (buffered pH 4) are required. However for proteins that can handle a short stay in buffered conditions at pH 4 our method is by far the most biocompatible procedure for Trp-modification. We hope this method will be applied for a range of protein modification applications.

8. Additional collaborative projects

8.1 Cysteine conjugation

The contents of this section were published in: Reviving old protecting group chemistry for site-selective peptide–protein conjugation, Smita B. Gunnoo, Abhishek Iyer, Willem Vannecke, Klaas W. Decoene, Tim Hebbrecht, Jan Gettemans, Mathias Laga, Stefan Loverix, Ignace Lasters and Annemieke Madder, *Chemical Communications*, **2018**, 54, 11929

My contribution to this project encompassed nanobody expression and purification in collaboration with Dr. Tim Hebbrecht. Nanobody peptide conjugation reactions, structural characterisation studies using circular dichroism (CD) and ELISA experiments were carried out in collaboration with Dr. Smita Gunnoo.

A new method was developed for conjugation of unprotected peptides to proteins bearing a solvent exposed cysteine. Old cysteine protecting group strategy is at the centre of this new approach. A cysteine protected with an acetamidomethyl (Acm) group can be converted on-resin to cysteine protected with *S*-carbomethoxysulfenyl (Scm) group. This protecting group conversion allows for the subsequent conjugation to proteins containing accessible cysteine amino acids via disulphide formation. In this work we demonstrated site-selective ligation to three different proteins with a free cysteine. The formed conjugate comprises of a crossed-disulfide between peptides and proteins. The methodology takes advantage of the facile introduction of the Acm group via SPPS. Moreover the Acm group is stable to a large range of conditions which may be required for on-resin peptide functionalization steps. Since the peptide functionalisation takes place on-resin, the reactions involved can be driven to completion, with swift removal of excess reagents via filtration. Another important feature if this methodology is that the final conjugation step can be carried out at ambient conditions (Tris buffer at pH 7.4 and room temperature), dimethylsulfoxide may be added to aid in the solubilisation of the (decorated)peptide. This versatile new approach should be applicable to many systems since a solvent accessible cysteine can easily be introduced into a protein of interest via site directed mutagenesis.

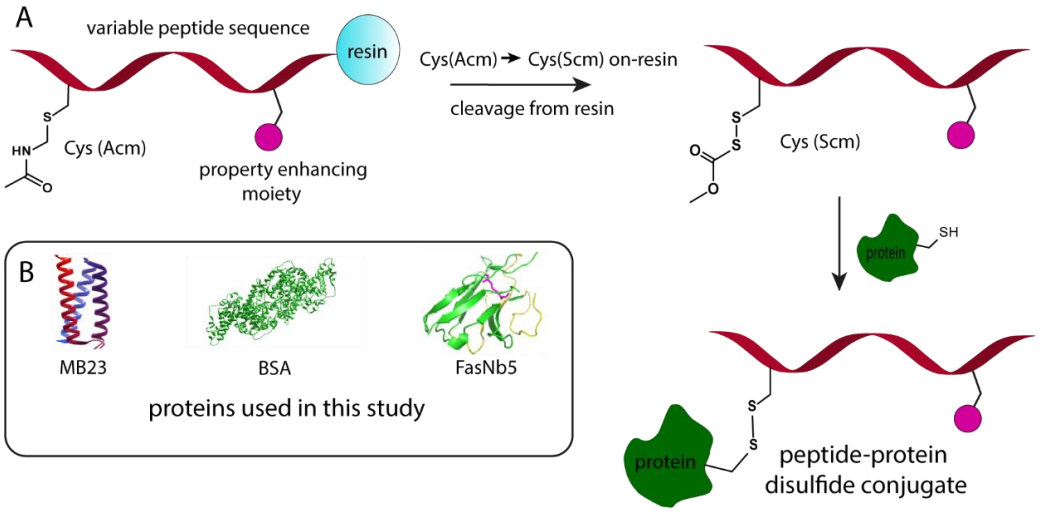


Figure 8.1 (A) Cys(Acm)-containing peptides can be reacted with proteins with a free Cys following on-resin Acm to Scm conversion, (B) proteins employed in this study: MB23 alphabody, BSA and FasNb5 nanobody.

8.1.1 Structural characterisation via circular dichroism (CD)

The secondary structure of the nanobody, alphabody and their conjugates were analysed via CD measurements. The proteins were dialysed into a 20mM phosphate buffer (pH 7) and diluted to a concentration of 0.2 mg/mL. The spectra were recorded on a Jasco (J-710) using a scanning speed of 100 nm/min and 2 s response time. The results were expressed as molar ellipticity (deg.cm2/dmol) figure 8.2. These results indicate that the overall structure is maintained after conjugation.

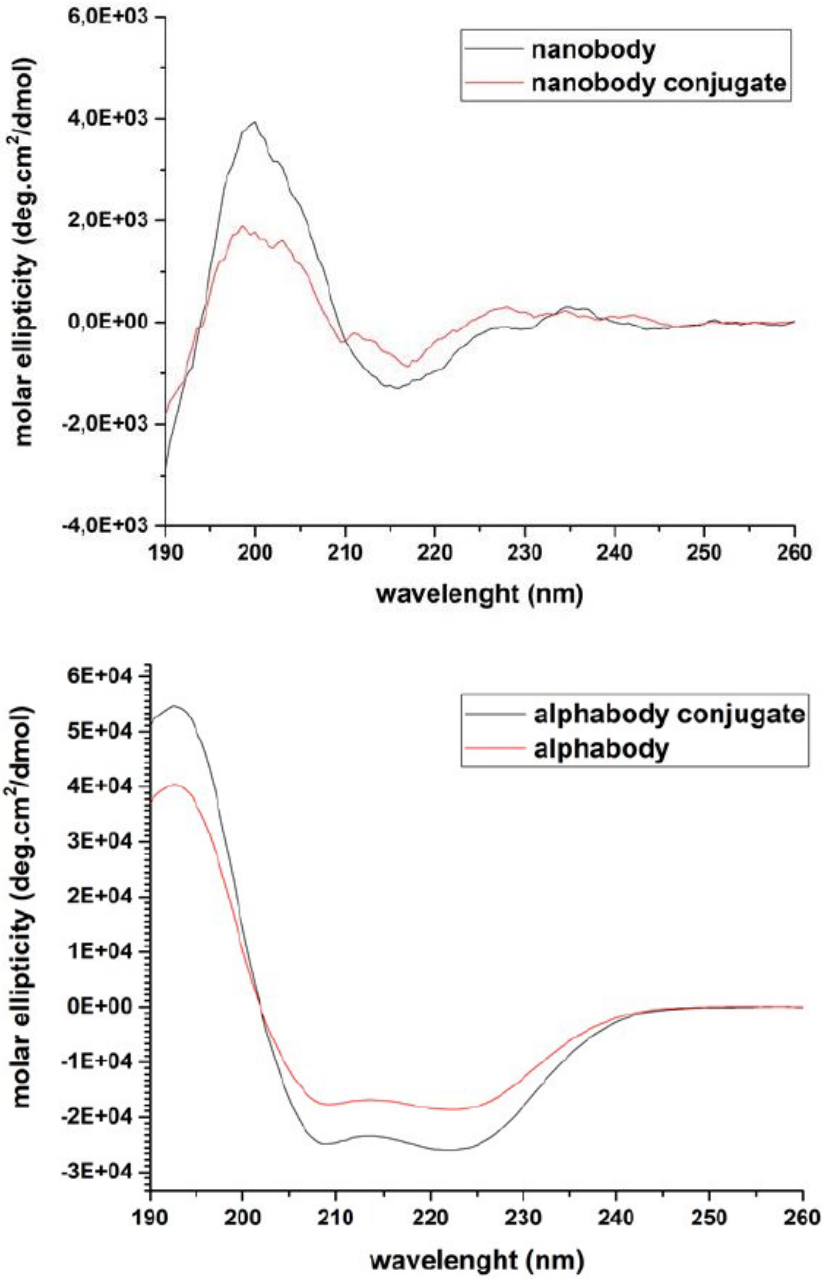


Figure 8.2. Top panel. Graph showing molar ellipticity (deg.cm2/dmol) vs wavelength for the nanobody and nanobody conjugate. Bottom panel. Graph showing molar ellipticity (deg.cm2/dmol) vs wavelength for the alphabody conjugate and the alphabody.

8.1.2 Binding studies via ELISA

To demonstrate that the conjugation reaction did not affect the binding properties of the FasNb5 nanobody an ELISA experiment was performed. The nanobody antigen, fascin was immobilized on a 96 well plate, as a negative control cortactin was used. After immobilisation, blocking and washing steps the FasNb5 and FasNb5 conjugate were incubated as an 8 fold dilution series. The FasNb5 nanobody that was used in this study contains a hemagglutinin derived tag (HA) tag. This HA tag was subsequently used for recognition with a rabbit anti-HA antibody. This rabbit anti-HA antibody was subsequently recognized with a anti rabbit HRP secondary antibody allowing for ELISA visualisation. The ELISA results (figure 8.3) demonstrate no significant changes in the nanobody (conjugate) antigen binding affinity after conjugation reaction.

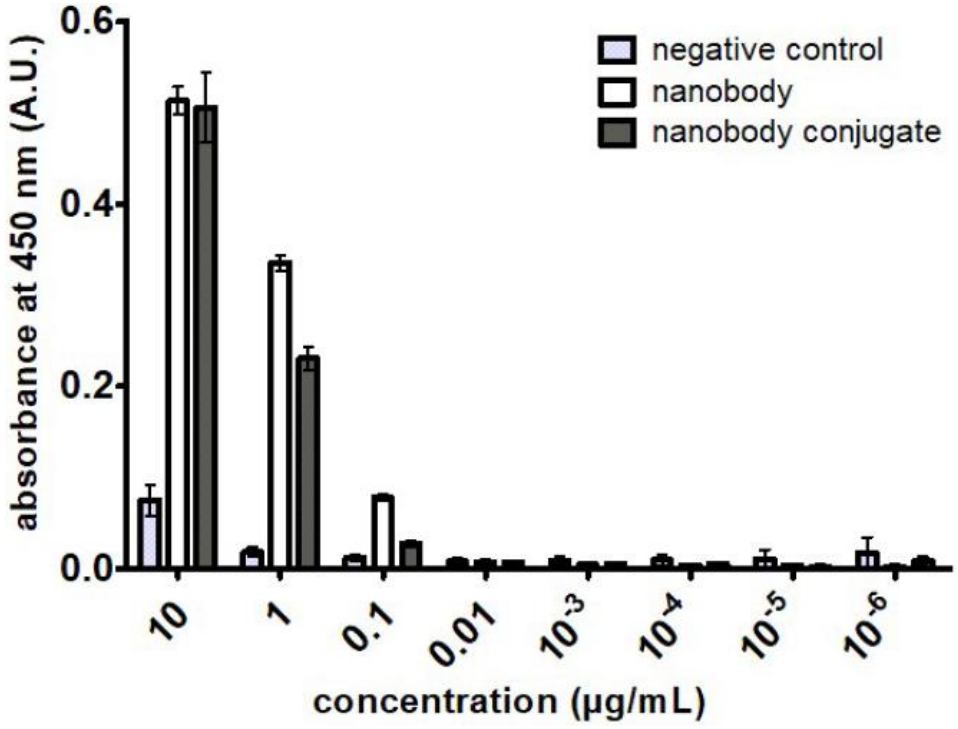


Figure 8.3 Graph showing absorbance at 450 nm (A.U.) vs concentration (µg/mL) for the negative control, nanobody and the nanobody-peptide conjugate.

8.1.3 Conclusion

In conclusion, we have shown that the Scm group can be used as an activating group in cysteine-containing peptides for crossed-disulfide formation with proteins bearing cysteine residues, thus adding a new strategy to the chemical protein modification toolbox. The methodology is facile, efficient and versatile, and takes advantage of solid phase peptide chemistry strategies for the versatile decoration of proteins with free Cys residues. The methodology should be generally applicable to many systems, as evidenced by our use of multiple proteins and peptides. Crucially, Cys can be introduced into proteins by simple site-directed mutagenesis. Structural integrity of the conjugates was demonstrated via comparative CD studies. The binding affinity of the nanobody after conjugation had not significantly changed as shown via ELISA experiments.

8.2 Nanobody click chemistry

The contents of this section were published in: Nanobody click chemistry for convenient site-specific fluorescent labelling, single step immunocytochemistry and delivery into living cells by photoporation and live cell imaging, Tim Hebbrecht, Jing Liu, Olivier Zwaenepoel, Gaëlle Boddin, Chloé Van Leene, Klaas Decoene, Annemieke Madder, Kevin Braeckmans, Jan Gettemans, *New Biotechnology*, **2020**, 59, 33-43

My contribution to this project encompassed the synthesis of pAzF ncAA and optimization of pAzF incorporation in response to an amber codon using genetic code expansion (GCE) during nanobody expression in collaboration with Dr. Tim Hebbrecht.

In this work, we reported the use of nanobody click chemistry in combination with photoporation for delivery into living cells for live cell imaging. A cortactin nanobody: cortactin Nb2 and two β-catenin nanobodies: β-catenin Nb77 and β-catenin Nb 86 were selected for this study. Cortactin is a cytoskeletal protein (*vide* supra) while β-catenin is localized in adherens junctions which are protein complexes that occur at cell–cell junctions in epithelial and endothelial tissues.[331] Two strategies of nanobody CuAAC chemistry were explored to fluorescently label nanobodies. A first strategy that was used involved ligation of synthetic peptide containing an alkyne group to the nanobody using sortase ligation (*vide supra*). A second strategy exploited the incorporation of the ncAA para azido phenylalanine (pAzF) into a nanobody using GCE via amber suppression (*vide supra*). Both strategies result in a nanobody bearing a bioorthogonal moiety suitable for specific reaction with a corresponding CuAAC label. In this work Alexa fluor 488 functionalized with either an azide or alkyne moiety for CuAAC reaction with a nanobody containing an alkyne (via sortase ligation of an alkyne bearing peptide) or azide (via GCE of the pAzF ncAA) respectively. Labelled nanobodies have excellent properties for imaging of target proteins like high specificity, small size and good stability. Unfortunately, nanobodies cannot spontaneously cross the cell membrane, hampering applications in intracellular targets. The site selectively labelled nanobodies were used in immunocytochemistry and cortactin and β-catenin were imaged in cells after fixation and permeation. However, fixation and permeation of cells can have an impact on the native protein structure and accessibility for antibody and nanobody binding. Laser induced photoporation can be used to deliver nanobodies inside living cells. Graphene quantum dots were used as photothermal sensitizers, these nanoparticles are able to interact with the cell membrane. After a short incubation time the fluorescent nanobodies are added to the cells and the cells are irradiated with pulsed laser light. Upon irradiation around the graphene quantum dots vapor nano bubbles (VNB) are formed which create temporary holes in the cell membrane, allowing the fluorescent nanobodies to diffuse into the cells (figure 8.4).[332][333] It was shown that photoporation allowed facilitated visualisation of β-catenin which was not possible in fixed and permeabilized cells.

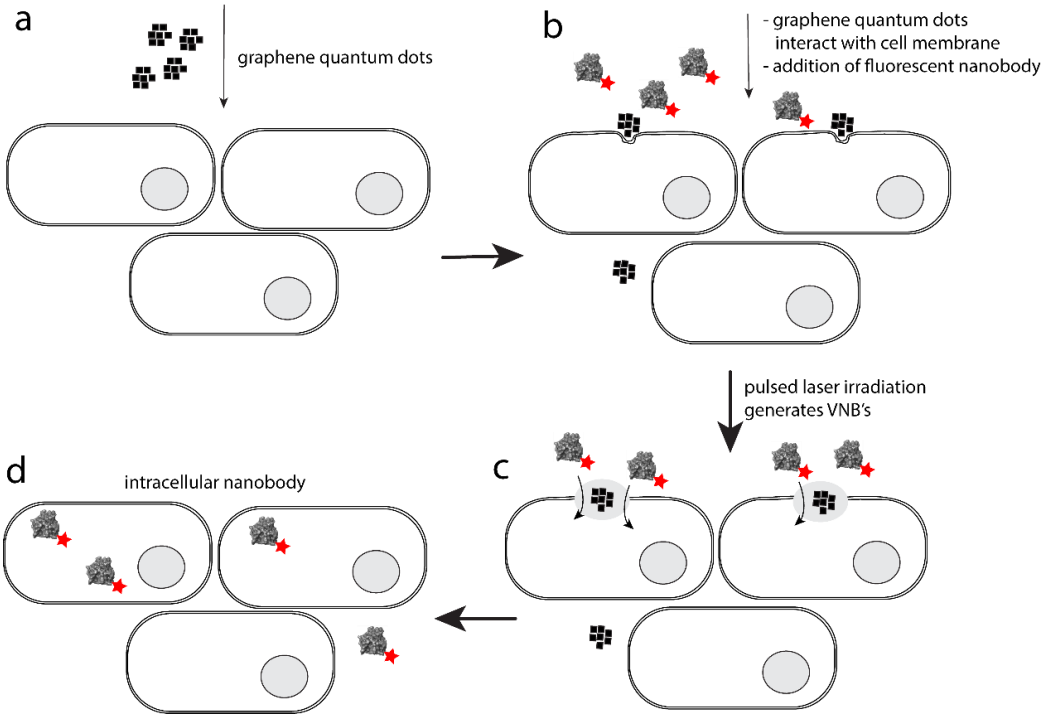
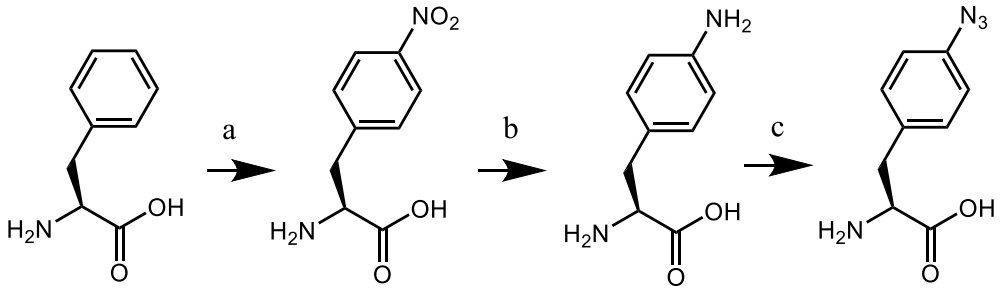


Figure 8.4 Schematic representation of laser induced photoporation for internalisation of fluorescent nanobodies. A: graphene quantum dots which act as photothermal sensitizers are added to the cells. B: the graphene quantum dots are incubated with the cells to allow them to interact with the cell membrane, fluorescent nanobody is added to the cells. C: the cells are irradiated with pulsed laser light, generating vapor nano bubbles (VNB's) allowing the fluorescent nanobody to diffuse into the cell. D: the fluorescent nanobody is present in the cells.

8.2.1 pAzF synthesis

We believed we could prepare a sufficiently large amount of pAzF for the incorporation experiments. Therefore we decided to initially perform the synthesis ourselves instead of buying the commercial pAzF. A slight mass difference with the expected mass was observed after incorporation into proteins due to N2 loss in the LCMS analysis. We observed this first with the synthesized variant so we decided to buy the commercial pAzF but the result was the same.



a: H_2SO_4/ HNO_3 5 hours 4 °C
b: H_2 Pd on $BaSO_4$ overnight
c: i) $NaNO_2$ in HCl ii) NaN_3

Figure 8.5 schematic representation of the pAzF synthesis.

pAzF was synthesized according to the synthetic procedure in figure 7.5.[334] The synthesis is started from phenylalanine and in a first step a nitration is performed in a mixture of sulphuric and nitric acid. In a second step, a hydrogenation reaction is performed to obtain the phenylamine. Lastly a diazonium ion is prepared from the arylamine followed by reaction with sodium azide to yield the final pAzF.

8.2.2 Conclusion

In conclusion, we report that a combination of photoporation with fluorescently labeled nanobodies allows developing a new method for the study of protein behaviour and function in mammalian cells. Two methods for introducing a fluorescent moiety were compared: genetic code expansion versus sortase peptide ligation. The genetic code expansion approach led overall to the most promising results. The procedure is straightforward, reliable, relatively fast and has a higher yield. It may become broadly applicable as it can be used for different nanobodies and coupling to a variety of moieties e.g. different fluorophores, magnetic beads, quantum dots, gold nanoparticles. The use of fluorescently-labelled nanobodies and the new avenue of introducing them into living cells by photoporation will allow study of short and medium range (24 h) dynamic cellular processes targeting a broad range of proteins. Moreover, this technique reduces linkage error and will enable the imaging of endogenous proteins at higher resolution and can be of interest for super resolution microscopy.

9. Experimental data

9.1 General discussion on MS-methods for analysis of modified proteins

In this doctoral dissertation, several methods for MS analysis of modified proteins were used. In this section an overview and critical discussion of these methods is given. An important first distinction to make is between methods for the analysis of intact proteins and methods for the MS/MS analysis of a protein digest. For the detailed characterisation of modified proteins both intact MS-analysis as well as MS/MS analysis of a protein digest are required. Intact protein MS-analysis provides a global picture, which is required to examine the total protein mass as well as the number and mass additions of potential modifications. The MS/MS analysis provides information on the exact location of modifications. Important to note is that the conditions used for the MS-analysis can have an impact on certain modifications. This was clearly demonstrated in section 6.3 where several MS/MS conditions have an important effect on the TAD-tryptophan modification.

9.1.1 Intact protein MS-analysis

In this work, we used LC-MS and MALDI-TOF for the analysis of intact proteins. MALDI-TOF intact protein analysis proved to be challenging for many protein samples. Typically, the spectra were less reproducible and broad peaks were obtained, hampering the analysis of small protein modifications such as oxidations (+16 Da). The experimental conditions such as the choice of the matrix, the type of spotting and instrument settings are crucial and can be fine-tuned for each experiment to obtain the best MALDI-TOF MS spectra. An advantage of MALDI-TOF is that only a very small amount of sample is required. On the other hand, LC-MS analysis was in my hands a more robust method capable of providing more accurate mass results, which allows swift detection of even small modifications. For LC-MS analysis the protein should be in solution, degradation of the protein of interest is a common issue. Provided the ionisation of the modified protein is not altered too much due to the modification (which is mostly the case), the intensity of the modified protein peak versus the unmodified protein peak gives a good estimation of the modification yield. Using intact protein analysis, it is not possible to identify the exact location of a modification.

9.1.2 MS/MS analysis of a protein digest

LC-MS/MS methods using three different fragmentation methods: CID, HCD and ETD were used in this thesis as well as MALDI-TOF/TOF. The impact of the different fragmentation methods on the TAD-tryptophan modification was clearly shown in section 6.3. LC-MS (CID/HCD) and MALDI-TOF/TOF all lead to a loss of the modification with a resulting poor fragmentation of the unmodified peptide precursor. Only ETD fragmentation allowed to preserve this TAD modification on tryptophan. For a confident assignment of the modification location, the amino acid selectivity should be first analysed on a small molecule/amino acid level via LC-MS analysis. Afterwards the selectivity can be tested on peptides using MS/MS methods. As the selectivity is already known, the MS/MS analysis can be used to investigate which MS/MS fragmentation methods preserve the modification on different residues. The combination of this information can be used for the analysis of protein digests.

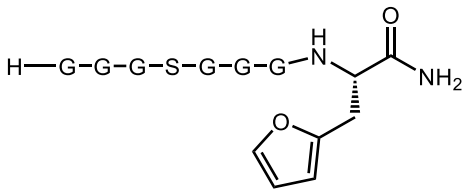
The results obtained in this thesis clearly demonstrate that a combination of intact protein analysis and MS/MS analysis of the digest is required to examine a modified protein sample. If only an intact protein MS-analysis is performed, there is no information about the exact location of the modification. On the other hand, if only MS/MS analysis is performed on a protein digest there is a lack of information on what is taking place at the level of the complete protein.

9.2 Experimental data for chapter 5

9.2.1 Experimental data for part 5.1.2

9.2.1.1 Data for synthesised compounds

Peptide 5.1



Chemical Formula: $C_{22}H_{33}N_9O_{10}$
Exact Mass: 583,24

Figure 9.1 Peptide **5.1** structure.

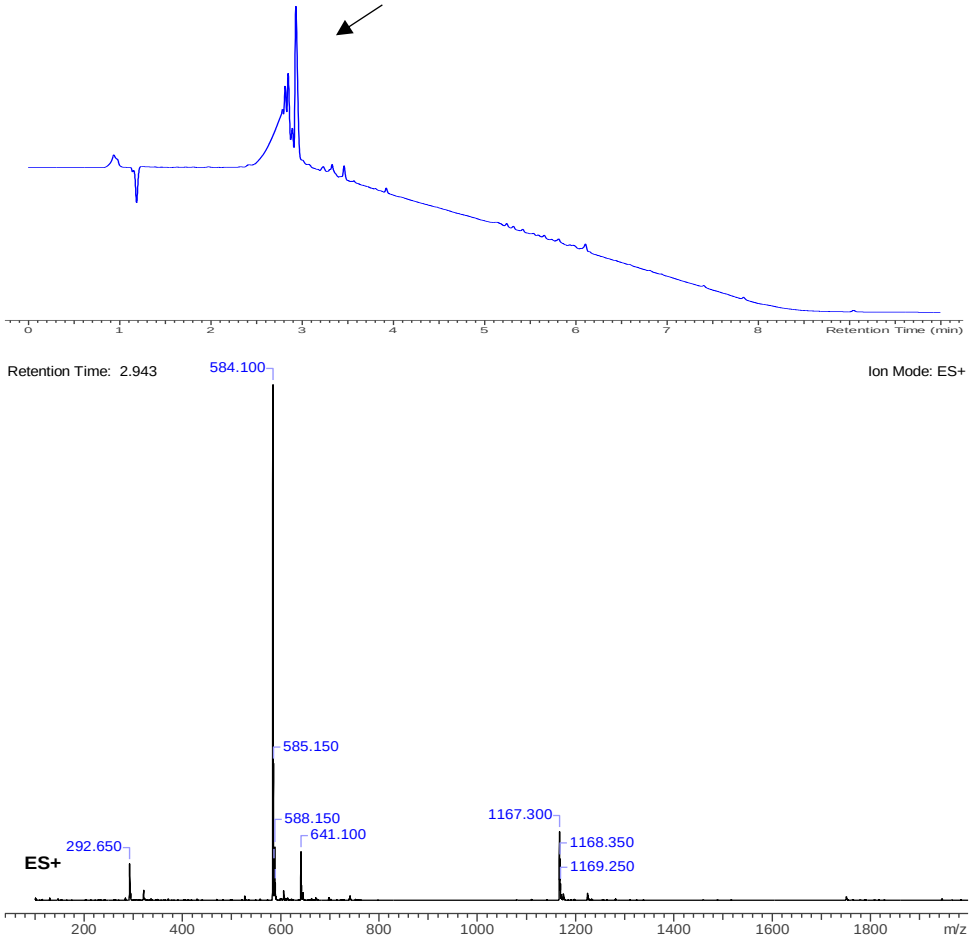
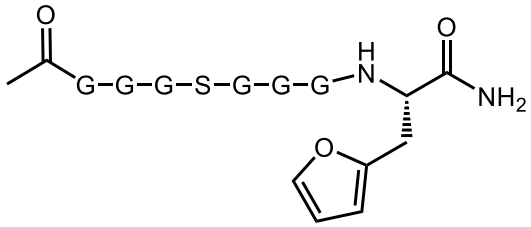


Figure 9.2 Crude LC chromatogram at 214 nm (top) with peptide **5.1** eluting at t_R: 2.943 min. ESI-MS spectrum (bottom) at t_R: 2.943 ($[M+H]^+$ = 584.10).

Peptide 5.2



Chemical Formula: $C_{24}H_{35}N_9O_{11}$
Exact Mass: 625,25

Figure 9.3 Peptide **5.2** structure.

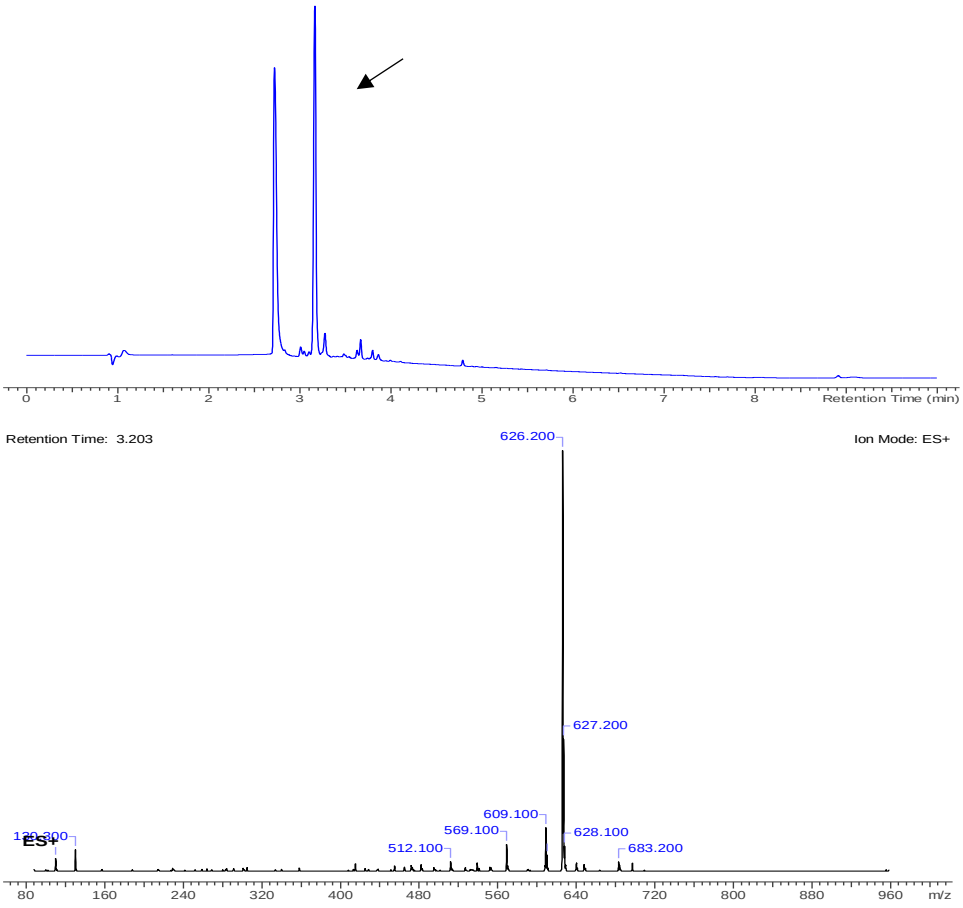
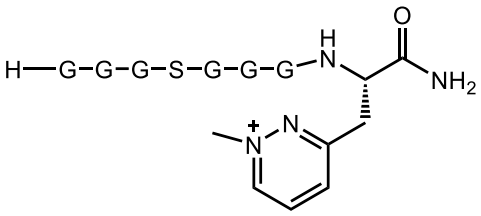


Figure 9.4 Crude LC chromatogram at 214 nm (top) with peptide **5.2** eluting at t_R: 3.203 min. ESI-MS spectrum (bottom) at t_R: 3.203 ($[M+H]^+$ = 626.20).

Peptide 5.1 labeled with metylhydrazine (MeH), photo-oxidation



Chemical Formula: $C_{23}H_{36}N_{11}O_9^+$
Exact Mass: 610.2692

Figure 9.5 Structure of peptide **5.1** labeled with methylhydrazine (MeH).

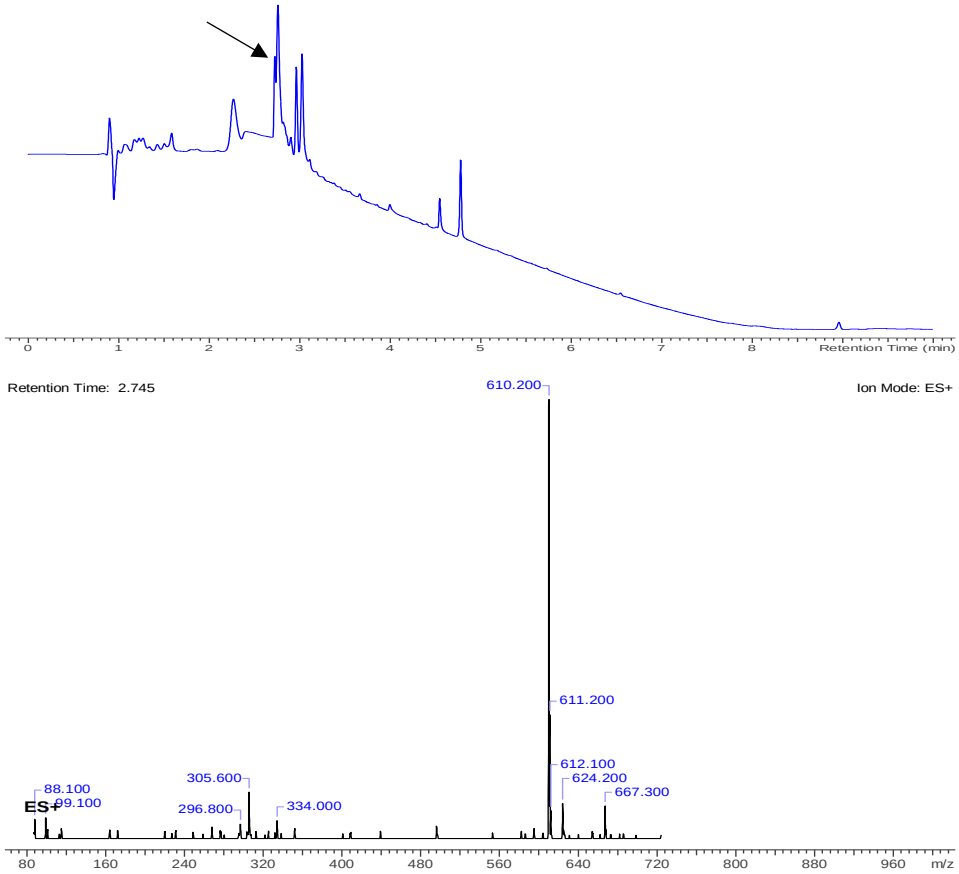


Figure 9.6 Crude LC chromatogram at 214 nm (top) with peptide **5.1** labeled with methylhydrazine (MeH) eluting at t_R: 3.203 min. ESI-MS spectrum (bottom) at t_R: 2.745 ($[M]^+$ = 610.20).

Peptide 5.1 labeled with lucifer yellow (LYCH), photo-oxidation

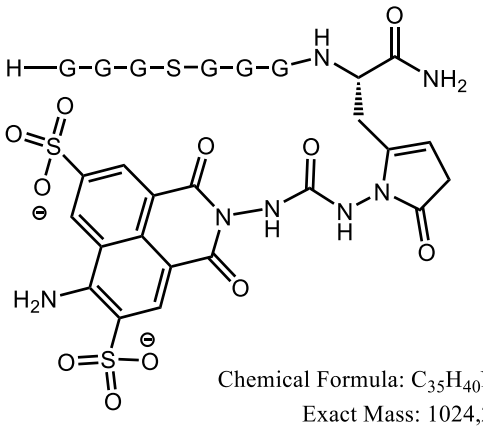


Figure 9.7 Peptide **5.1** labeled with lucifer yellow (LYCH).

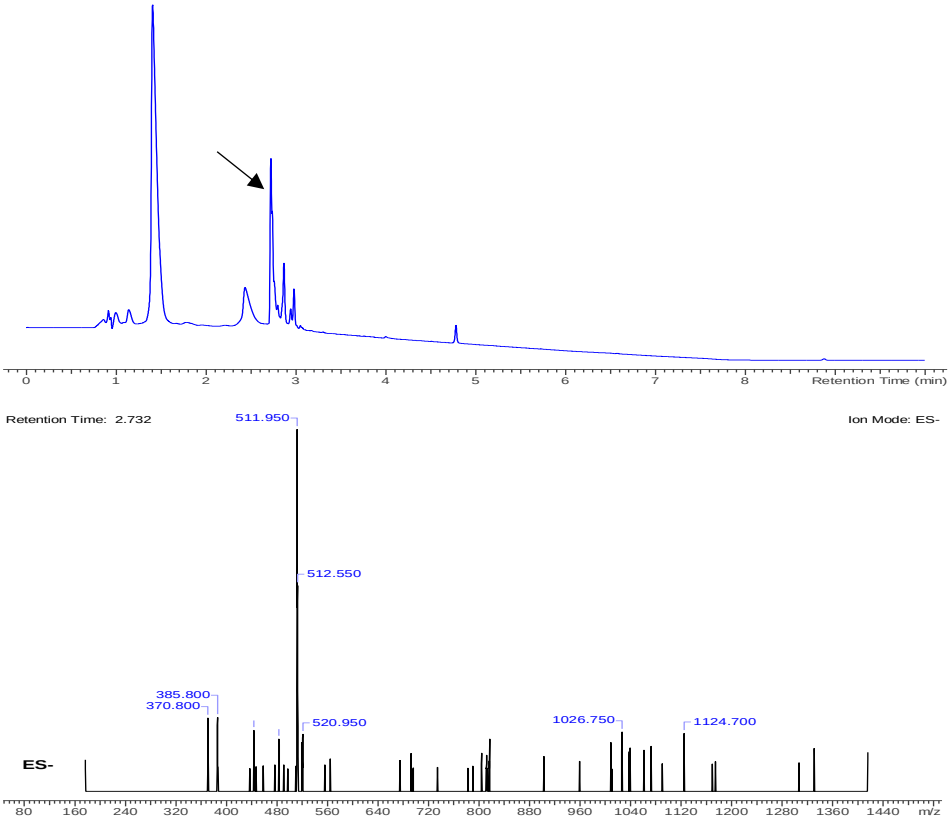


Figure 9.8 Crude LC chromatogram at 214 nm (top) with peptide **5.1** labeled with lucifer yellow (LYCH) eluting at t_R: 2.732 min. ESI-MS (negative mode) spectrum (bottom) at t_R: 2.732 ($[M]^{2-}$ = 511.95).

9.2.1.2 Data nanobodies

FasNb5 amber (failed ncAA incorporation)

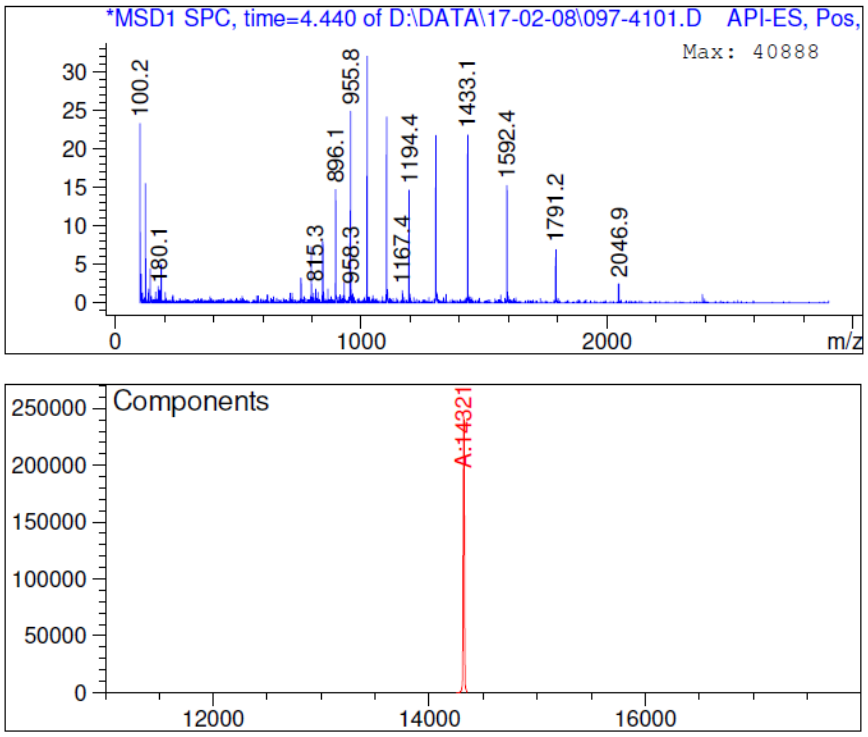


Figure 9.9 LC-MS analysis of FasNb5 amber nanobody expressed for incorporation of FurLys in response to the amber codon. In the expression of this nanobody the tRNA synthetase was not induced causing termination of protein translation at the amber stop codon. Expected mass for the truncated nanobody 14324,01 Da.

FasNb5 amber GG amber (failed ncAA incorporation)

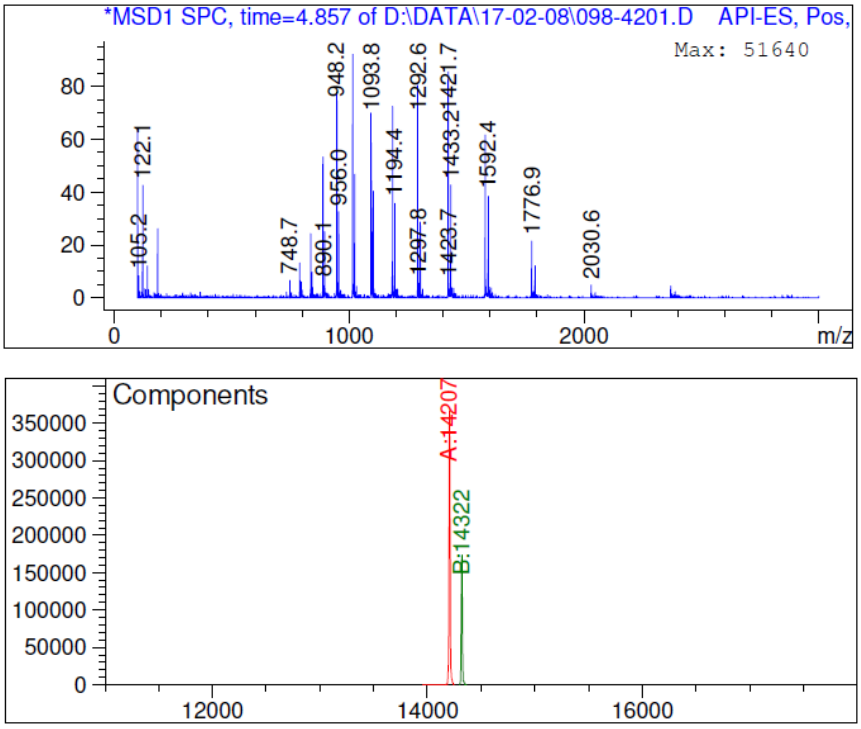


Figure 9.10 LC-MS analysis of FasNb5 amber GG amber nanobody expressed for incorporation of FurLys in response to the amber codon. In the expression of this nanobody the tRNA synthetase was not induced causing termination of protein translation at the first amber stop codon. Expected mass for the truncated nanobody 14209,91 Da. Additionally there is a second peak (14322 Da) present in the deconvoluted MS spectrum. This mass originates from the FasNb5 amber nanobody with failed incorporation which was analysed directly before this nanobody without extra washing of the column.

FasNb5 amber FurLys incorporation

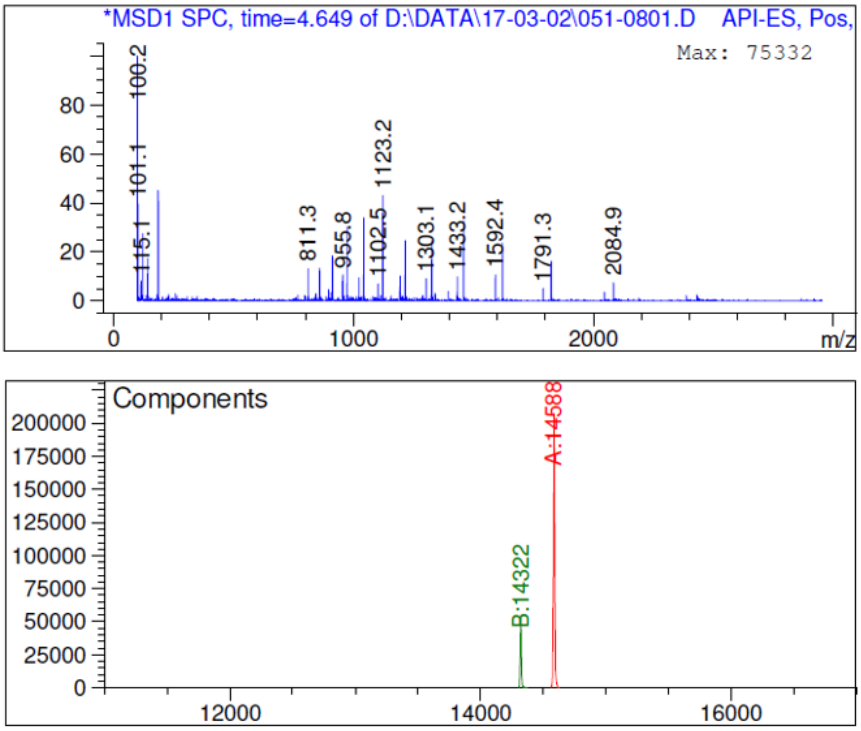


Figure 9.11 LC-MS analysis of FasNb5 amber nanobody expressed for incorporation of furan lysine in response to the amber codon. Expected mass for the nanobody with FurLys incorporated 14590.18 Da. Additionally a second mass is detected in the deconvoluted MS spectrum corresponding with the mass of the nanobody where translation is terminated at the amber stop codon.

FasNb5 amber GG amber FurLys incorporation

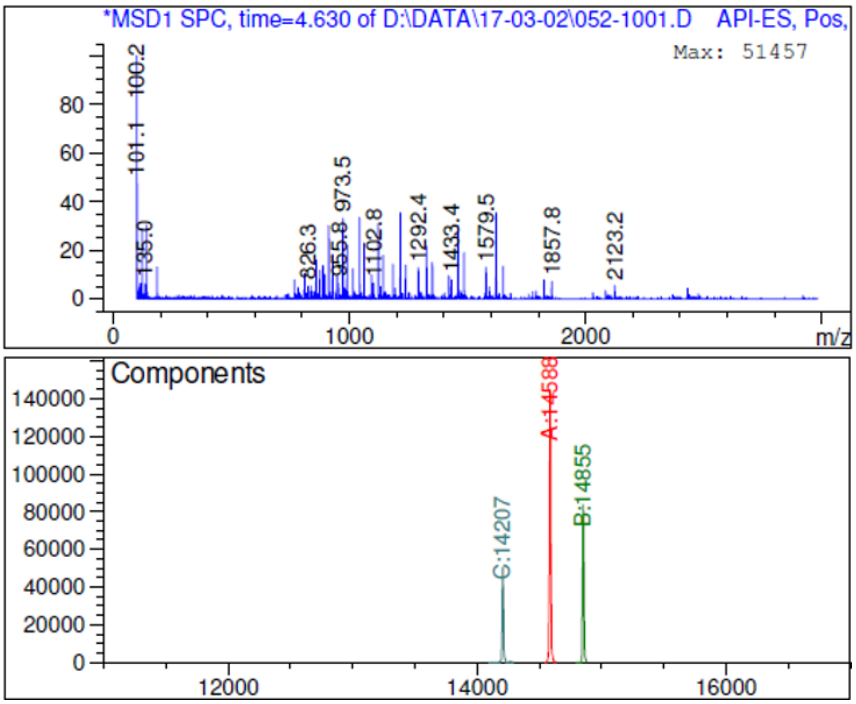


Figure 9.12 LC-MS analysis of FasNb5 amber GG amber nanobody expressed for incorporation of two FurLys residues in response to the 2 amber codons. Expected mass for the nanobody with both FurLys residues incorporated 14856.39 Da (peak B). Additionally peak A is detected in the deconvoluted MS spectrum corresponding with the mass of the nanobody where one FurLys is incorporated in response to the first amber codon and translation is terminated at the second amber stop codon expected mass: 14590.18 Da. Peak C corresponds with a nanobody where translation is terminated at the first amber codon, expected mass: 14209.91.

CortNb2 amber (no incorporation)

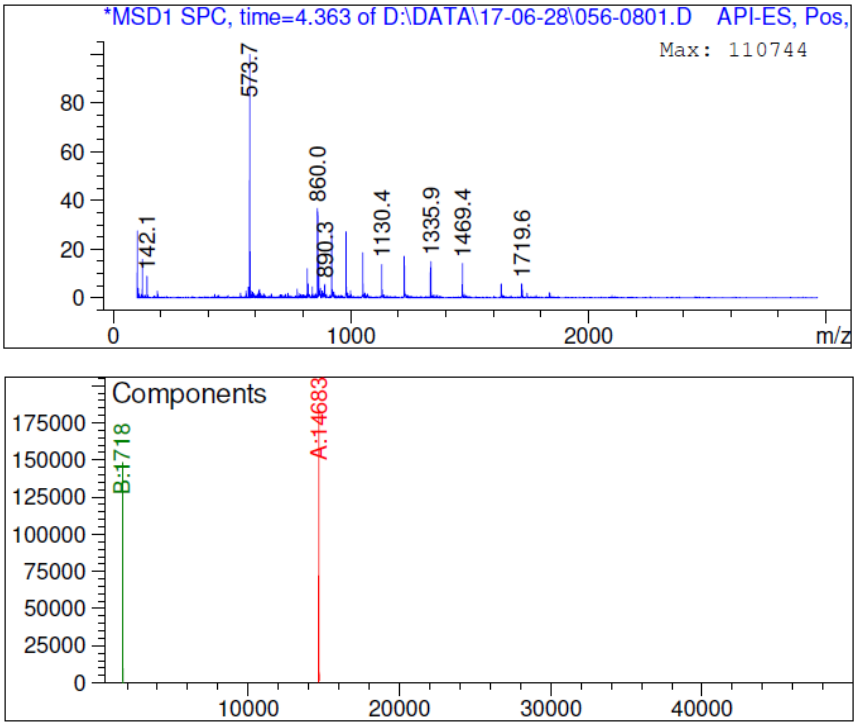


Figure 9.13 LC-MS analysis of CortNb2 amber nanobody expressed in absence of FurLys. Expected mass for the truncated nanobody 14686.28 Da.

CortNb2 amber FurLys incorporation

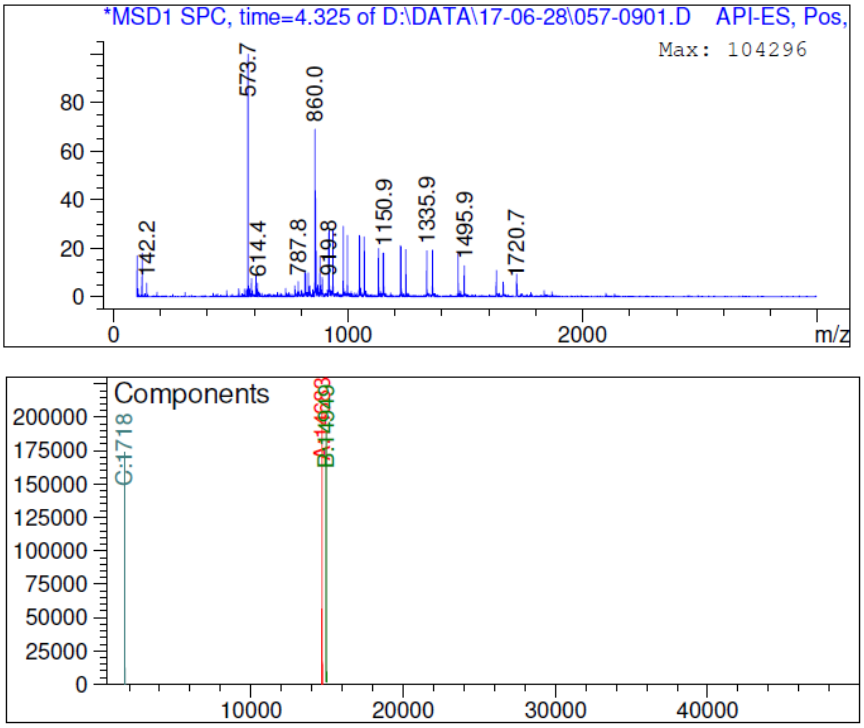


Figure 9.14 LC-MS analysis of CortNb2 amber nanobody expressed for incorporation of FurLys in response to the amber codon. Expected mass for the nanobody with FurLys incorporated 14952.59 Da, observed mass: 14949 Da. Additionally a second mass is detected in the deconvoluted MS spectrum corresponding with the mass of the nanobody where translation is terminated at the amber stop codon 14683 Da.

9.2.1.3 Oxidative damage

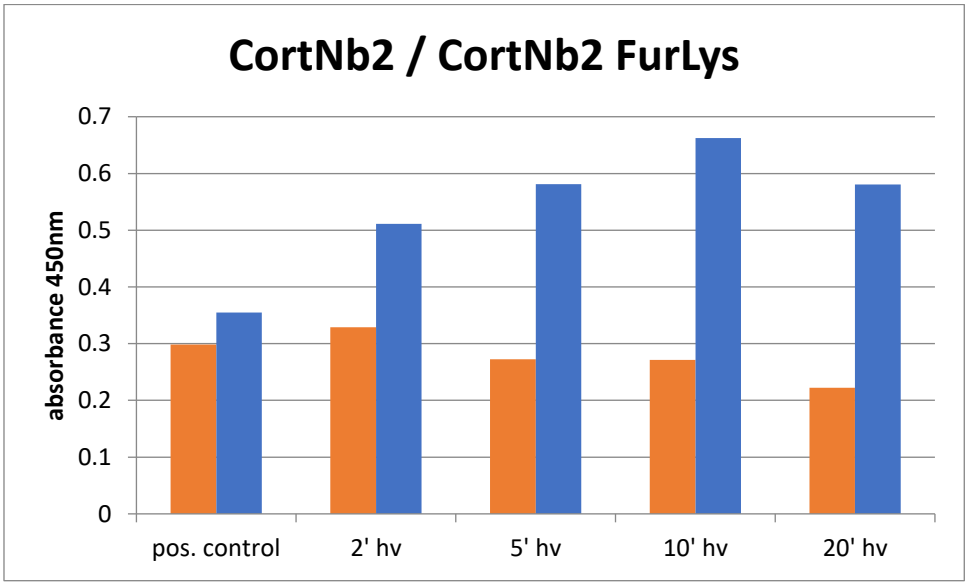


Figure 9.15 Bar graph for the oxidative damage ELISA response at 450 nm for CortNb2 (orange) and FurLys modified CortNb2 amber (blue). 15 equivalents of cortactin. Positive control: not irradiated CortNb2; 2' hv: 2 min irradiation; 5' hv: 5 min irradiation; 10' hv: 10 min irradiation; 20' hv: 20 min irradiation.

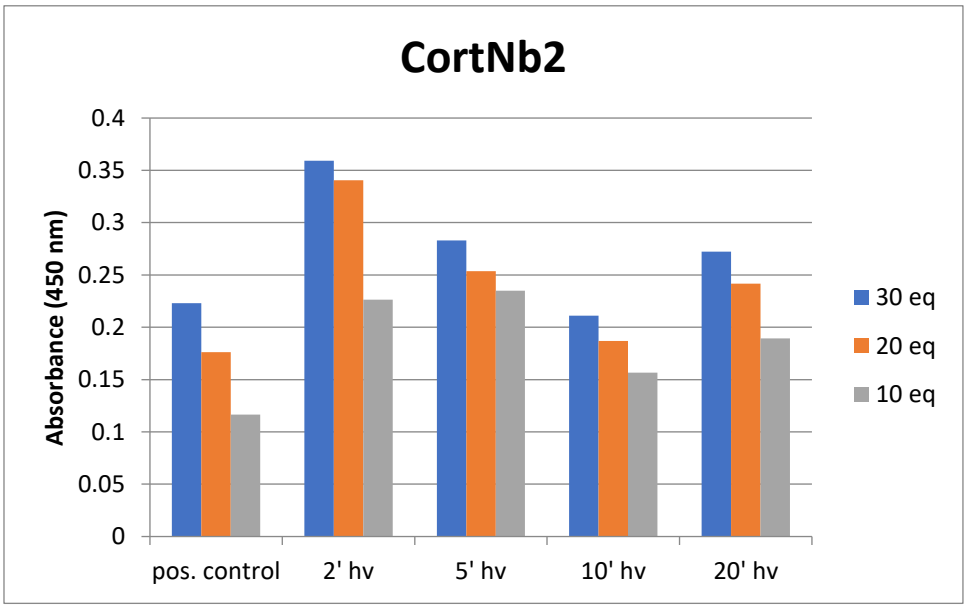


Figure 9.16 Bar graph for the oxidative damage ELISA response at 450 nm for CortNb2 with different equivalents of cortactin: blue: 30 eq.; orange: 20 eq.; gray: 10 eq.. Positive control: not irradiated CortNb2; 2' hv: 2 min irradiation; 5' hv: 5 min irradiation; 10' hv: 10 min irradiation; 20' hv: 20 min irradiation.

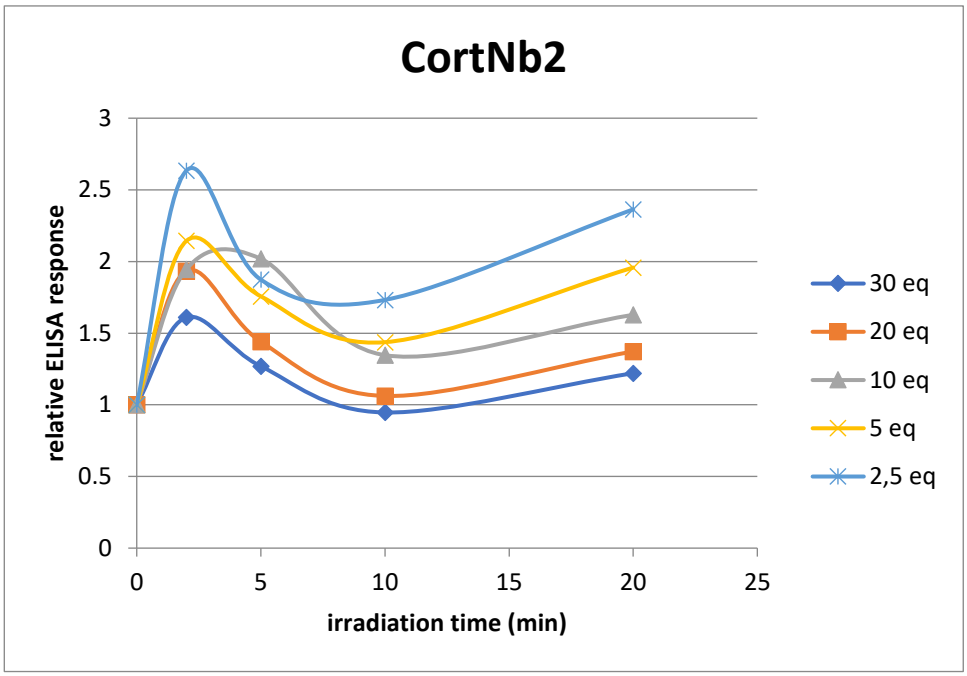


Figure 9.17 Graph for the relative oxidative damage ELISA response at 450 nm. The relative ELISA response is calculated as the response for a given sample divided by the response for the positive control. CortNb2 with different equivalents of cortactin: blue: 30 eq.; orange: 20 eq.; gray: 10 eq.; yellow: 5 eq.; light blue: 2.5 eq.. Samples with were irradiated for 0 min, 2 min, 5 min, 10 min and 20 min.

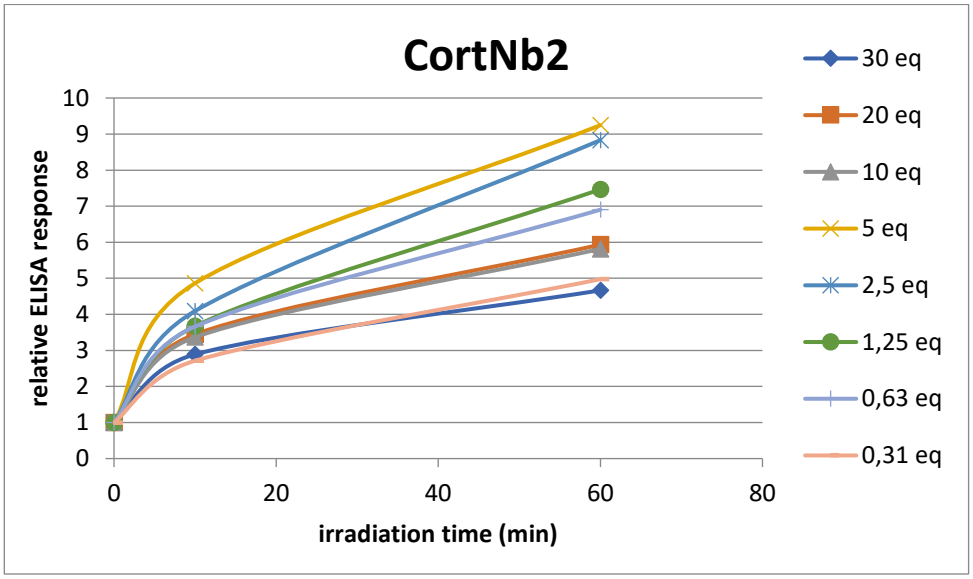


Figure 9.18 Graph for the relative oxidative damage ELISA response at 450 nm. The relative ELISA response is calculated as the response for a given sample divided by the response for the positive control. CortNb2 with different equivalents of cortactin: dark blue: 30 eq.; orange: 20 eq.; gray: 10 eq.; yellow: 5 eq.; blue: 2.5 eq.; green: 1.25 eq.; light blue: 0.63 eq.; pink: 0.31 eq.. Samples with were irradiated for 0 min, 10 min, 60 min.

CortNb2 photo-oxidation

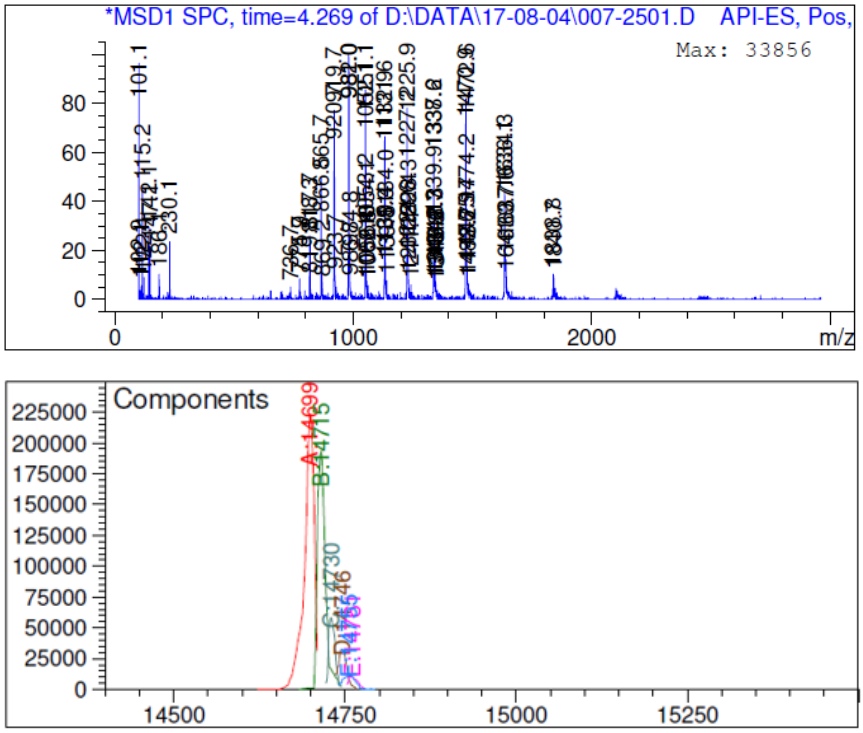


Figure 9.19 LC-MS analysis of CortNb2 truncated at the amber codon after a photo-oxidation reaction with 3 µM Rose Bengal (RB) and 10 minutes irradiation with white light at 10 °C. The observed mass for the nanobody before oxidation was 14683 Da, the largest peak observed in this deconvoluted spectrum is the first oxidation peak +16 Da, additionally 4 other oxidation products can be observed in this figure.

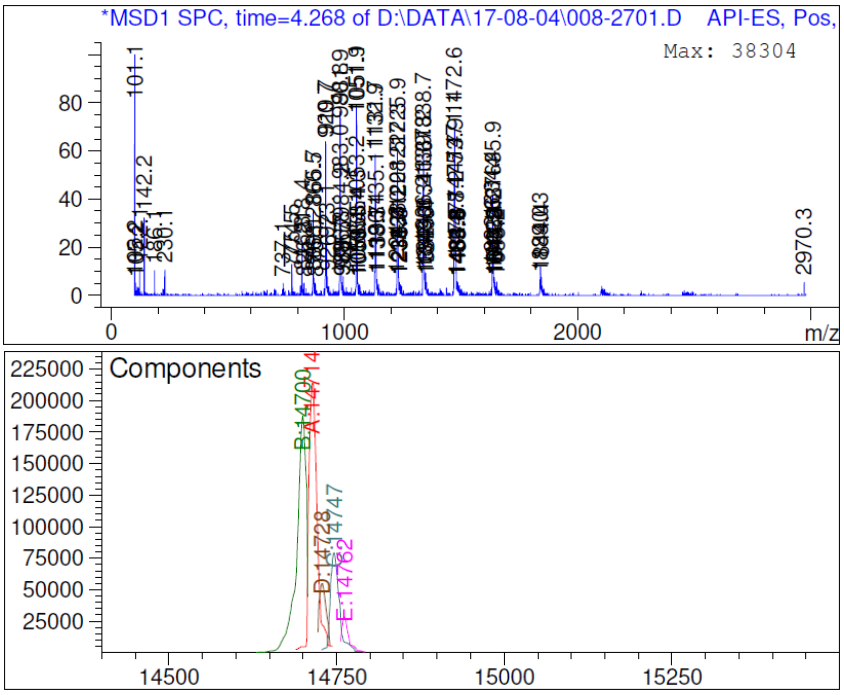


Figure 9.20 LC-MS analysis of CortNb2 truncated at the amber codon after a photo-oxidation reaction with 3 µM rose bengal and 20 minutes irradiation with white light at 10 °C. The observed mass for the nanobody before oxidation was 14683 Da, the largest peak observed in this deconvoluted spectrum is the second oxidation peak +32 Da, additionally 5 other oxidation products can be observed in this figure.

CortNb2 FurLys photo-oxidation

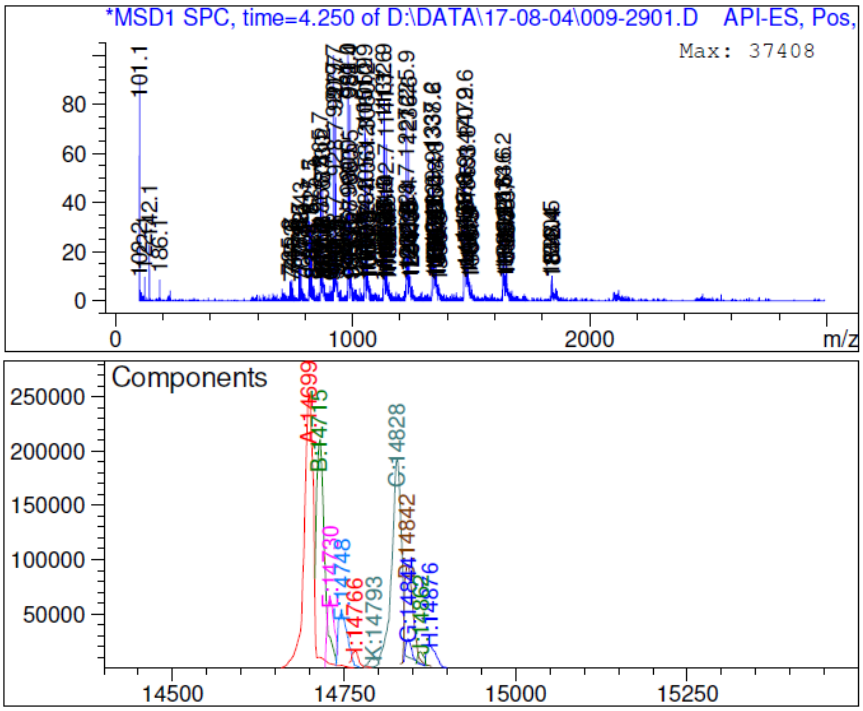


Figure 9.21 LC-MS analysis of FurLys CortNb2 amber after a photo-oxidation reaction with 3 µM rose bengal and 10 minutes irradiation with white light at 10 °C. For the truncated nanobody a similar oxidation pattern is observed as before. Strangely the FurLys CortNb2 amber nanobody with a mass of 14952.59 Da is not detected or any of it's oxidation products. A second species with oxidation pattern is however observed with a lower mass as the FurLys CortNb2 amber.

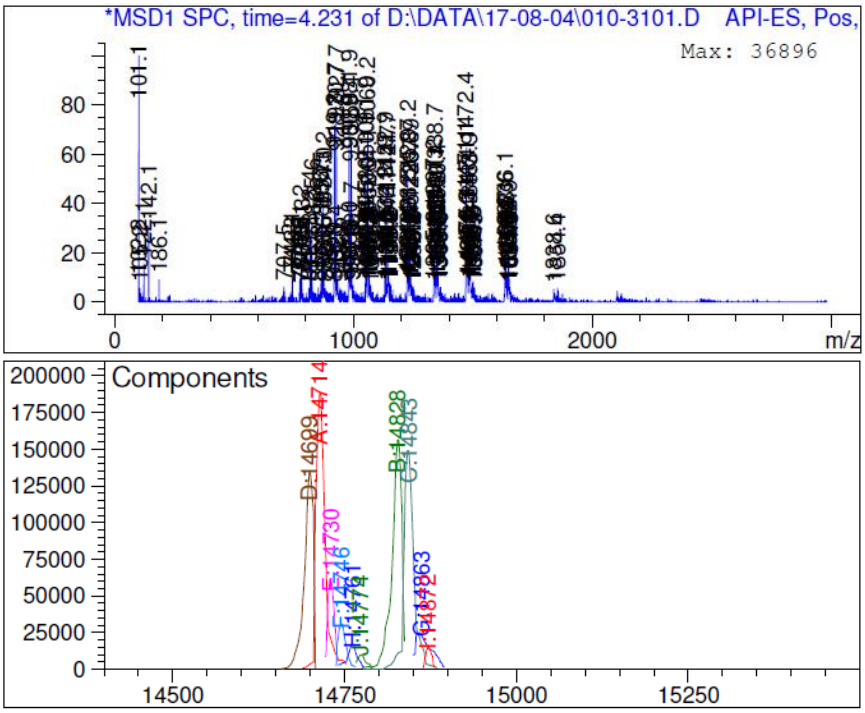


Figure 9.22 LC-MS analysis of FurLys CortNb2 amber after a photo-oxidation reaction with 3 µM rose bengal and 20 minutes irradiation with white light at 10 °C. For the truncated nanobody a similar oxidation pattern is observed as before. Strangely the FurLys CortNb2 amber nanobody with a mass of 14952.59 Da is not detected or any of its oxidation products. A second species with oxidation pattern is however observed with a lower mass as the FurLys CortNb2 amber.

FasNb5 FurLys conjugation with PTAD

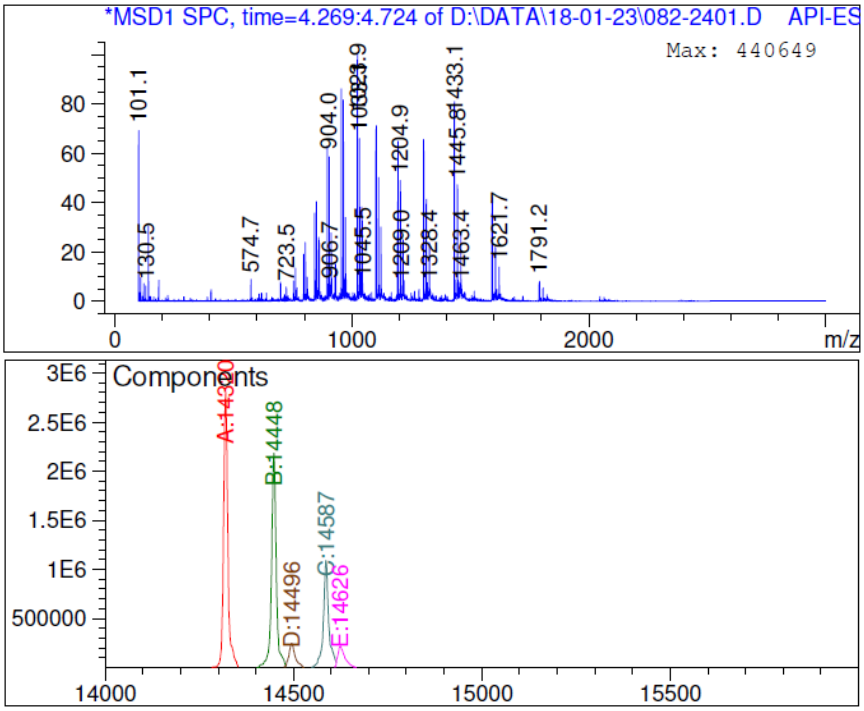


Figure 9.23 LC-MS analysis of FurLys FasNb5 amber after conjugation with 5 equivalents of PTAD. Both the FurLys FasNb5 amber and the truncated nanobody are observed as expected. However no species is detected with a higher mass than the FurLys modified nanobody corresponding with the desired conjugated nanobody. Additionally a third compound is detected with a mass lower than the FurLys modified nanobody.

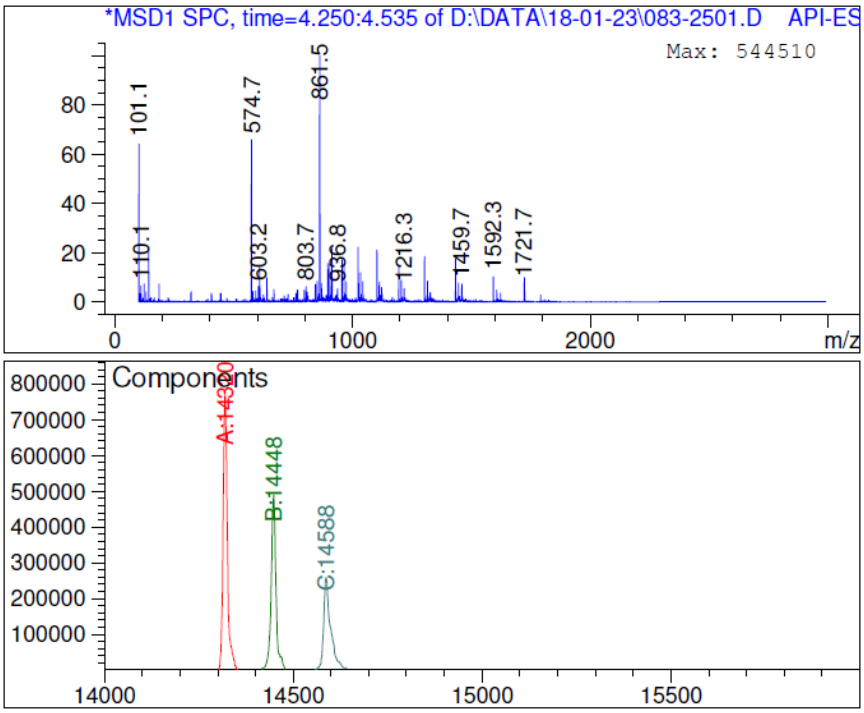


Figure 9.24 LC-MS analysis of FurLys FasNb5 amber after conjugation with 20 equivalents of PTAD. Both the FurLys FasNb5 amber and the truncated nanobody are observed as expected. However no species is detected with a higher mass than the FurLys modified nanobody corresponding with the desired conjugated nanobody. Additionally a third compound is detected with a mass lower than the FurLys nanobody.

9.2.2 Experimental data for part 5.1.3

9.2.2.1 Data for synthesised compounds

Peptides WT, WT-3, WT-6

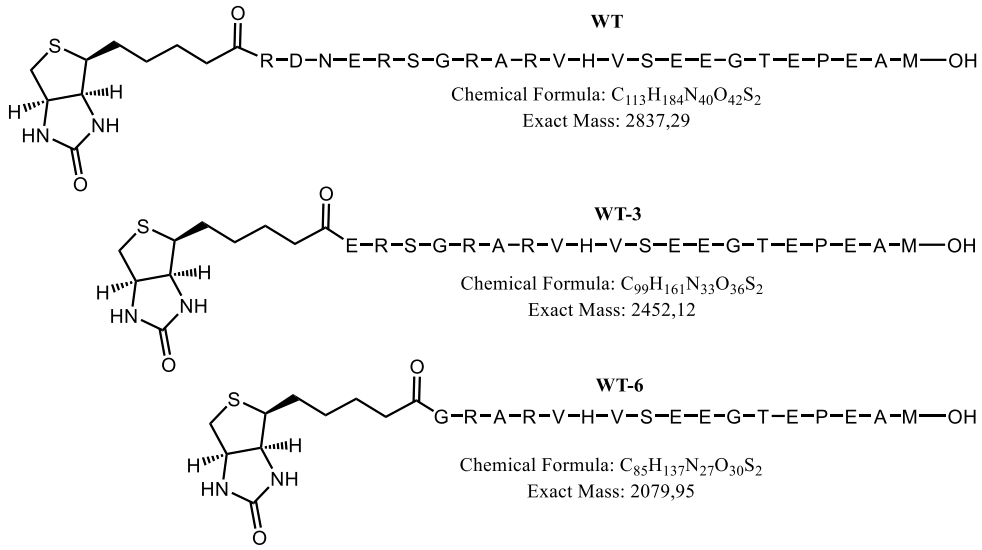


Figure 9.25 Peptides **WT, WT-3, WT-6** structures.

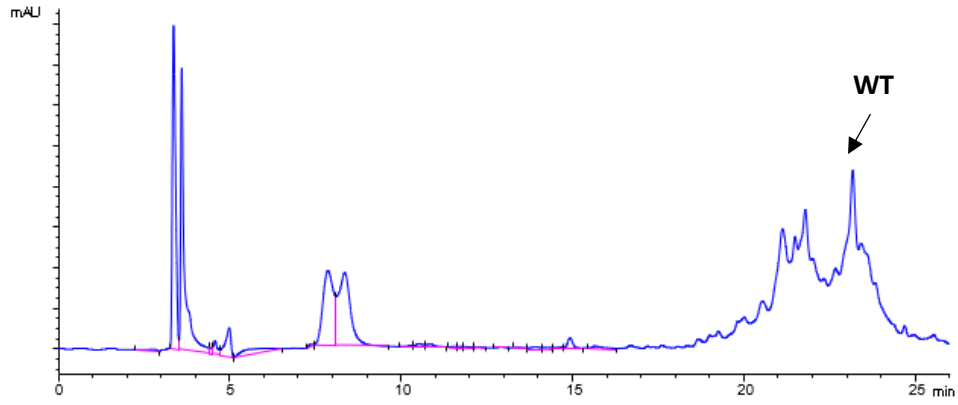


Figure 9.26 Crude HPLC chromatogram at 214 nm, peptide **WT** indicated with an arrow. The peptide was purified using reversed phase preparative HPLC with a 0-30 % ACN gradient in 20 min. The product peptide fraction was checked via MALDI-TOF.

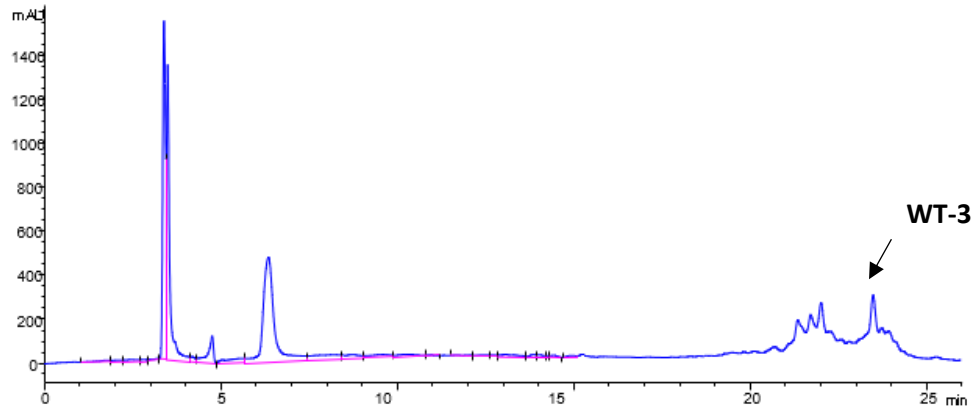


Figure 9.27 Crude HPLC chromatogram at 214 nm, peptide **WT-3** indicated with an arrow. The peptide was purified using reversed phase preparative HPLC with a 0-30 % ACN gradient in 20 min. The product peptide fraction was checked via MALDI-TOF.

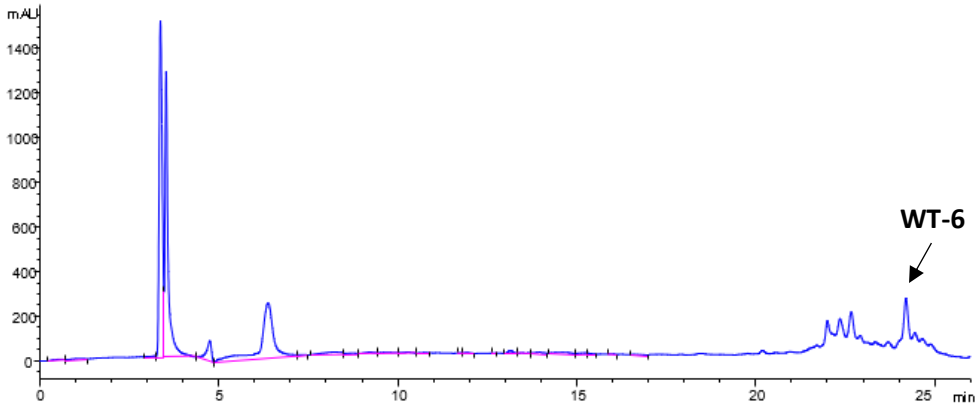
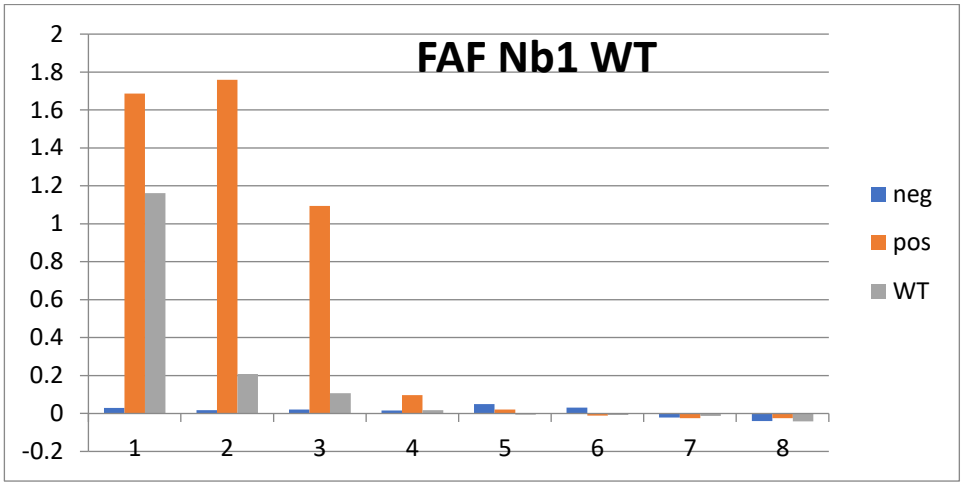
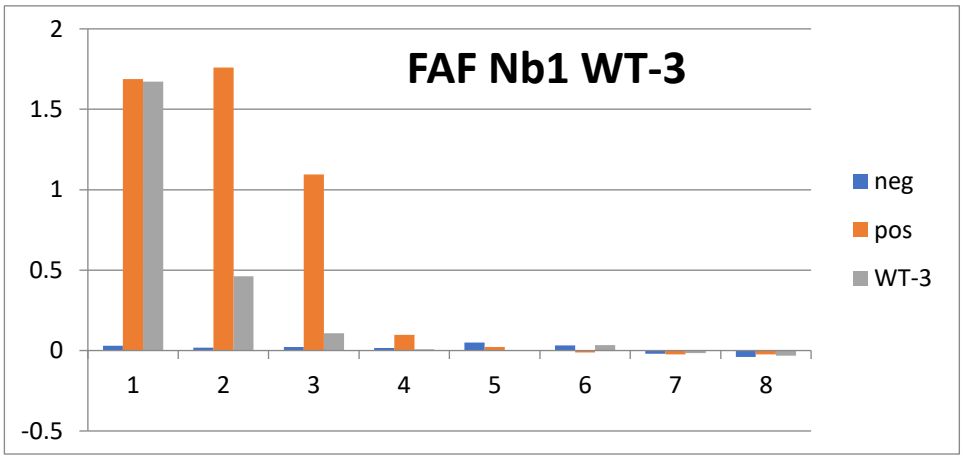


Figure 9.28 Crude HPLC chromatogram at 214 nm, peptide **WT-6** indicated with an arrow. The peptide was purified using reversed phase preparative HPLC with a 0-30 % ACN gradient in 20 min. The product peptide fraction was checked via MALDI-TOF.

9.2.2.2 ELISA of gelsolin peptides





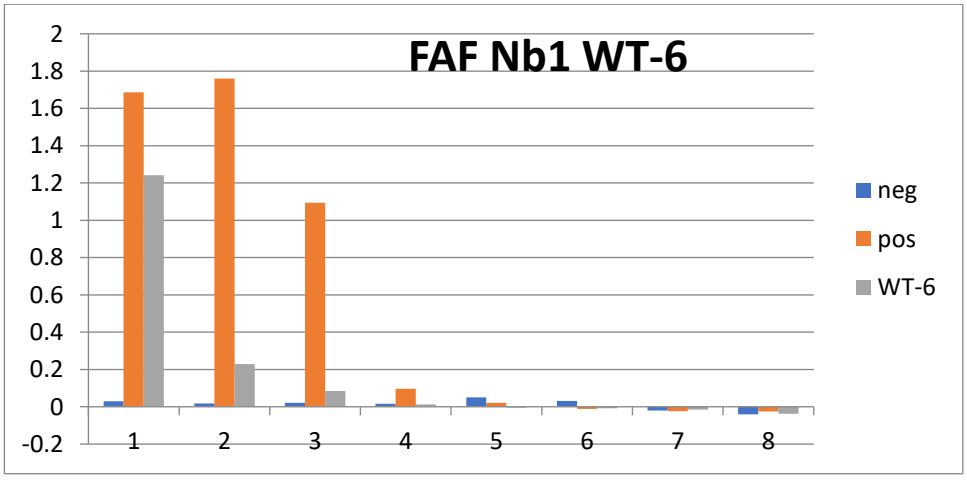


Figure 9.29 ELISA results for peptides **WT, WT-3** and **WT-6** and FAF nanobody (FAF Nb1). Fluorescence is plotted against 8 tenfold dilutions of the FAF nanobody. Negative control (blue) consists of a gelsolin fragment with no affinity for FAF nanobody whereas the positive control (orange) consists of the wildtype gelsolin fragment.

9.2.2.3 DLS analysis of gelsolin peptides (WT, WT-3 and WT-6)

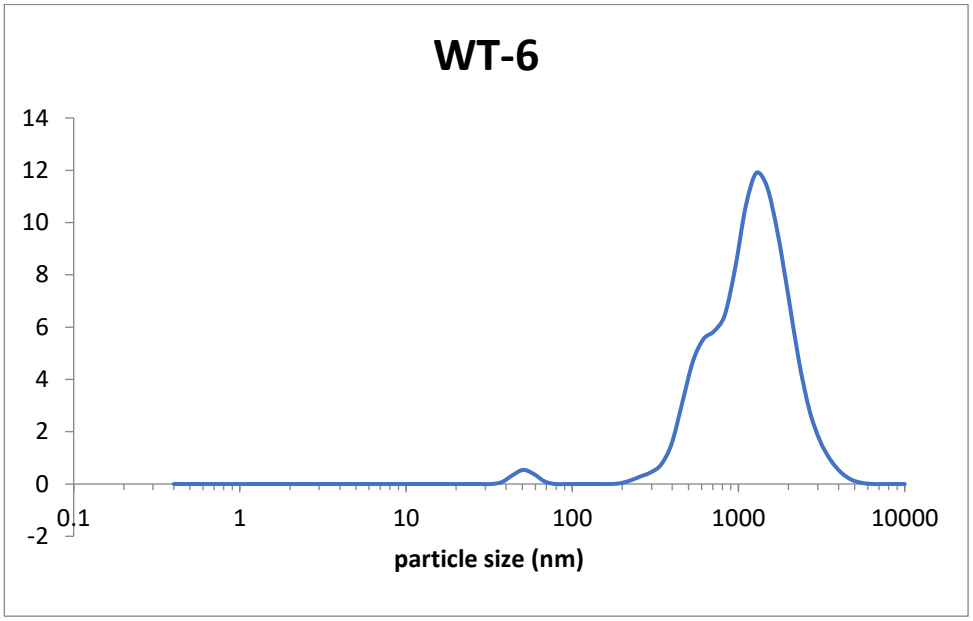


Figure 9.30 Dynamic light scattering (DLS) graph of peptide **WT-6**.

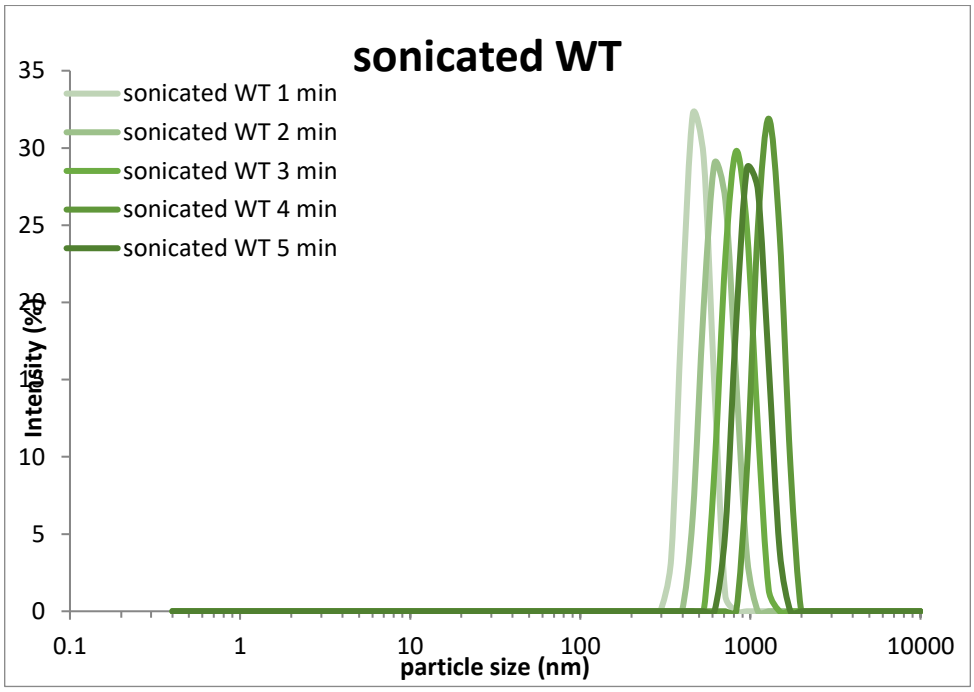


Figure 9.31 Dynamic light scattering (DLS) graph of peptide **WT**. Peptide **WT** was sonicated for 15 minutes and afterwards the sample was measured after 1, 2, 3, 4 and 5 minutes after sonication.

9.2.2.4 EgA1 crosslinking nanobody

Nanobody sequences for EgA1 nanobody mutants, **X** indicates the amber position.

Nanobody WT

QVQLQESGGGLVQPGGSLRLSCAASGRTFSSYAMGWFRQAPGKQREFVAAIRWSGGYTYY
TDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAATYLSSDYSRYALPQRPLDYDYWGQ
GTQVTVSS

R107 amber

QVQLQESGGGLVQPGGSLRLSCAASGRTFSSYAMGWFRQAPGKQREFVAAIRWSGGYTYY
TDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAATYLSSDYS**X**YALPQRPLDYDYWGQ
GTQVTVSS

S103 amber

QVQLQESGGGLVQPGGSLRLSCAASGRTFSSYAMGWFRQAPGKQREFVAAIRWSGGYTYY
TDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAATYLS**X**DYSRYALPQRPLDYDYWGQ
GTQVTVSS

S102 amber

QVQLQESGGGLVQPGGSLRLSCAASGRTFSSYAMGWFRQAPGKQREFVAAIRWSGGYTYY
TDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAATYL**X**SDYSRYALPQRPLDYDYWGQ
GTQVTVSS

Y100 amber

QVQLQESGGGLVQPGGSLRLSCAASGRTFSSYAMGWFRQAPGKQREFVAAIRWSGGYTYY
TDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAAT**X**LSSDYSRYALPQRPLDYDYWGQ
GTQVTVSS

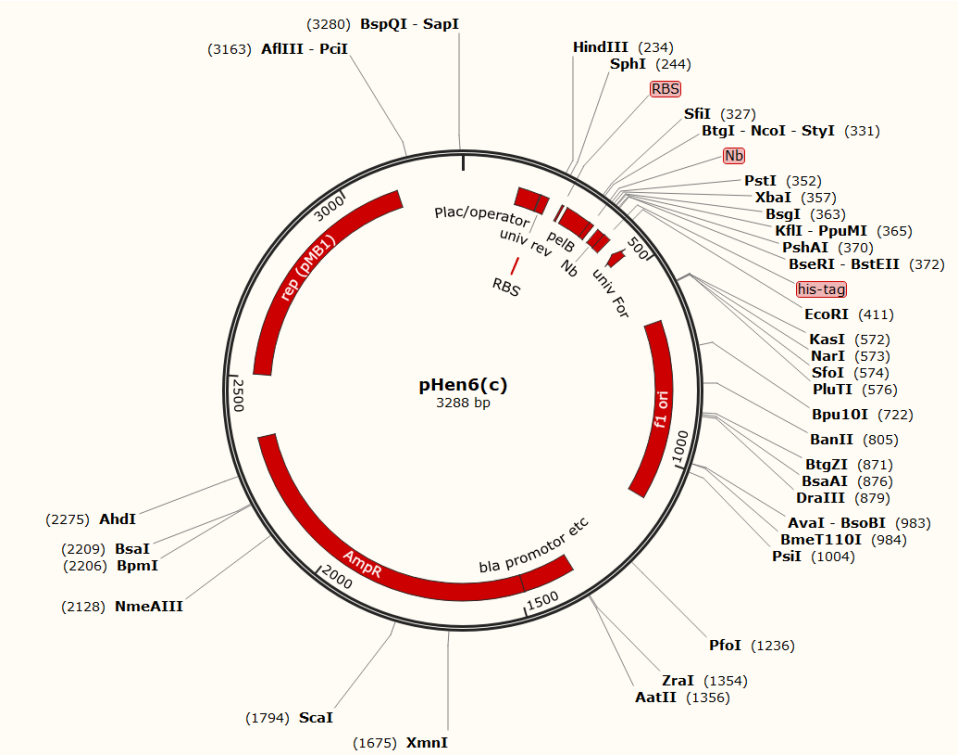


Figure 9.32 Plasmid map for pHen6, the nanobody genes are cloned into this vector via the PstI and BstEII sites. The nanobodies have a N-terminal PelB sequence which directs the nanobodies to the periplasm, afterwards the PelB sequence is cleaved off. Additionally the nanobodies have a C-terminal His-tag for purification and visualisation purposes.

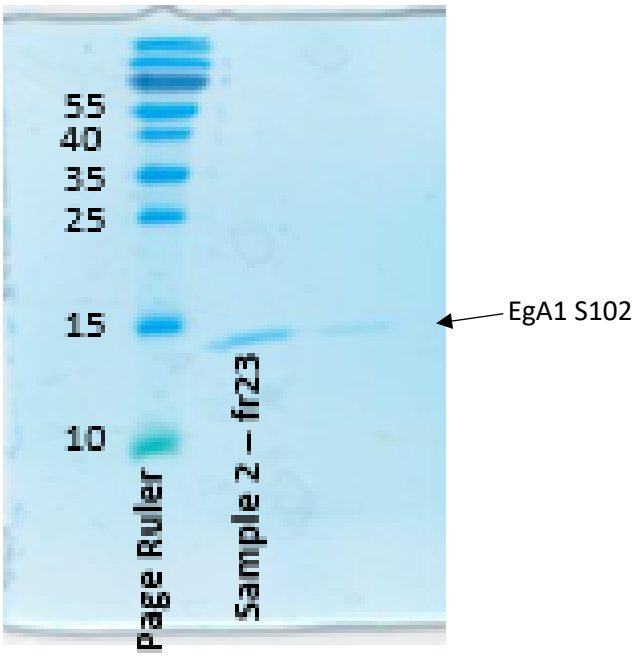


Figure 9.33 Image of the SDS-PAGE analysis of fractions 23 from the gel filtration of the EgA1 S102 amber FurLys sample. Note that the amount of protein was very low for this nanobody mutant and not sufficient for LC-MS analysis.

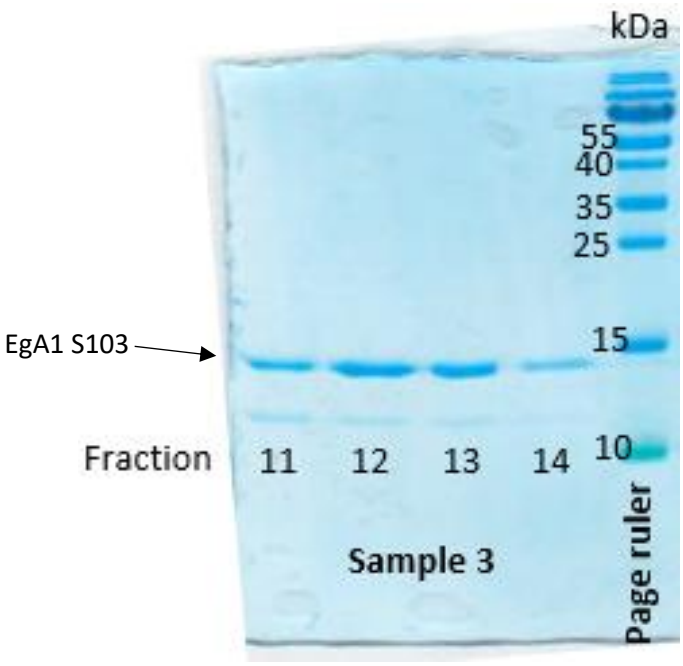


Figure 9.34 Image of the SDS-PAGE analysis of fractions 11-14 from the gel filtration of the EgA1 S103 amber FurLys sample.

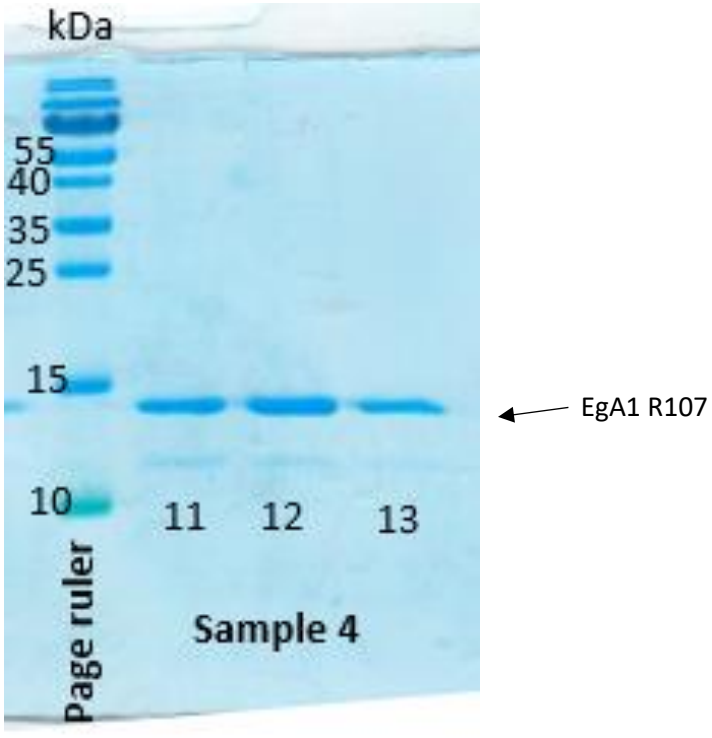


Figure 9.35 Image of the SDS-PAGE analysis of fractions 11-13 from the gel filtration of the EgA1 R107 amber FurLys sample.

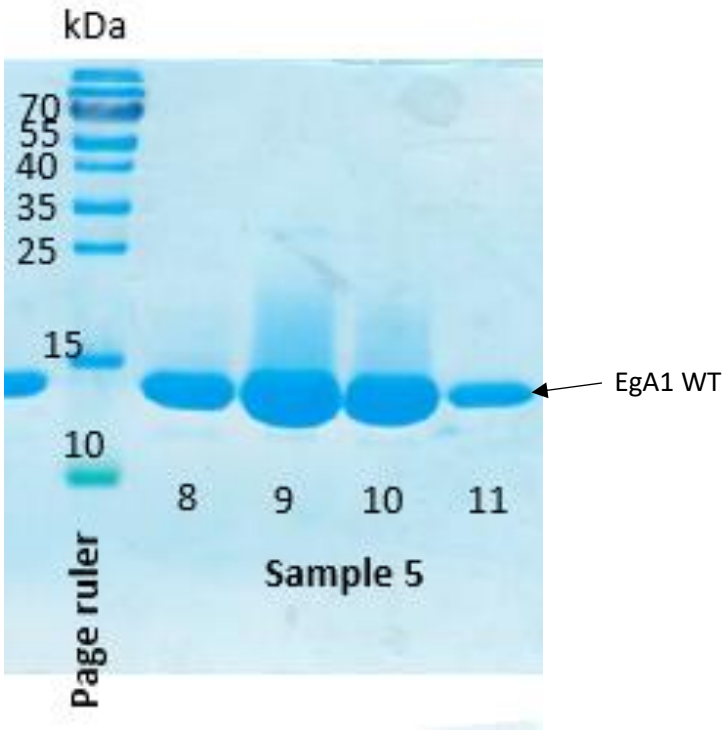


Figure 9.36 Image of the SDS-PAGE analysis of fractions 11-14 from the gel filtration of the EgA1 wildtype sample.

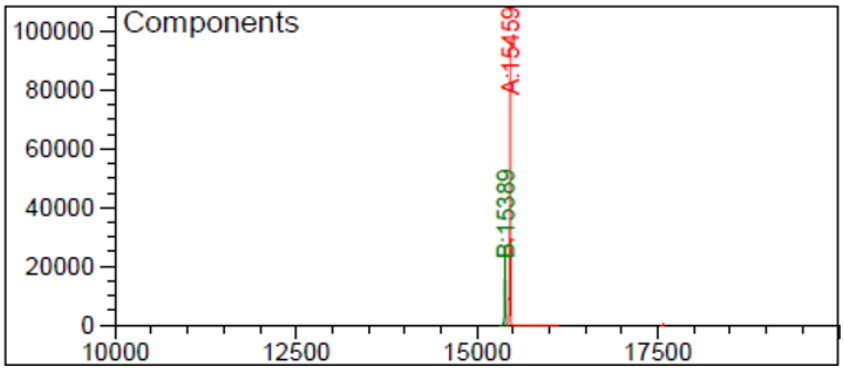


Figure 9.37 Deconvoluted MS spectrum of EgA1 S103 amber FurLys. The calculated mass is 15464 Da, observed 15459 Da.

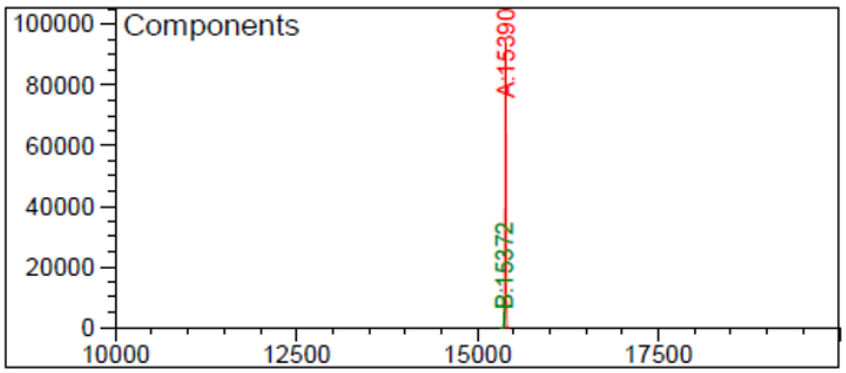


Figure 9.38 Deconvoluted MS spectrum of EgA1 R107 amber FurLys. The calculated mass is 15395 Da, observed 15390 Da.

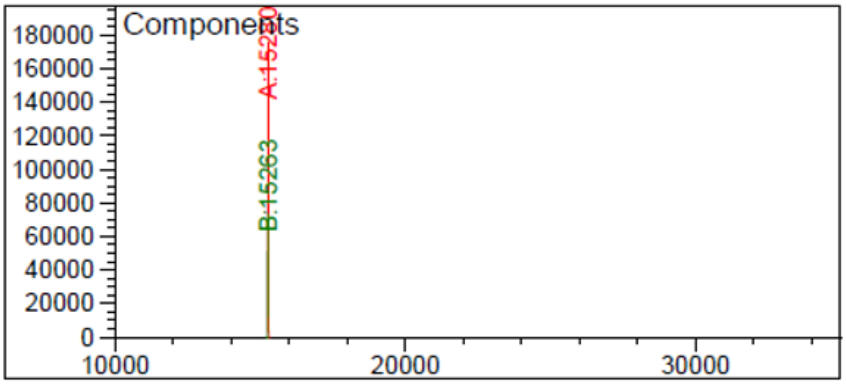


Figure 9.39 Deconvoluted MS spectrum of EgA1 wildtype. The calculated mass is 15284 Da, observed 15280 Da.

9.2.3 Experimental data for part 5.2.2

9.2.3.1 Furylalanine-CME synthesis

The synthesis of furylalanine-CME synthesis was performed by Dr. Willem Vannecke. dicyclohexylammonium (DCHA, 1 part) was suspended in 5–10 volumes of cold t-butyl methyl ether. 10% of phosphoric acid was added under stirring until the DCHA salt was completely dissolved and two clear phases appeared. The pH of the lower aqueous phase was around 2–3. The organic phase was isolated and washed once with 2 volume parts of phosphoric acid 10% and 3 times with 2 volume parts of water. The pH of the aqueous phase was ≥4. TLC (2:1 hexane/EtOAc) was used to check whether the amino acid was liberated from its DCHA salt. The organic phase was dried over anhydrous sodium sulfate, filtered off, and evaporated to dryness in vacuo to obtain the free amino acid. Boc-L-2-furylalanine was dissolved in 300 µL of chloroacetonitrile, 1,2 eq of DIPEA was added, and the mixture was reacted overnight at room temperature. Completion of the reaction was followed by TLC (2:1 hexane/EtOAc). The reaction mixture was transferred to a separation funnel of 100 mL and extracted with 50 mL of Et2O. The organic phase was washed with 3 × 50 mL of 1M HCl (aq), 2 × 50 mL with saturated NaHCO3 (aq), 2 × 10 mL with saturated NaCl (aq). The organic phase was dried over Na2SO4 and evaporated to dryness. Flash chromatography was used for purification of the compound (2:1 hexane/EtOAc). The collected fractions were evaporated to dryness. The purified compound was then dissolved in 4 M HCl/dioxane and stirred for 10 min on ice, followed by 30 min at room temperature. The reaction was followed by TLC (2:1 hexane/EtOAc). The reaction mixture was evaporated to almost complete dryness. Cold Et2O was added to precipitate the amino acid. The Boc-group was removed by dissolving the peptide in 4 M HCl/dioxane and keeping the solution on ice for 30 min. After reaction, the HCl/dioxane is partially removed by evaporation, with subsequent addition of cold diethylether to precipitate the amino acid. After centrifugation, the supernatant was discarded and the amino acid was obtained after drying in vacuo.

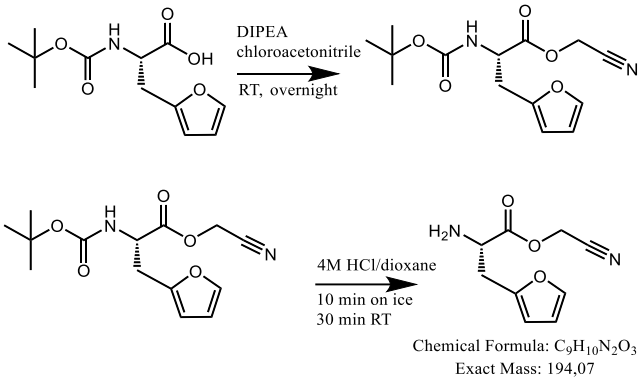


Figure 9.40 Synthesis scheme for the synthesis of furylalanine-CME.

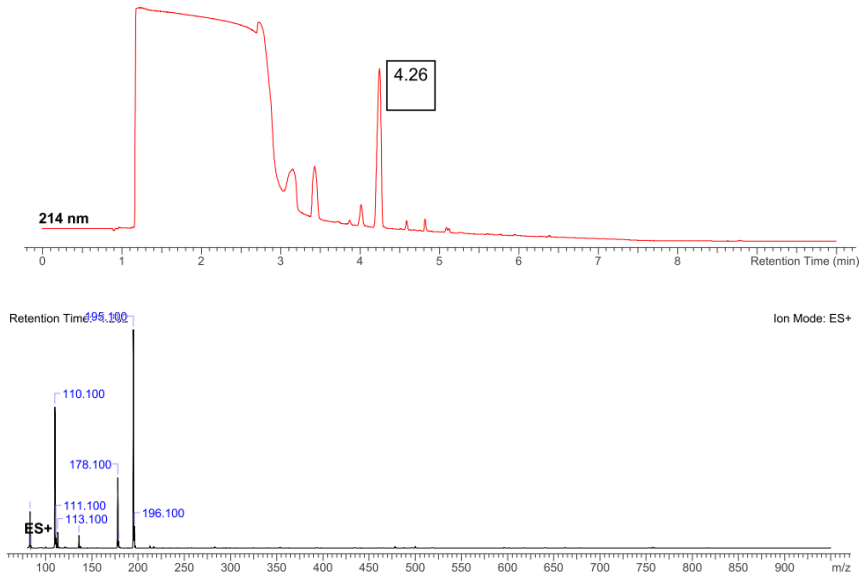


Figure 9.41 Top: LC-MS chromatogram showing the furylalanine-CME product at t_R= 4.26 min. Bottom: ESI-MS from LC-MS at t_R = 4.26 min ($[M+H]^+$ = 195.10 Da).

9.2.3.2 DNA template

The sequence of the translated template is AcFAGAFuaGPG**X**AGA (with **X** = C, H, K, R, S, Y). Both residues indicated in red were incorporated in response to the ATG codon.

Forward primer (5' to 3'):

TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATG

Reverse primer (5' to 3'):

GCTAGCTTAAGCACCCGC

Final DNA: Coding sequence + 5' adaptor:

1C CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**TGC**GCGGGTGCTTAA

1H CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**CAC**GCGGGTGCTTAA

1K CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**AAG**GCGGGTGCTTAA

1R CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**CGC**GCGGGTGCTTAA

1S CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**AGC**GCGGGTGCTTAA

1K CAAGGAGAAAAACATGGCCGGTGCGATGGGTCCGGGC**TAC**GCGGGTGCTTAA

The template was amplified by PCR. Master mix (1 mL) was prepared by adding 100 µL PCR buffer (10 x) to 820 µL MQ H_2O, 10 µL $MgCl_2$ 250 mM, 50 µL5 mM of each dNTP mix, 5 µL 100 µM forward primer, 5 µL 100 µM reverse primer and 10 µL (100 x) Taq polymerase. This master mix was divided in 6 PCR tubes and different test oligo's were added (0.5 µL 1 µM solution) to each tube. The PCR reaction was set up: 95 °C 40 sec., 50 °C 40 sec., 72°C 40 sec. and repeated for 15 cycles. And agarose gel was run with 2µL sample mixed with 2 µL of 2x DNA loading buffer. The gel was a 3% agarose gel run at 135 V for 10 min.

9.2.3.3 Flexible in vitro translation of peptide library

Flexizymes and tRNAs were synthesized as previously described[69]. Aminoacylation of the elongator tRNA with furylalanine, 1 µL 500 mM HEPES-KOH buffer of pH 7.5, and 250 µM flexizyme (eFx) and 250 µM tRNA were added to 3 µL of milliQ H2O. The mixture (6 µL total volume) was placed in a heat block at 95 ◦C for 2 min. Afterwards, 2 µL 3M MgCl was added and the tube was mixed well and placed at room temperature for 5 min. The tube was placed on ice for several minutes and subsequently 2 µL of 25 mM furylalanine-CME (DMSO solution) was added. The solution was mixed and incubated on ice for 6 h. For the aminoacylation of initiator tRNA with N-acetylated phenylalanine, the same protocol was used except for the final incubation time on ice, which was 2 h. After EtOH precipitation of aminoacyl-tRNA, the mixture was divided in two (5 µL each). Then, 20 µL of 0.3 M NaOAc and 50 µL of EtOH were added and the solution was mixed well. The solution was centrifuged at 13 krpm, 25 ◦C for 15 min, and the supernatant was discarded. Next, 30 µL of 70% EtOH containing 0.1 M NaOAC (pH 5.2) was added to the pellet and vortexed for 10 s, centrifuged for 5 min, and the supernatant was discarded. This last step was repeated and finally 20 µL of 70% EtOH was added and the sample was centrifuged for 3 min. The supernatant was discarded, the lid was opened, and the tube was covered with a clean paper towel and the sample was left to dry for 5–10 min. Genetically reprogrammed translation was performed in a FIT reaction at a final scale of 2.5 µL at 37 ◦C for 30 min as previously described.[69] After translation, MALDI-TOF MS analysis was performed as previously described[69] using an UltraFlex instrument (Bruker, Billerica, MA, USA).

9.2.3.4 MALDI-TOF MS data translated peptides

Template: AcFAGAFuaGPG**X**AGA (**X**= C, H, K, R, S, Y)

Peptide 1C

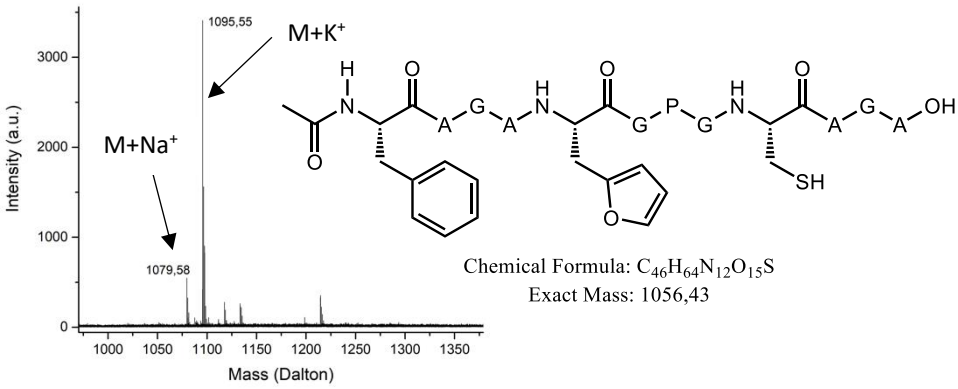


Figure 9.42 MALDI-TOF spectrum of translated peptide **1C** and the structure with exact mass. The $M+K^+$ peak: 1056.43 Da + 39.10 Da = 1095.53 Da and $M+Na^+$ peak:1056.43 Da + 22.99 Da = 1079.42 Da peak are indicated with arrows in the mass spectrum.

Peptide 1H

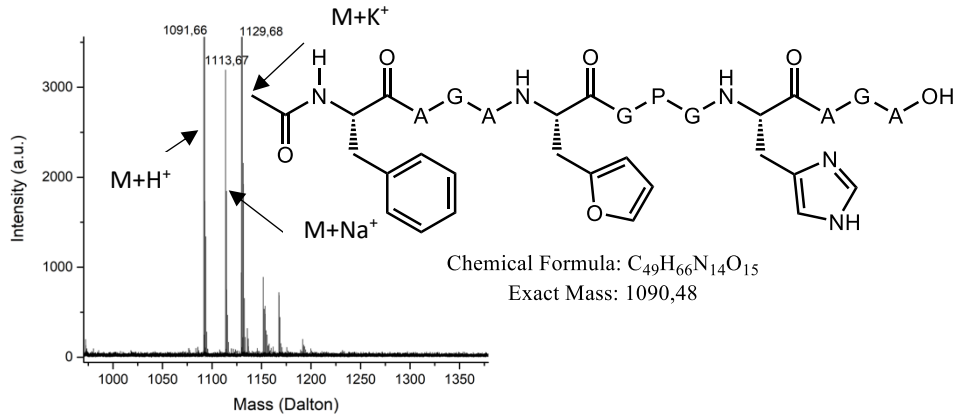


Figure 9.43 MALDI-TOF spectrum of translated peptide **1H** and the structure with exact mass. The $M+H^+$ peak: 1090.48 Da + 1.01 Da = 1091.49, $M+K^+$ peak : 1090.48 Da + 39.10 Da = 1129.58 Da and $M+Na^+$ peak :1090.48 Da + 22.99 Da = 1113.47 Da are indicated with arrows in the mass spectrum.

Peptide 1K

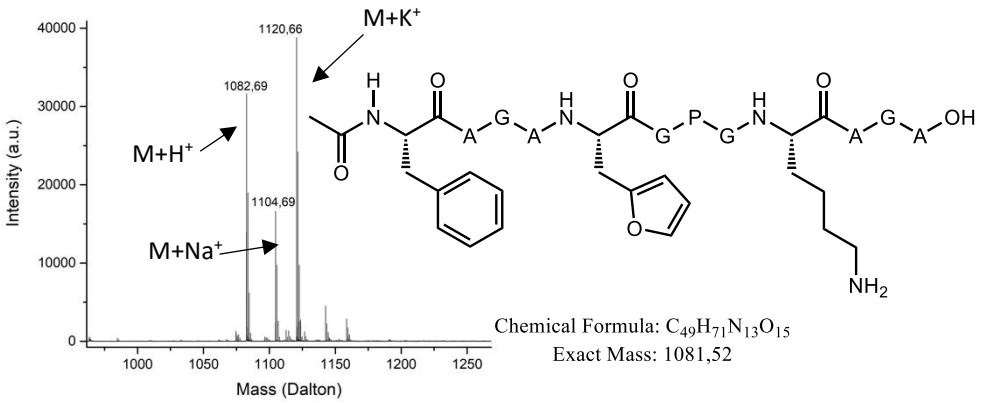


Figure 9.44 MALDI-TOF spectrum of translated peptide **1K** and the structure with exact mass. The $M+H^+$ peak: 1081.52 Da + 1.01 Da = 1082.53, $M+K^+$ peak: 1081.52 Da + 39.10 Da = 1120.62 Da and $M+Na^+$ peak: 1081.52 Da + 22.99 Da = 1104.51 Da are indicated with arrows in the mass spectrum.

Peptide 1R

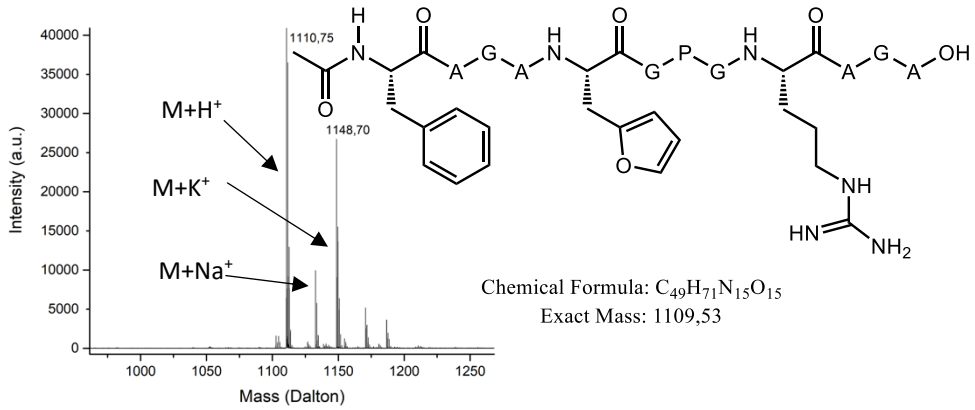


Figure 9.45 MALDI-TOF spectrum of translated peptide **1R** and the structure with exact mass. . The $M+H^+$ peak: 1109.53 Da + 1.01 Da = 1110.54, $M+K^+$ peak: 1109.53 Da + 39.10 Da = 1148.63 Da and $M+Na^+$ peak :1109.53 Da + 22.99 Da = 1132.52 Da are indicated with arrows in the mass spectrum.

Peptide 1S

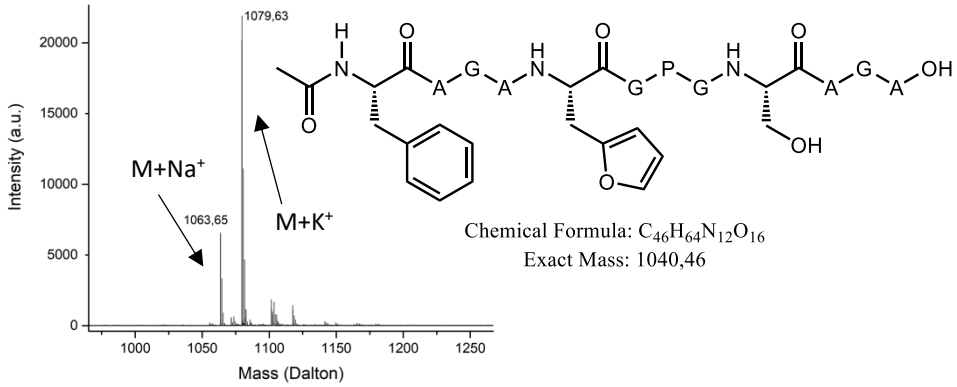


Figure 9.46 MALDI-TOF spectrum of translated peptide **1S** and the structure with exact mass. The $M+K^+$ peak: 1040.46 Da + 39.10 Da = 1079.56 Da and $M+Na^+$ peak: 1040.46 Da + 22.99 Da = 1063.45 Da are indicated with arrows in the mass spectrum.

Peptide 1Y

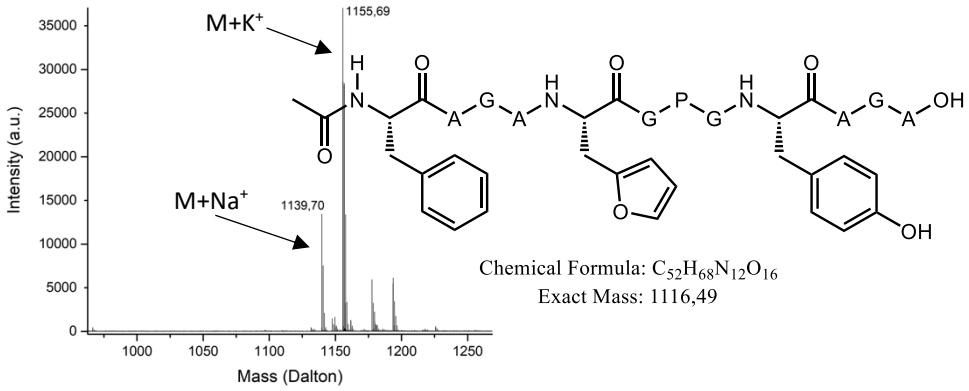


Figure 9.47 MALDI-TOF spectrum of translated peptide **1Y** and the structure with exact mass. The $M+K^+$ peak : 1116.49 Da + 39.10 Da = 1155.59 Da and $M+Na^+$ peak :1116.49 Da + 22.99 Da = 1139.48 Da are indicated with arrows in the mass spectrum.

9.2.3.5 MALDI-TOF MS data translated peptides after NBS oxidation

Peptide 2C

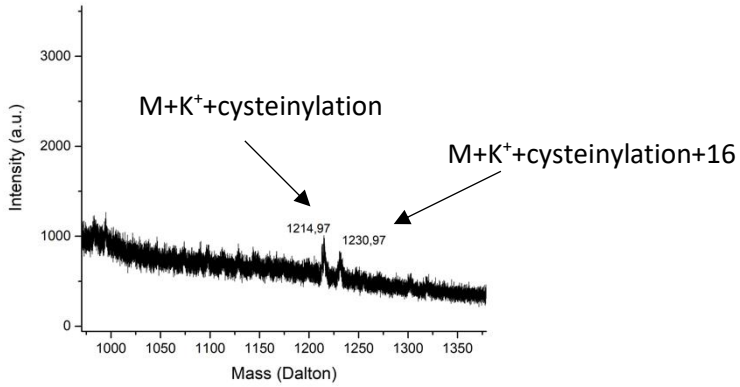


Figure 9.48 MALDI-TOF spectrum of translated and NBS oxidized peptide **2C**. The cysteinylation peak: 1056.43 Da + 39.10 Da + 119 Da = 1214.54 Da and the oxidized cysteinylation peak: 1056.43 Da + 39.10 Da + 119 Da + 16 Da = 1230.54 Da are indicated in the spectrum.

Peptide 2H

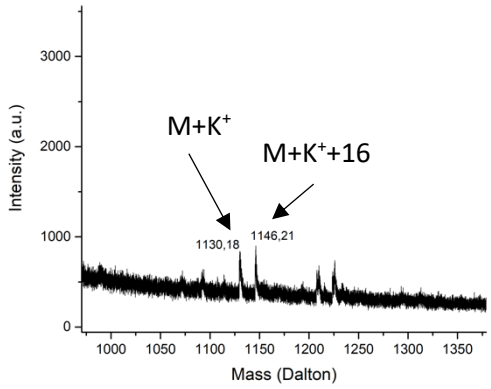


Figure 9.49 MALDI-TOF spectrum of translated and NBS oxidized peptide **2H**.The $M+K^+$ peak: 1090.48 Da + 39.10 Da = 1129.20 Da and $M+K^+$+16 peak: 1090.48 Da + 39.10 Da + 16 Da = 1145.58 Da are indicated in the spectrum.

Peptide 2K

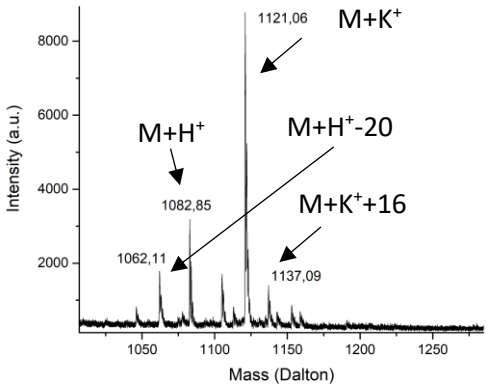


Figure 9.50 MALDI-TOF spectrum of translated and NBS oxidized peptide **2K** . The $M+H^+$-20 peak: 1081.52 Da + 1.01 Da -20 Da = 1062.53 Da, $M+H^+$ peak: 1081.52 Da + 1.01 Da = 1082.53 Da, $M+K^+$ peak: 1081.52 Da + 39.10 Da = 1120.62 Da, $M+K^+$+16 peak: 1081.52 Da + 39.10 Da +16 Da = 1136.62 Da are indicated in the spectrum.

Peptide 2R

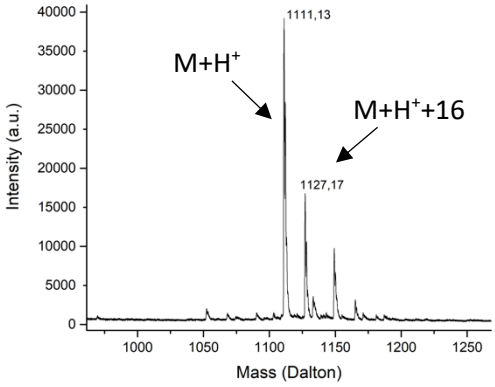


Figure 9.51 MALDI-TOF spectrum of translated and NBS oxidized peptide **2R** . $M+H^+$ peak: 1109.53 Da + 1.01 Da = 1110.54 Da, $M+H^+$+16 peak: 1109.53 Da + 1.01 Da +16 Da = 1126.62 Da are indicated in the spectrum.

Peptide 2S

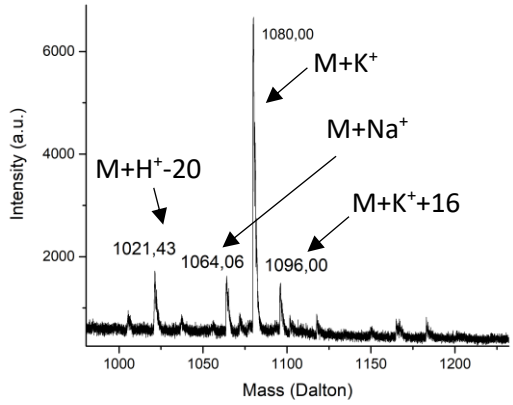


Figure 9.52 MALDI-TOF spectrum of translated and NBS oxidized peptide **2S**. The $M+H^+$-20 peak: 1040.46 Da + 1.01 Da -20 Da = 1021.47 Da, $M+H^+$ peak: 1040.46 Da + 1.01 Da = 1041.47 Da, $M+K^+$ peak: 1040.46 Da + 39.10 Da = 1079.56 Da, $M+K^+$+16 peak: 1040.46 Da + 39.10 Da +16 Da = 1095.56 Da are indicated in the spectrum.

Peptide 2Y

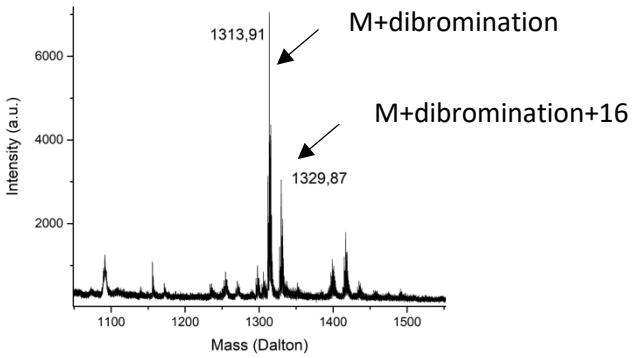


Figure 9.53 MALDI-TOF spectrum of translated and NBS oxidized peptide **2Y**. The dibromination peak: 1116.49 Da + 158 Da +39.10 Da = 1313.59 Da and the oxidized dibromination peak: 1116.49 Da + 158 Da + 39.10 Da + 16 Da = 1329.59 Da are indicated in the spectrum.

9.2.3.6 LC-MS data on peptides synthesized via SPPS

Peptide 1G

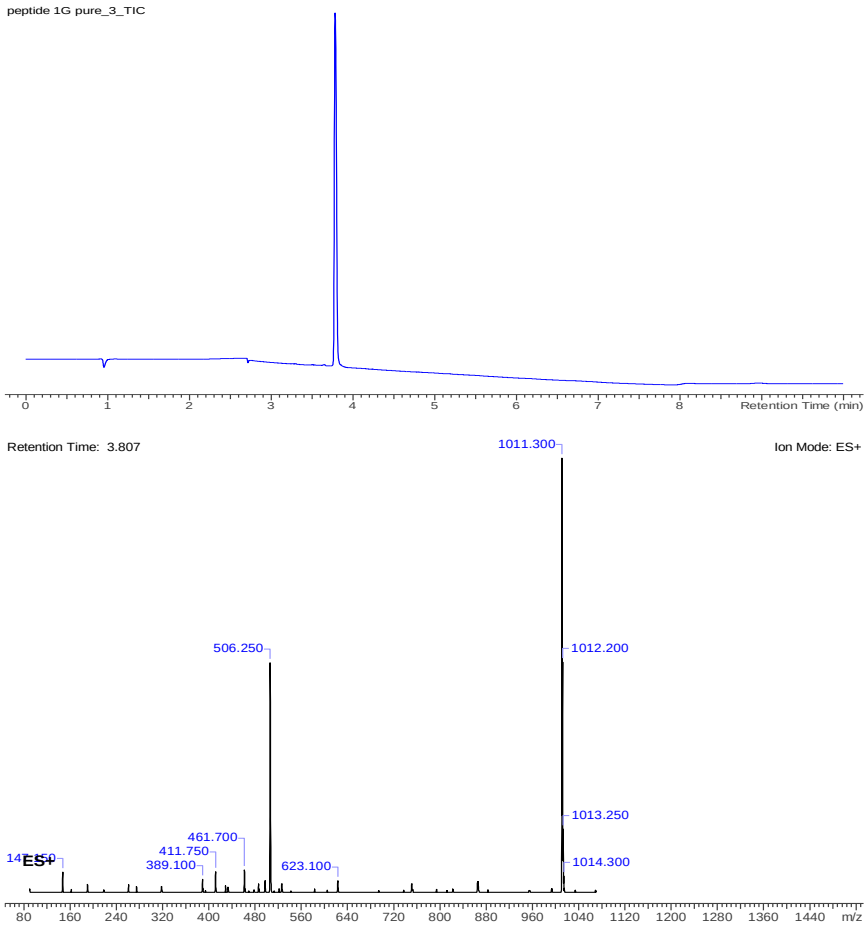


Figure 9.54 Pure LC chromatogram at 214 nm (top) with peptide **1G** eluting at t_R: 3.607 min. ESI-MS spectrum (bottom) at t_R: 3.607 $[M+H]^+$ = 1011.3.

Peptide 1K

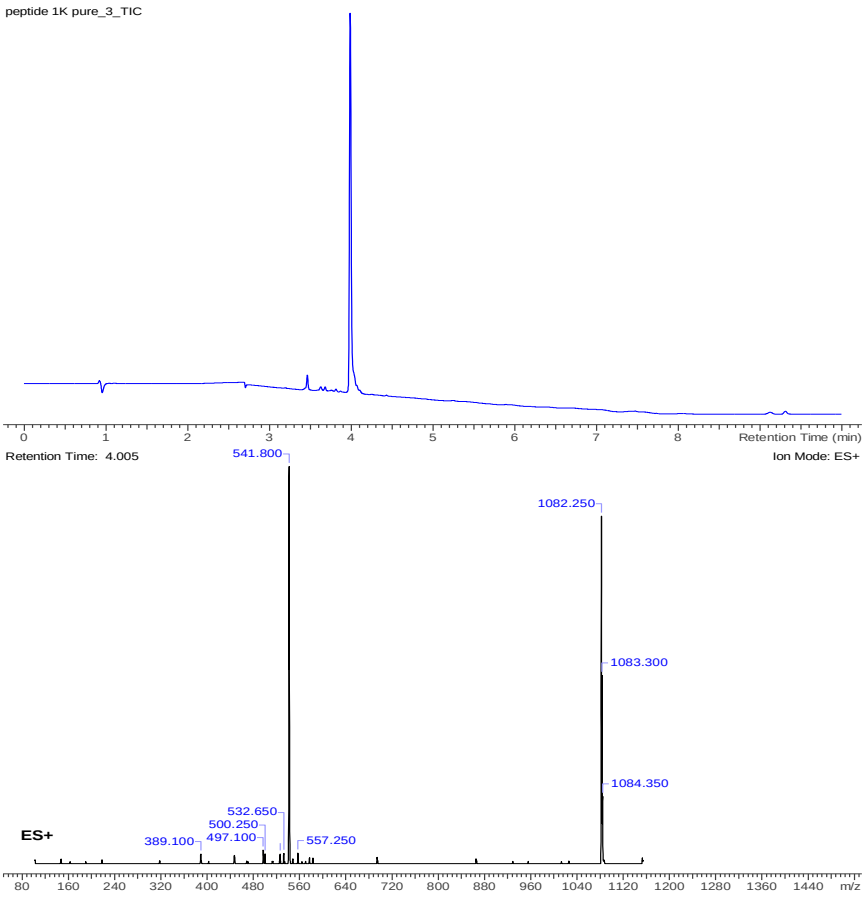


Figure 9.55 Pure LC chromatogram at 214 nm (top) with peptide **1K** eluting at t_R: 4.006 min. ESI-MS spectrum (bottom) at t_R: 4.006 $[M+H]^+$ = 1082.3.

Peptide 1S

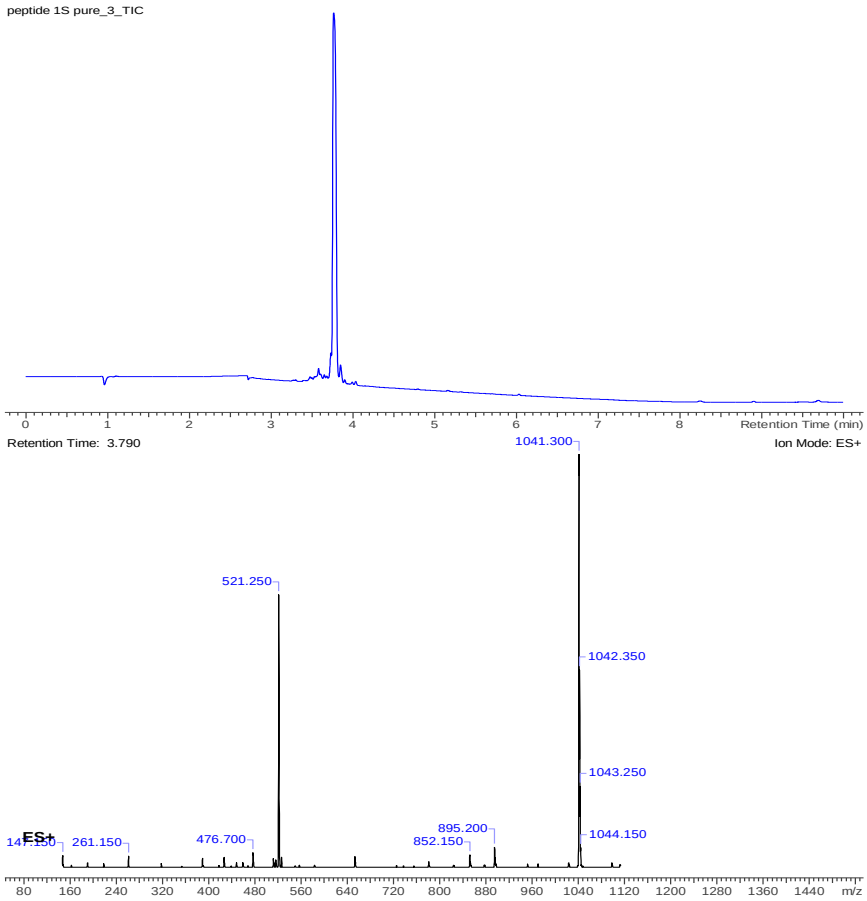


Figure 9.56 Pure LC chromatogram at 214 nm (top) with peptide **1S** eluting at t_R: 3.790 min. ESI-MS spectrum (bottom) at t_R: 3.790 $[M+H]^+$ = 1041.3.

Peptide 5K

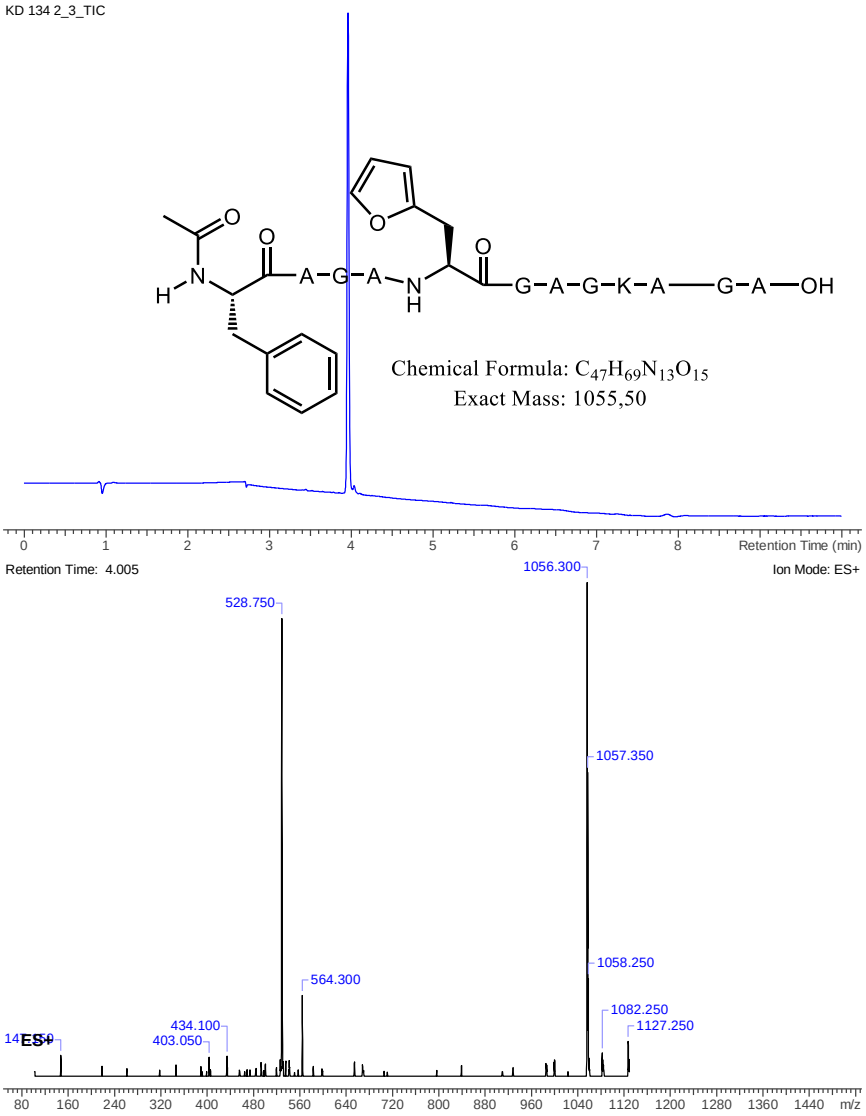


Figure 9.57 Pure LC chromatogram at 214 nm (top) with peptide **5K** eluting at t_R: 4.005 min. ESI-MS spectrum (bottom) at t_R: 4.005 $[M+H]^+$ = 1056.3.

Peptide 4K

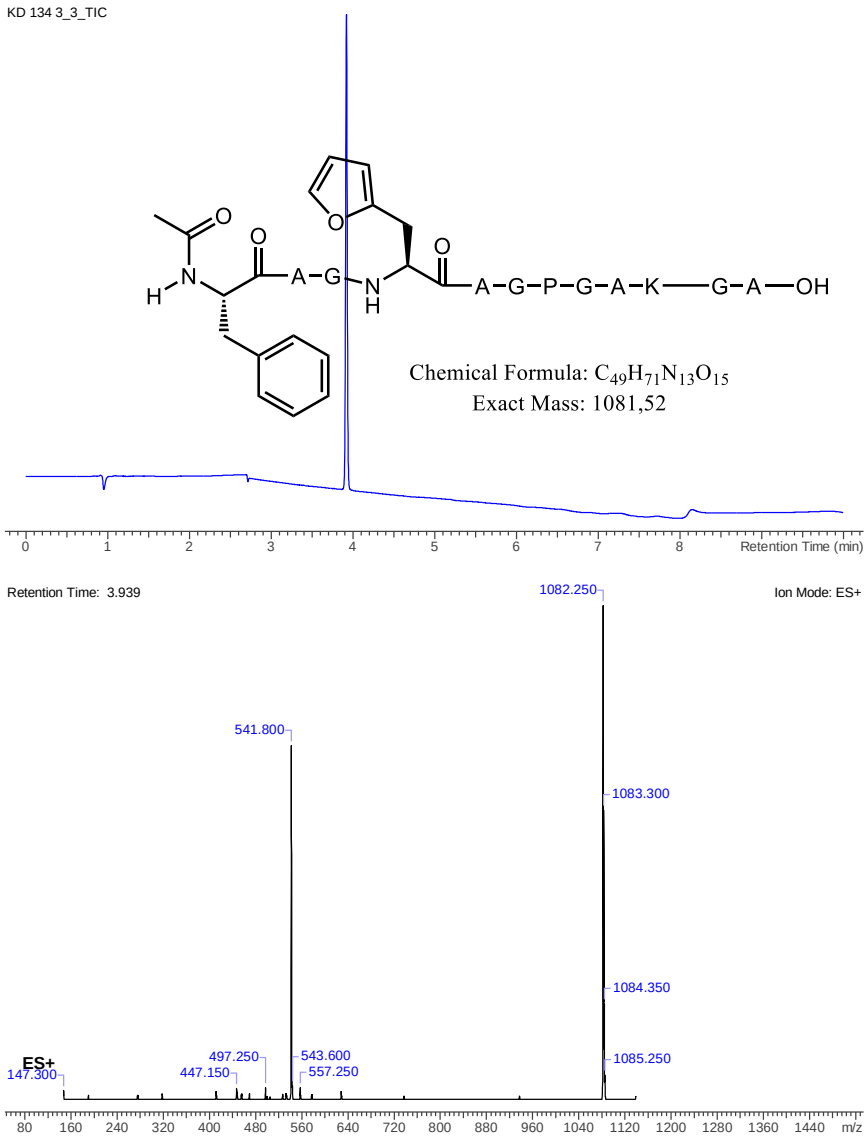


Figure 9.58 Pure LC chromatogram at 214 nm (top) with peptide **4K** eluting at t_R: 3.939 min. ESI-MS spectrum (bottom) at t_R: 3.939 $[M+H]^+$ = 1082.3.

9.2.3.7 LC-MS data on SPPS peptides after NBS oxidation

Peptides were solubilized in MQ H_2O with a concentration of 2.5 mM. 10 µL of this peptide solution was added to 10 µL NaOAc buffer (pH 5.2) and 3 eq of NBS was added for the oxidation reaction. In the case of **2S** the amount of peptide used was higher and this resulted in incomplete peptide oxidation. The MS spectrum of the oxidation peak shows a mass which corresponds to $[M+H]^+$-2 Da this is can be caused by the oxidation reaction (+16 Da) followed by a subsequent loss of water (-18 Da) during the ionisation. To prove this hypothesis we compared the extracted ion chromatograms of the $[M+H]^+$-2 and $[M+H]^+$+16 ions. In both **2G** and **2S** the chromatograms indicate that both ions are resulting from the same compound in the LC, while for **2K** it is clear that the $[M+H]^+$-2 ion comes from another compound in the LC. This is an expected result as the amine side chain of the lysine residue can form an imine with the oxidized furan moiety leading to a mass loss of -2 Da.

Doing the oxidation experiments in presence of NaOAc buffer always leads to 2 large peaks in the LC chromatogram at 214 nm. Those peaks were present in all experiments and were not peptide related.

Peptide 2G

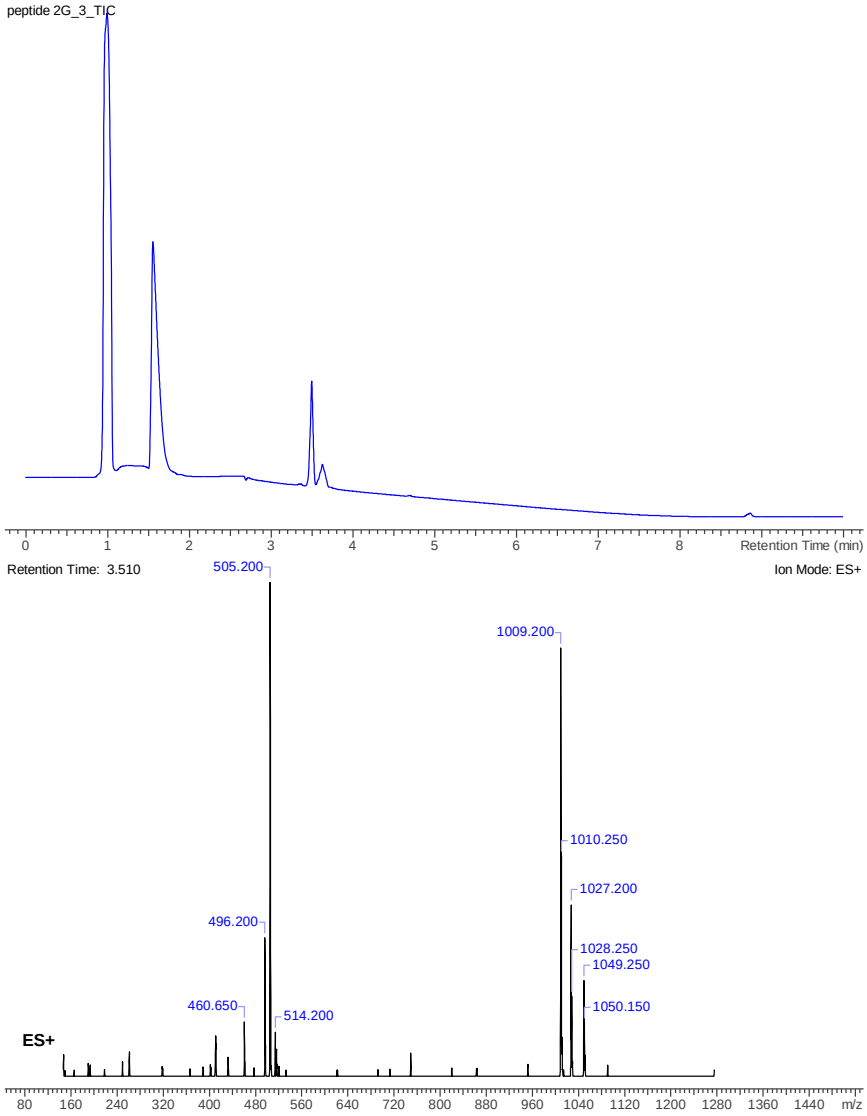


Figure 9.59 LC-MS chromatogram at 214 nm (top) of oxidized peptide **2G** eluting at t_R: 3.510 min. ESI-MS spectrum (bottom) at t_R: 3.510 $[M+H]^+$+16 = 1027.2; $[M+H]^+$+16-18 = 1009.2.

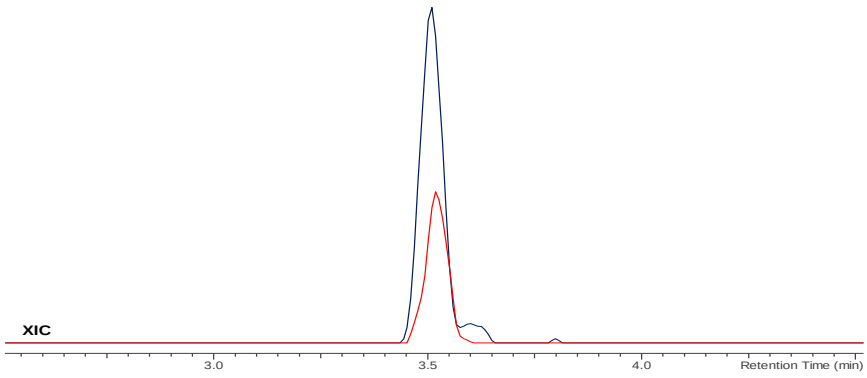


Figure 9.60 Extracted ion chromatogram (XIC) for $[M+H]^+$+16 = 1027.2 (red); $[M+H]^+$+16-18 = 1009.2 (blue).

Peptide 2K

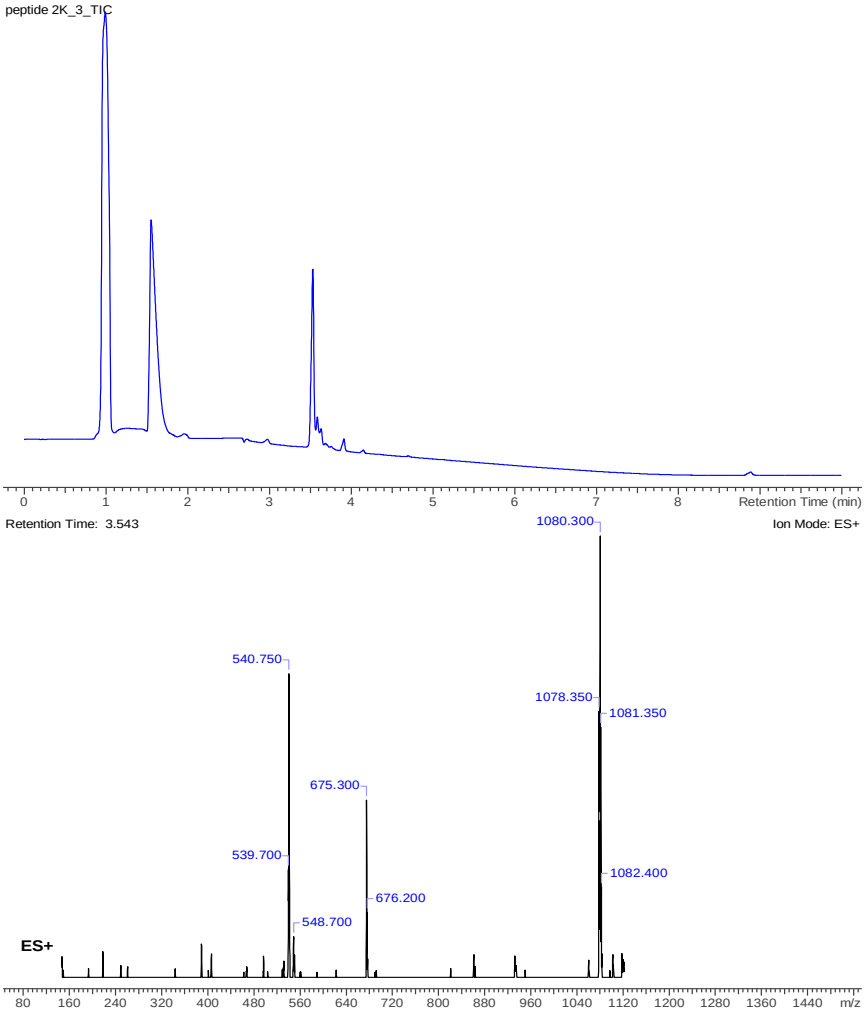


Figure 9.61 LC-MS chromatogram at 214 nm (top) of oxidized peptide **2K** eluting at t_R: 3.543 min. ESI-MS spectrum (bottom) at t_R: 3.543 $[M+H]^+$+16-18 = 1080.3.

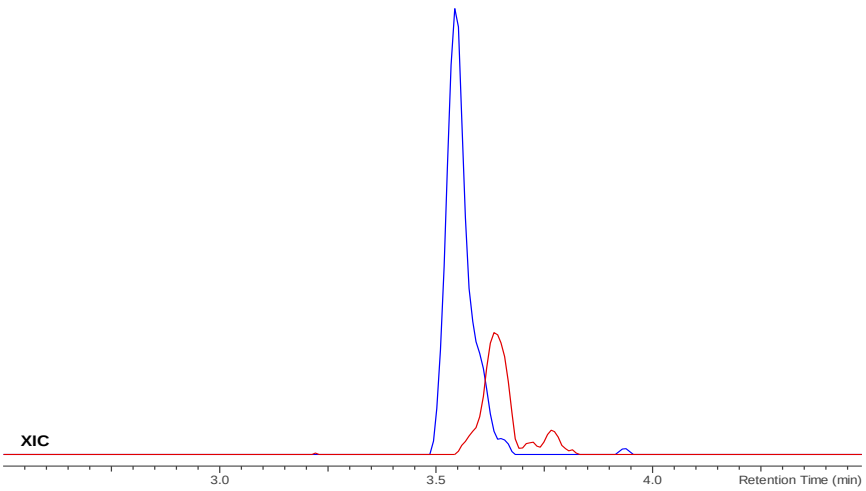


Figure 9.62 Extracted ion chromatogram (XIC) for $[M+H]^+$+16 = 1098.3 (red); $[M+H]^+$+16-18 = 1080.3 (blue).

Peptide 2S

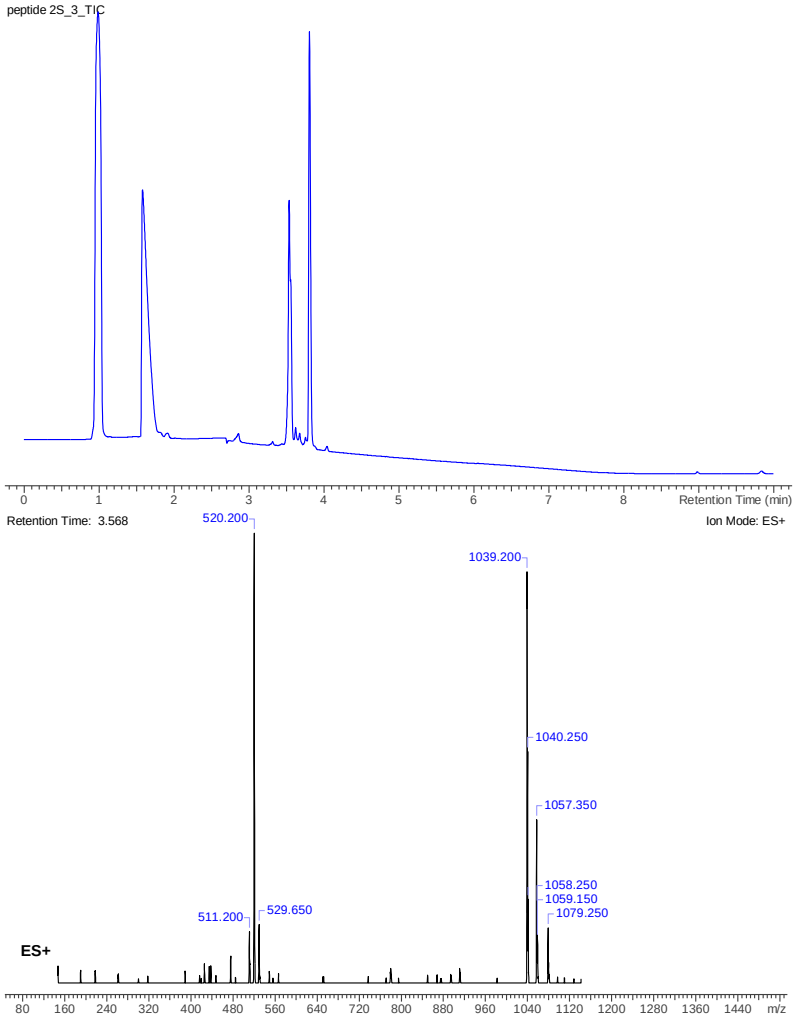


Figure 9.63 LC-MS chromatogram at 214 nm (top) of oxidized peptide **2S** eluting at t_R: 3.568 min. ESI-MS spectrum (bottom) at t_R: 3.538 $[M+H]^+$+16 = 1057.4 ; $[M+H]^+$+16-18 = 1039.3.

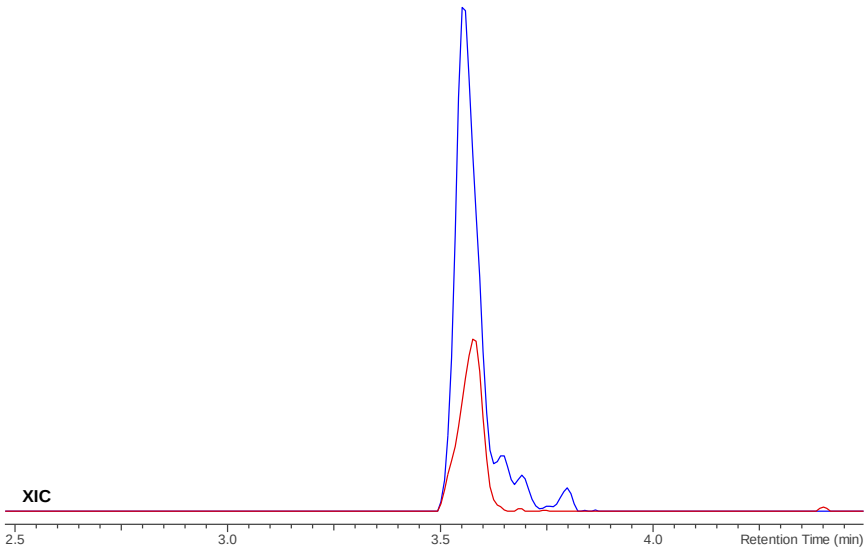


Figure 9.64 Extracted ion chromatogram (XIC) for $[M+H]^+$+16 = 1057.3 (red); $[M+H]^+$+16-18 = 1039.3 (blue).

9.2.3.8 LC-MS data after one pot oxidation reduction

The setup for the one pot oxidation and reduction reactions of peptides **1K**, **1G** and **1S** was very similar to the oxidation reaction. The same peptide stock solutions of 2.5 mM were used, and the oxidation reaction was done exactly the same as in 9.1.3.7. Following the addition of NBS the sample was mixed well and afterwards 100 eq of $NaCNBH_3$ was added to reduce the formed imine. The samples were analysed by LC-MS.

Peptide 3G

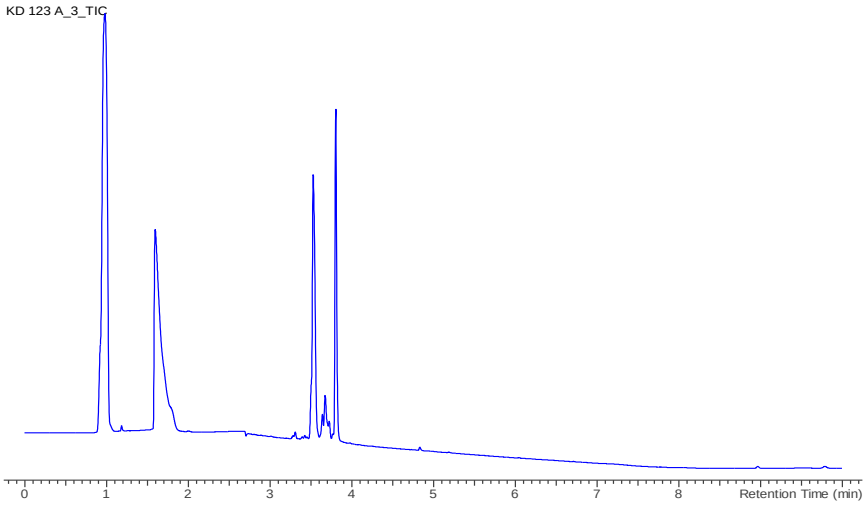


Figure 9.65 LC-MS chromatogram at 214 nm for **3G**, starting peptide **1G** eluting at t_R: 3.832 and oxidized peptide **2G** eluting at t_R: 3.551.

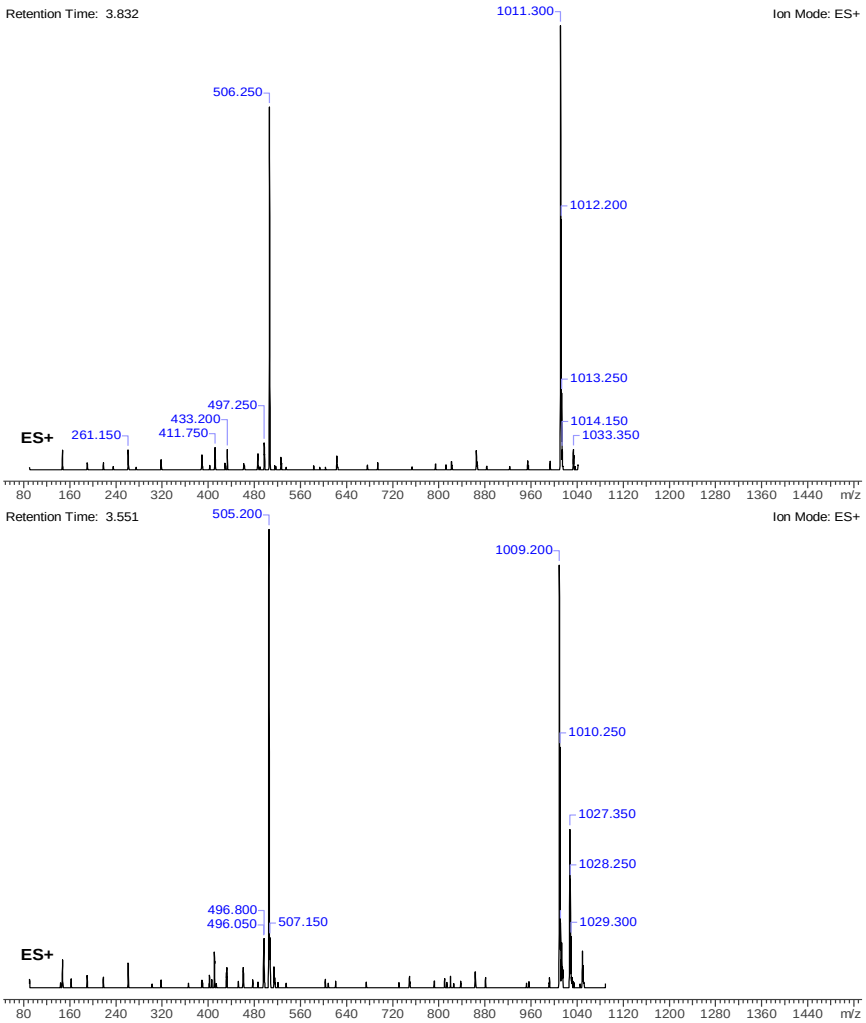


Figure 9.66 ESI-MS spectrum at t_R: 3.832 (top) and t_R: 3.551 (bottom).

peptide 3S

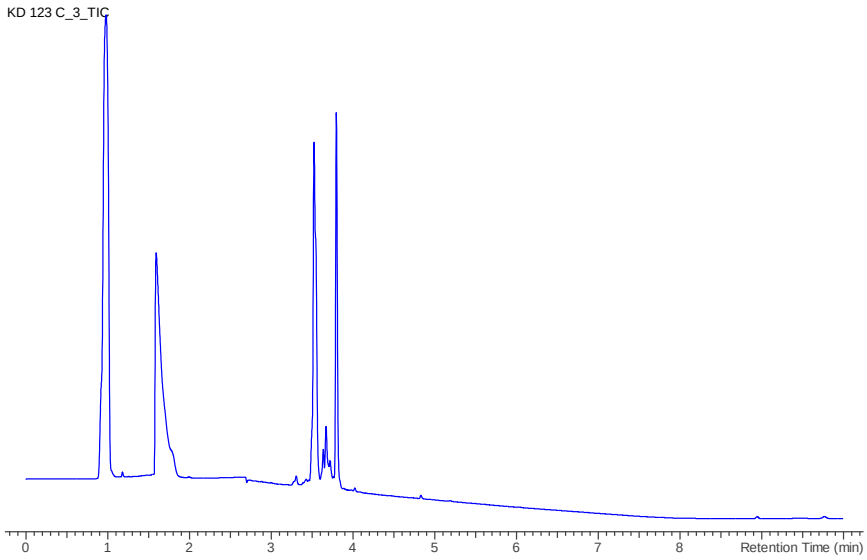


Figure 9.67 LC-MS chromatogram at 214 nm for **3S**, starting peptide **1S** eluting at t_R: 3.815 and oxidized peptide **2S** eluting at t_R: 3.552.

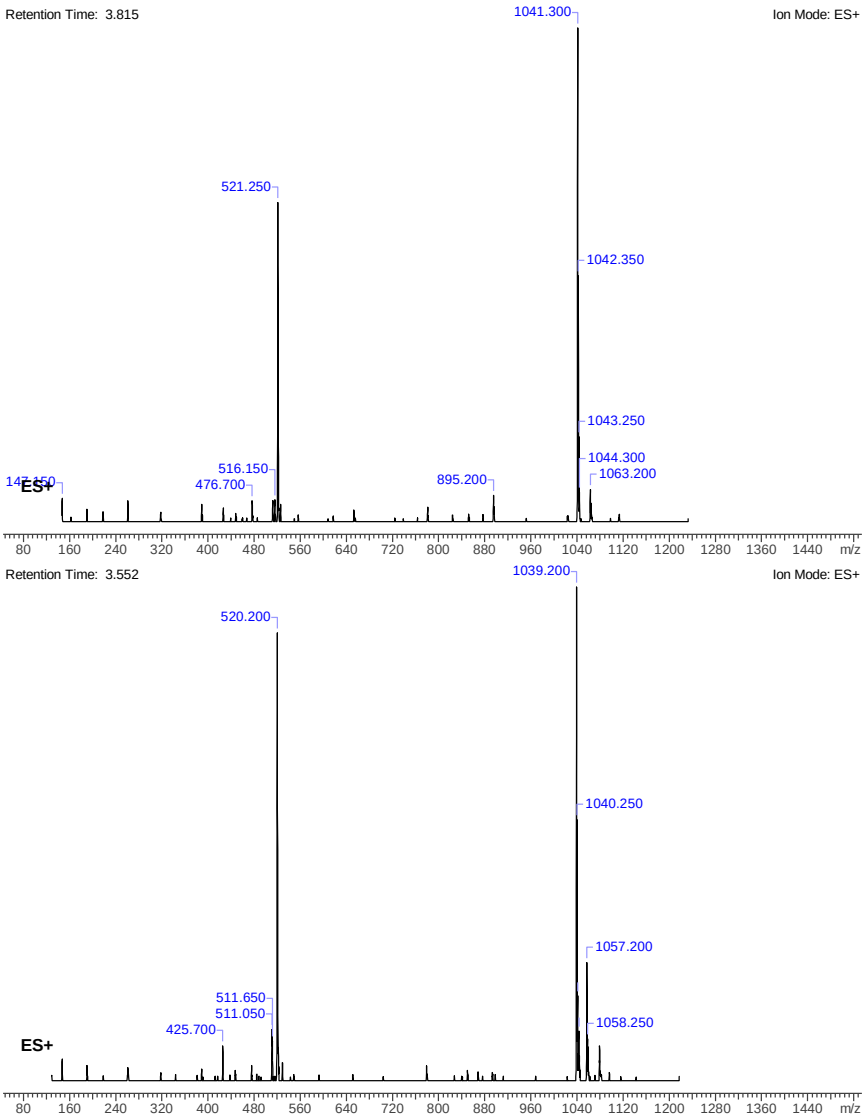


Figure 9.68 ESI-MS spectrum at t_R: 3.815 (top) and t_R: 3.552 (bottom).

Peptide 3K

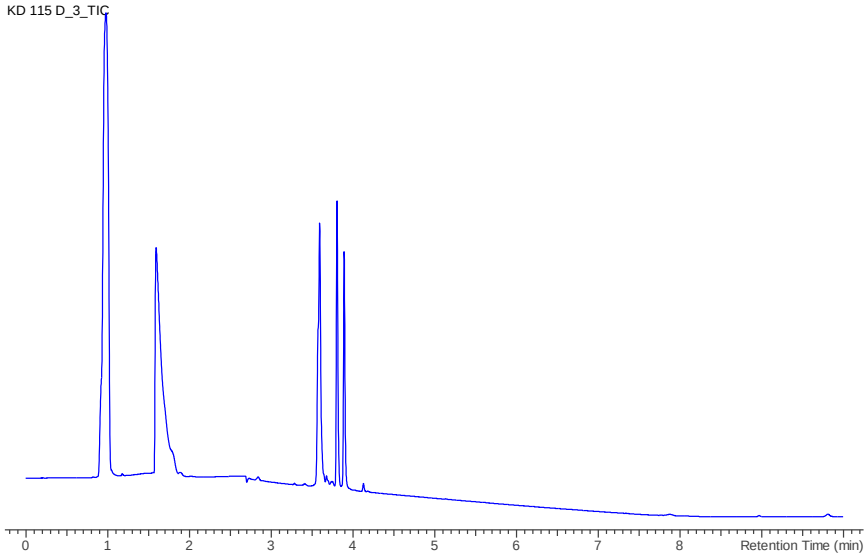


Figure 9.69 LC-MS chromatogram at 214 nm for **3K**, starting peptide **1K** eluting at t_R: 3.914 and oxidized peptide **2K** eluting at t_R: 3.617 and final product **3K** eluting at t_R: 3.831.

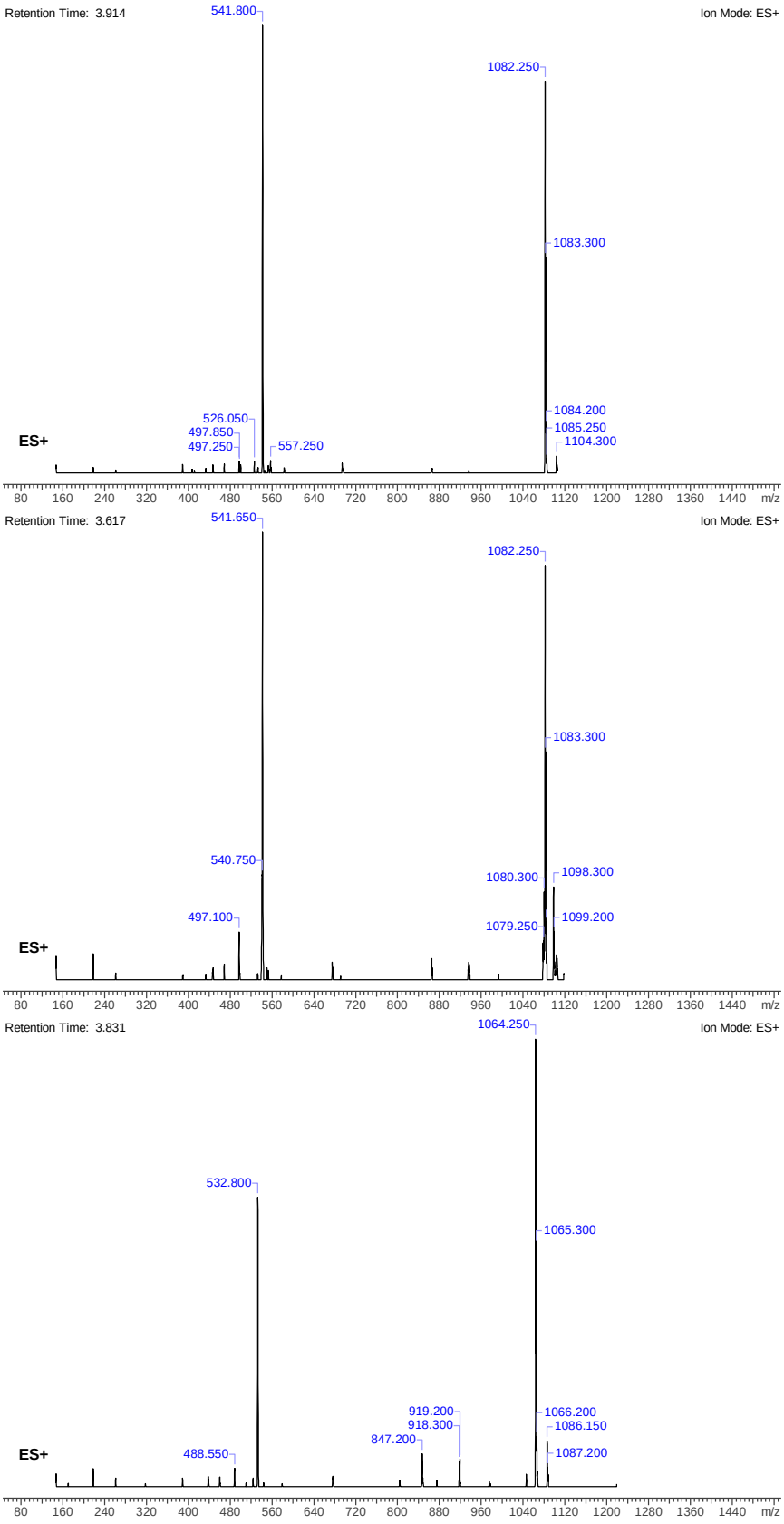


Figure 9.70 ESI-MS spectrum at t_R: 3.914 (top), t_R: 3.617 (middle) and t_R:3.831 (bottom).

Peptide 5K

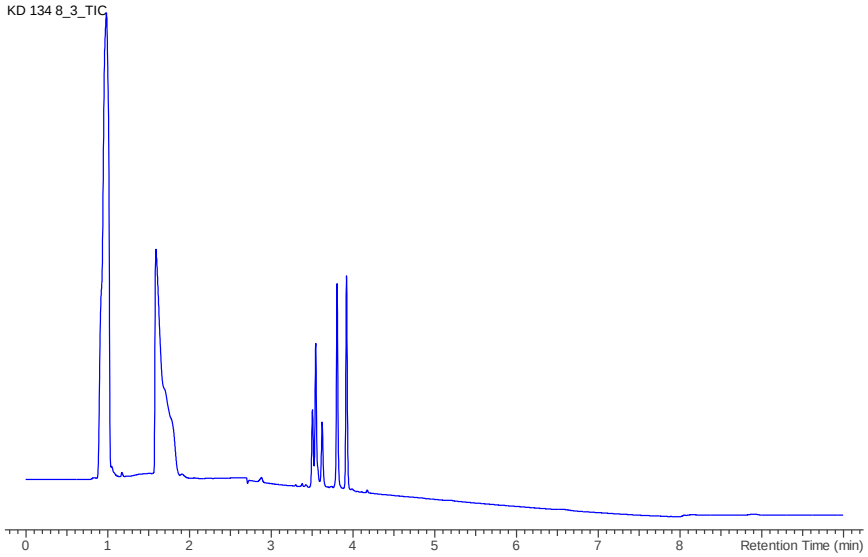


Figure 9.71 LC-MS chromatogram at 214 nm for **5K**, starting peptide **5K** eluting at t_R: 3.939 and oxidized peptide eluting around t_R: 3.568 and final product **7K** eluting at t_R: 3.824.

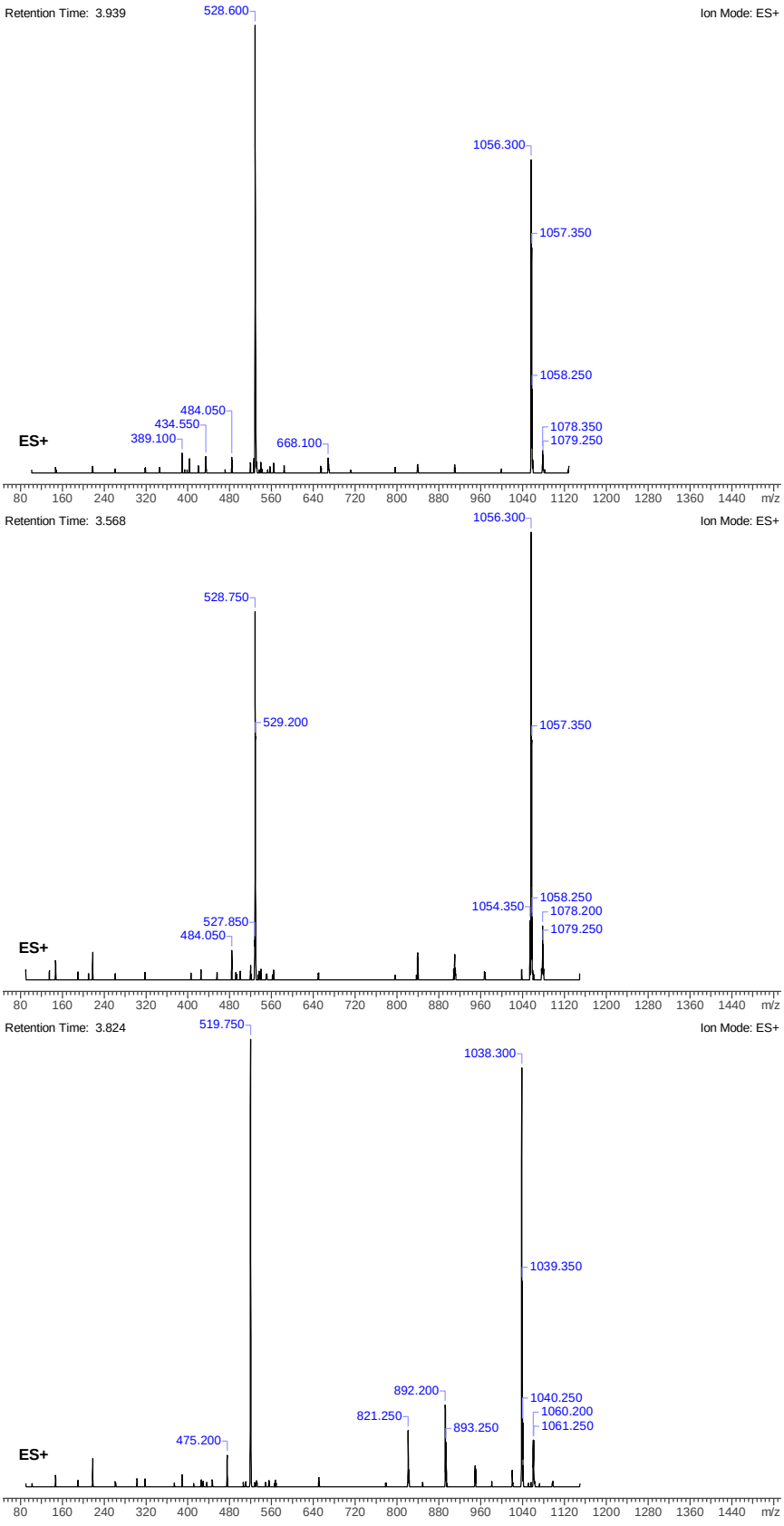


Figure 9.72 ESI-MS spectrum at t_R: 3.939 (top), t_R: 3.568 (middle) and t_R:3.824 (bottom).

Peptide 4K

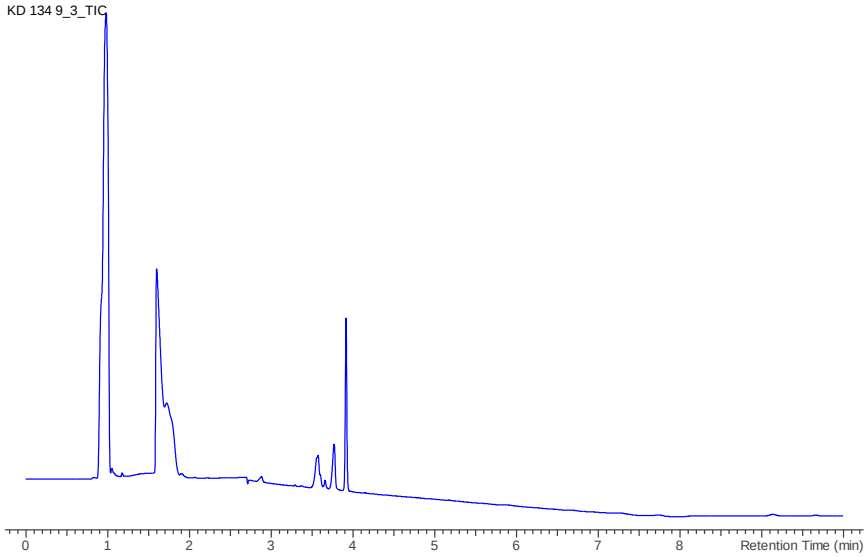


Figure 9.73 LC-MS chromatogram at 214 nm for **4K**, starting peptide **4K** eluting at t_R: 3.930 and the oxidized peptide eluting around t_R: 3.584 and final product **6K** eluting at t_R: 3.782.

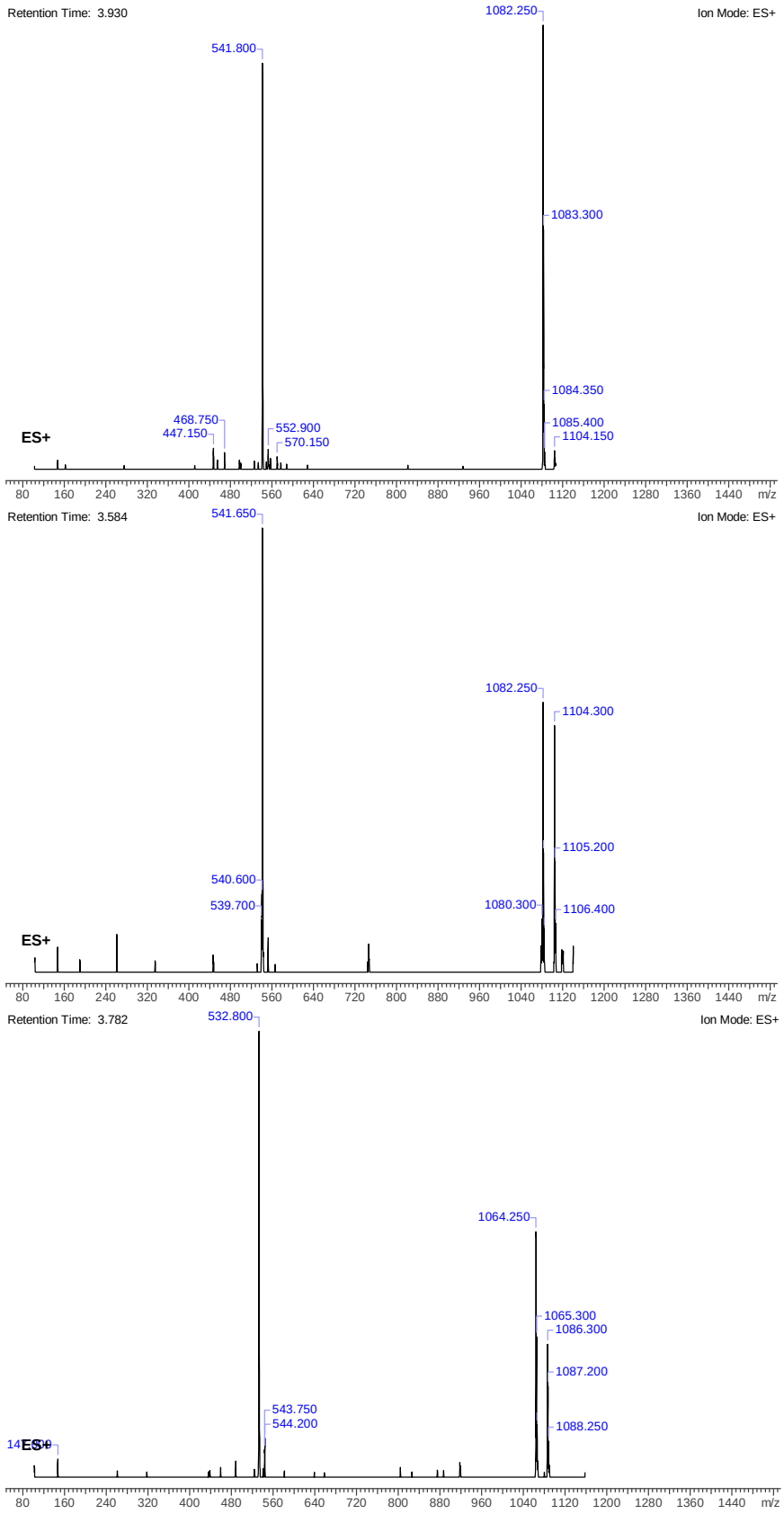


Figure 9.74 ESI-MS spectrum at t_R: 3.930 (top), t_R: 3.584 (middle) and t_R:3.782 (bottom).

9.2.3.9 NMR analysis of peptide 6K

To confirm the proposed pyrrole linkage for the cyclisation NMR analysis was performed. Larger scale reaction was carried out and purified via Semi-PrepHPLC Luna C18 column, 0-100 ACN in 15 min.

The 1D ^{1}H and 2D ^{1}H-^{1}H, ^{1}H-^{13}C experiments were recorded on a Bruker 700.13MHz Avance II spectrometer equipped with a 5 mm $^{1}H/^{19}F$ ^{13}C ^{15}N Prodigy TCI cryoprobe. For each sample measurement, the sample temperature was set at 25° (and controlled within ± 0.1 °c with a Eurotherm 2000 VT controller. The spectra recorded on the samples included 1D ^{1}H, ^{1}H-{^{1}H} TOCSY (80ms spinlock), ^{1}H-{^{13}C} HSQC and ^{1}H-{^{13}C} HMBC spectra with an 8Hz long range coupling constant and ^{1}H-{^{1}H} off-resonance ROESY (200ms spinlock). The 2D 1H spectra were recorded with excitation sculpting as the method of choice for the suppression of the residual ^{1}H H_2O signal. No ^{13}C spectra were recorded. All spectra were processed using TOPSPIN 3.5 pl7. All spectra were referenced with respect to the internal TSP standard signal, residing at 0ppm. Standard pulse programs from the Bruker pulse program library were used throughout. The dry sample (±1mg) were dissolved in ± 600 µl 90%/10% H_2O/D_2O (99,96% D) solvent mixture + 0,05mM TSP (Trimethylsilylpropanoic acid) for internal chemical shift referencing.

Assignment of the gathered NMR data (Table 9.1) comfirmed the proposed pyrrole structure. ROESY signals indicating that the proton on C5 of the pyrrole unit is in close proximity (<5 Å) to two CH_2 protons of the lysine side chain. On the other hand TOCSY cross peaks of low intensity were identified between the CH_2 of the former furylalanine and the proton on the C3 of the pyrrole unit. These findings confirm the covalent connection of the pyrrole unit to the peptide backbone on positions 4 and 10.

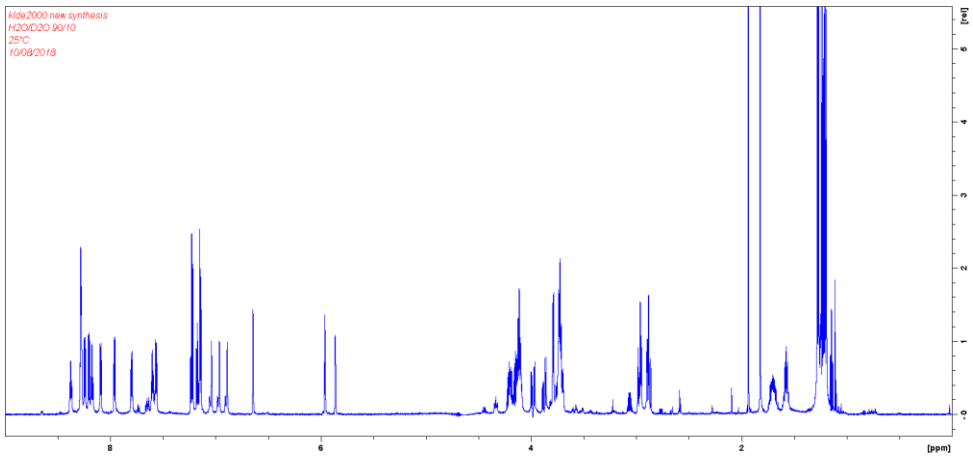


Figure 9.75 ^{1}H NMR spectrum of peptide **6K** dissolved in 90/10 H_2O/D_2O.

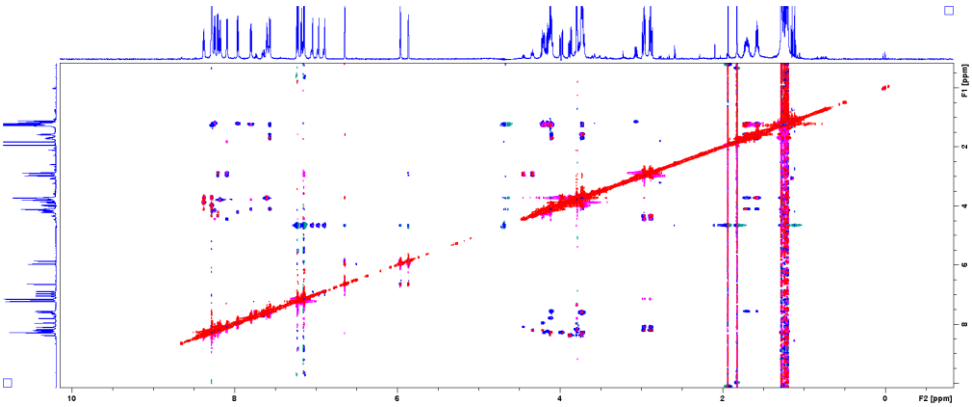


Figure 9.76 TOCSY (blue/green) and ROESY (red/pink) overlay 2D NMR spectrum of peptide **6K** dissolved in 90/10 H_2O/D_2O.

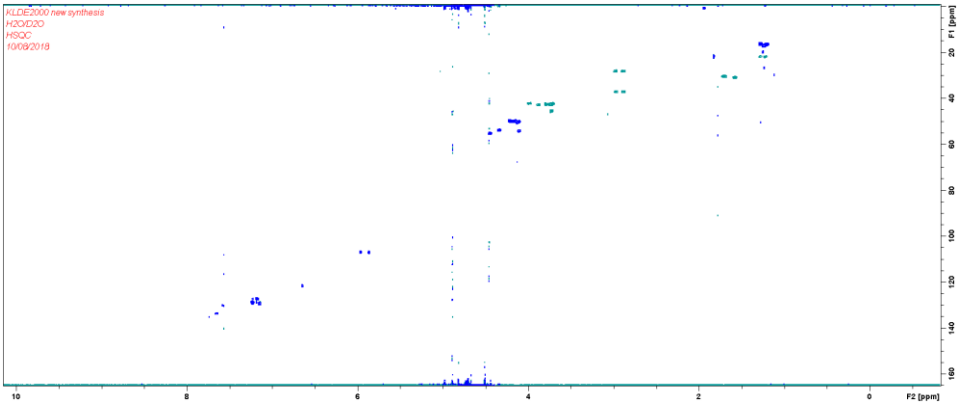


Figure 9.77 HSQC 2D NMR spectrum of peptide **6K** dissolved in 90/10 H_2O/D_2O

Table 9.1 ^{1}H and ^{13}C assignment of peptide **6K** (dissolved in 90/10 H_2O/D_2O, 700 MHz). Residues 4 and 10, where the pyrrole unit is connected to the peptide backbone, are marked with an asterisk.

residue		^{1}H δ (ppm)	^{13}C δ (ppm)	residue		^{1}H δ (ppm)	^{13}C δ (ppm)
Phe 1	Cα	4.3311	53.7453	Pro 7	$_5CH_2$	3.0691	46.7821
	CH_2	2.9273	36.9783				
	NH	8.2					
				Gly 8	Cα	3.7929	42.3807
Ala 2	Cα	4.1196	50.3152		NH	8.17	
	CH_3	1.2759	16.2085				
	NH	8.28		Ala 9	Cα	4.1090	50.1739
					CH_3	1.2381	17.0170
Gly 3	Cα	3.9873/3.7146	42.09		NH	7.8	
	NH	8.28					
				Lys * 10	Cα	4.1019	54.1592
Pyrrole * 4	Cα	4.4465	55.1382		$C\beta H_2$	1.2205/1.2798	21.7659
	CH_2	2.9766/2.8801	27.9836		$C\gamma/\delta H_2$	1.5753	30.8094
	$_5CH$	6.6454	121.5040		$C\gamma/\delta H_2$	1.6998	30.3496
	$_4CH$	5.9622	106.9262		NH	7.56	
	$_3CH$	5.8646	106.9940				
	NH	8.09		Gly 11	Cα	3.7259	45.4140
					NH	7.6	
Ala 5	Cα	4.1528	49.7401				
	CH_3	1.1962	16.3453	Ala 12	Cα	4.2177	49.6695
	NH	8.24			CH_3	1.2101	16.4448
					NH	7.96	
Gly 6	Cα	3.7327/3.8788	42.4232				
	NH	8.37					

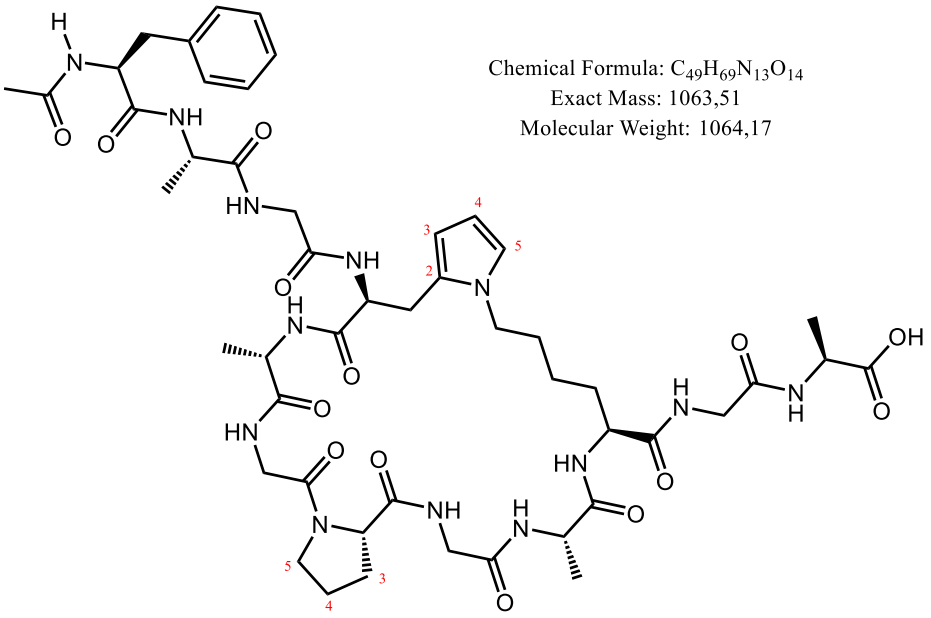


Figure 9.78 Structural representation of peptide **6K**.

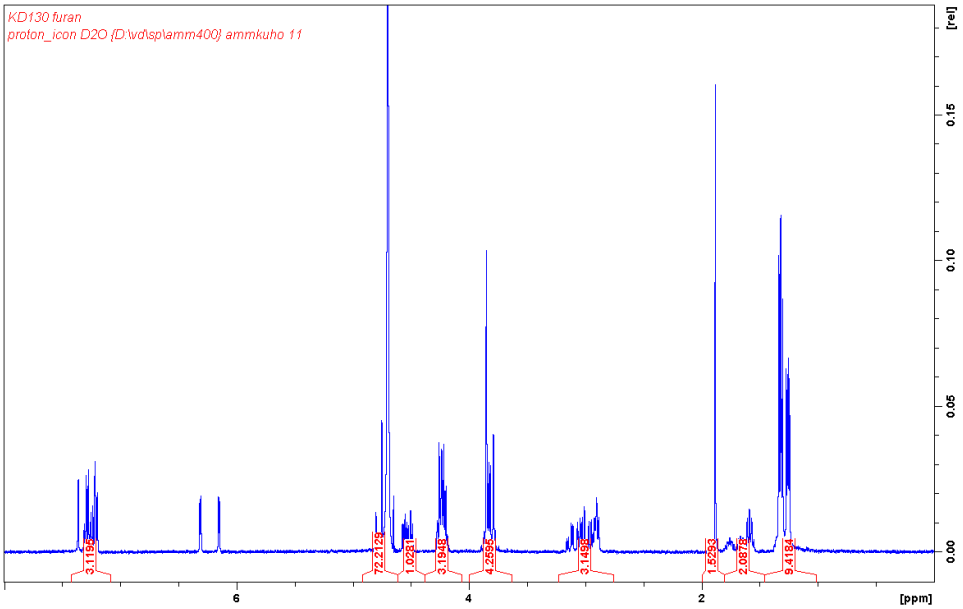


Figure 9.79 ^{1}H NMR spectrum of peptide **6K** dissolved in D_2O. Measured on 400 MHz Bruker equipment.

9.3 Experimental data for chapter 6

9.3.1 Experimental data for part 6.1

9.3.1.1 Cp lysine synthesis

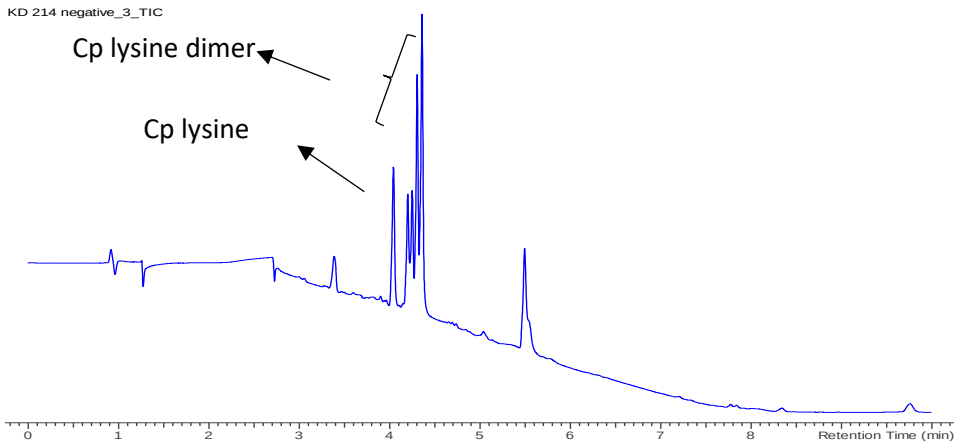


Figure 9.80 LC-MS analysis of Cp lysine synthesis. LC chromatogram 214 nm.

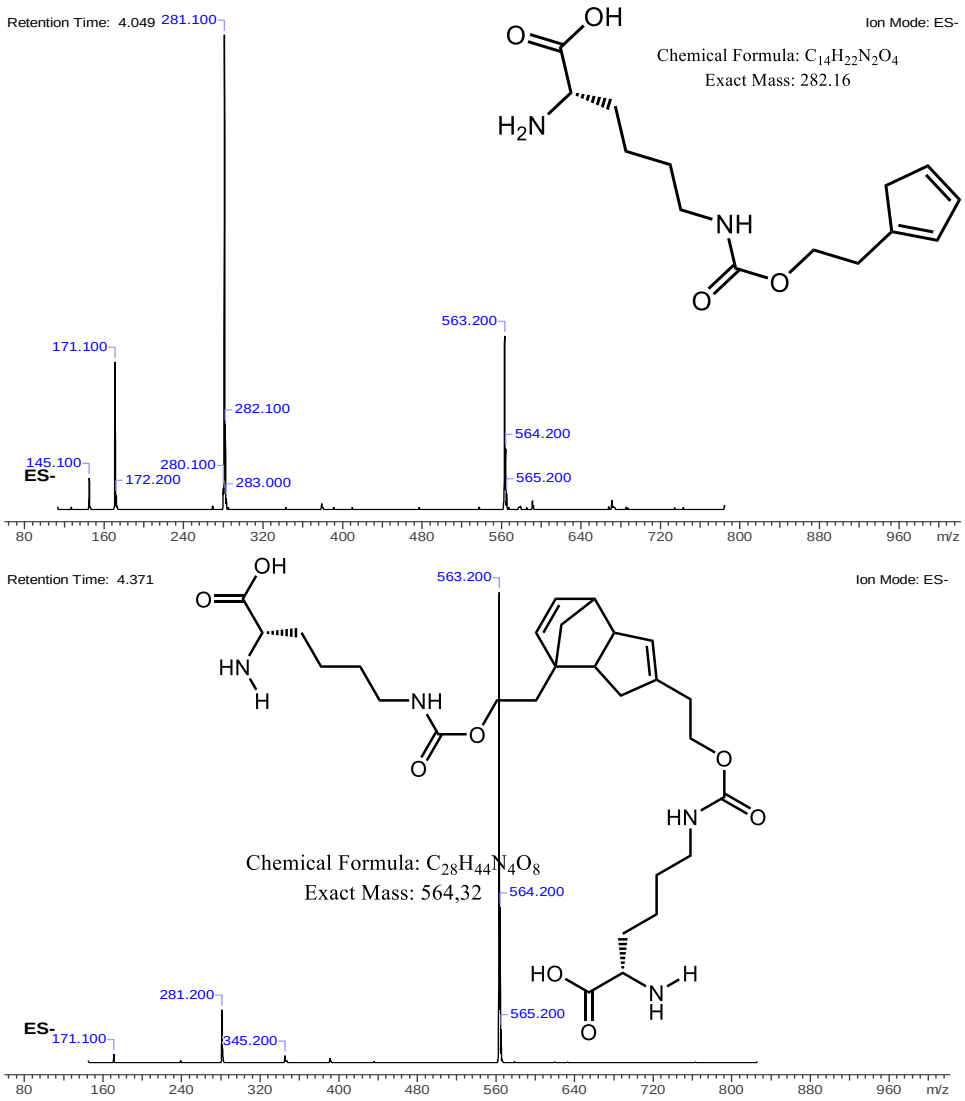


Figure 9.81 ESI (negative mode) MS spectrum at t_R: 4.049 min (top), ESI (negative mode) MS spectrum at t_R: 4.371 min (bottom). Briefly heating up during rotavapor removal of DMF caused a lot of dimer formation.

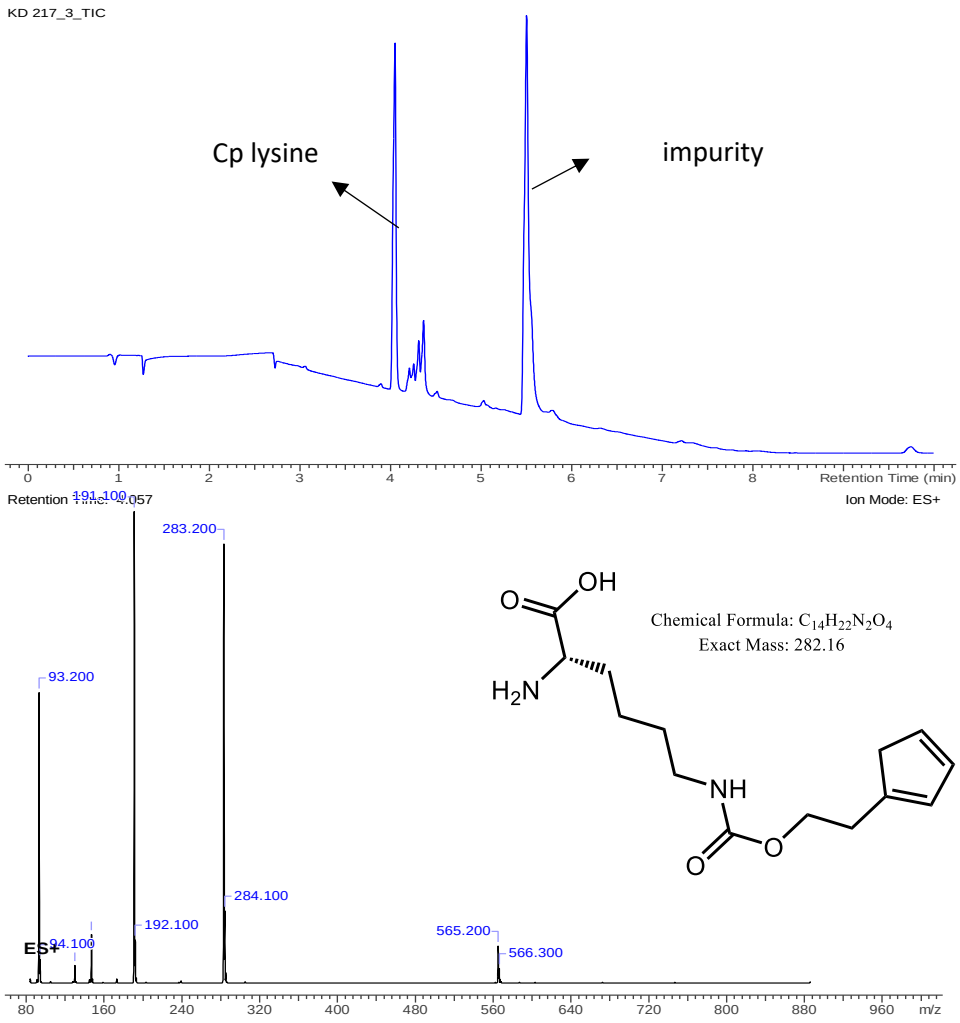


Figure 9.82 LC-MS analysis of Cp lysine synthesis. LC chromatogram 214 nm (top), ESI (positive mode) MS spectrum at t_R: 4.047 min (bottom). An unidentified impurity is abundantly present in the 214 nm chromatogram.

9.3.2 Experimental data for part 6.2.3

9.3.2.1 pH study of the tyrosine TAD reaction

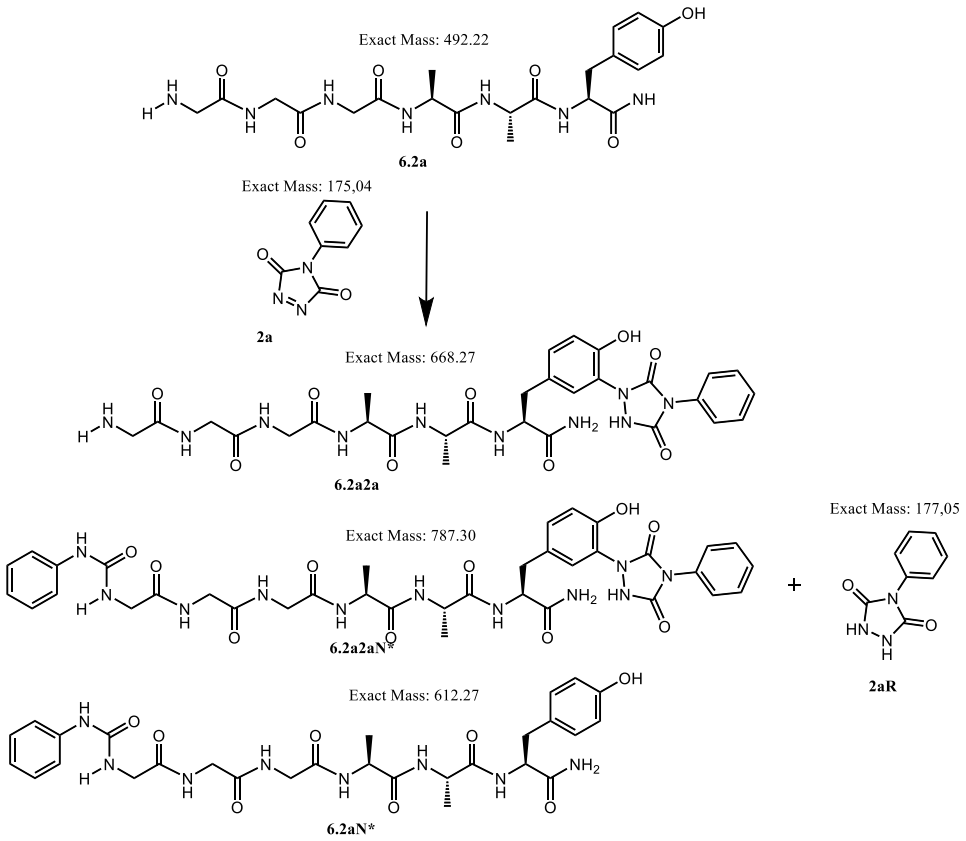


Figure 9.83 structural representation of the reaction products of GGGAAY **6.2a** reaction with PTAD **2a**.

For the reaction of GGGAAY **6.2a** with PTAD **2a** in PBS buffer at pH 7 and pH 9 we observe some side product formation on the N-terminal amide. In earlier reports the formation of a side product on the lysine amine was observed with PTAD in PBS buffers caused by the reaction of a degradation product of PTAD and the free amine. The same side product formation but on the N-terminus is taking place in our experiments.

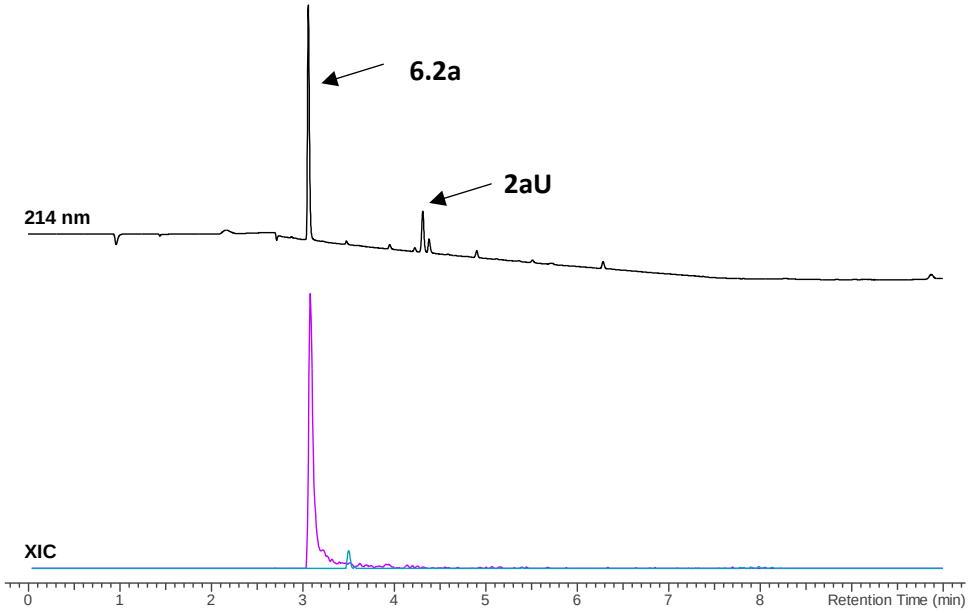


Figure 9.84 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with PTAD **2a** (2eq.) in milliQ H_2O. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2a** (669.2 Da) (bottom).

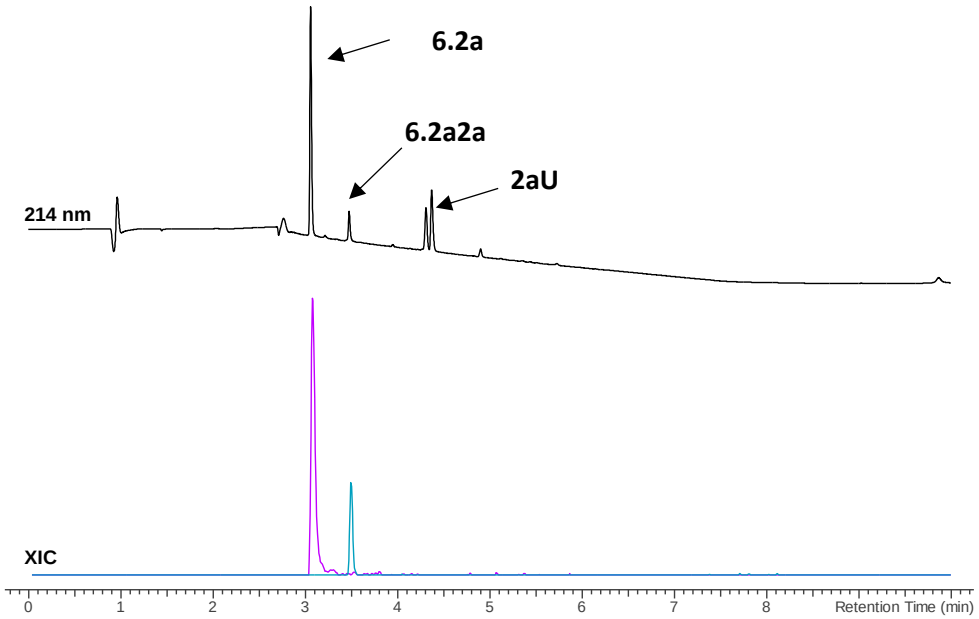


Figure 9.85 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with PTAD **2a** (2eq.) in PBS pH 5. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2a** (669.2 Da) (bottom).

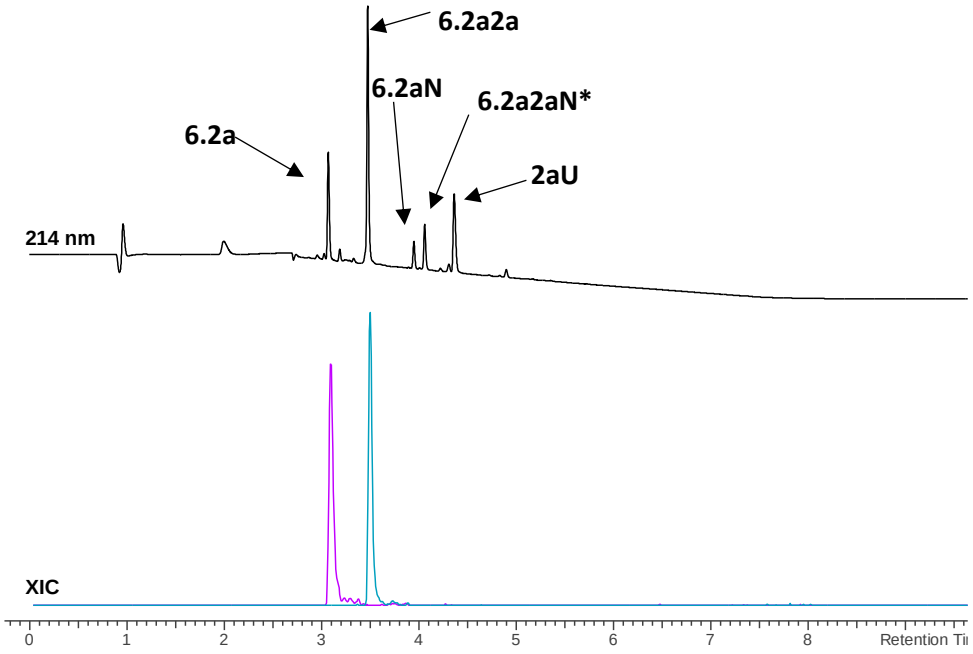


Figure 9.86 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with PTAD **2a** (2eq.) in PBS pH 7. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2a** (669.2 Da) (bottom).

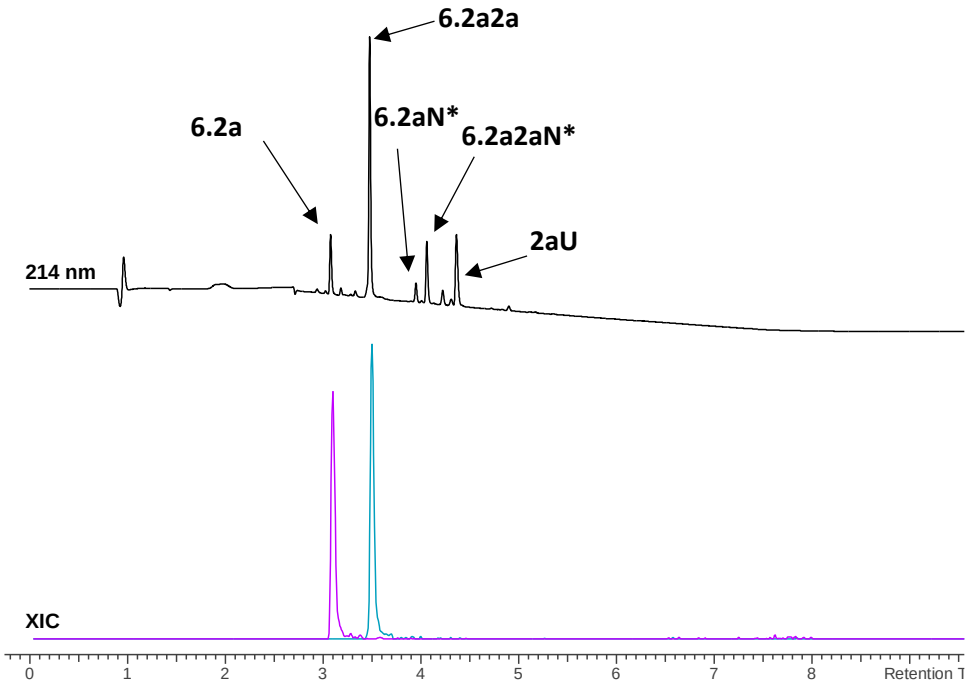


Figure 9.87 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with PTAD **2a** (2eq.) in PBS pH 9. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2a** (669.2 Da) (bottom).

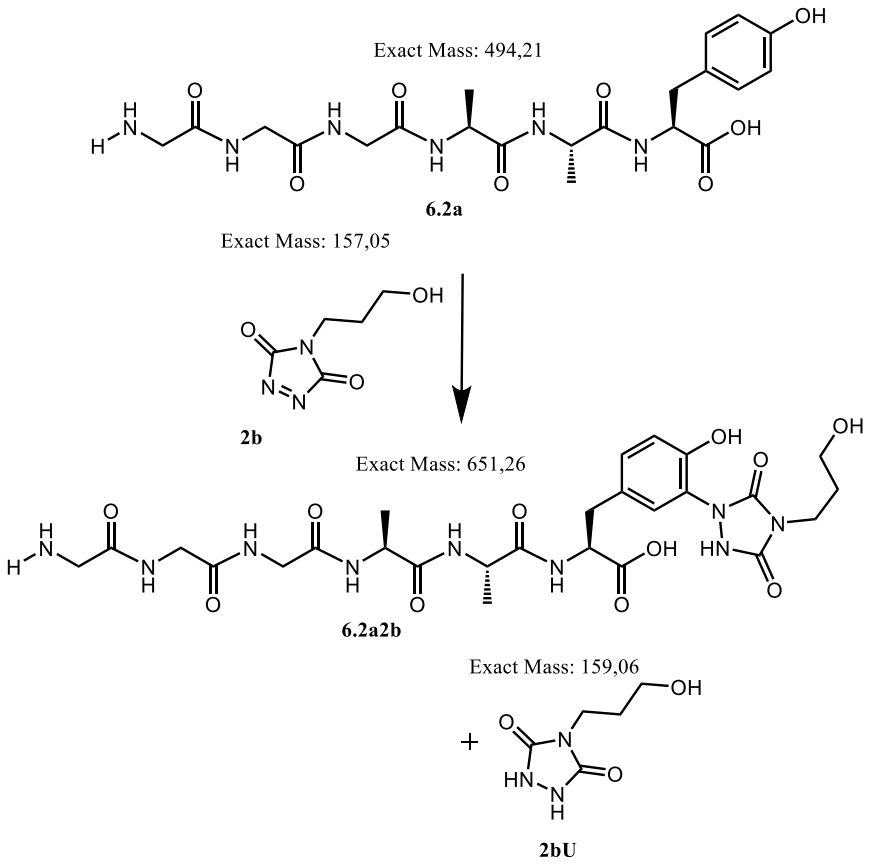


Figure 9.88 structural representation of the reaction products of GGGAAY **6.2a** reaction with TAD-propanol **2b**.

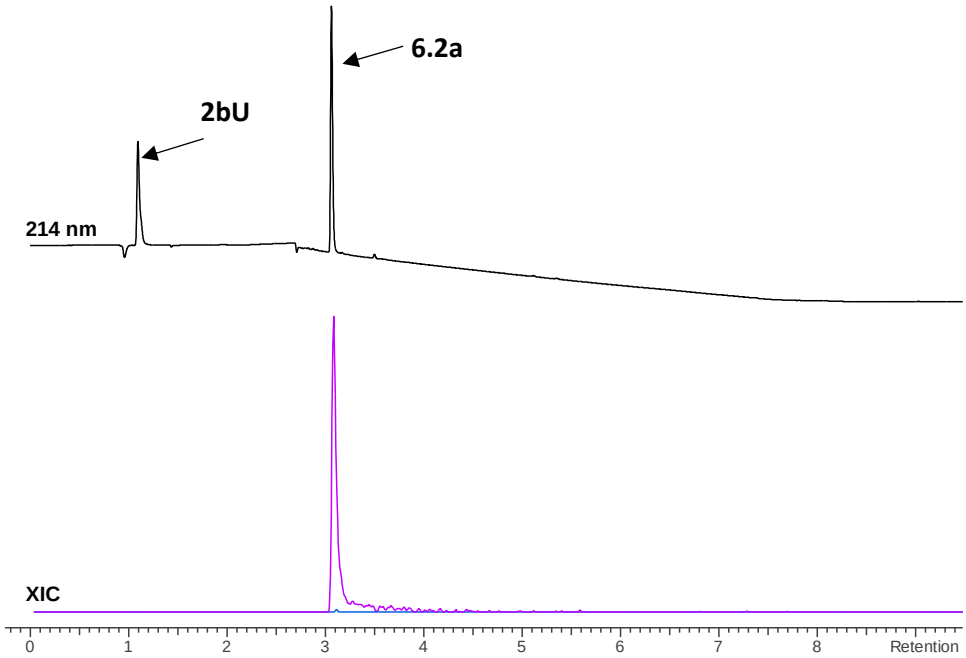


Figure 9.89 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with TAD-propanol **2b** (2eq.) in milliQ H_2O. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2b** (651.2 Da) (bottom).

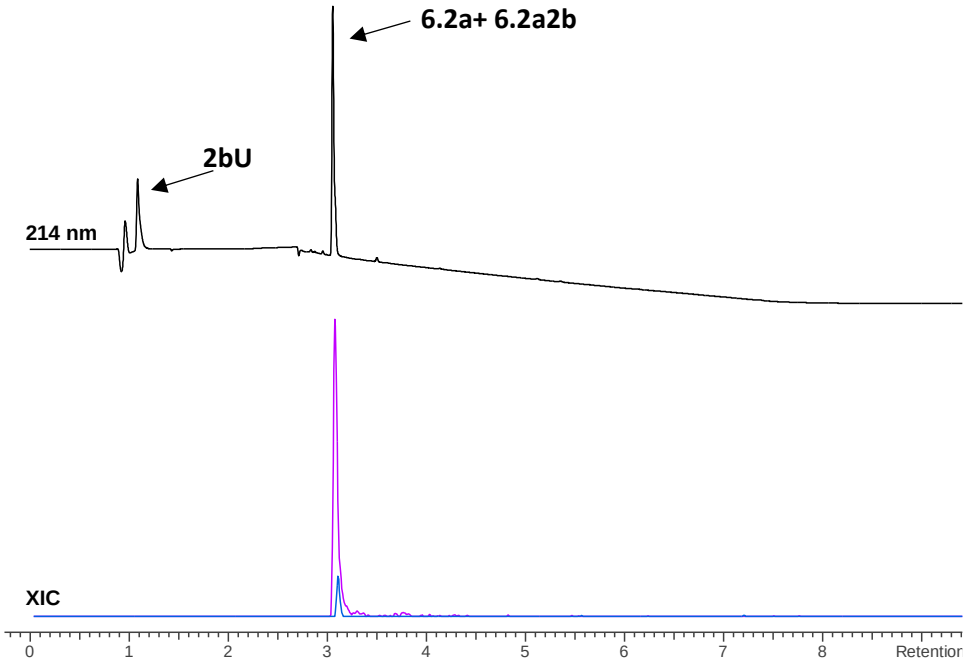


Figure 9.90 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with TAD-propanol **2b** (2eq.) in PBS pH 5. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2b** (651.2 Da) (bottom).

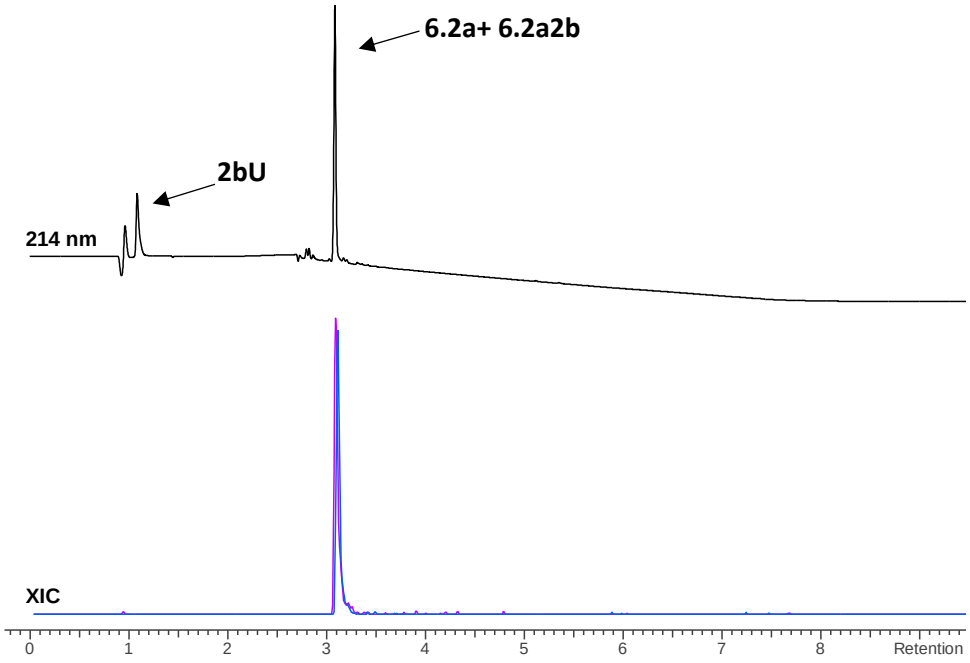


Figure 9.91 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with TAD-propanol **2b** (2eq.) in PBS pH 7. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2b** (651.2 Da) (bottom).

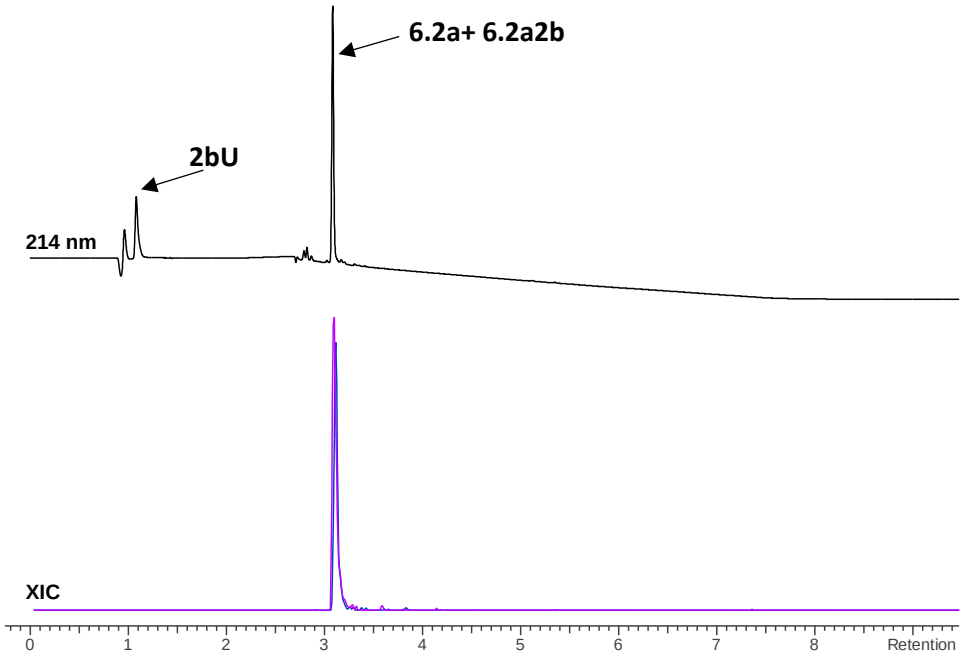


Figure 9.92 Analysis of the reaction of GGGAAY **6.2a** (1 eq.) with TAD-propanol **2b** (2eq.) in PBS pH 9. LC chromatogram 214 nm with annotated peaks (top), XIC of **6.2a** (494.2 Da) and **6.2a2b** (651.2 Da) (bottom).

9.3.2.2 Chemical farnesylation

To a cooled (0 °C) solution of farnesol (mixture of isomers) 1.0 eq. , 1.4 M in THF was added PBr3(0.4 eq.) dropwise. It was stirred for 60 min at 0 °C and the reaction mixture was poured into ice-water. The aqueous phase was extracted three times with hexane. The combined organic layers were dried over MgSO4and concentrated under reduced pressure. The farnesylbromide was obtained as a yellow oil that were used in the next step without purification. A volume of 100 mM Zn(OAc)2 stock solution was prepared by dissolving 22.0 mg Zn(OAc)2 in 1.0 mL of 0.1% aqueous TFA. 1.0 equivalents of peptide GGCAAS **6.2b1** (1.44 mg, 2.77 µmol) was dissolved in 4.8 mL solvent (DMF⁄1-butanol⁄0.1% aqueous TFA, 2:1:1) to which 160 µL of the 100 mM stock $Zn(OAc)_2$ solution (5.77 eq., 16 µmol) was added. To this solution, 4.4µL (4.0 eq.) of farnesyl bromide was added. The solution was stirred at room temperature, in dark, overnight and the product **6.2b2** was purified by RP-HPLC on a C18 column.

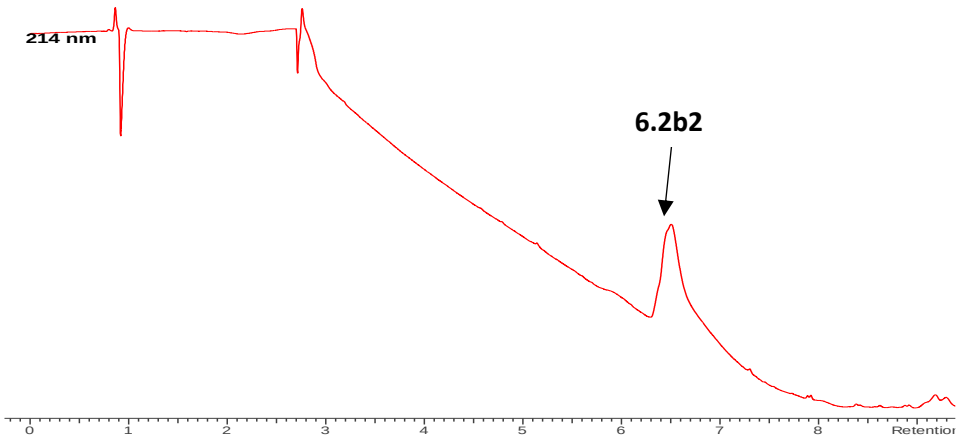


Figure 9.93 LC-MS chromatogram (214 nm) of purified peptide **6.2b2**

The attachment of the farnesyl moiety causes a large shift in retention time t_R ± 6.5 min and a tailing peak. Additionally the farnesyl isomers cause a broad peak.

9.3.2.3 Sortase ligation

The transpeptidase Sortase A His-tag was added to the fascin nanobody FasNb5-sort-GG-HIS along with peptide **6.2b2** (50 µM peptide **6.2b2** / 4 µM Sortase A pentamutant / 20 µM nanobody) in sortase reaction buffer (50 mM Tris-HCl, 150 mM NaCl and 10 mM $CaCl_2$ (Sigma-Aldrich), pH 8.0). Samples were taken for SDS-PAGE analysis at 0h and after the reaction. The reaction was allowed to continue for three hours. The C-terminal GG-HIS tag is cleaved from the fascin nanobody and replaced by **6.2b2** by the Sortase A. In the affinity purification steps TALON® beads (Co^{2+}) remove all proteins with a His-Tag except for the farnesylated nanobody. The initial nanobody and the sortase have a His tag and will stick to the affinity beads as will the GG-HIS fragment that is cleaved off.

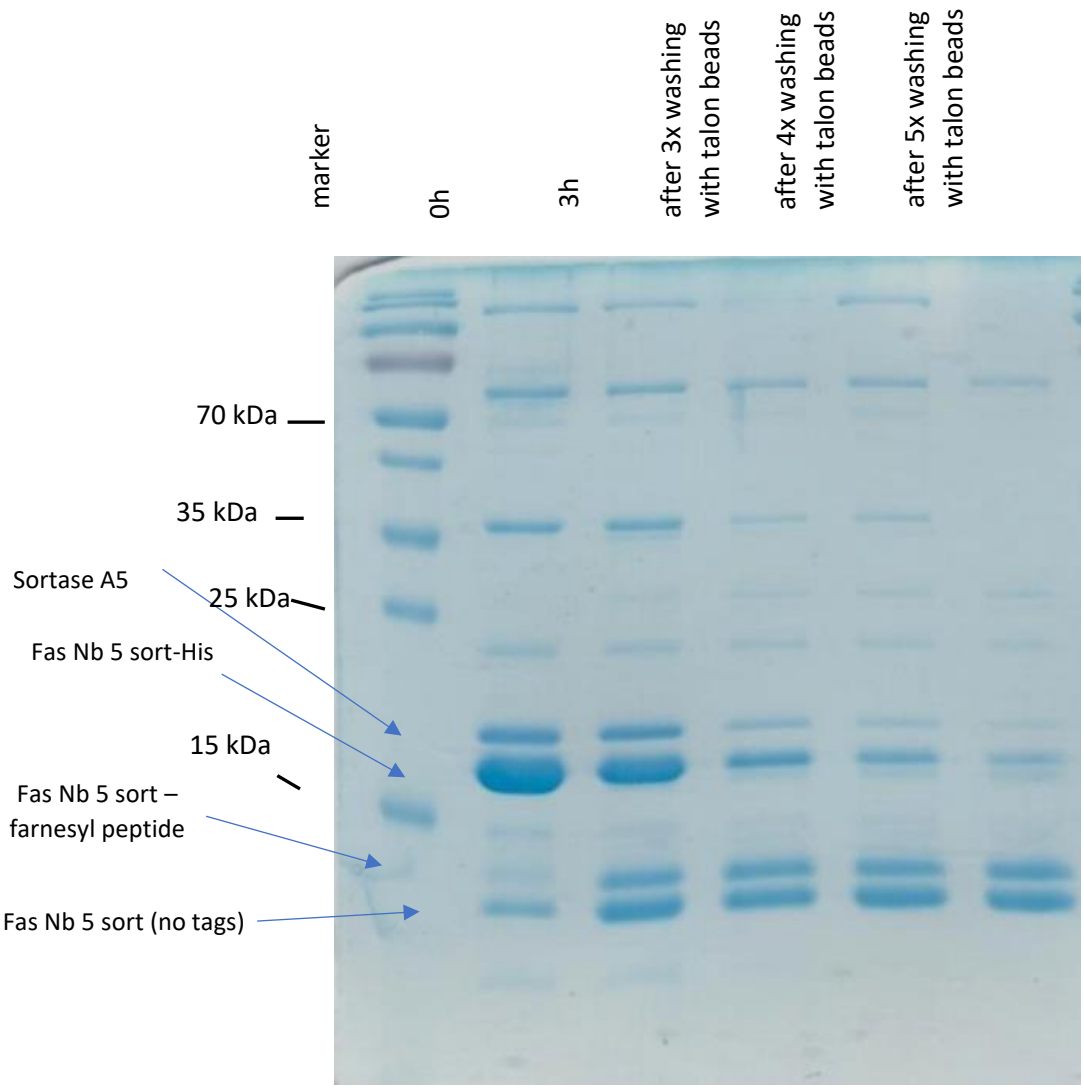


Figure 9.94 Sortase mediated ligation of peptide **6.2b2** to FasNb5-sort-GG-HIS. SDS-PAGE gel (15 %), PageRuler™ prestained protein ladder. lane 1: marker, lane 2: before sortase, lane 3: after sortase, lane 4: after 3x washing with TALON® beads, lane 5: after 4x washing with TALON® beads, lane 6: after 5x washing with TALON® beads.

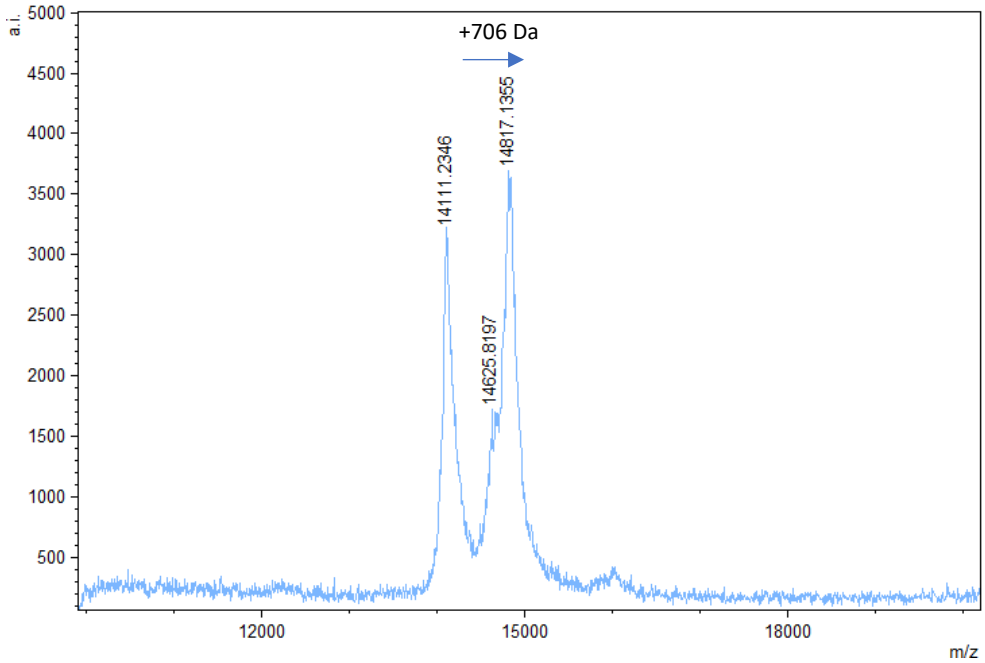


Figure 9.95 MALDI-TOF MS spectrum of a reaction mixture after sortase ligation of peptide **6.2b2** to FasNb5-sort-GG-HA/His. The reaction mixture was dialysed in milliQ H2O. The two main peaks correspond with FasNb5-sort (14111 Da, calcd: 14154 Da) and FasNb5 farnesylated (14817 Da, calcd: 14861). The mass difference between these peaks is 706 Da, the expected mass difference for ligation of **6.2b2** is 707 Da. A less intense third peak is observed (14625 Da) this peak was not present in all nanobody farnesylation experiments and it is believed to be an artefact of the enzymatic ligation reaction.

An experiment with TAD-propanol **2b** and the farnesylated nanobody in milliQ water was performed. The mixture was desalted using Micro Bio-Spin™ 6 Columns - Bio-Rad. These columns were loaded with milliQ H_2O by flushing 6x with milliQ H_2O. After reaction with TAD-propanol we observe TAD-modifications on both the non farnesylated nanobody (2x) and the farnesylated nanobody (3x). It is worth noting that a large amount of farnesylated nanobody is lost after the desalting process (figure 9.96). We presume the hydrophobic properties of the farnesylgroup cause the farnesylated nanobody to stick to the desalting columns.

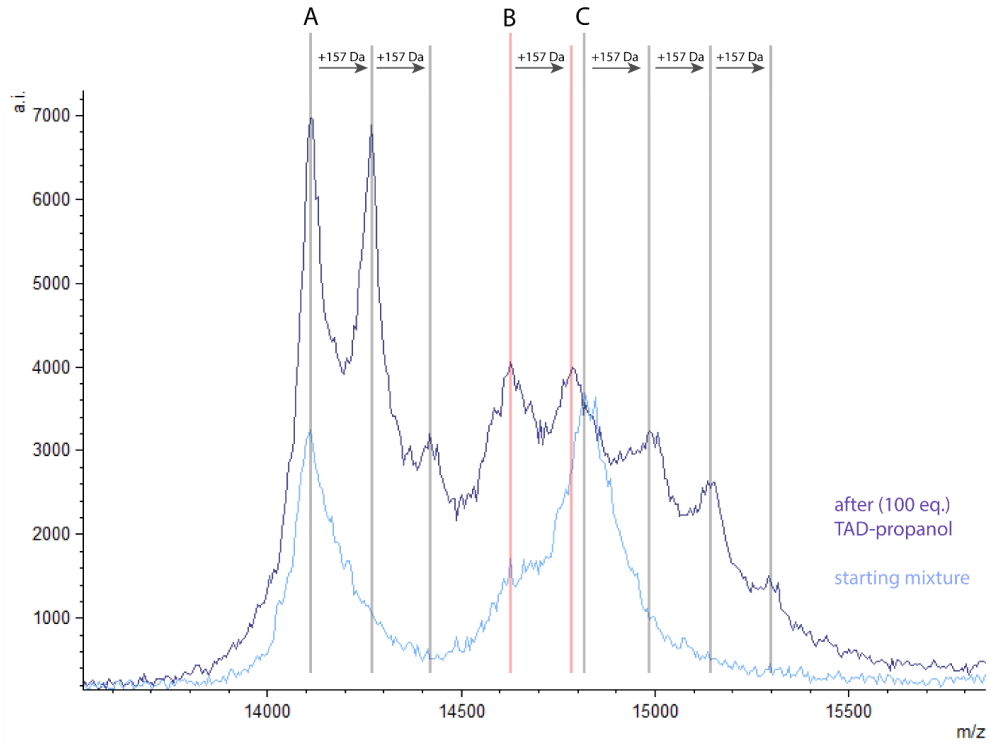


Figure 9.96 MALDI-TOF MS spectrum of a reaction mixture after sortase ligation of peptide **6.2b2** to FasNb5-sort-GG-His, the reaction mixture was dialysed in milliQ H_2O (starting mixture, blue). MALDI-TOF MS spectrum of the same reaction mixture after reaction with 100 eq. TAD-propanol 2b (after (100 eq. TAD-propanol, purple). Peak A corresponds with FasNb5-sort (14111 Da, calcd: 14154 Da) and peak C with FasNb5 farnesylated (14817 Da, calcld: 14861). Peak B is an artefact from the ligation reaction. After reaction with **2b**, 2 TAD modification are detected for peak A, 3 TAD modifications for peak C as well as a TAD modification for peak B.

Additionally an experiment with TAD-propanol was performed on the FasNb5-sort. This nanobody does not contain a farnesylgroup and was obtained via desalting in milliQ H2O using Micro Bio-Spin™ 6 Columns (vide supra). Using formic acid the mixture was further acidified to further prevent tyrosine residues in the nanobody from reacting with TAD-propanol. Subsequently TAD-propanol (100 eq. was added and the mixture was analysed with LC-MS (figure S6.2.2). Two TAD modifications (+157 Da) were found on the FasNb5-sort.

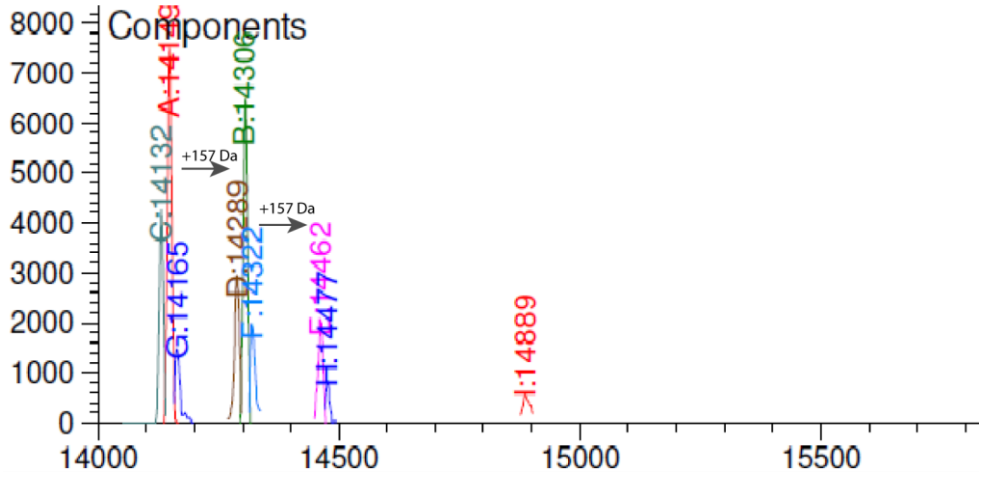


Figure 9.97 LC-MS deconvolution mass spectrum from reaction of FasNb5-sort with TAD-propanol **2b** in milliQ H_2O further acidified with formic acid. A double TAD modification can be observed on the FasNb5-sort.

In all TAD modification reactions on FasNb5-sort or FasNb5 farnesylated mixtures, TAD-modifications were observed on the non farnesylated nanobody even in milliQ H_2O that was further acidified with formic acid. This result was not expected as we saw in peptide experiments (section 6.3) that tyrosine is quasi not able to react in buffers with a pH < 5 and completely inhibited in milliQ H_2O. We demonstrate that tryptophan residues are actually kinetically favoured to react with TAD-reagents in most cases. The data we present here can be explained satisfactory by the reaction with tryptophan residues.

9.3.2.4 Crystal structure analysis, framework tryptophans

The general structure of nanobodies consists of framework regions building up the core structure and additionally several CDR loops responsible for antigen recognition. Framework sequences of nanobodies are highly conserved. There are two tryptophan residues present in the sequence of the FasNb5: W36 and W114 and both are located in framework regions. One tryptophan is located in the beginning of framework region 2 (W36) and the other in the beginning of framework region 4 (W114). Even though the X-ray structure of our nanobody of interest is not known we can predict the solvent accessibility of both tryptophan residues in our nanobody of interest by comparing X-ray structures of other nanobodies. Looking at the X-ray structure of a Gelsolin nanobody[335] (PDB ID: 2X1P) and EgA1 nanobody[336] (PDB ID: 4KRN) both with the same 2 framework tryptophan residues we observed that one of these tryptophans is exposed and the other is buried. We can conclude that the tryptophan W36 is likely buried in the folded protein structure and tryptophan W114 is likely exposed to the surface.

Gelsolin nanobody (framework W's)

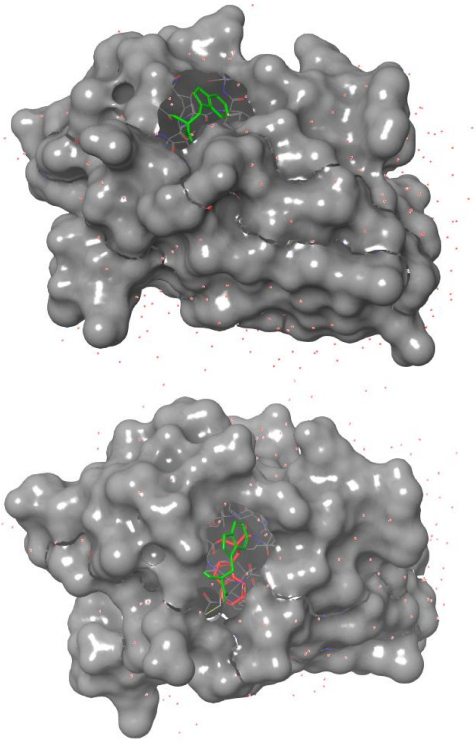


Figure 9.98 Two views of the X-ray structure of the Gelsolin nanobody with protein surface and indicated tryptophan residues. Tryptophan at the beginning of framework region 4: green and tryptophan at the beginning of framework region 2: red. On the left view a tryptophan (green) residue can be seen on the surface of the nanobody. On the right view a buried tryptophan (red) can be observed underneath the exposed tryptophan (green).

EgA1 nanobody (framework W's)

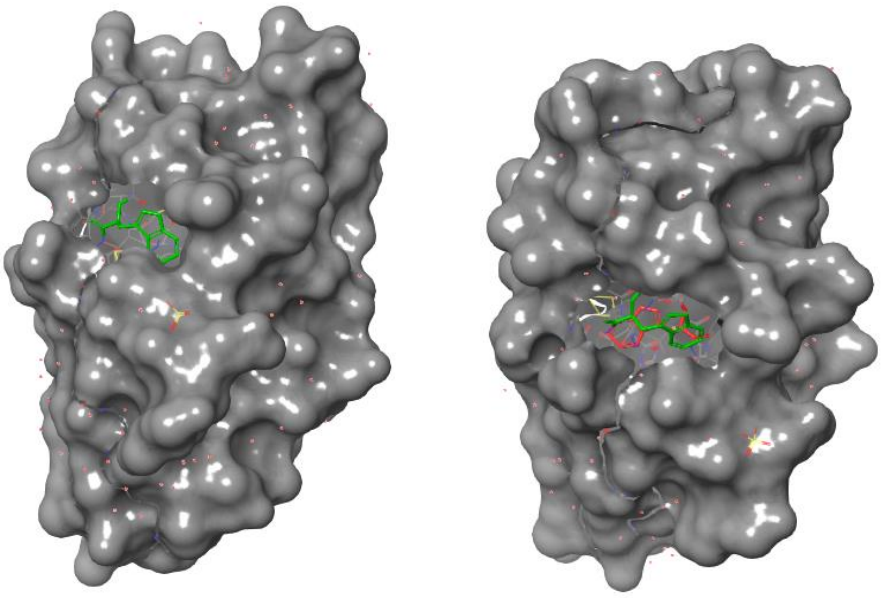


Figure 9.99 Two views of the X-ray structure of the EgA1 nanobody with protein surface and indicated tryptophan residues. Tryptophan at the beginning of framework region 4: green and tryptophan at the beginning of framework region 2: red. On the left view a tryptophan (green) residue can be seen on the surface of the nanobody. On the right view a buried tryptophan (red) can be observed underneath the exposed tryptophan (green).

9.3.2.5 TAD propanol conjugation to farnesylated H-Ras

Ras protein family amino acid sequences

K-Ras (sequence)

MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGET
60 70 80 90 100
CLLDILDTAG QEEYSAMRDQ YMRTGEGFLC VFAINNTKSF EDIHHYREQI
110 120 130 140 150
KRVKDSEDVP MVLVGNKCDL PSRTVDTKQA QDLARSYGIP FIETSAKTRQ
160 170 180
RVEDAFYTLV REIRQYRLKK ISKEEKTPGC VKIKKCIIM

H-Ras (sequence)

MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGET
60 70 80 90 100
CLLDILDTAG QEEYSAMRDQ YMRTGEGFLC VFAINNTKSF EDIHQYREQI
110 120 130 140 150
KRVKDSDDVP MVLVGNKCDL AARTVESRQA QDLARSYGIP YIETSAKTRQ
160 170 180
GVEDAFYTLV REIRQHKLRK LNPPDESGPG CMSCKCVLS

N-Ras (sequence)

MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGET
60 70 80 90 100
CLLDILDTAG QEEYSAMRDQ YMRTGEGFLC VFAINNSKSF ADINLYREQI
110 120 130 140 150
KRVKDSDDVP MVLVGNKCDL PTRTVDTKQA HELAKSYGIP FIETSAKTRQ
160 170 180
GVEDAFYTLV REIRQYRMKK LNSSDDGTQG CMGLPCVVM

Conjugation experiments with farnesylated H-Ras

Two control experiments were designed to investigate the selectivity of TAD-propanol for farnesylated H-Ras. The setup of control reaction A was exactly the

same as the one described in figure 6 except for the fact that there was no enzymatic farnesylation performed preceding the reaction with TAD-propanol **2b**. This resulted in a mixture where only the H-Ras protein (without farnesylgroup) is present in HEPES buffer at pH 7.2 and reaction is done with 200 eq. of TAD-propanol **2b**. In the second control reaction (B), both the enzymatic farnesylation and the subsequent reaction with TAD-propanol **2b** were performed as described in the manuscript except for the fact that no farnesyl pyrophosphate FPP was added during the enzymatic farnesylation. This resulted in a mixture with H-Ras (without farnesylgroup and farnesyltransferase FTase in HEPES buffer at pH 7.2 and reaction with 200 eq. TAD-propanol **2b**. Both control samples were analysed via LC-MS. For reference the complete deconvoluted mass spectrum of the TAD-propanol with farnesylated H-Ras described in figure 6 is included (figure 9.100). Besides the expected masses of the farnesylated nanobody (21421 Da) and the farnesylated nanobody after reaction with TAD-propanol **2b** (21578 Da) several other low abundant protein peaks are detected. The deconvoluted mass spectrum of control reaction A at t_R = 4.400 min. (figure 9.101) demonstrates several low abundant protein peaks, compared to the intensity of the main peak in figure 9.100. None of these peaks correspond with the expected mass for the nonfarnesylated starting nanobody (14219 Da) or with the expected mass for a mono (+ 157 Da) or multi TAD-modified compound. The deconvoluted mass spectrum of control reaction B at t_R = 4.438 min. was obtained with a lower noise cut-off, 500 compared to the default value of 1000. This deconvoluted mass spectrum is extremely crowded caused by the lower noise cut-off value, this demonstrates that peaks with an intensity of around 5000 are actually barely more intense than background peaks. Similarly as for control reaction A no expected mass for the nonfarnesylated starting nanobody (14219 Da) or the expected mass for a mono (+ 157 Da) or multi TAD-modified compound was observed.

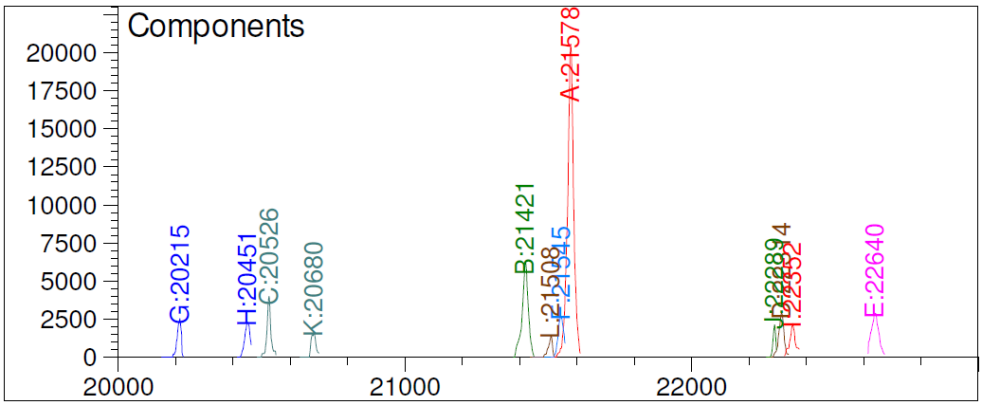


Figure 9.100 LC-MS deconvoluted mass spectrum of the TAD-propanol **2b** (200 eq.) reaction with farnesylated H-Ras in HEPES buffer at pH 7.2.

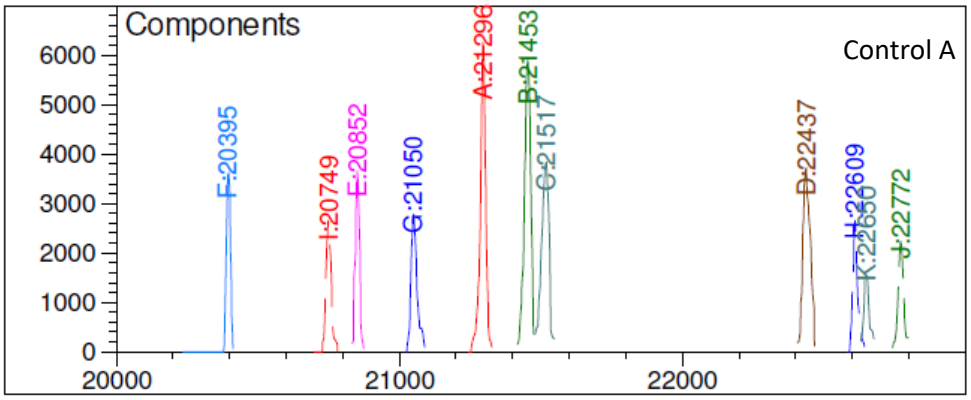


Figure 9.101 LC-MS deconvoluted mass spectrum at t_R=4.400 min. of control reaction A.

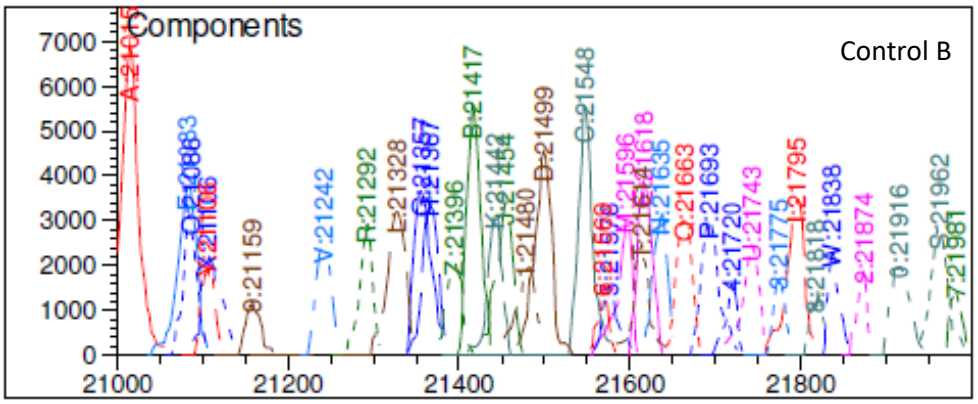


Figure 9.102 LC-MS deconvoluted mass spectrum at t_R=4.438 min. of control reaction B.

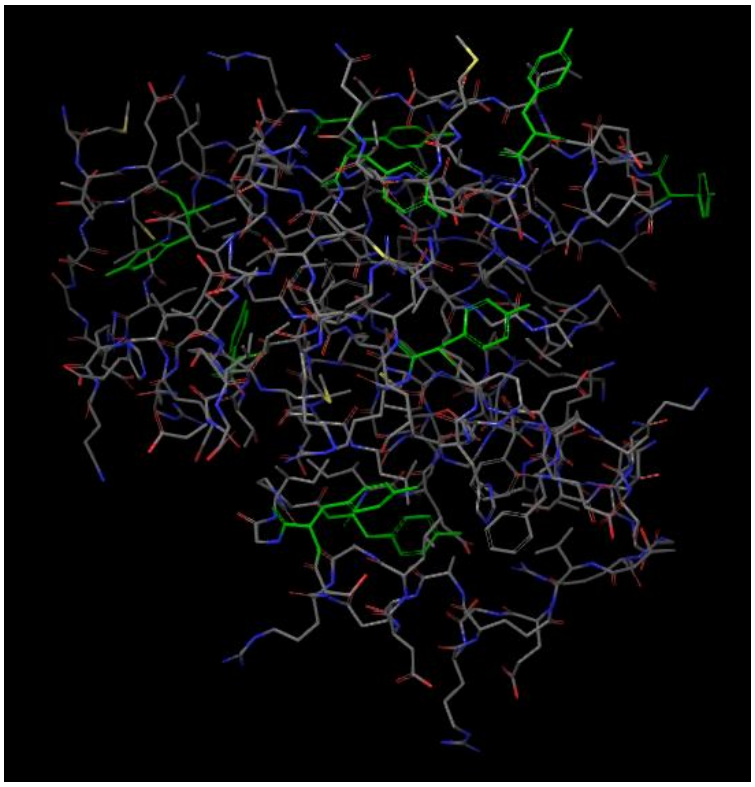


Figure 9.103 X-ray structure of H-Ras (PDB ID: 4Q21) tyrosine residues are indicated in green.

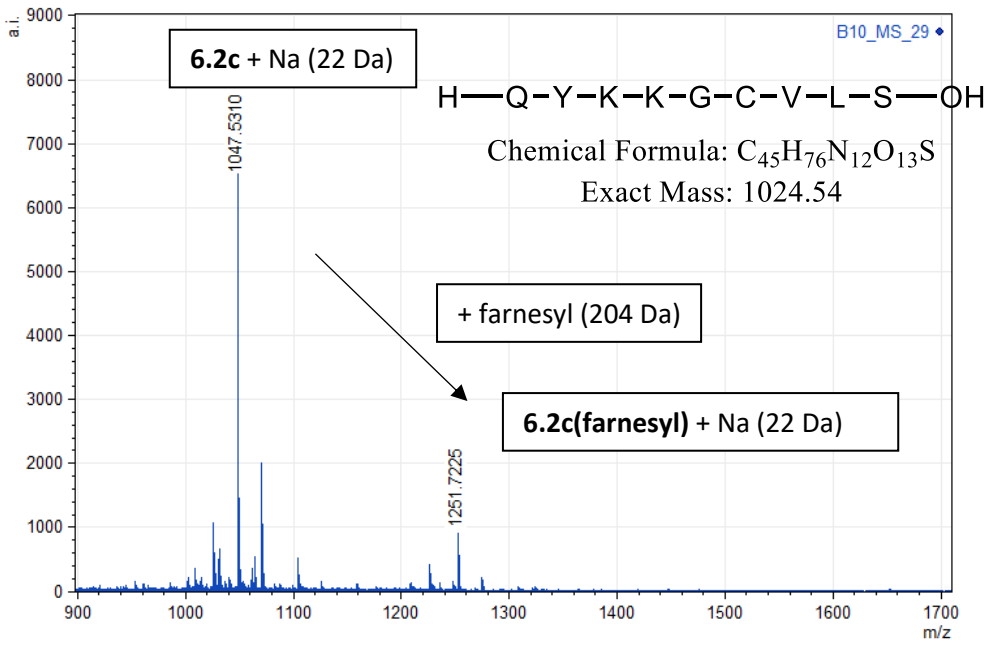


Figure 9.104 MALDI-TOF MS spectrum from *in vitro* farnesylation of peptide **6.2c** after 2h enzymatic reaction. The successful addition (+204 Da) of the farnesyl moiety is already clearly visible in the MS spectrum

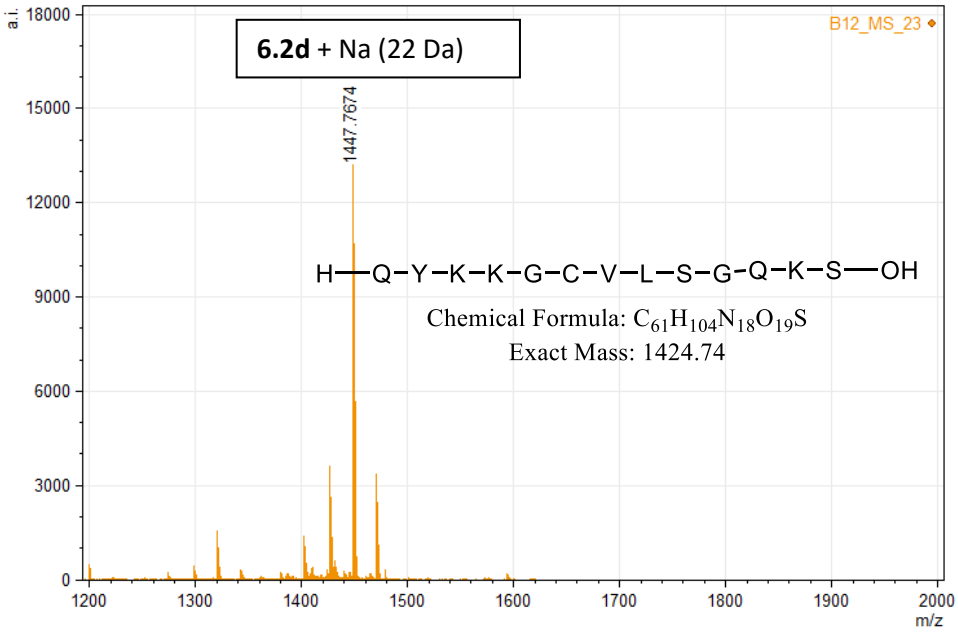


Figure 9.105 MALDI-TOF MS spectrum from *in vitro* farnesylation of peptide **6.2d** after 2h enzymatic reaction. No farnesylation is visible.

9.3.3 Experimental data for part 6.3

9.3.3.1 General experimental procedures

Solid phase peptide synthesis (SPPS)

Peptides were synthesized using standard Fmoc-chemistry on polystyrene resin with 2-chlorotrityl chloride linker, loading 1.57 mmol/g. The resin was swollen for 20 min in dry DCM, after which the solvent was filtered off. The first (C-terminal) amino acid was coupled manually using 4 equivalents of N,N diisopropylethylamine and 1.1 equivalents of the first (C-terminal) Fmoc protected amino acid in dry DCM. The reaction mixture was added to resin and shaken for 2 hours at room temperature. Afterwards the reaction mixture is filtered off and the resin is washed with DCM (3x). Unreacted linker functionalities are capped using a DCM/MeOH/DIPEA (17/2/1) mixture, 30 min shaking at room temperature. Subsequently the capping mixture was filtered off and the resin was washed with DCM (3x) and DMF (3x). The remaining amino acids are incorporated using automated peptide synthesis performed on either a MultiPep (Intavis, Tübingen, Germany) or Syro (Multisyntech, Witten, Germany) device. Every cycle in the automated peptide synthesis starts with the Fmoc deprotection of the Fmoc-protected current N-terminus followed by coupling the next Fmoc-protected amino acid. Fmoc deprotection was carried out with piperidine (40%) in DMF, 3 minutes with 40% piperidine in DMF followed by 12 minutes with 20% piperidine in DMF. Synthesis with double couplings was performed with 5 equivalents HBTU (0.5 M) as coupling reagent, 5 equivalents amino acid (0.5 M) and 10 equivalents DIPEA in NMP (2M) as base. Coupling time was 40 minutes. For longer peptides an capping extra step was performed after the double coupling steps of each amino acid. A solution of DMF/acetic anhydride/N-methylmorpholine (90/5/5) was used for capping, 5 minutes reaction time. A mixture of 95% trifluoro acetic acid (TFA), 2.5%

tri-isopropyl silane, and 2.5% H_2O was used to cleave the peptide from the resin, 3 hour reaction time. Peptides were subsequently precipitated in cold methyl tert-butyl ether (2x). Purification of the synthesized peptides was done on an semiprep HPLC instrument equipped with a Phenomenex Luna C18 column at 35 °C with a flow rate of 16 mL/min. The column was eluted with a gradient starting at 100% H_2O containing 0.1% TFA to 100% acetonitrile in 20 min. All materials were obtained via commercial suppliers (Sigma Aldrich, Iris Biotech GmbH, Novabiochem, Acros, Fujifilm Wako chemicals) unless otherwise stated. LC-MS analysis was done on a Agilent LC-MS system. MALDI TOF and MALDI TOF/TOF spectra were obtained with a 4800 MALDI TOF/TOF analyses (Applied Biosystems).

300 fmol of synthetic peptide was injected for LC-MS/MS analysis on an Ultimate 3000 RSLCnano system in-line connected to a Fusion Lumos mass spectrometer (Thermo). Trapping was performed at 10µl/min for 4 min in loading solvent A (0.1% TFA 98/2 H_2O/ACN) on a trapping column (made in-house, 100 µm I.D. x 20 mm, 5µm beads, C18 Reprosil-HD, Dr.Maisch, Germany). The peptides were eluted from the trapping column and further separated on a 200cm µPAC™ column (C18-endcapped functionality, 300 µm wide channels, 5µm porous-shell pillars, inter pillar distance of 2.5 µm and a depth of 20 µm, Pharmafluidics, Belgium). Peptides were eluted by a gradient starting at 1% solvent B (0.1% FA in H_2O/ACN 80/20), reaching 9% solvent B in 15 min , 33% solvent B after 30 min, 55% solvent B after 35 min and eventually reaching 99% solvent B after 38 min followed by a 10 min wash and re-equilibration with 99% solvent A (0.1% FA in H_2O). The flow rate was increased to 750 nl/min for the first 15 min, after which it was decreased to 300nl/min and kept constant. The column temperature is set to a constant value of 50°C.

The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition in TopSpeed mode using cycle time of 3s. Full-scan MS spectra (300-2000 m/z) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target AGC value of 400,000 with a maximum injection time of 50 ms. The precursor ions were filtered for charge states (2-7 required), MIPS set to peptide and a minimum intensity of 5E4. The precursor ions were selected in the ion routing multipole with an isolation window of 1.6 Da and accumulated to an AGC target of 1E4 or a maximum injection time of 35 ms. Three different activation types were applied for each selected precursor. A first activation was done using HCD fragmentation (NCE 34%). A second parallel activation was done using CID fragmentation (35% CE) with an activation time of 10 ms. A third parallel activation was done by ETD fragmentation using the calibrated charge dependent ETD parameters combining 3 microscans. Fragments of the different fragmentation modes were each analyzed in the ion trap analyzer at normal scan rate.

9.3.3.2 Preparation of TAD-compounds

PTAD, **2a**

Synthesized as described elsewhere.[337]

Urazole precursor: TAD-propanol **2b**

5 g diphenyl carbonate (1 eq., 0.023 mol) was melted in a 250 mL flask and 5 g ethyl carbazate (2 eq., 0.047 mol) was added. The mixture stirred under inert atmosphere for 1 h at 90 °C. 2 ml 3-Amino-1-propanol (1,1 eq., 0,026 mol) was added into this molten mixture and stirred for additional 15 minutes without removing oil bath. Subsequently, the product was precipitated with addition of 150 ml ethyl acetate to bulk mixture while it was cooling down. 4,61 g (97%) pure white powder was obtained via filtration and dried in overnight under vacuum

1H-NMR (300MHz, DMSO-d6): δ (ppm) = 1.17 (t, 3H, O-CH2-CH3), 1.52 (quint, 2H, CH2-CH2-CH2), 3.05 (q, 2H, CH2-NH), 3.38 (q, 2H, HO-CH2), 4.03 (q, 2H, O-CH2-CH3), 4.41 (t, 1H, HO-CH2), 6.31 (s, 1H, CH2-NH),7.68 (s, 1H, O-CO-NH), 8.75 (s, 1H, NH-NH-CO-NH).

2 g of the semicarbazide (1 eq., 0,0097 mol), 60 ml ethanol and 6,80 g potassium carbonate (5 eq., 0,0487 mol) were added into 250 ml round bottom flask and the mixture was refluxed overnight. he reaction mixture was cooled to room temperature and acidified to pH 1 with HCl in propanol (5-6 N). The precipitate was filtered off, the reaction mixture was concentrated in vacuo. 1,55 g (95%) propanol urazole was obtained as pure white powder and dried in vacuum oven.

1H-NMR (300MHz, DMSO-d6): δ (ppm) = 1.67 (quint, 2H, CH2-CH2-CH2), 3.40 (m, 4H, CH2-CH2-CH2), 10.02 (s, 2H, NH-NH).

Urazole precursor: PTAD-alkyne **2c**

Purchased from Sigma-Aldrich and used without further purification.

DMEQ-TAD, **2d**

Purchased from Fujifilm Wako Chemicals and used without further purification. DMEQ-TAD is received in the TAD form and does not require oxidation.

Trichloroisocyanuric acid oxidation

Urazole precursors of TAD-propanol **2b** and PTAD-alkyne **2c** are oxidized using trichloroisocyanuric acid (TCCA). To this end 0.33 equivalents of trichloroisocyanuric acid is added to a slight excess: 1.x equivalents of the urazole precursor in acetonitrile, the reaction mixture is shaken for 2 hours at room temperature. Upon the addition of TCCA to the urazole MeCN suspension the mixture turns pink (TAD-propanol, **2b**) or red (PTAD-alkyne, **2c**). After completion the reaction mixture is centrifuged shortly to spin down all insoluble compounds. Afterwards the desired TAD derivative solution in MeCN is pipetted directly out of the supernatant. The TAD derivative is used without further purification.

It is worth noting that there can be trace amounts of HCl left in the TAD solution. This is due to oxidation with trichloroisocyanuric acid. This causes a lower effective

pH value in peptide-TAD conjugation samples using no or weak buffers. This effect is minimal in 10x PBS buffers or 100 mM Tris-HCl buffers.

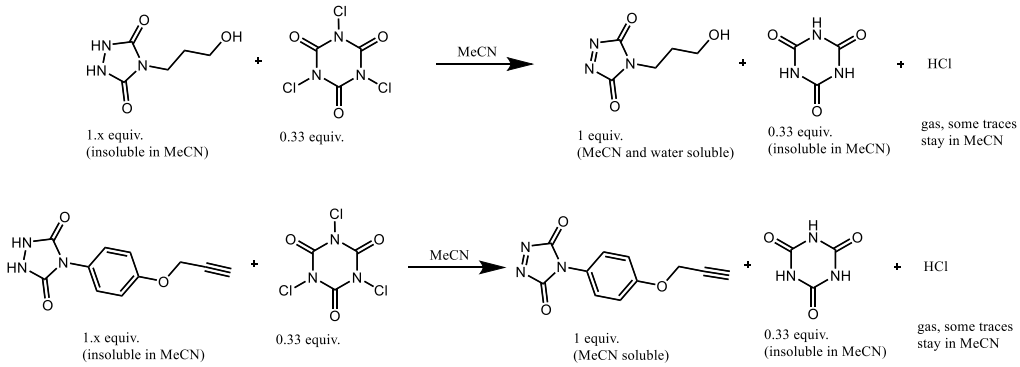


Figure 9.106 Synthesis of TAD-propanol **2b** and PTAD-alkyne **2c**.

9.3.3.3 Peptide TAD modification

The purified peptides were dissolved in milliQ H2O to make stock solutions of 3 mM. Typically, 4 mM TAD solutions are prepared in MeCN. 2 µL (6 nmol) of the peptide stock solution is added to 16 µL buffer solution (10 X PBS) with a known pH value. To this mixture 2 µL of the 4 mM TAD solution (8 nmol) is added mixed immediately by pipetting up and down several times. In competition experiments 2 µL of both peptide 3 mM stock solutions are added to 14 µL buffer solution (10 X PBS). To this mixture 2 µL of the 4 mM TAD solution (8 nmol) is added mixed immediately by pipetting up and down several times. The reaction of TAD with both tyrosine and tryptophan containing peptides is completed in a matter of seconds, that is why it is important to mix the reaction mixture immediately after addition of TAD reagents. The resulting reaction mixture is analyzed via HPLC and/or LC-MS and/or MALDI-TOF(/TOF) and/or LC (ESI) (ETD/CID/HCD) MS/MS.

9.3.3.4 Analytical data: peptide conjugation

Characterization of starting peptides

LC-MS data (reversed phase) were recorded on an Agilent 1100 Series instrument with diode array detector (set to 214, 254, 280, 310, 360 nm), equipped with a Phenomenex Kinetex C18 100 Å (150 x 4.6 mm, 5 µm, at 35 °C), hyphenated to an Agilent ESI-single quadrupole MS detector type VL. Mass detection operated in the positive mode. Linear gradient elutions were performed by flushing 0.5 min with A followed by 0 to 100% buffer B in 6 minutes and finally by a 2 min flushing with B using a binary solvent system composed of buffer A: 5 mM NH_4OAc (**6.3a**-**6.3f**) or 0.1% HCOOH in H2O (**6.3i**, **6.3j)** and B: MeCN with a flow of 1.5 mL/min at 35 °C. A solution of 4-5 mg αcyano-4-hydroxycinnamic acid in 500µL MeCN, 490µL mQ, 10µL 1 M ammonium citrate, 1µL TFA was used as a matrix for MALDI-TOF-MS.

Asn-Trp-Ala-Ser-OH (NWAS, 6.3a). MS (ESI): m/z 477.1 (calcd $[M+H]^+$ 477.2) retention time 3.32 min. Purity: >95 % (HPLC analysis at 214 nm)

Asn-Tyr-Ala-Ser-OH (NYAS, 6.3b). MS (ESI): m/z 454.1 (calcd $[M+H]^+$ 454.2) retention time 2.86 min. Purity: >95 % (HPLC analysis at 214 nm)

Asn-Ser-Ala-Trp-OH (NSAW, 6.3c). MS (ESI): m/z 477.1 (calcd $[M+H]^+$ 477.2) retention time 3.26 min. Purity: >95 % (HPLC analysis at 214 nm)

Asn-Ser-Ala-Tyr-OH (NSAY, 6.3d). MS (ESI): m/z 454.1 (calcd $[M+H]^+$ 454.2) retention time 2.81 min. Purity: >95 % (HPLC analysis at 214 nm)

Trp-Ser-Ala-Asn-OH (WSAN, 6.3e). MS (ESI): m/z 477.1 (calcd $[M+H]^+$ 477.2) retention time 3.23 min. Purity: >95 % (HPLC analysis at 214 nm)

Tyr-Ser-Ala-Asn-OH (YSAN, 6.3f). MS (ESI): m/z 454.1 (calcd $[M+H]^+$ 454.2) retention time 2.86 min. Purity: >95 % (HPLC analysis at 214 nm)

Lys-Lys-Ser-Tyr-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH (KKSYLSPRTALINFLV, 6.3g). MS (ESI): m/z 1849.8 (calcd $[M+H]^+$ 1850.1) retention time: 3.95 min. Purity: >95 % (HPLC analysis at 214 nm)

Lys-Lys-Ser-Trp-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH (KKSWLSPRITALINFLV, 6.3h). MS (ESI): m/z 1872.8 (calcd $[M+H]^+$ 1873.1) retention time 4.02 min. Purity: >95 % (HPLC analysis at 214 nm)

Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH (VWSNRHFY, 6.3i). MS (ESI): m/z 1108.4 (calcd $[M+H]^+$ 1108.5) retention time: 3.40 min. Purity: >95 % (HPLC analysis at 214 nm)

Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH (VYSNRHFW, 6.3j). MS (ESI): m/z 1108.4 (calcd $[M+H]^+$ 1108.5) retention time 3.48 min. Purity: >95 % (HPLC analysis at 214 nm)

Val-Trp-Ser-Gln-Lys-Arg-His-Phe-Gly-Tyr-OH (VWSQKRHFGY, 6.3k). MS (MALDI-TOF): m/z 1307.7 (calcd $[M+H]^+$ 1307.7)

Lys-Asp-Tyr-Trp-Glu-Cys-Ala-OH (KDYWECA, 6.3l). MS (MALDI-TOF): m/z 914.4 (calcd $[M+H]^+$ 914.4)

LC (MS) data of peptide-TAD conjugates

Intermolecular Tyr versus Trp competition

The peptide-TAD conjugates are prepared according to the general protocol for peptide-TAD conjugation reported above for the competitive reactions. We found that the reaction of TAD with both tyrosine and tryptophan in buffer is fast (finished in a few seconds). In order to examine the amino acid selectivity of TAD-compounds we performed reactions in a slight excess of TAD (1.25 eq.). Since in buffer there is always background hydrolysis of TAD compounds only a portion of the peptides can react with TAD.

In the previous section the pH dependency of the TAD tyrosine reaction was demonstrated. The generation of TAD-propanol **2b** and PTAD-alkyne **2c** is done via TCCA oxidation and some HCl remains in the MeCN solution with the freshly prepared TAD compound. This HCl can have an impact on the pH of the reaction.

The following experiments were carried out in 10 X PBS buffers to ensure the pH is not altered upon addition of the TAD. The LC-chromatogram (214 nm) is provided along with a more detailed zoom of the relevant elution window. Since the conjugated peptide ions often have similar retention times an extraction ion chromatogram (XIC) for the relevant peptide starting material and peptide-TAD conjugate of the same elution window is shown.

Conjugation of Asn-Trp-Ala-Ser-OH (NWAS, 6.3a) with TAD-propanol 2b (6.3a2b)
MS (ESI): m/z 634.2 (calcd $[M+H]^+$ 634.3) retention time 2.85 min.

Conjugation of Asn-Tyr-Ala-Ser-OH (NYAS, 6.3b) with TAD-propanol 2b (6.3b2b)
MS (ESI): m/z 611.1 (calcd $[M+H]^+$ 611.2) retention time 2.83 min.

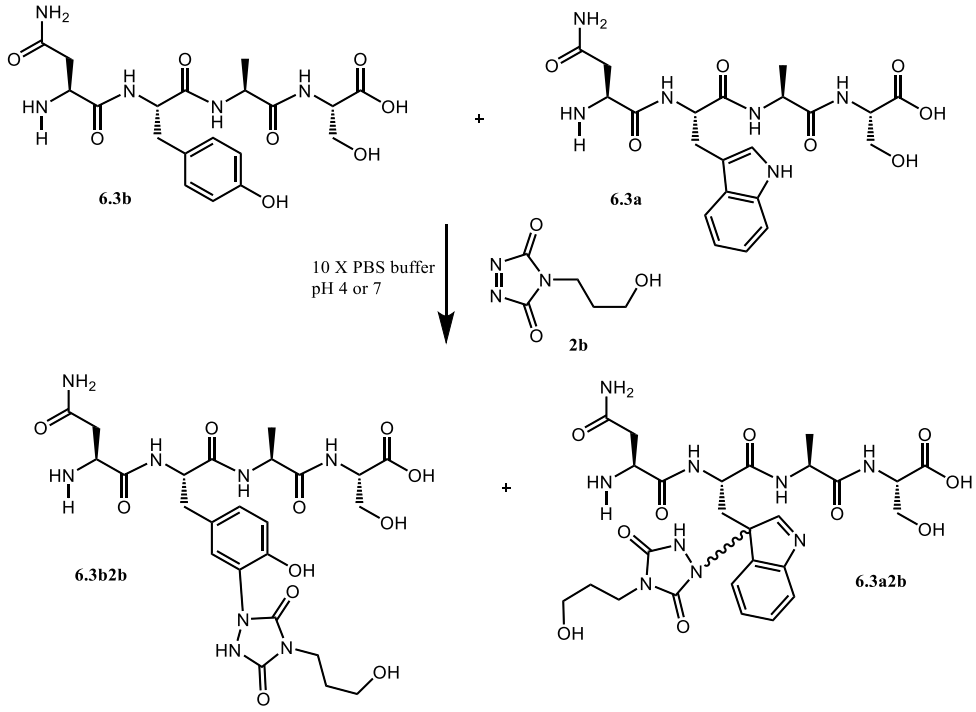


Figure 9.107 Schematic representation of the reaction of peptides **6.3b** and **6.3a** with TAD-propanol **2b** in 10 X PBS buffer pH 4 or 7.

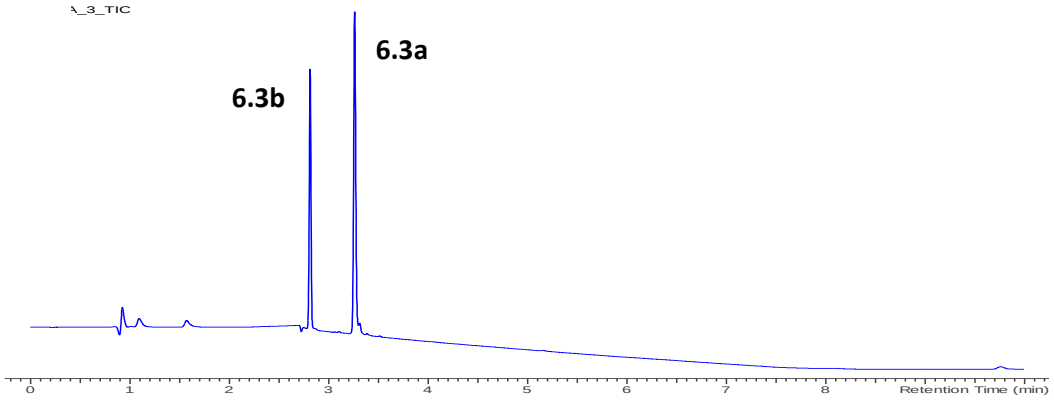


Figure 9.108 LC chromatogram (214 nm) of a mixture of pure peptides **6.3a** and **6.3b**.

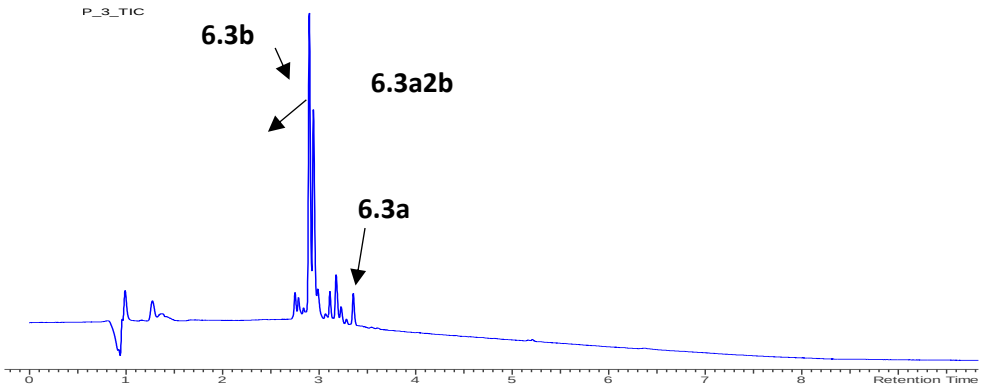


Figure 9.109 LC chromatogram (214 nm) of **2b** conjugation of a **6.3a** and **6.3b** peptide mixture. In this experiment 10 X PBS pH 4 was used as a buffer.

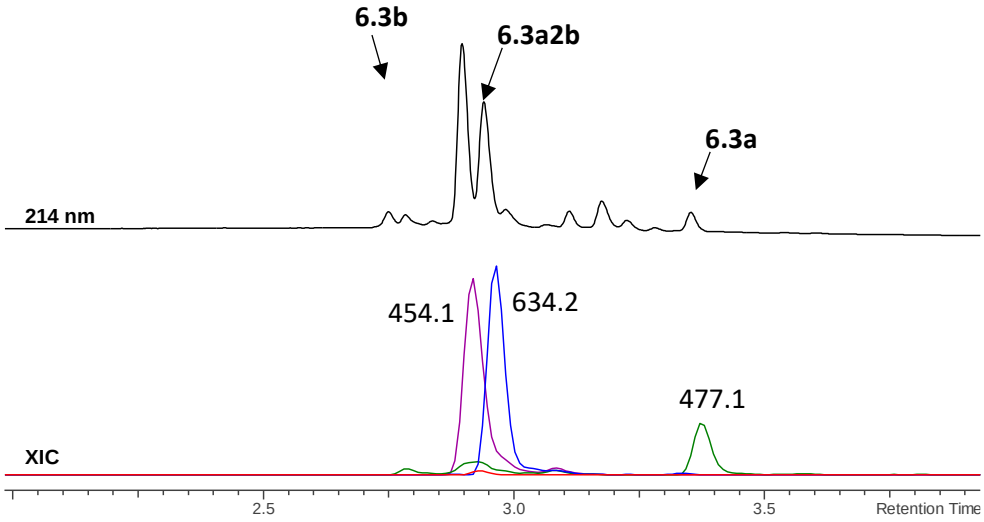


Figure 9.110 LC chromatogram (214 nm) between t_R = 2 min - 4 min (top) of **2b** conjugation of a **6.3a** and **6.3b** peptide mixture in 10 X PBS pH 4. Extracted ion chromatogram (XIC) between t_R = 2 min - 4 min (bottom) for the peptide ions **6.3a** (477.1 Da, green) **6.3b** (454.1 Da, purple) and the conjugated peptide ions **6.3a2b** (634.2 Da, blue) and **6.3b2b** (611.1 Da, red). Product **6.3b2b** is not formed at pH 4.

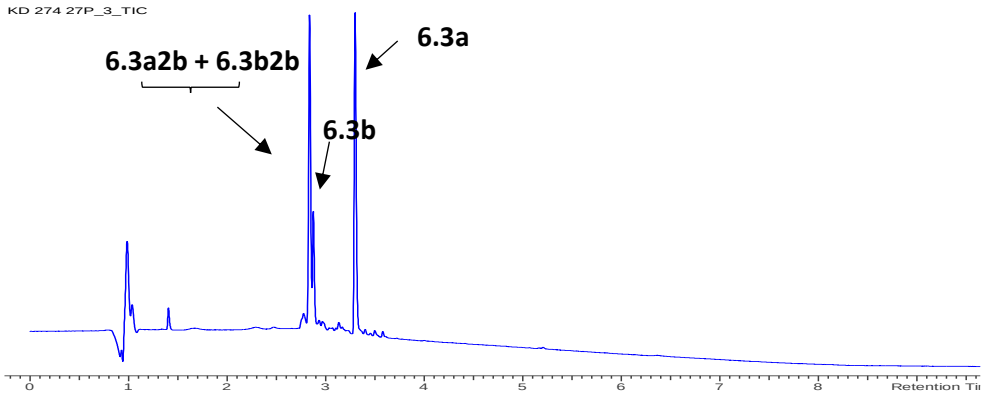


Figure 9.111 LC chromatogram (214 nm) of **2b** conjugation of a **6.3a** and **6.3b** peptide mixture. In this experiment 10 X PBS pH 7 was used as a buffer.

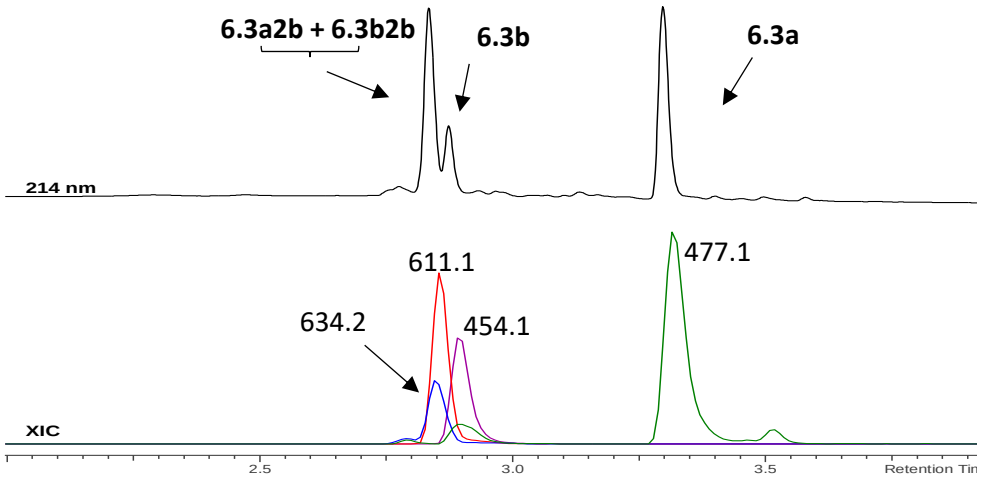


Figure 9.112 LC chromatogram (214 nm) between t_R = 2 min - 4 min (top) of **2b** conjugation of a **6.3a** and **6.3b** peptide mixture in 10 X PBS pH 7. Extracted ion chromatogram (XIC) between t_R = 2 min - 4 min (bottom) for the peptide ions **6.3a** (477.1 Da, green) **6.3b** (454.1 Da, purple) and the conjugated peptide ions **6.3a2b** (634.2 Da, blue) and **6.3b2b** (611.1 Da, red). Both products **2ab** and **2bb** are formed at pH 7.

Conjugation of Asn-Ser-Ala-Trp-OH (NSAW, 6.3c) with PTAD-alkyne 2c (6.3c2c)
MS (ESI): m/z 706.2 (calcd $[M+H]^+$ 706.3) retention time 3.42 min + 3.61 – 6.68min.

Conjugation of Asn-Ser-Ala-Tyr-OH (NSAY, 6.3d) with PTAD-alkyne 2c (6.3d2c)
MS (ESI): m/z 683.1 (calcd $[M+H]^+$ 683.2) retention time 3.44 min.

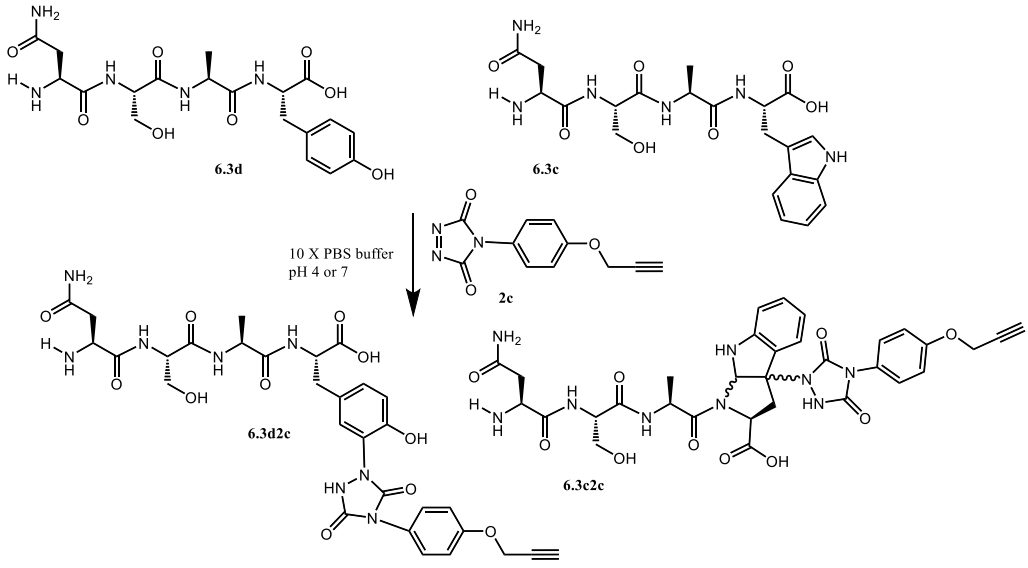


Figure 9.113 Schematic representation of the reaction of peptides **6.3c** and **6.3d** with PTAD-alkyne **2c** in 10 X PBS buffer pH 4 or 7.

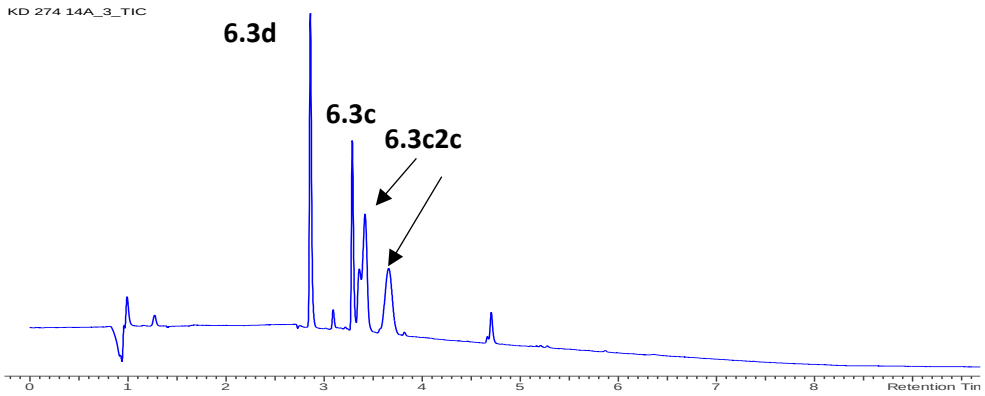


Figure 9.114 LC chromatogram (214 nm) of **2c** conjugation of a **6.3c** and **6.3d** peptide mixture. In this experiment 10 X PBS pH 4 was used as a buffer.

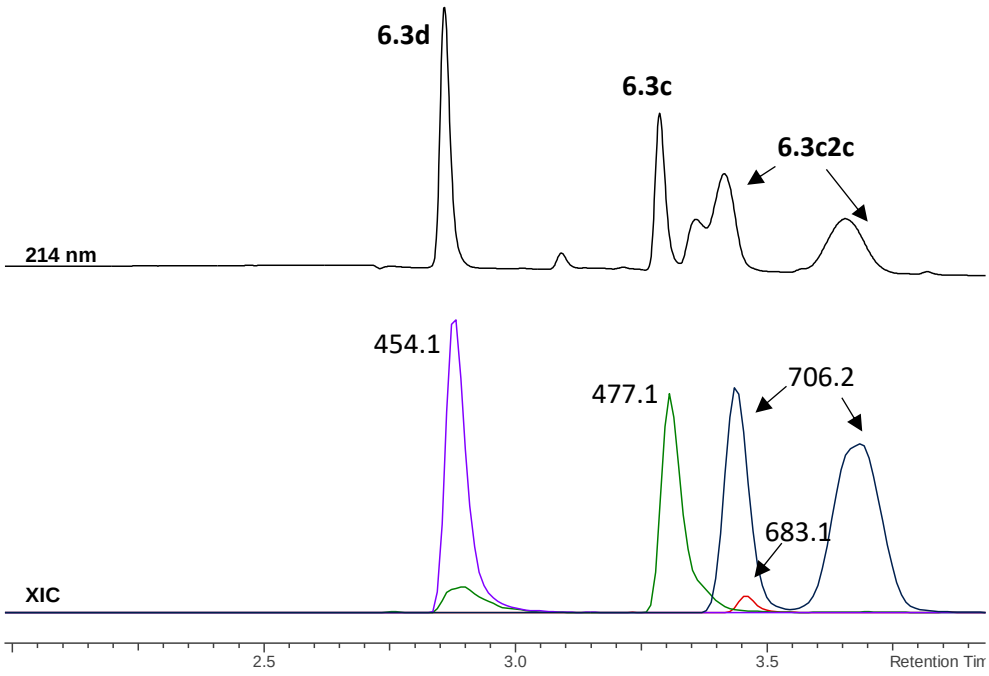


Figure 9.115 LC chromatogram (214 nm) between t_R = 2 min - 4 min (top) of **2c** conjugation of a **6.3c** and **6.3d** peptide mixture in 10 X PBS pH 4. Extracted ion chromatogram (XIC) between t_R = 2 min - 4 min (bottom) for the peptide ions **6.3c** (477.1 Da, green) **6.3d** (454.1 Da, purple) and the conjugated peptide ions **2cc** (706.2 Da, blue) and **6.3d2c** (683.1 Da, red). Almost no product **6.3d2c** is observed at pH 4.

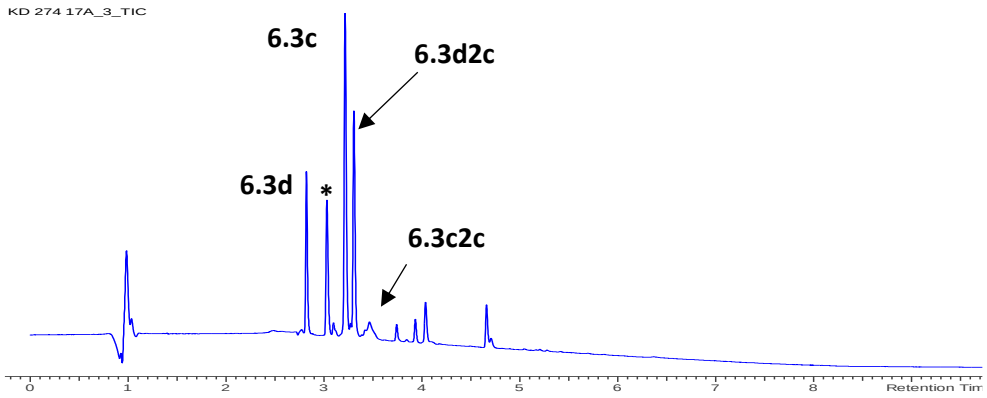


Figure 9.116 LC chromatogram (214 nm) of **2c** conjugation of a **6.3c** and **6.3d** peptide mixture. In this experiment 10 X PBS pH 7 was used as a buffer. The peak annotated with * corresponds with a mass of 232 Da corresponding to the reduced form of **2c**.

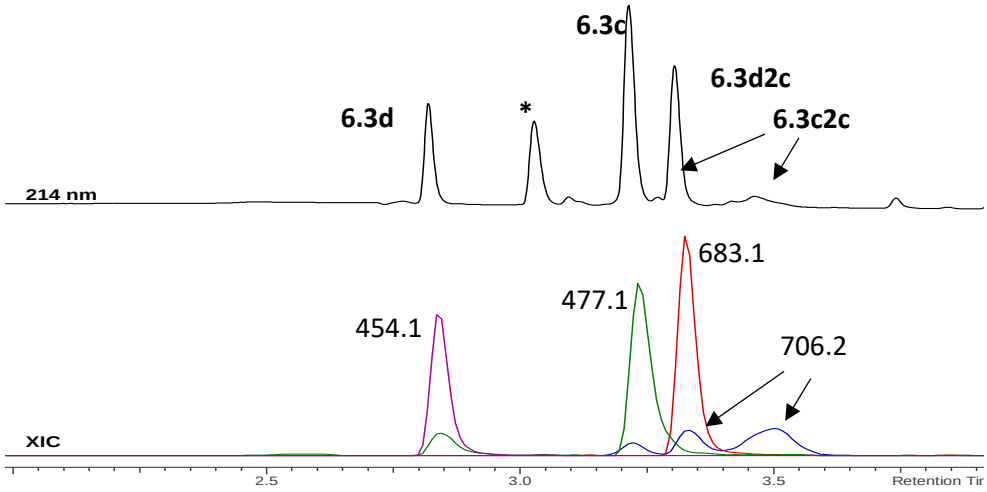


Figure 9.117 LC chromatogram (214 nm) between t_R = 2 min - 4 min (top) of **2c** conjugation of a **6.3c** and **6.3d** peptide mixture in 10 X PBS pH 7. Extracted ion chromatogram (XIC) between t_R = 2 min - 4 min (bottom) for the peptide ions **6.3c** (477.1 Da, green) **6.3d** (454.1 Da, purple) and the conjugated peptide ions **2cc** (706.2 Da, blue) and **6.3d2c** (683.1 Da, red). Both products **6.3d2c** and **6.3c2c** are formed with **6.3d2c** as the main product at pH 7.

Remarkably for peptide with a C-terminal tryptophan amino acid NSAW **6.3c** the reaction mixture after reaction with TAD-propanol **2b** or PTAD-alkyne **2c** contained several reaction product peaks in the LC-MS analysis. These peaks correspond to diastereoisomers of reaction product **6.3c2c** or **6.3c2b**, this product has undergone an additional annulation caused by the reaction of the lone pair on the backbone nitrogen with the indole C2 after reaction of TAD with the indole C3. These structural findings were confirmed via NMR analysis of Boc-Trp-OH and N-Ac-Trp-OMe adducts with TAD-propanol 2b (SI 4). These findings are in agreement with the results reported by Baran et al.[325] on non-peptide related TAD-indole reactions.

Conjugation of Lys-Lys-Ser-Tyr-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH (KKSYLSPRTALINFLV 6.3g) with DMEQ-TAD 2d (6.3g2d). MS (ESI): (M+3)/3: 732.4 (calcd $[M+3H]^{+3}$ 732.4) retention time: 4.00 min.

Conjugation of Lys-Lys-Ser-Trp-Leu-Ser-Pro-Arg-Thr-Ala-Leu-Ile-Asn-Phe-Leu-Val-OH (KKSWLSPRTALINFLV 6.3h) with DMEQ-TAD 2b (6.3h2d). MS (ESI): (M+3)/3: 740.1 (calcd $[M+3H]^{+3}$ 740.0) retention time 4.05 min

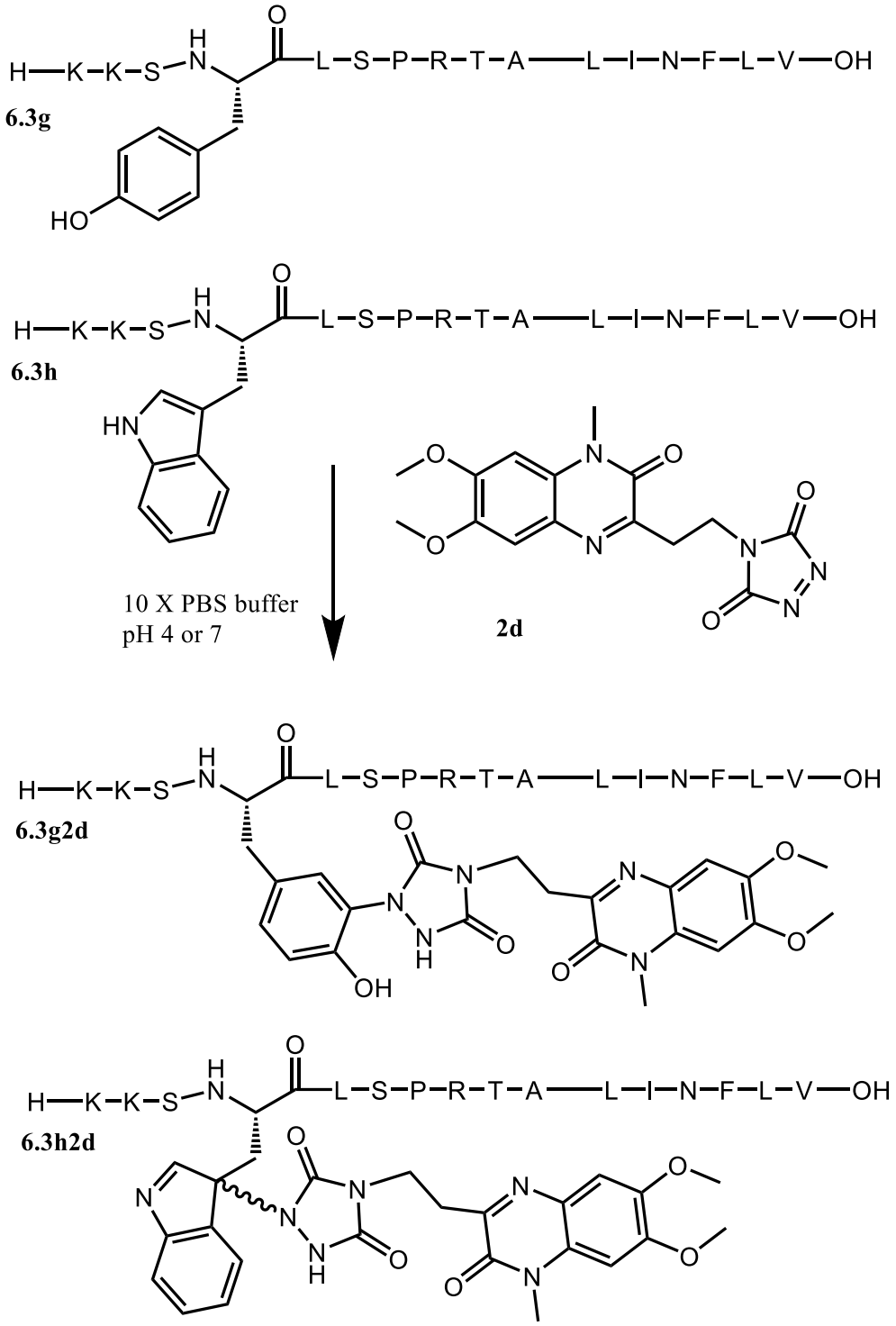


Figure 9.118 Schematic representation of the reaction of peptides **6.3g** and **6.3h** with DMEQ-TAD **2d** in 10 X PBS pH buffer.

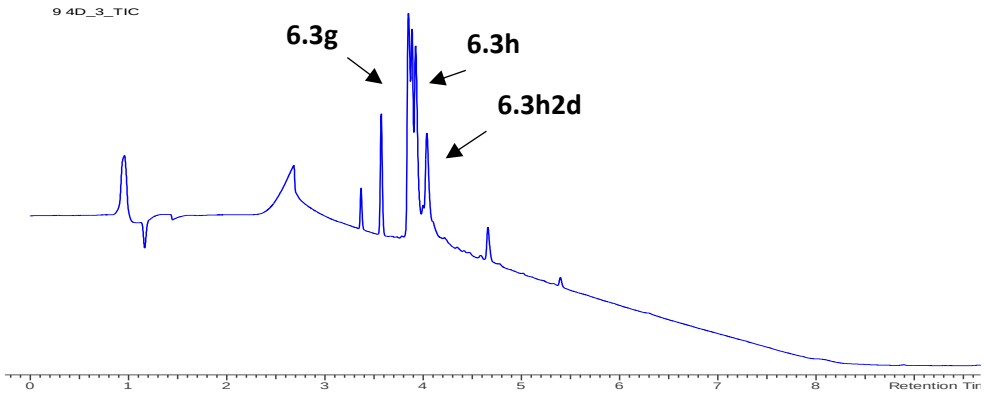


Figure 9.119 LC chromatogram (214 nm) of **2d** conjugation of a **6.3g** and **6.3h** peptide mixture. 10 X PBS pH 4 is used as buffer.

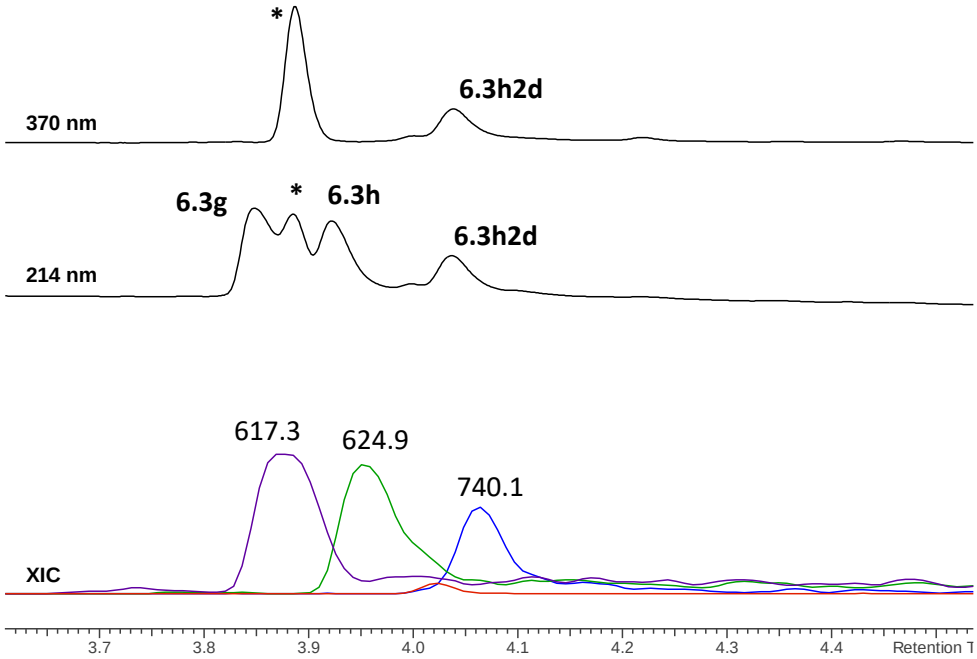


Figure 9.120 LC chromatograms (370 nm / 214 nm) between t_R = 3.6 min – 4.6 min (top) of **2d** conjugation of a **6.3g** and **6.3h** peptide mixture. Extracted ion chromatogram (XIC) between t_R = 3.6 min – 4.6 min (bottom) for the $[M+3H]^{+3}$ peptide ions of **6.3g** (617.3 Da, purple) and **6.3f** (624.9 Da, green) and the conjugated peptide $[M+3H]^{+3}$ ions **6.3g2d** (732.4 Da, red) and **6.3h2d** (740.1 Da, blue). 10 X PBS pH 4 is used as buffer. The peak annotated with * corresponds to DMEQ-urazole with a mass of 348 Da.

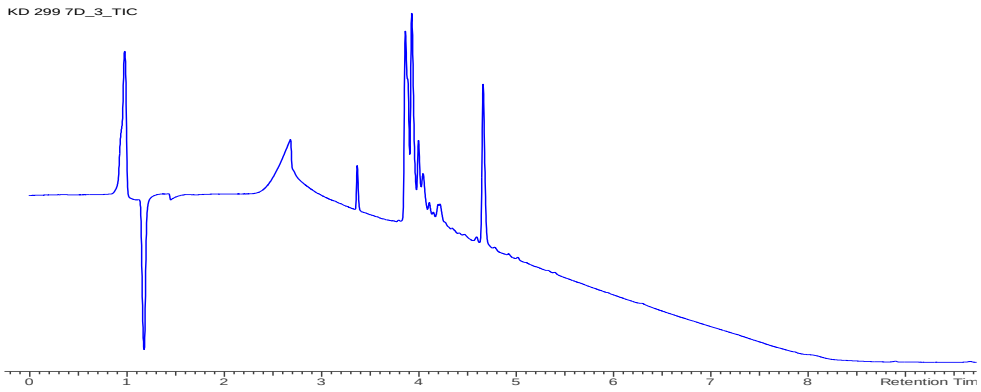


Figure 9.121 LC chromatogram (214 nm) of **2d** conjugation of a **6.3g** and **6.3h** peptide mixture. 10 X PBS pH 7 is used as buffer.

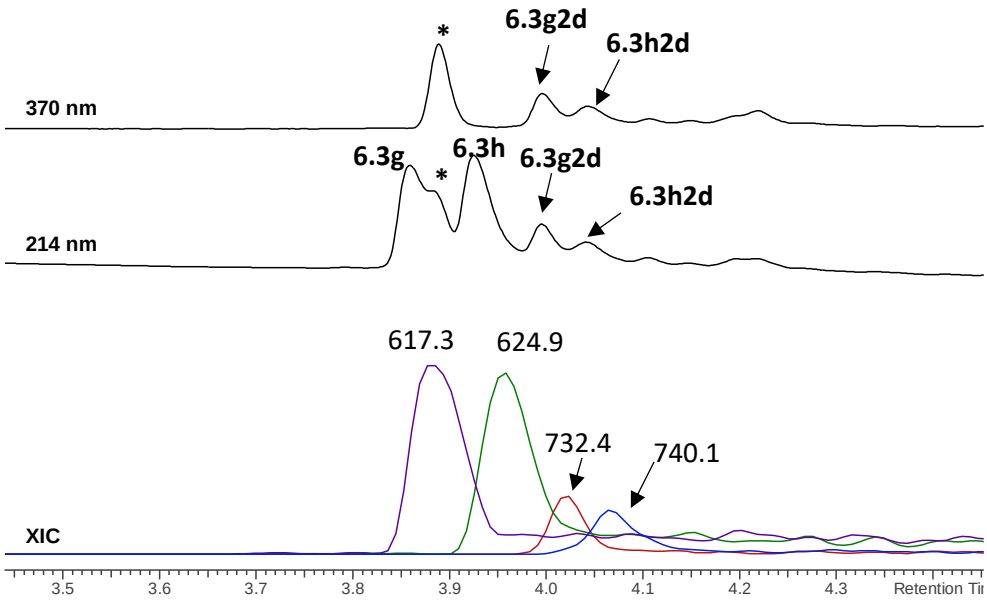


Figure 9.122 LC chromatograms (370 nm / 214 nm) between t_R = 3.6 min – 4.6 min (top) of **2d** conjugation of a **6.3g** and **6.3h** peptide mixture. Extracted ion chromatogram (XIC) between t_R = 3.6 min – 4.6 min (bottom) for the $[M+3H]^{+3}$ peptide ions of **6.3g** (617.3 Da, purple) and **1f** (624.9 Da, green) and the conjugated peptide $[M+3H]^{+3}$ ions **6.3g2d** (732.4 Da, red) and **6.3h2d** (740.1 Da, blue). 10 X PBS pH 7 is used as buffer. The peak annotated with * corresponds to DMEQ-urazole with a mass of 348 Da.

9.3.3.5 MALDI TOF/TOF data of peptide conjugates

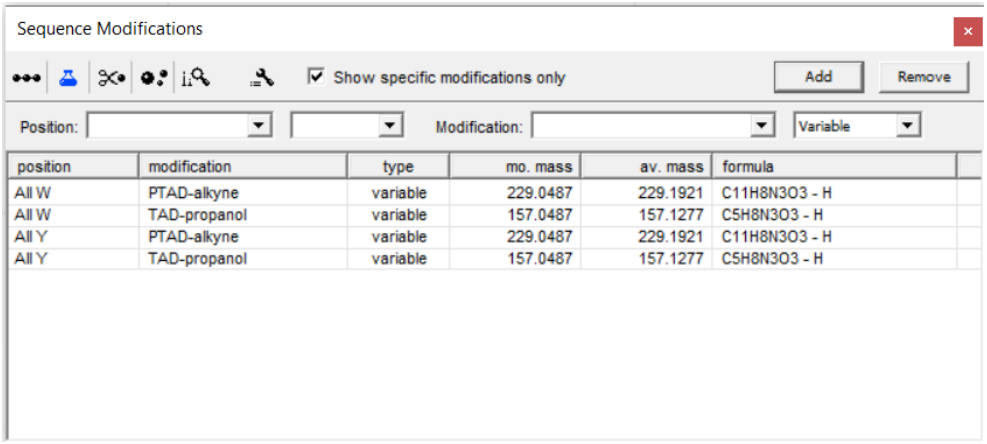
pH dependency of the intramolecular Tyr versus Trp competition: experimental and software setup

For peptides containing both a tyrosine and tryptophan residue, conjugation reactions were performed with different TAD compounds (TAD-propanol **2b,** PTAD-alkyne **2c**). Samples were analyzed via MALDI-TOF/TOF analysis on a 4800+ MALDI TOF/TOF analyser (Applied Biosystems) after HPLC purification of the peptide conjugate peaks. Peptide conjugations were performed as described in the general protocol for peptide conjugation.

Peptides **6.3i** and **6.3j** were used in conjugation reactions with TAD-propanol **2b** or PTAD-alkyne **2c**, conjugation reactions were done in 10x PBS buffers at pH 4 and 7. Peptide **6.3k** was used in conjugation reactions with TAD-propanol **2b** or PTAD-alkyne **2c**, conjugation reactions were done in 10 x PBS buffer pH 3 or pH 7. Finally peptide **6.3l** was used in a conjugation reaction with TAD-propanol **2b** in H_2O.

The TAD-peptide conjugate precursor ions were fragmented using the 1 kV positive method and without gas in the collision chamber (CID off). These settings are post source decay (PSD) MS/MS settings. Using these soft MS/MS settings, still a lot of TAD-modification on tryptophan falls off during MS/MS analysis (vide infra).

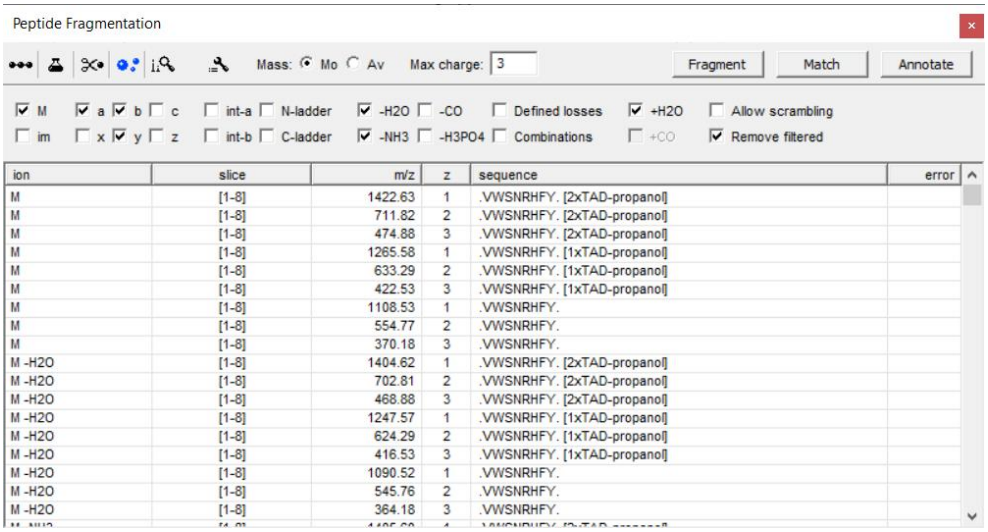
The resulting MS/MS spectra are annotated using the mMass software, TAD modifications are programmed in the software both for tyrosine and tryptophan residues (figure 9.123).



position	modification	type	mo. mass	av. mass	formula
All W	PTAD-alkyne	variable	229.0487	229.1921	C11H8N3O3 - H
All W	TAD-propanol	variable	157.0487	157.1277	C5H8N3O3 - H
All Y	PTAD-alkyne	variable	229.0487	229.1921	C11H8N3O3 - H
All Y	TAD-propanol	variable	157.0487	157.1277	C5H8N3O3 - H

Figure 9.123 Print screen of the TAD-propanol **2b** and PTAD-alkyne **2c** modifications that are taken into account as variable modifications on both tryptophans and tyrosines. A variable modification means that it can either be present or not present.

Subsequently all possible fragment ions are calculated for the fragment ions of choice, which are M, a, b and y ions with possible loss of water and ammonia and addition of water (figure 9.124). Additionally a maximum charge can be indicated, 3 was selected since typically only 1+ and 2+ ions are found in MALDI TOF/TOF spectra.



ion	slice	m/z	z	sequence	error
M	[1-8]	1422.63	1	.VWSNRHFY. [2xTAD-propanol]	
M	[1-8]	711.82	2	.VWSNRHFY. [2xTAD-propanol]	
M	[1-8]	474.88	3	.VWSNRHFY. [2xTAD-propanol]	
M	[1-8]	1265.58	1	.VWSNRHFY. [1xTAD-propanol]	
M	[1-8]	633.29	2	.VWSNRHFY. [1xTAD-propanol]	
M	[1-8]	422.53	3	.VWSNRHFY. [1xTAD-propanol]	
M	[1-8]	1108.53	1	.VWSNRHFY.	
M	[1-8]	554.77	2	.VWSNRHFY.	
M	[1-8]	370.18	3	.VWSNRHFY.	
M -H2O	[1-8]	1404.62	1	.VWSNRHFY. [2xTAD-propanol]	
M -H2O	[1-8]	702.81	2	.VWSNRHFY. [2xTAD-propanol]	
M -H2O	[1-8]	468.88	3	.VWSNRHFY. [2xTAD-propanol]	
M -H2O	[1-8]	1247.57	1	.VWSNRHFY. [1xTAD-propanol]	
M -H2O	[1-8]	624.29	2	.VWSNRHFY. [1xTAD-propanol]	
M -H2O	[1-8]	416.53	3	.VWSNRHFY. [1xTAD-propanol]	
M -H2O	[1-8]	1090.52	1	.VWSNRHFY.	
M -H2O	[1-8]	545.76	2	.VWSNRHFY.	
M -H2O	[1-8]	364.18	3	.VWSNRHFY.	

Figure 9.124 Print screen of the calculated peptide fragmentation.

The calculated peptide fragments are matched with the peaks in the MS/MS spectra. The tolerance for deviation between the calculated and detected fragments can be indicated. In this work we selected 0.5 Da as tolerance. The matches are detected fragments that are within the tolerance from the calculated fragment. The error of all the matches is plotted as function of the m/z. We found that for experiments on the MALDI-TOF/TOF with charge +1, the error typically increases slightly from values below 0.1 Da for 300 m/z to around 0.3 Da for 1000 m/z and above. If two calculated fragments incidentally match with one detected fragment the error plot is helpful to select the most suitable fragment ion. In figure 9.125 two calculated fragments match to a detected fragment around 400 m/z.

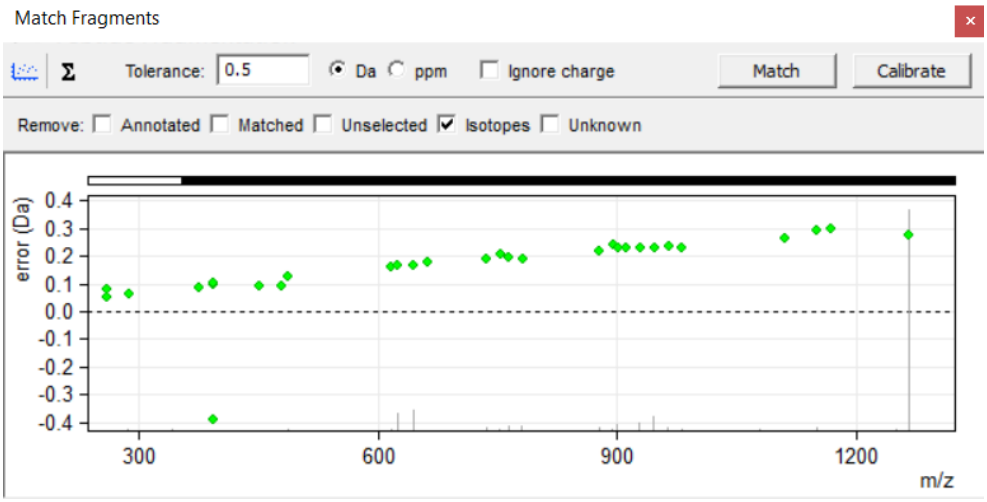


Figure 9.125 Print screen of the error vs m/z plot for the matched peptide fragments.

The annotated MS/MS spectra are provided along with the chemical structure of the peptide with the detected fragments indicated. a and b ions are displayed below the structure and start from the N-terminus. y ions are displayed above the structure and start from the C-terminus.

The analysis of these MS/MS data demonstrates that the TAD modification shifts from being selective on tryptophan, in conditions of pH 4 not a single fragment ion with a TAD modification on tyrosine was detected, to both modification on tyrosine

and tryptophan in higher pH conditions (pH 7). Additionally the data show that the TAD modification on tryptophan is not stable during the MALDI TOF/TOF analysis. This is shown by the large abundance of the ion corresponding to the complete peptide without TAD modification and this effect is seen in experiments that were done in pH 4 (tryptophan conjugation). Due to the fact that the TAD modification on tryptophan is not stable during MALDI TOF/TOF analysis it seems that the TAD modification is selectively taking place on the tyrosine residue in conditions of pH 7. This is however not the case as we have shown that both tyrosine and tryptophan will react at pH 7, and sometimes TAD modified peptide fragments encompassing the tryptophan residue are still observed in MALDI TOF/TOF analysis when conjugation was done at pH 7.

MS/MS spectra of unmodified peptides

Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH (VWSNRHFY, 6.3i). MS (MALDI-TOF) precursor ion: m/z 1108.8 (calcd $[M+H]^+$ 1108.5)

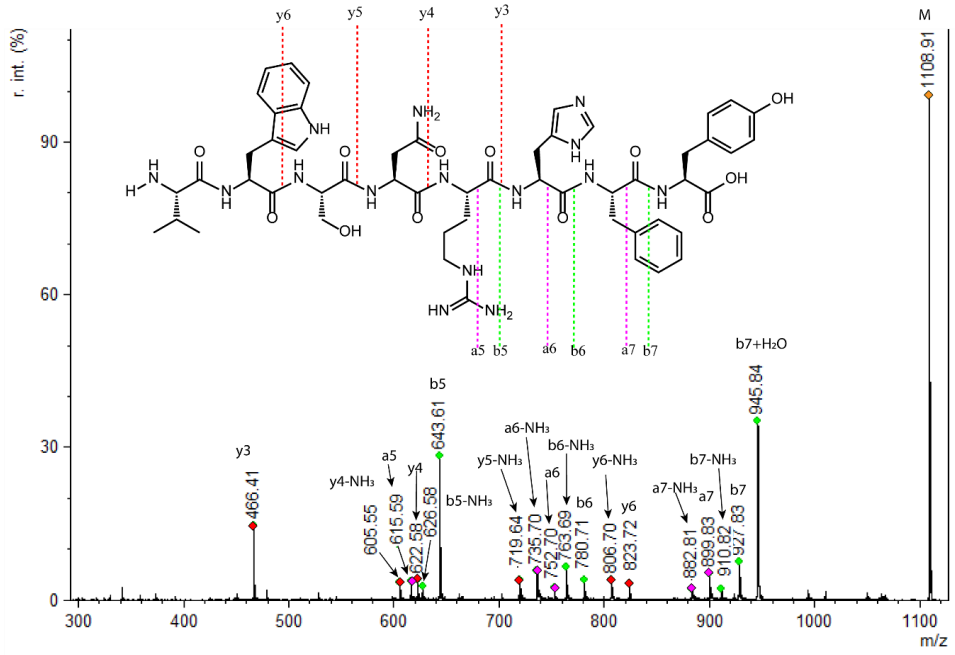


Figure 9.126 MALDI TOF/TOF spectrum of unmodified peptide **6.3i**.

Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH (VYSNRHFW, 6.3j). MS (MALDI-TOF) precursor ion: m/z 1108.8 (calcd $[M+H]^+$ 1108.5)

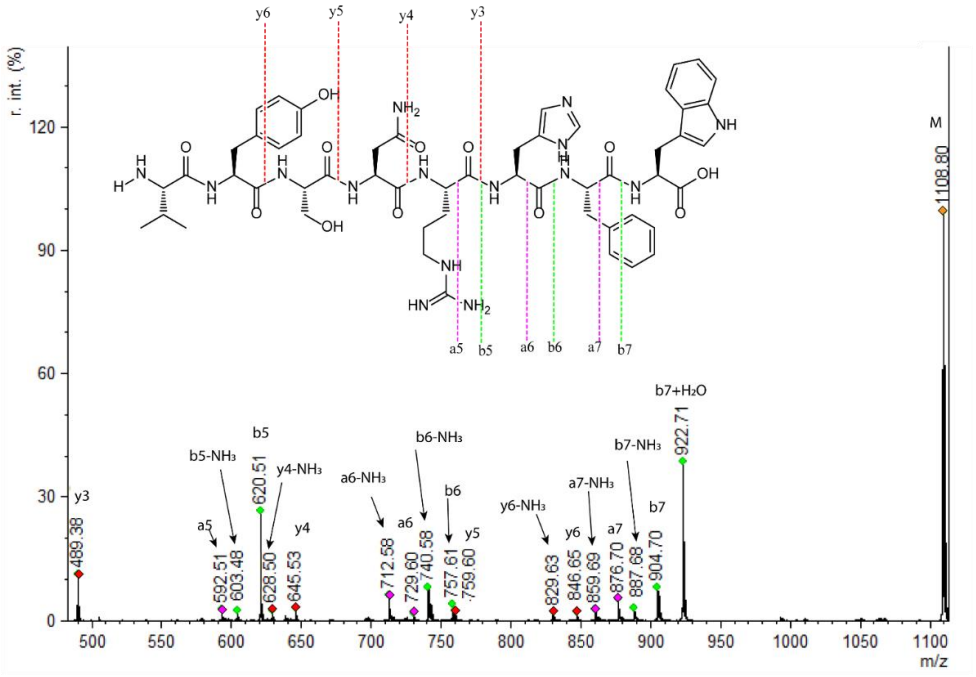


Figure 9.127 MALDI TOF/TOF spectrum of unmodified peptide **6.3j**.

Val-Trp-Ser-Gln-Lys-Arg-His-Phe-Gly-Tyr-OH (VWSQKRHFGY, 6.3k). MS (MALDI-TOF) precursor ion: m/z 1307.6 (calcd $[M+H]^+$ 1307.7)

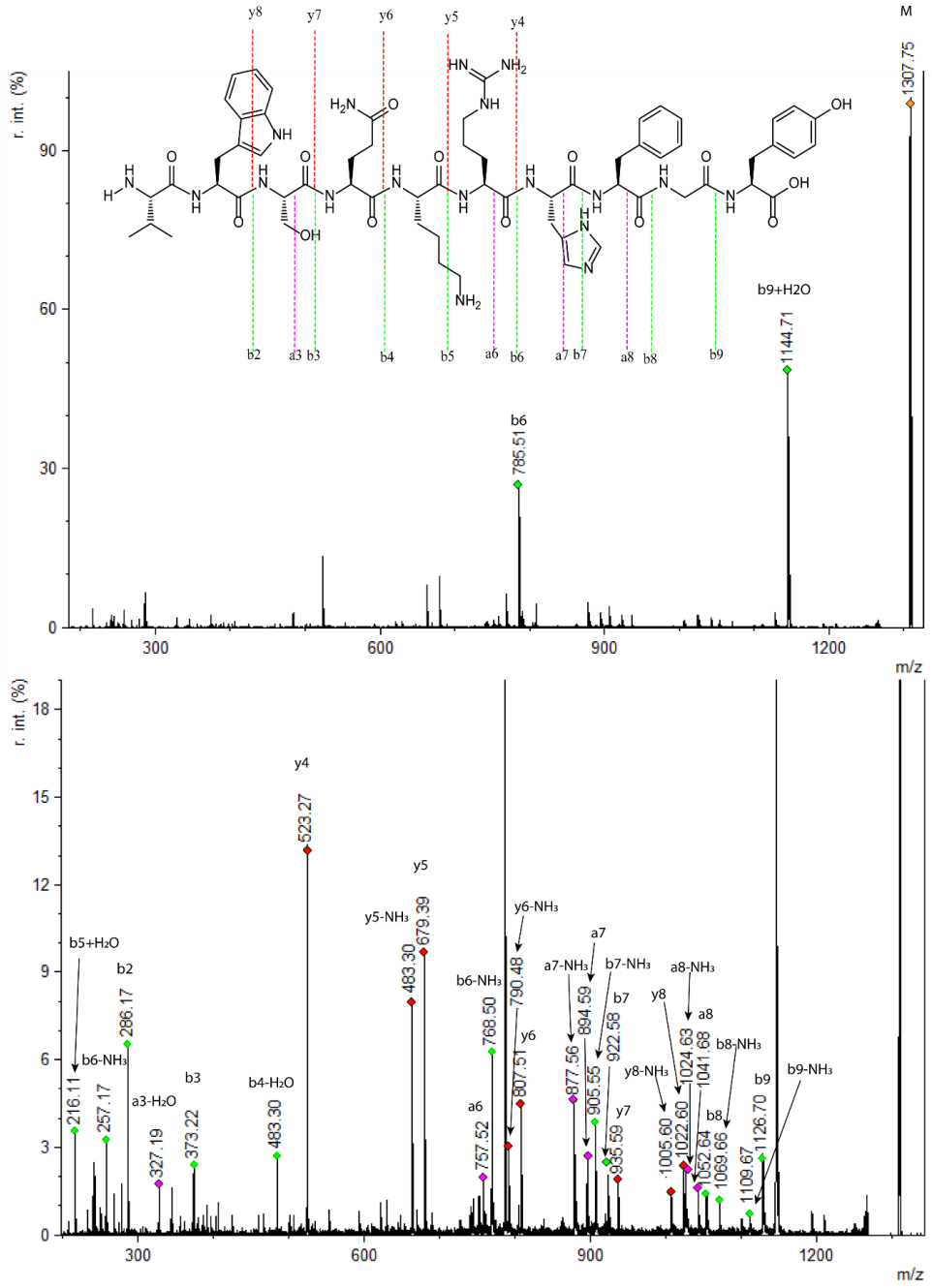


Figure 9.128 MALDI TOF/TOF spectrum of unmodified peptide **6.3k**.

Lys-Asp-Tyr-Trp-Glu-Cys-Ala-OH (KDYWECA, 6.3l). MS (MALDI-TOF) precursor ion: m/z 914.4 (calcd $[M+H]^+$ 914.4)

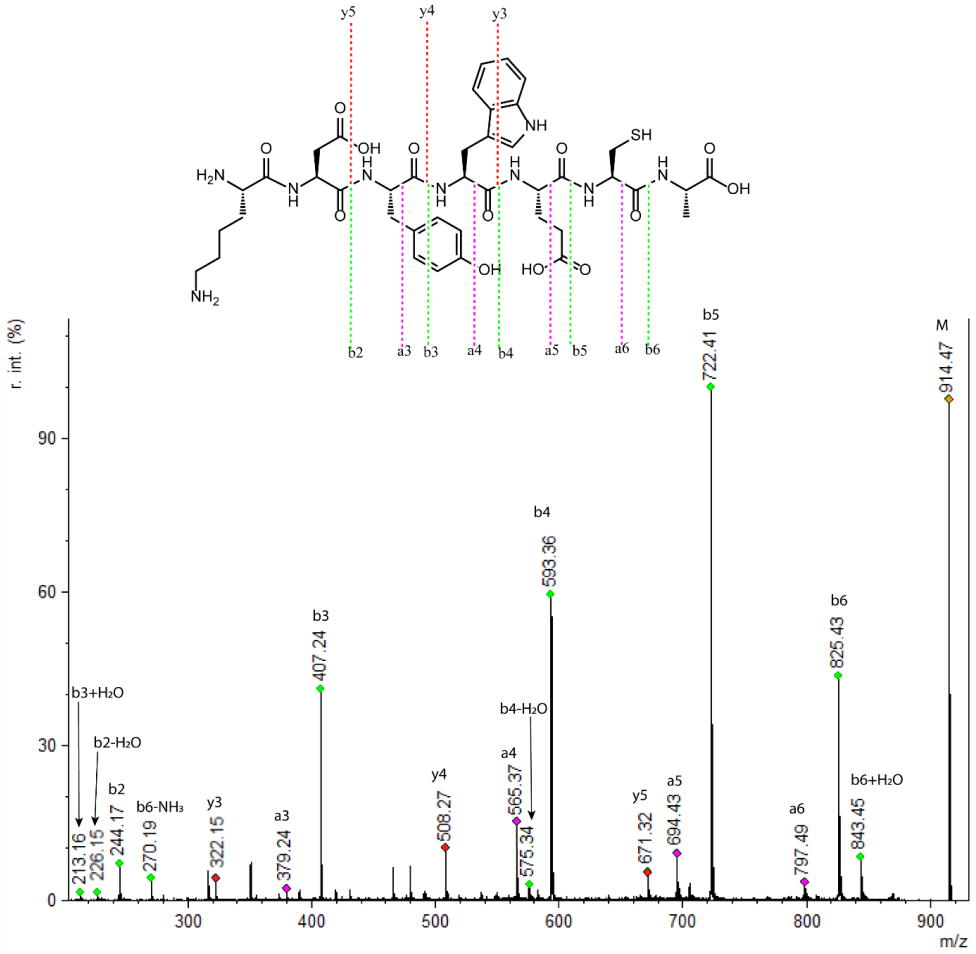


Figure 9.129 MALDI TOF/TOF spectrum of unmodified peptide **6.3l**.

MS/MS spectra of modified peptides

Conjugation product of Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH (VWSNRHFY, 6.3i) with TAD-propanol 2b (6.3i2b). MS (MALDI-TOF) precursor ion: m/z 1265.8 (calcd $[M+H]^+$ 1265.6)

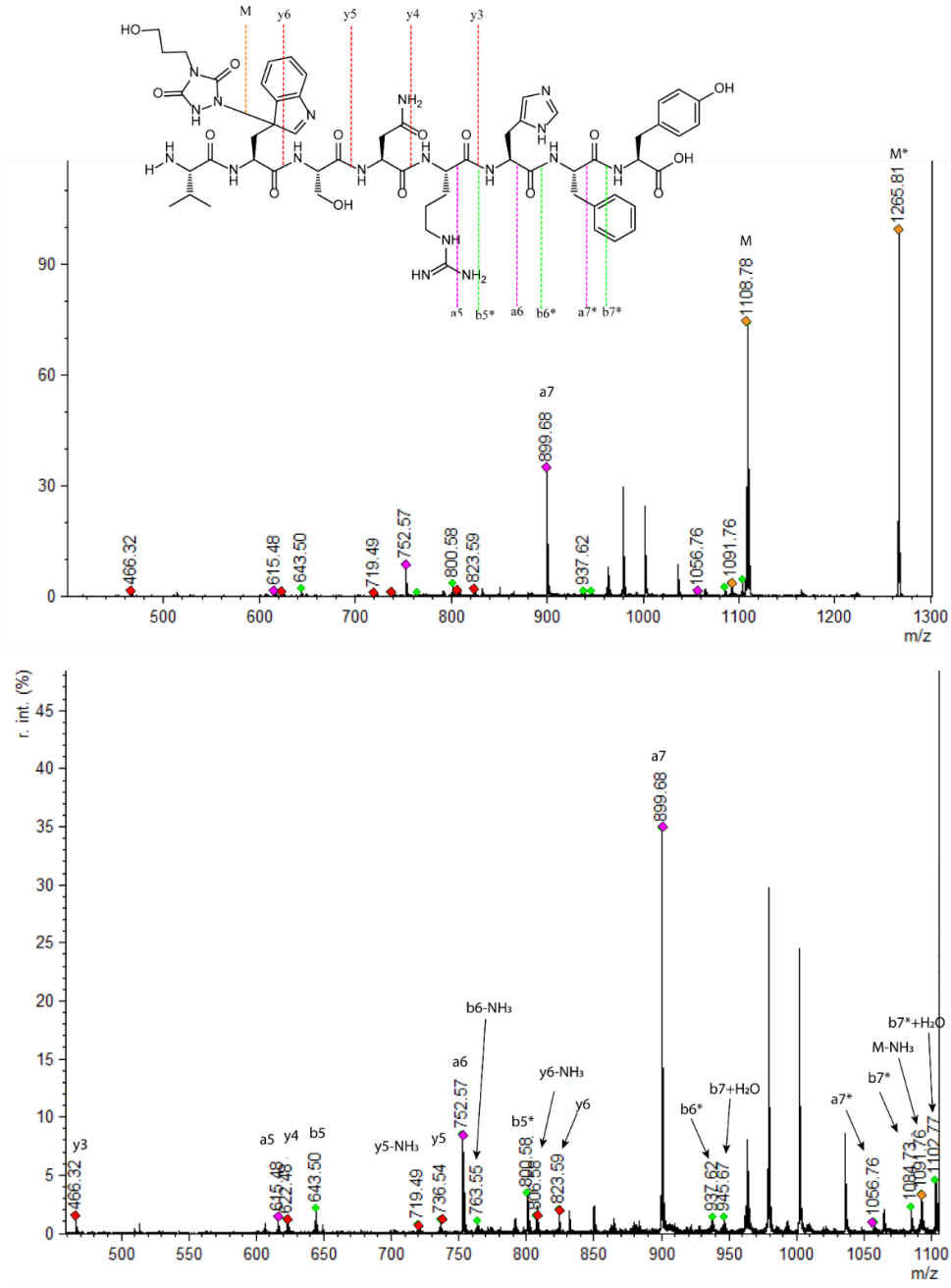


Figure 9.130 MALDI-TOF/TOF spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 1265.8 Da). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 4.

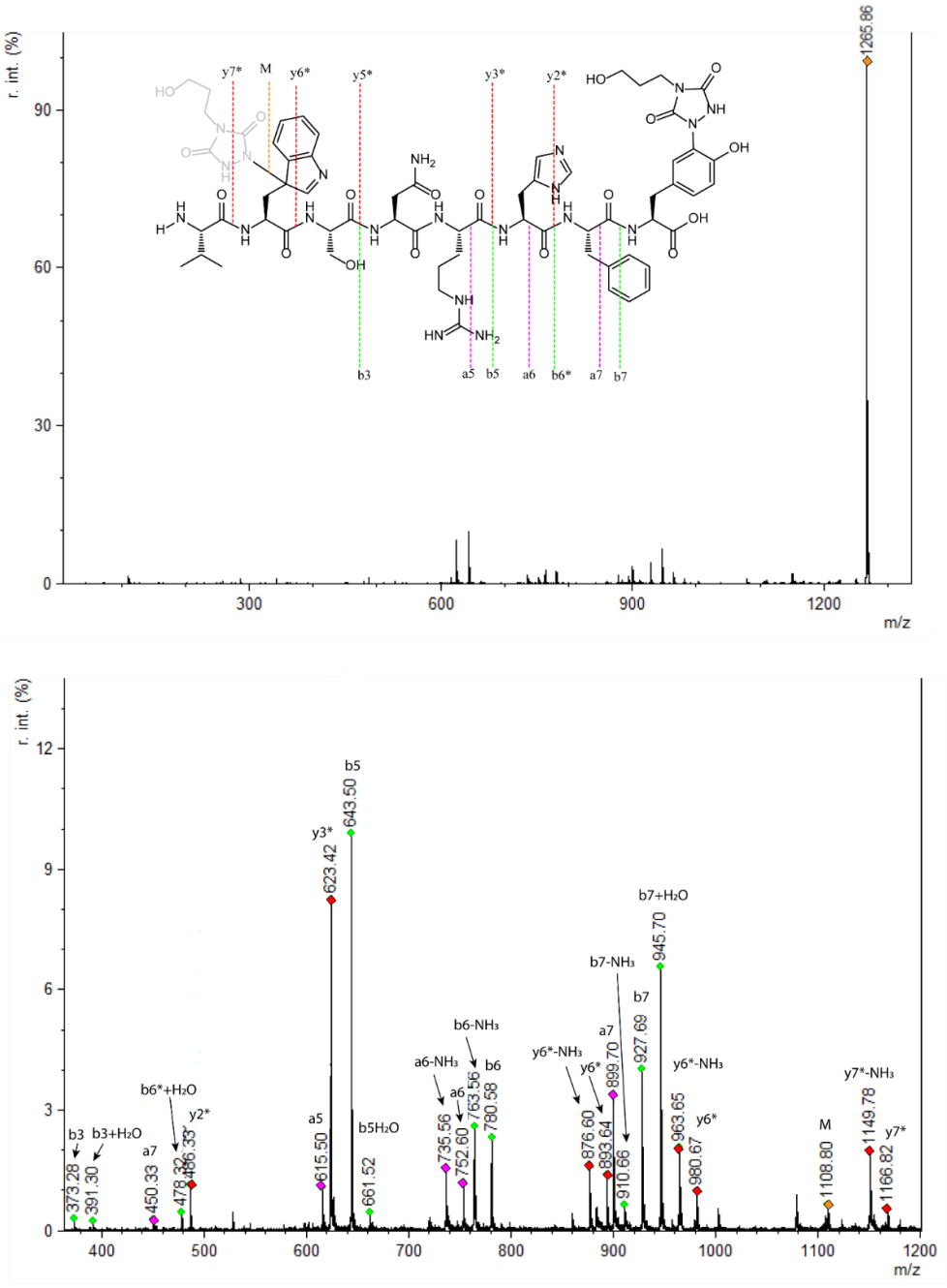


Figure 9.131 MALDI-TOF/TOF spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 1265.8 Da). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 7.

Conjugation product of Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH (VWSNRHFY, 6.3i) with PTAD-alkyne 2c (6.3i2c). MS (MALDI-TOF) precursor ion: m/z 1337.8 (calcd $[M+H]^+$ 1337.5)

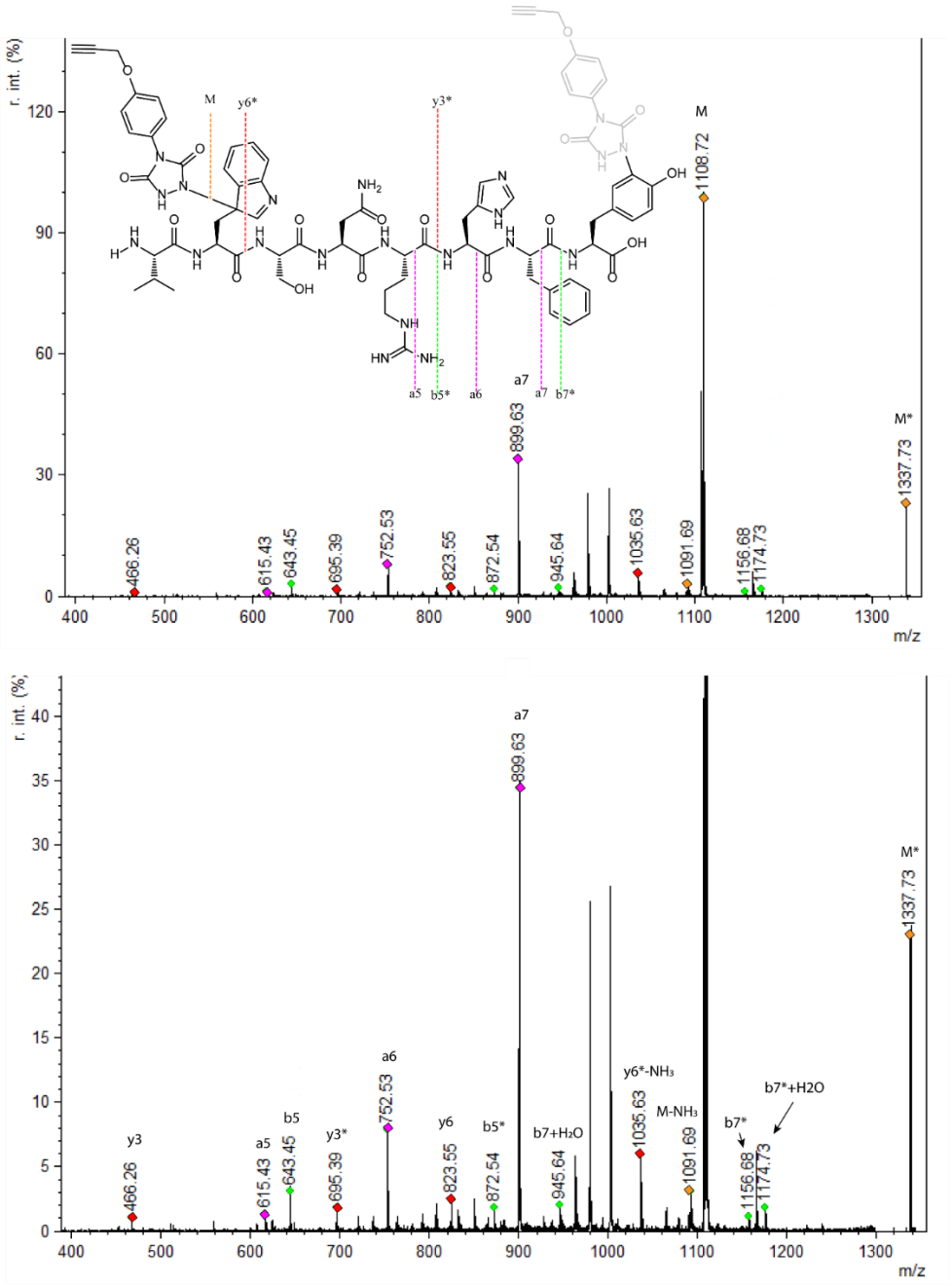


Figure 9.132 MALDI-TOF/TOF spectrum of peptide **6.3i** conjugated with **2c** (precursor ion: 1337.8 Da). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 4.

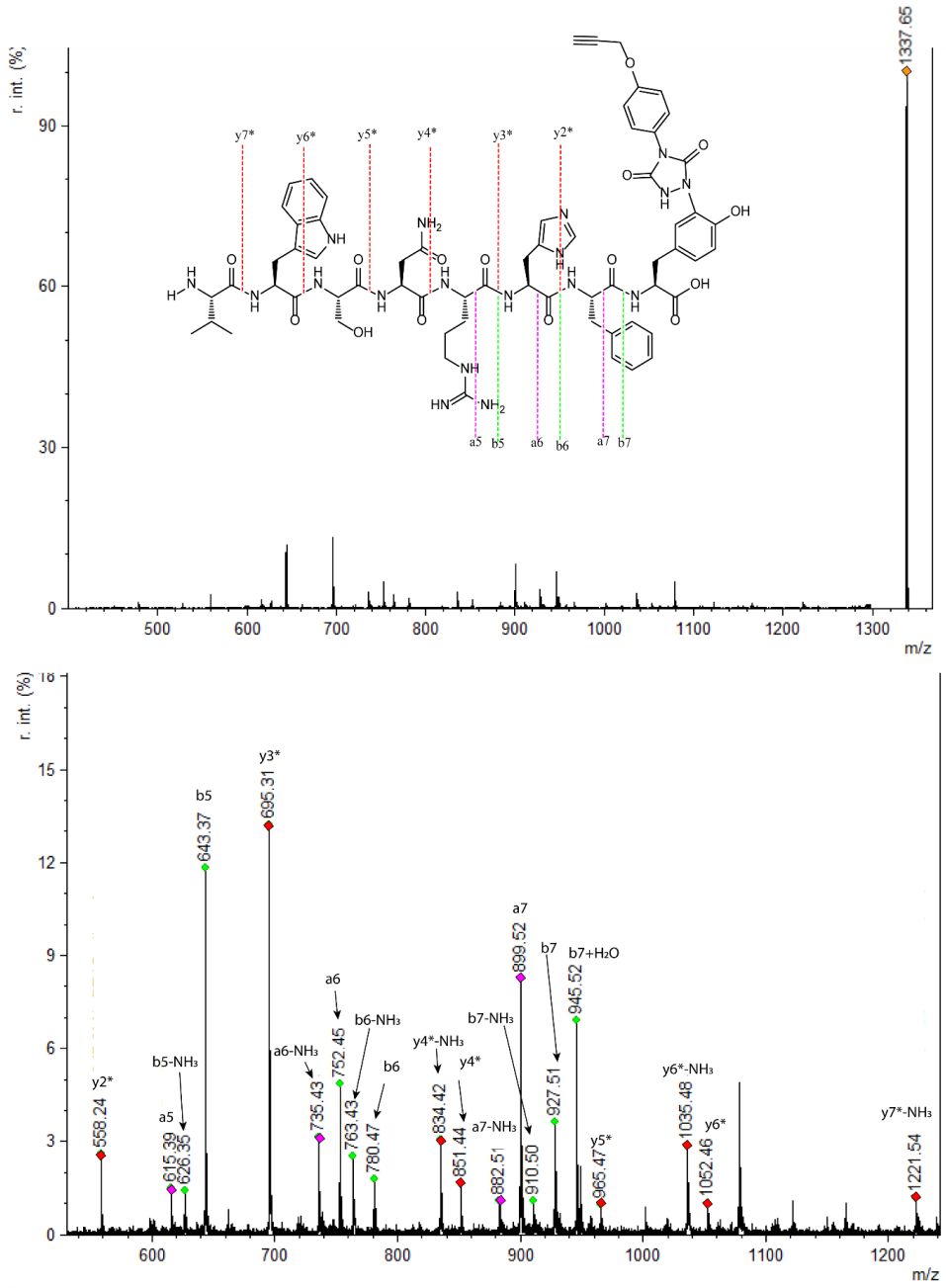


Figure 9.133 MALDI-TOF/TOF spectrum of peptide **6.3i** conjugated with **2c** (precursor ion: 1337.8 Da). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 7.

Conjugation product of Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH (VYSNRHFW, 6.3j) with TAD-propanol 2b (6.3j2b). MS (MALDI-TOF) precursor ion: m/z 1265.8 (calcd $[M+H]^+$ 1265.6)

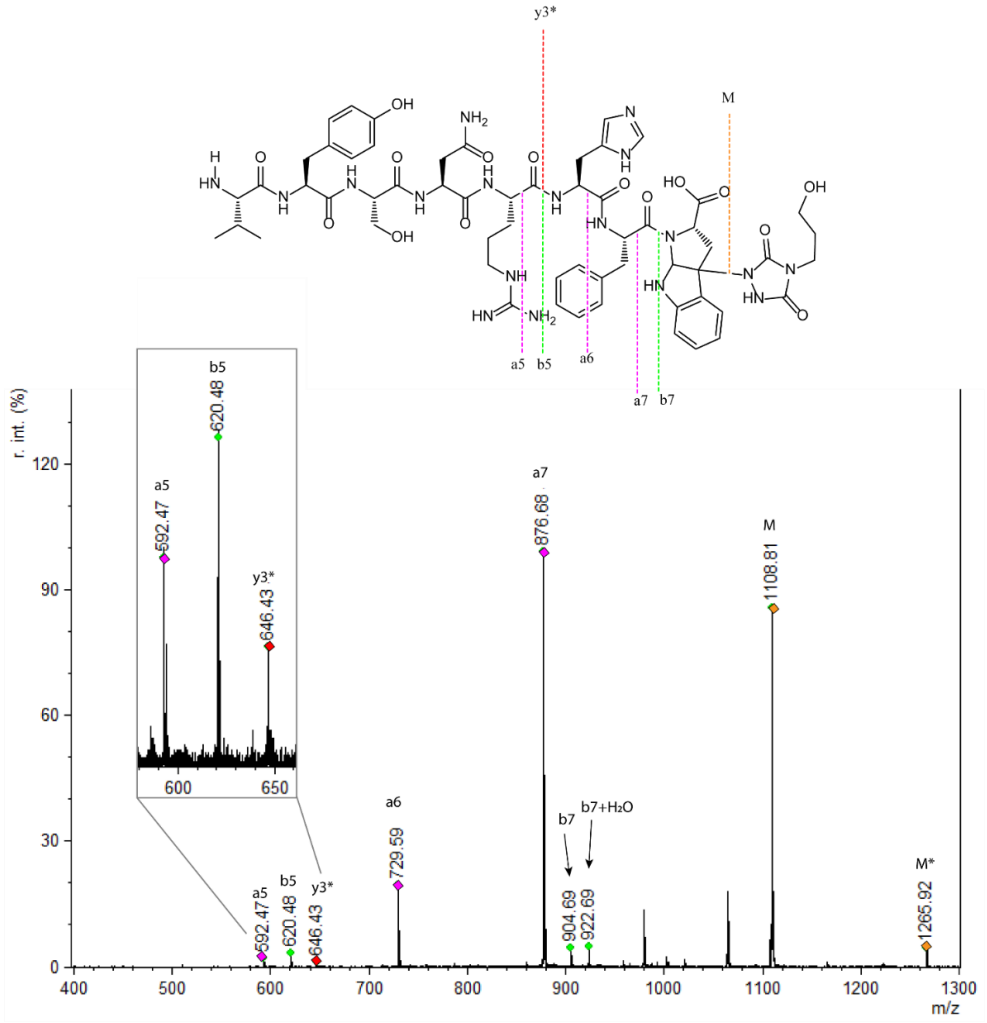
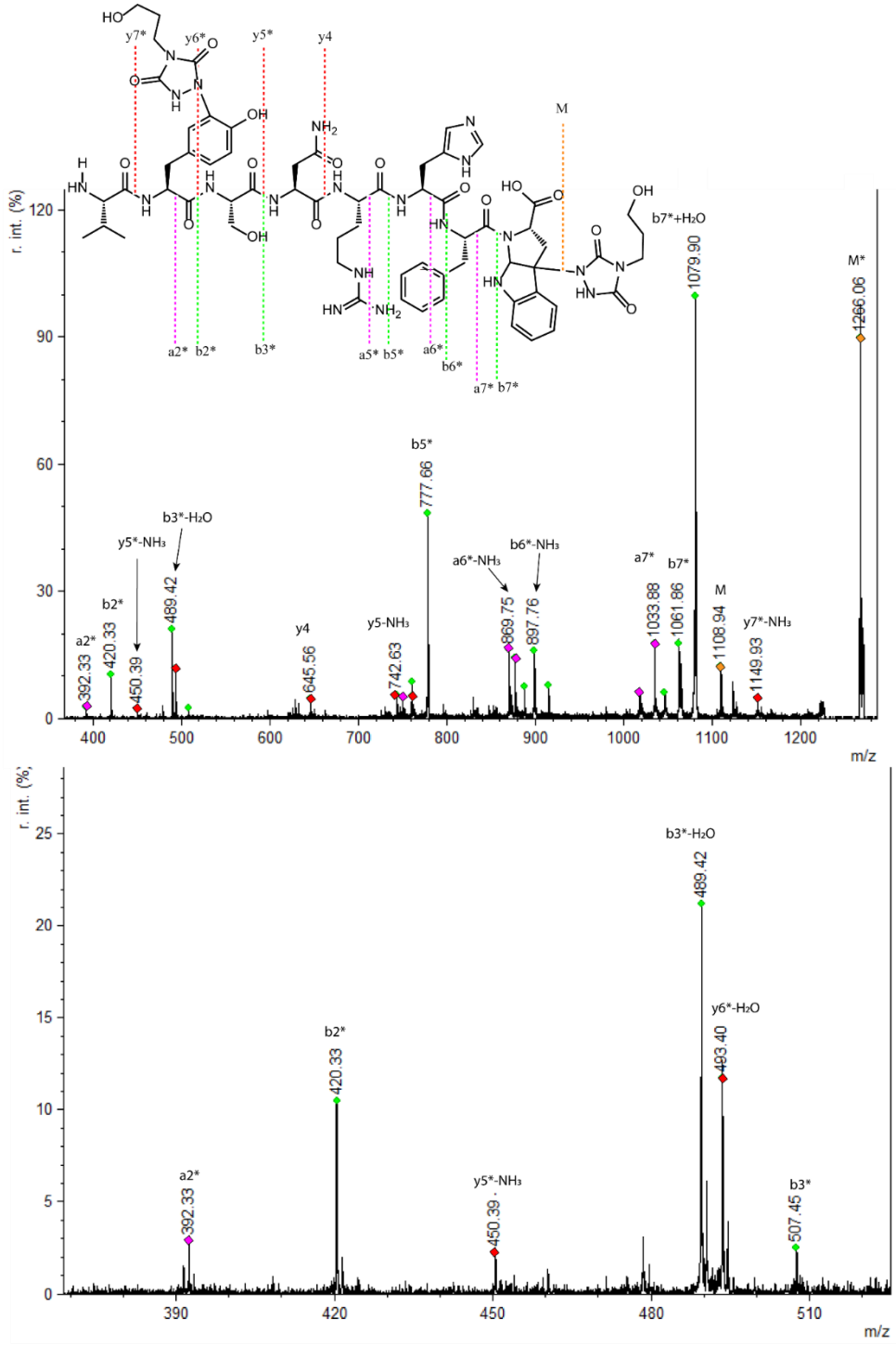


Figure 9.134 MALDI-TOF/TOF spectrum of peptide **6.3j** conjugated with **2b** (precursor ion: 1265.8 Da). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 4.


r. int. (%)
m/z
y7*
y6*
y5*
y4
M
b7*+H₂O
M*
1266.06
1079.90
a2*
b2*
b3*
a5*
b5*
a6*
b6*
a7*
b7*
b5*
777.66
b3*-H₂O
489.42
y5*-NH₃
b6*-NH₃
a6*-NH₃
a7*
b7*
M
b2*
a2*
420.33
392.33
450.39
y4
645.56
y5-NH₃
742.63
869.75
897.76
1033.88
1061.86
1108.94
y7*-NH₃
1149.93
b3*-H₂O
489.42
y6*-H₂O
493.40
b2*
420.33
a2*
392.33
y5*-NH₃
450.39
b3*
507.45

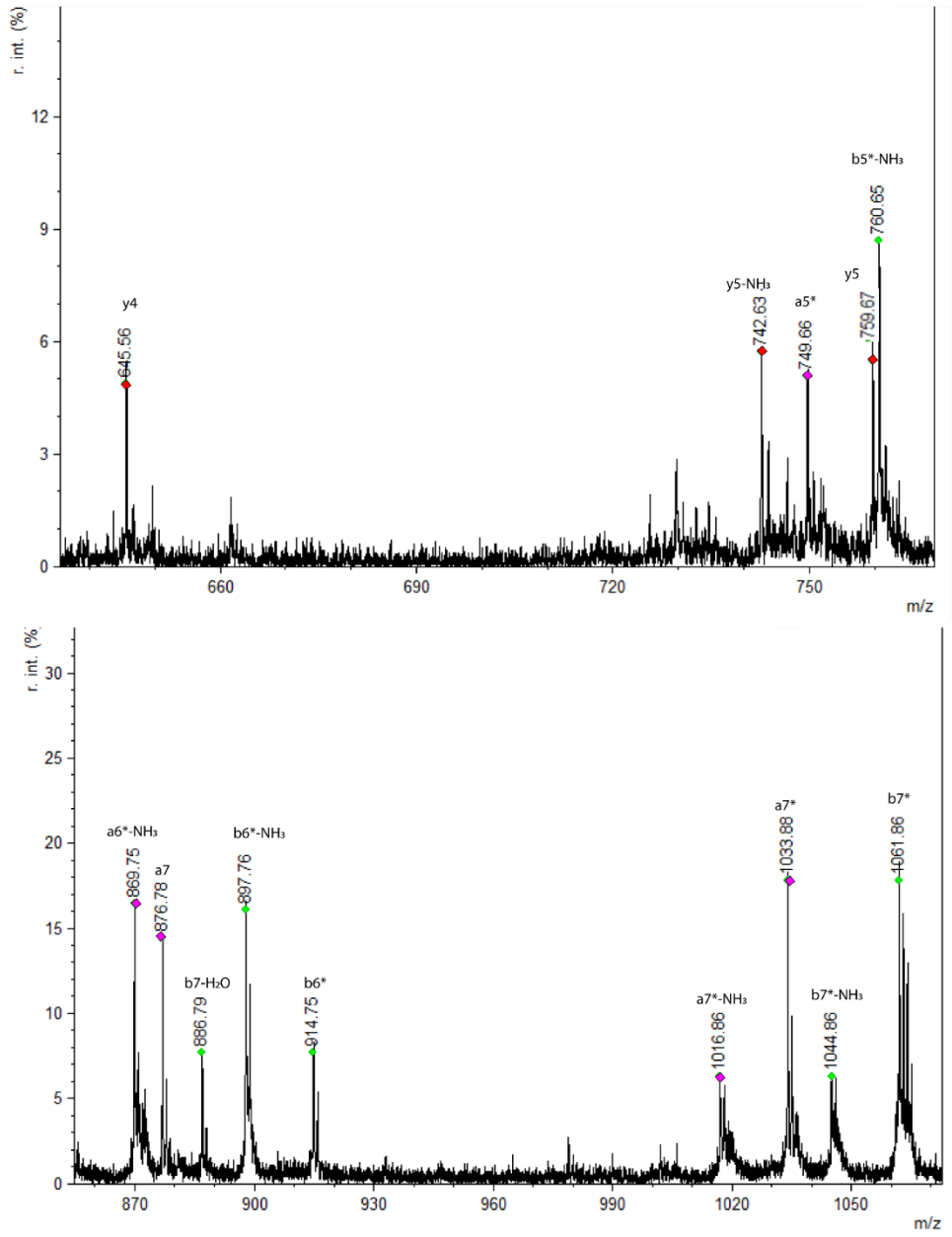


Figure 9.135 MALDI-TOF/TOF spectrum of peptide **6.3j** conjugated with **2b** (precursor ion: 1265.8 Da). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 7.

Conjugation product of Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH (VYSNRHFW, 6.3j) with PTAD-alkyne 2c (6.3j2c). MS (MALDI-TOF) precursor ion: m/z 1337.8 (calcd $[M+H]^+$ 1337.5)

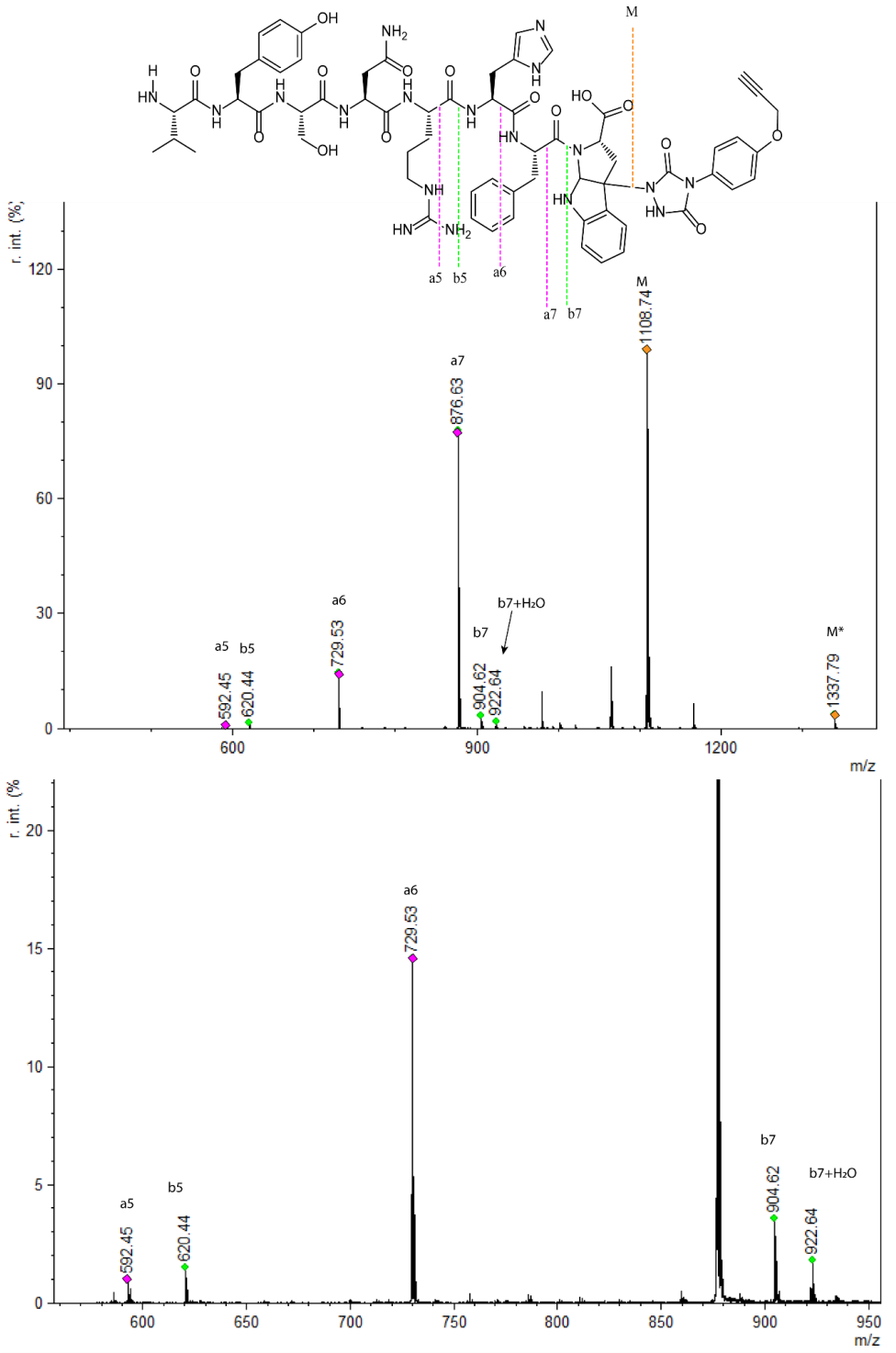


Figure 9.136 MALDI-TOF/TOF spectrum of peptide **6.3j** conjugated with **2c** (precursor ion: 1337.8 Da). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 4.

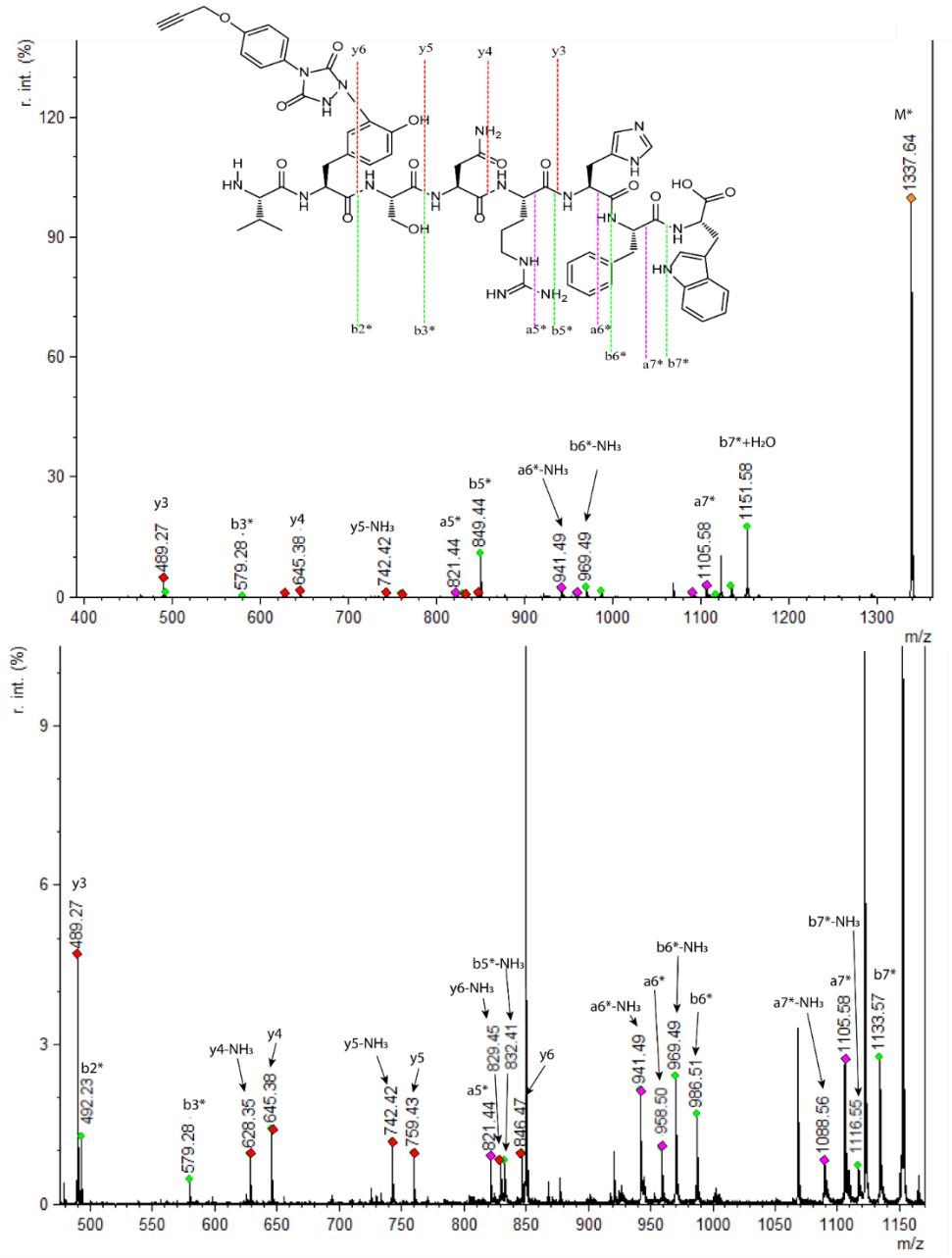
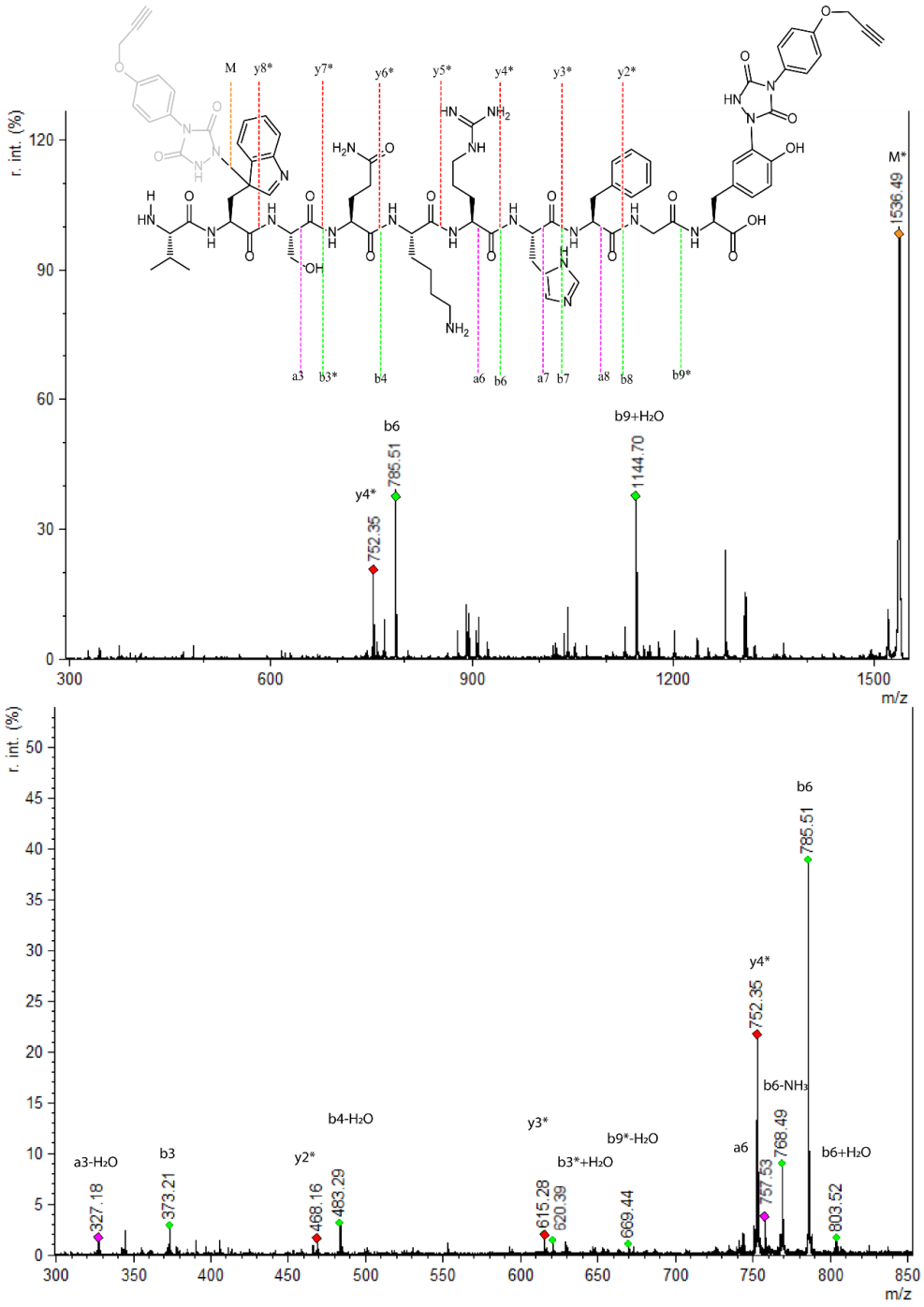


Figure 9.137 MALDI-TOF/TOF spectrum of peptide **6.3j** conjugated with **2c** (precursor ion: 1337.8 Da). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 7.

Conjugation product of Val-Trp-Ser-Gln-Lys-Arg-His-Phe-Gly-Tyr-OH (VWSQKRHFGY, 6.3k) with PTAD-alkyne 2c (6.3k2c). MS (MALDI-TOF) precursor ion: m/z 1536.6 (calcd $[M+H]^+$ 1536.7)



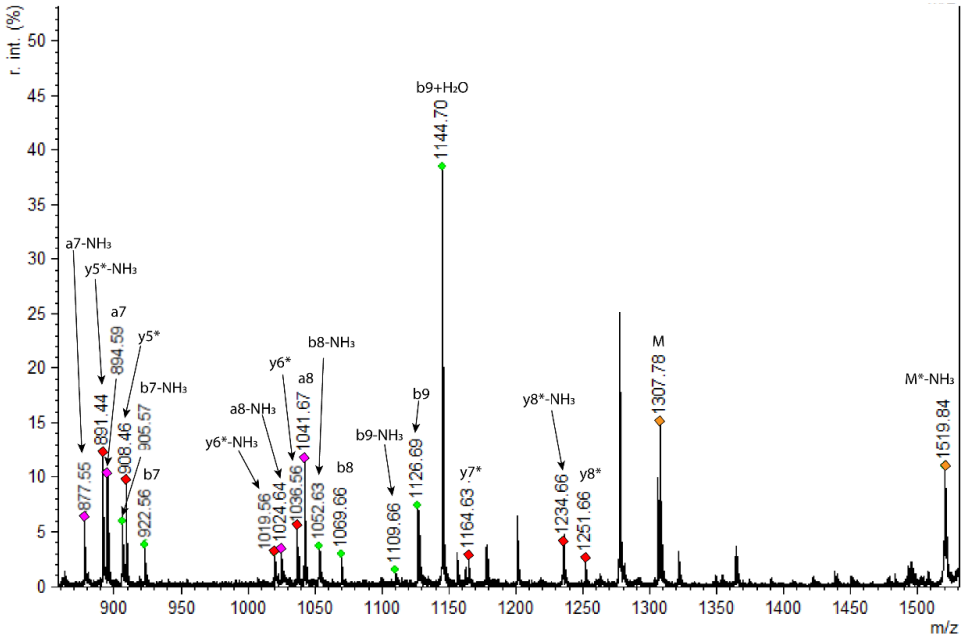


Figure 9.138 MALDI-TOF/TOF spectrum of peptide **6.3k** conjugated with **2c** (precursor ion: 1536.6 Da). Chemical structure of peptide **6.3k** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10x PBS pH 7.

Conjugation product of Lys-Asp-Tyr-Trp-Glu-Cys-Ala-OH (KDYWECA, 1l) with TAD-propanol 2b (2lb). MS (MALDI-TOF) precursor ion: m/z 1071.4 (calcd $[M+H]^+$ 1071.4)

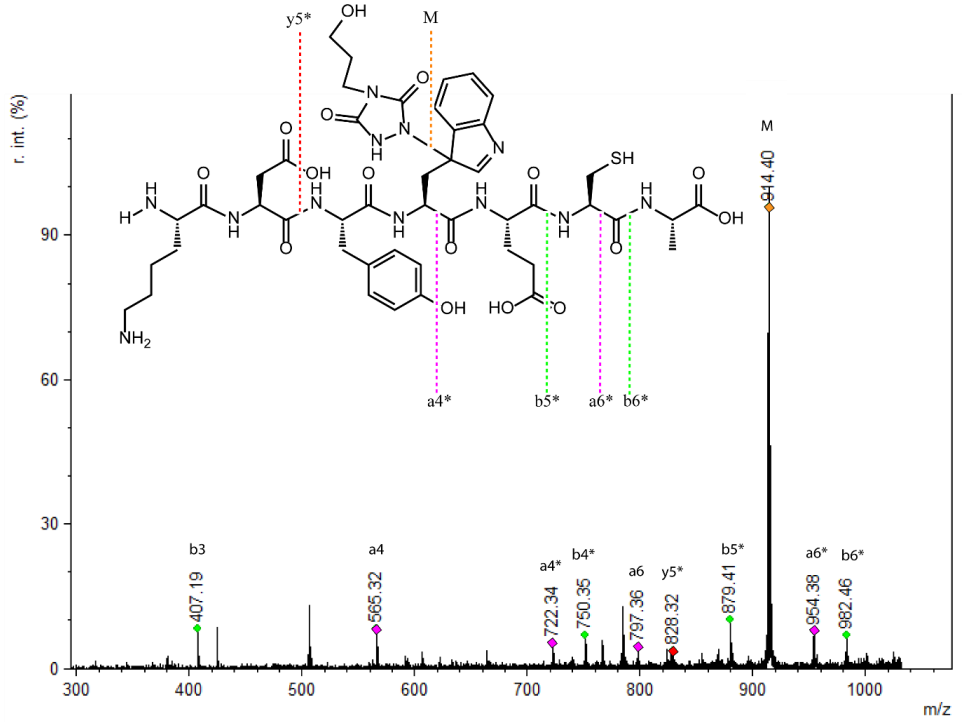


Figure 9.139 MALDI-TOF/TOF spectrum of peptide **6.3l** conjugated with **2b** (precursor ion: 1071.4 Da). Chemical structure of peptide **6.3l** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in MilliQ H_2O.

MS/MS tryptophan-TAD fragmentation patterns

In this section several recurring peaks in the MS/MS spectra of TAD conjugated peptides are investigated in more detail. In the MS/MS spectra of peptides containing a tryptophan residue a peak was observed after reaction with TAD corresponding with the cleavage of the C_α-C_β bond of the labeled tryptophan residue. This mass cannot be attributed to a TAD modified fragment since the same mass is found irrespective of the nature of the TAD reagent used. For peptide **6.3j** with a C-terminal tryptophan an additional alternative fragment was observed in the MS/MS spectra. The bond between C3 of the indole and TAD-nitrogen is cleaved as well as a decarboxylation taking place. The same mass is found in MS/MS spectra of peptides modified with different TAD reagents.

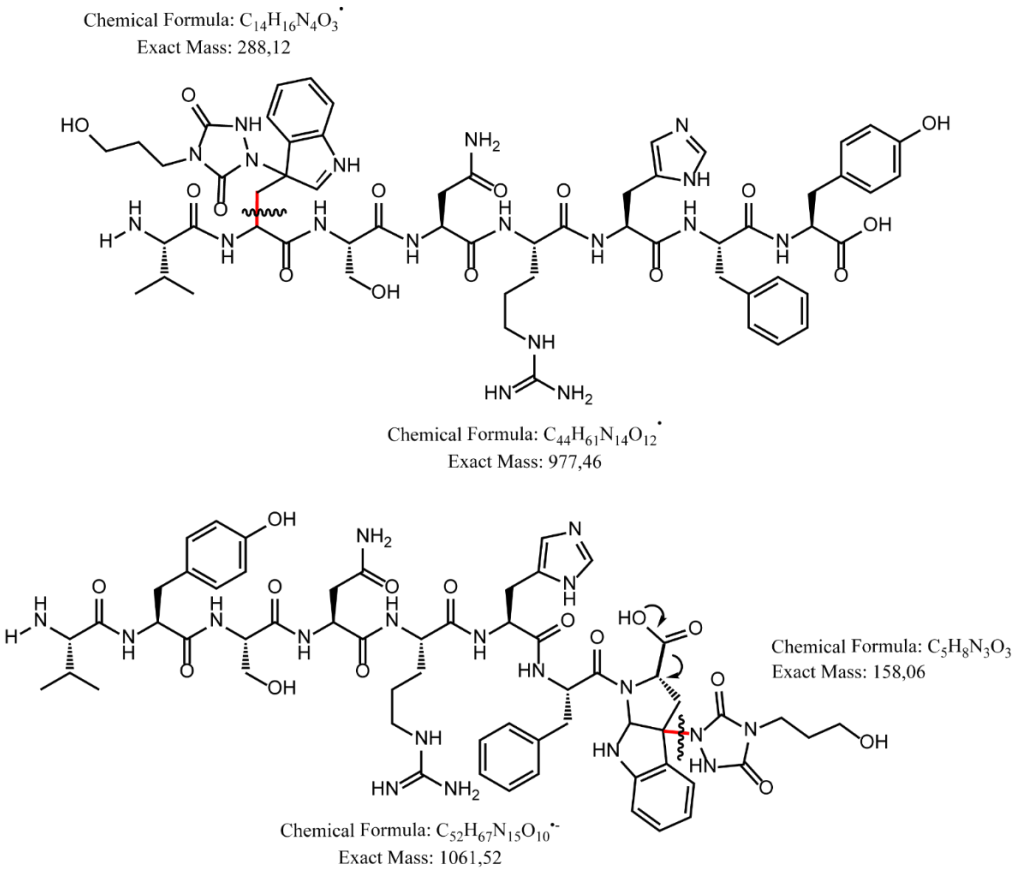


Figure 9.140 Structural representation of VWSNRHFY, **6.3i** modified with TAD propanol, **2b**. The Cα-Cβ bond of the tryptophan residue is cleaved and the resulting mass for the radicals is depicted (top). The expected mass for the radical is 977.5 Da and this mass can be related to a peak in the MS/MS spectrum $[M+H]^{+\cdot}$ of 978.5 Da (peak not picked up by the mMass software). This peak appears in the MS/MS spectra of **6.3i** conjugated with **2b** or **2c** especially when the modification was performed at lower pH (higher Trp modification). Structural representation of VYSNRHFW, **6.3j** modified with TAD propanol, **2b** (bottom). The bond between C3 of the indole and TAD-nitrogen is cleaved as well as a decarboxylation. The expected mass for the fragment ion $[M]^{-\cdot}$ is 1061.5 Da and can be related to a peak in the MS/MS spectrum $[M+2H]^{+\cdot}$ of 1063.5 Da (peak not picked up by the mMass software). This peak appears in the MS/MS spectra of **6.3j** conjugated with **2b** or **2c** especially when the modification was performed at lower pH (higher Trp modification).

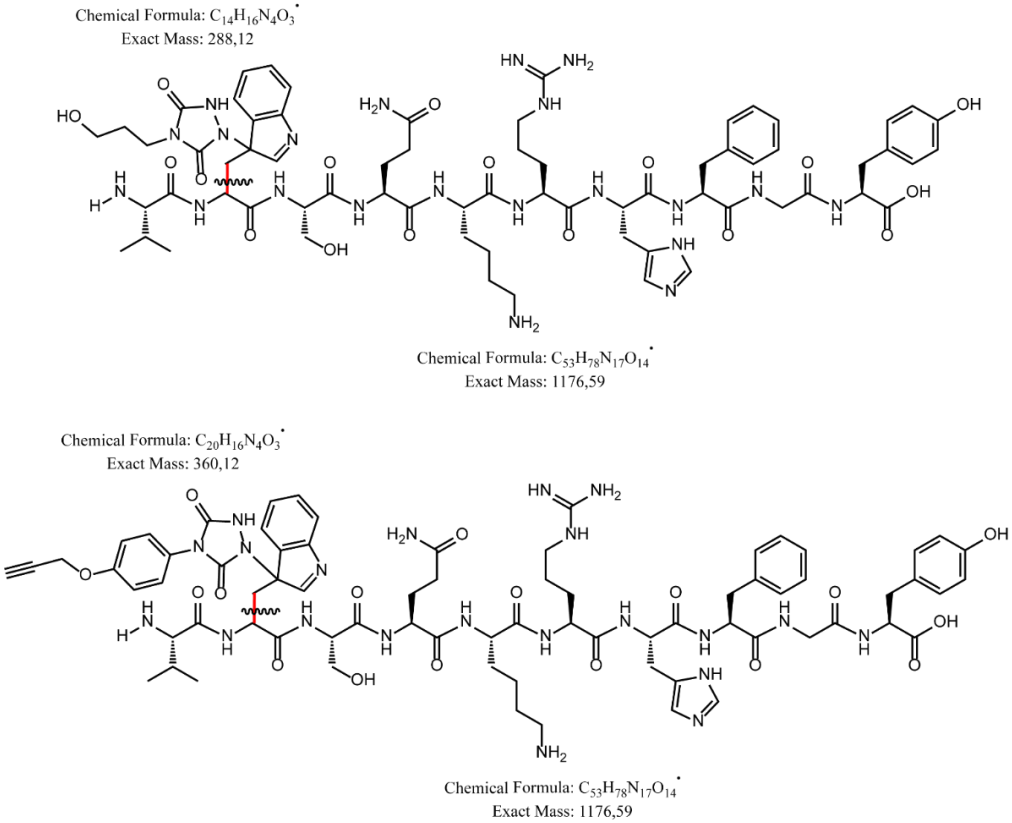


Figure 9.141 Structure representation of VWSQKRHFGY, **6.3k** with TAD-propanol, **2b**. The Cα-Cβ bond of the tryptophan residue is cleaved (top). Structure representation of VWSQKRHFGY, **6.3k** with PTAD-alkyne, **2c**. The Cα-Cβ bond of the tryptophan residue is cleaved (bottom).The expected mass for the radical is 1176.6 Da and this mass can be related to a peak in the MS/MS spectrum of 1177.7 Da (peak not picked up by the mMass software). This peak appears in the MS/MS spectra of **6.3k** conjugated with **2b** or **2c** when the modification was performed at lower pH (Trp modification).

9.3.3.6 LC (ESI) MS Fusion Lumos

VWSQKRHFGY peptide 1k with TAD-propanol 2b.

The peptide modification reaction was performed according to the general protocol but with 3 equivalents of TAD-propanol. The mixture was analysed on Fusion Lumos as described above.

The graphs below show the intensity of the double, triple and quadruple charged single TAD propanol modified peptide ion as a function of the retention time (Skyline software).

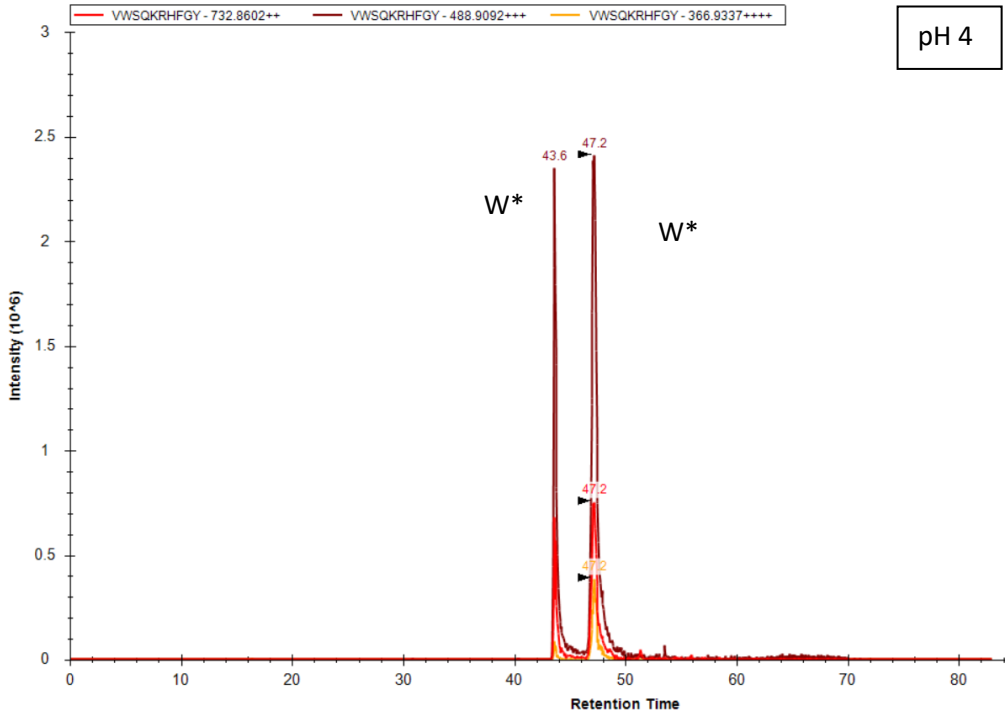


Figure 9.142 Ion chromatogram showing the double, triple and quadruple charged single TAD propanol modified peptide ions. The conjugation reaction of peptide **6.3k** with TAD-propanol **2b** was performed in 10 X PBS pH 4. 2 peaks corresponding with the double, triple and quadruple charged modified peptide ions at t_R = 43.6 min and at t_R = 47.2 min are observed. The ETD analysis of the triple charged modified peptide ion (488.91+++) in both peaks correspond to modification on tryptophan (see figure 9.143 and 9.144). These two peaks correspond with both diastereomers of the TAD modified tryptophan product.

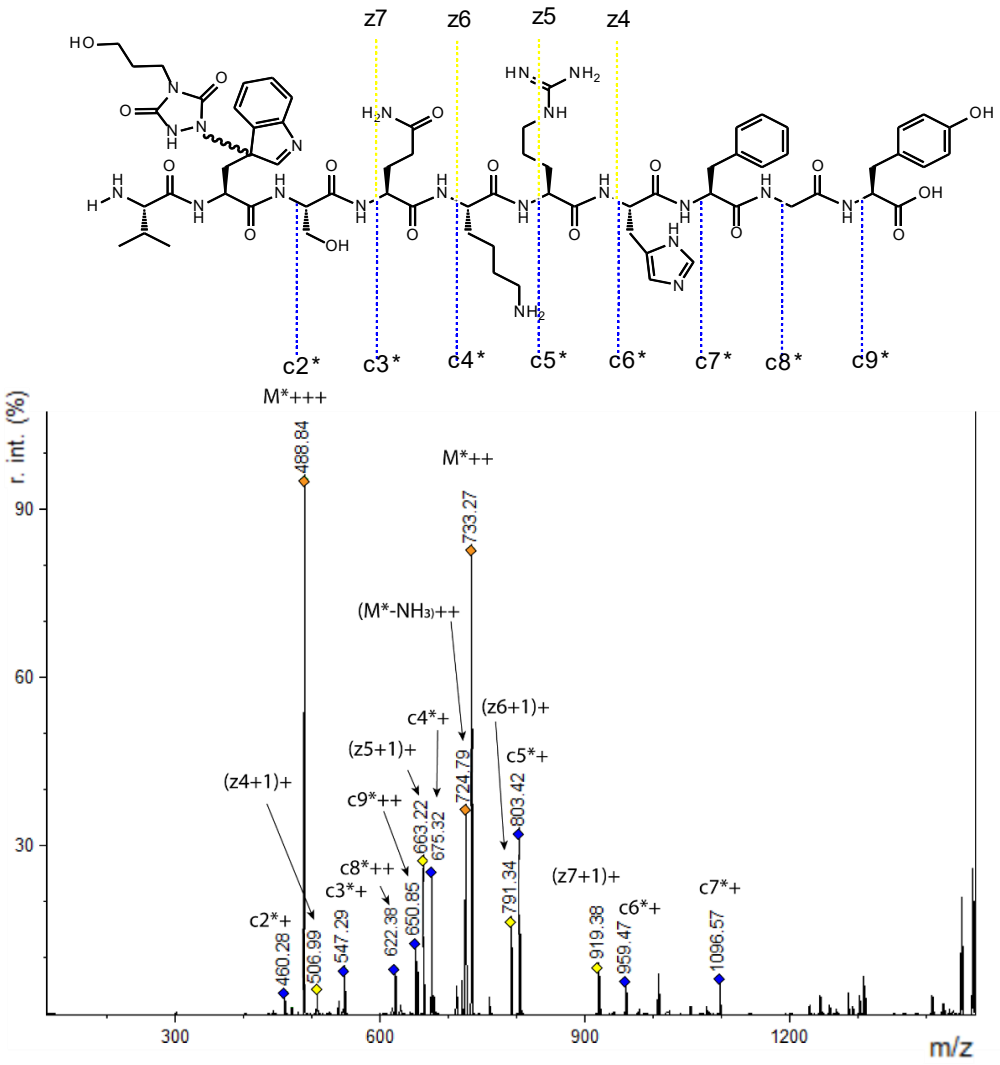


Figure 9.143 ETD MS/MS analysis at t_R = 43.62 min, precursor: TAD propanol modified peptide **6.3k** (488.91+++). Modification was done in 10 X PBS pH 4.

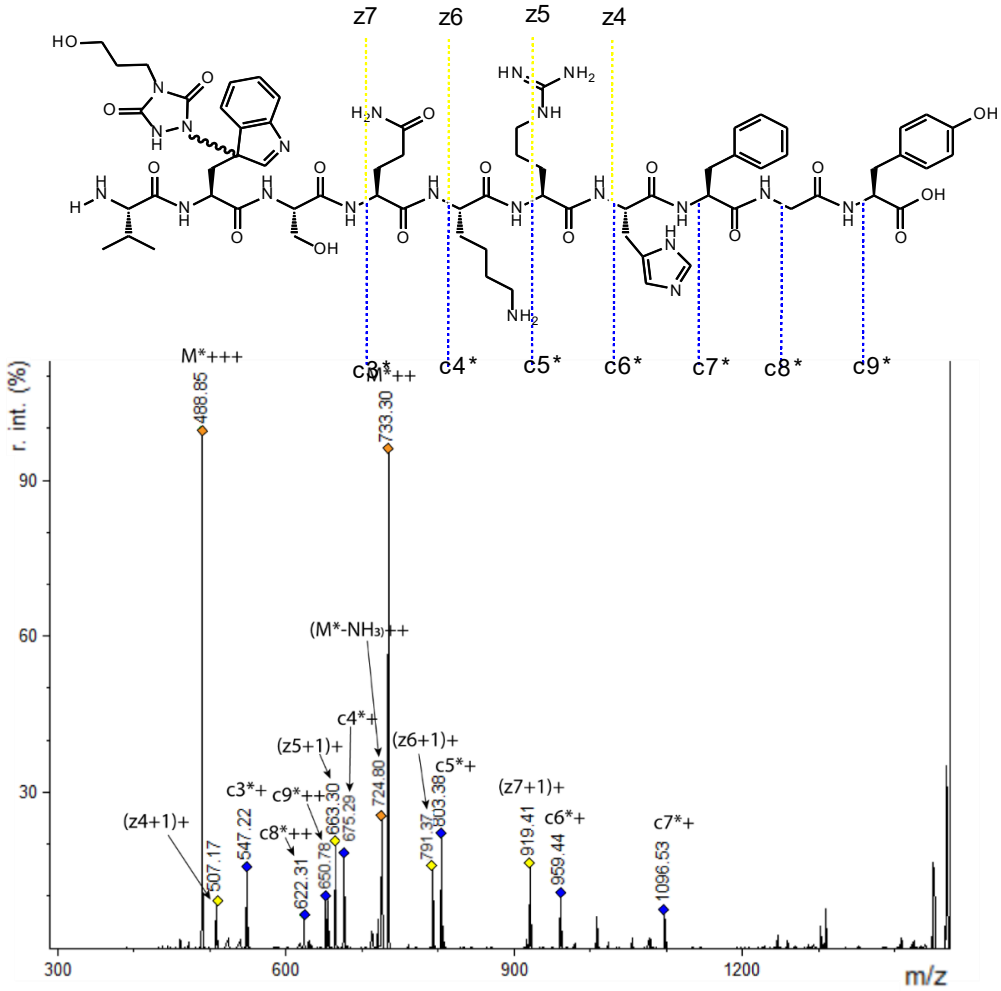


Figure 9.144 ETD MS/MS analysis at t_R = 47.16 min, precursor: TAD propanol modified peptide **6.3k** (488.91+++). Modification was done in 10 X PBS pH 4.

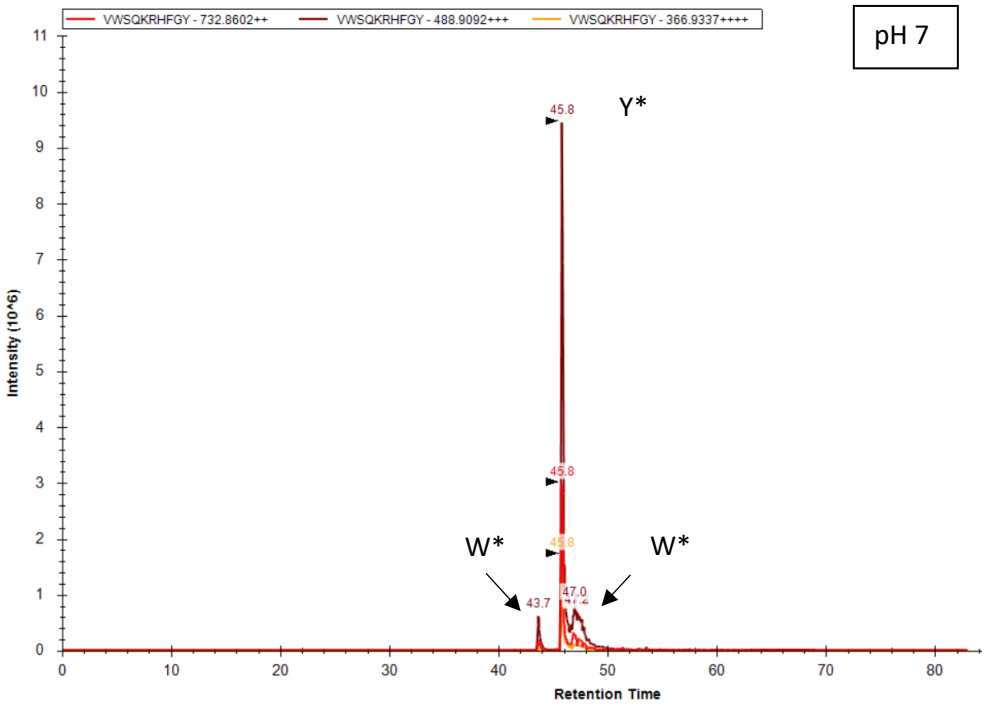


Figure 9.145 Ion chromatogram showing the double, triple and quadruple charged single TAD propanol modified peptide ions. The conjugation reaction of peptide **6.3k** with TAD-propanol **2b** was performed in 10 X PBS pH 7. The main peak is observed at t_R = 45.8 min and alongside this main peak 2 smaller peaks are observed at t_R = 43.7 min and 47.0 min. The ETD analysis of the triple charged modified peptide ion (488.91+++) in the main peak demonstrates tyrosine modification (see figure 9.146 below), however the smaller peaks t_R = 43.7 min and 47.0 min indicate modification on tryptophan.

The absence of a peak at t_R = 45.8 min in figure 9.142 (modification at pH 4) together with the ETD analysis pinpointing the TAD moiety to tryptophan for the peptides eluting at RT 43.62 and 47.16, confirms the chemo selectivity towards tryptophan at pH 4.

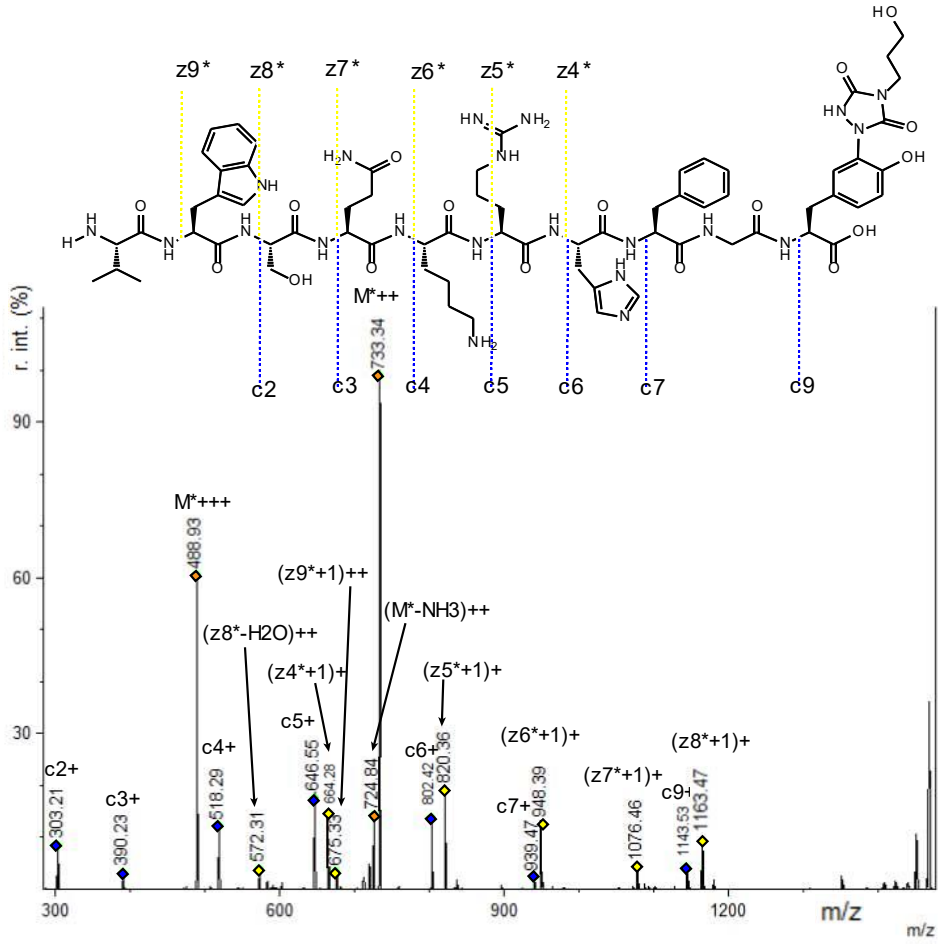


Figure 9.146 ETD MS/MS analysis at t_R = 45.82 min, precursor: TAD propanol modified peptide **1k** (488.91+++). Modification was done in 10 X PBS pH 7.

VWSNRHFY peptide 6.3i and VYSNRHFW peptide 6.3j with TAD-propanol 2b

The peptide modification reaction was performed according to the general protocol but with 3 equivalents of TAD-propanol. The mixture was analyzed on Fusion Lumos as described above.

The graphs below show the intensity of the double and triple charged single TAD propanol modified peptide ion as a function of the retention time (Skyline software).

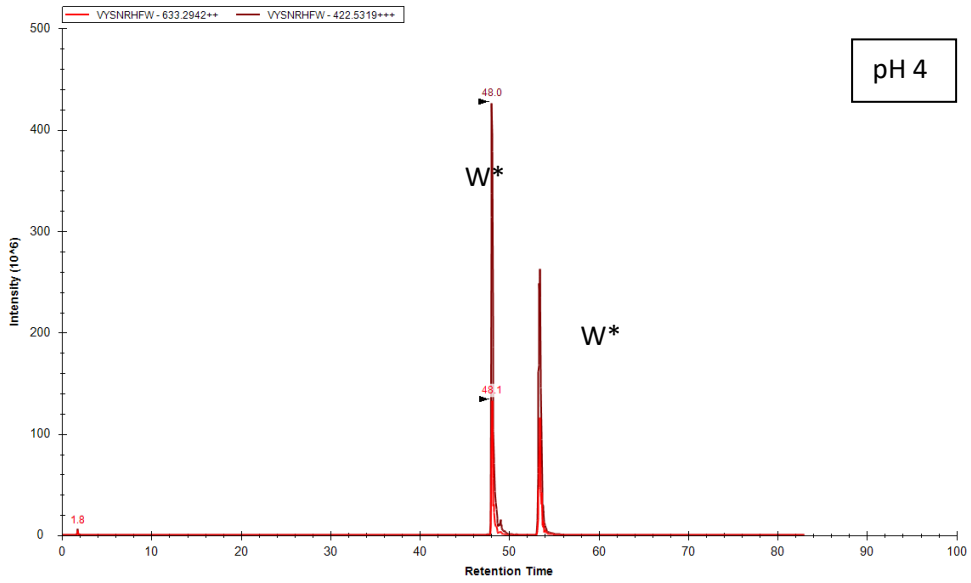


Figure 9.147 Ion chromatogram showing the double and triple charged single TAD propanol modified peptide ions. The conjugation reaction of peptide **6.3j** with TAD-propanol **2b** was performed in 10 X PBS pH 4. 2 peaks corresponding with the double and triple charged modified peptide ions at t_R = 48.0 min and at t_R = 53.5 min are observed. These two peaks correspond with both diastereomers of the TAD modified tryptophan product.

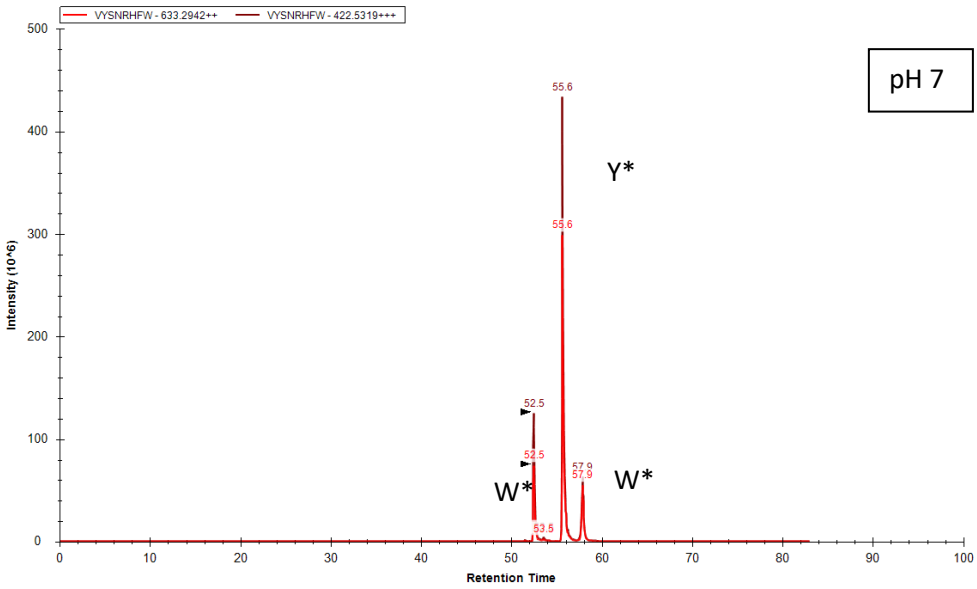


Figure 9.148 Intensity of the double and triple charged single TAD propanol modified peptide ion. The conjugation reaction of peptide **6.3j** with TAD-propanol **2b** was performed in 10 X PBS pH 7. 3 peaks are observed, 2 peaks correspond with the diastereomers of the TAD modified tryptophan product t_R = 52.5 min and at t_R = 57.9 min. the most intense peak corresponds with the TAD propanol modification located on tyrosine t_R = 55.6 min. Note that the retention times are slightly shifted due to the fact that the samples of pH 4 and pH 7 were measured on different columns.

The histogram below shows the peak area for the **doubly TAD propanol modified peptide ions** for peptide **6.3i** and peptide **6.3j**. The double (711.82++) and triple (474.88+++) charged ions of these peptides with 2 TAD-propanol moieties are the same for peptide **6.3i** and **6.3j** (only the relative position of Y and W is different for **6.3i** and **6.3j**). The peptide modification reaction was performed in either 10 X PBS pH 4 or 10 X PBS pH 7, according to the general protocol but with 3 equivalents of TAD-propanol.

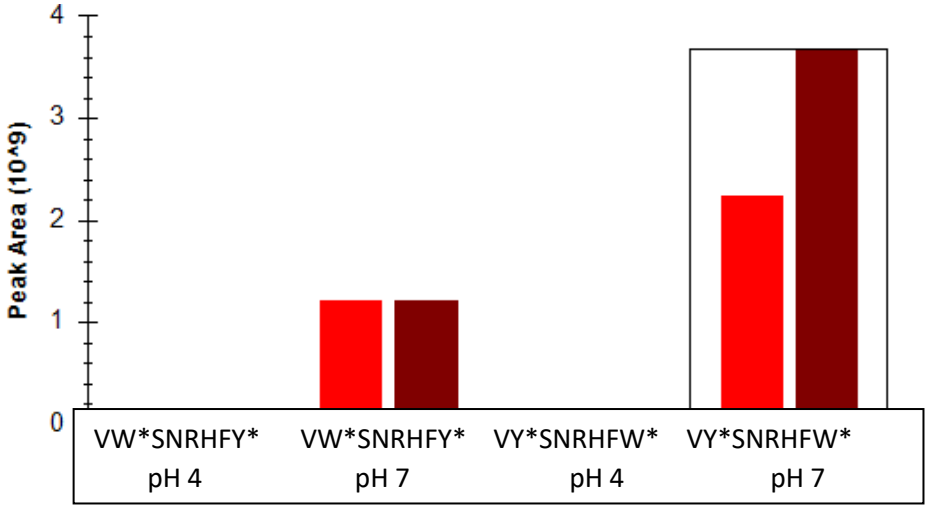


Figure 9.149 Histogram showing peak area for the double (711.82++, red) and triple (474.88+++, dark red) charged ions of these peptides with 2 TAD-propanol moieties.

To obtain a double TAD modified peptide both the tryptophan and the tyrosine in the peptide have to react with TAD. The fact that double TAD-propanol modified peptides are nearly absent at pH 4, and their presence at pH 7 further demonstrates the chemo selectivity for tryptophan at pH 4.

9.3.3.7 LC (ESI) ETD MS/MS

Previous experiments have shown that the TAD-tryptophan modification is not stable in MALDI-TOF/TOF experiments. Electron transfer dissociation (ETD) is a soft dissociation method that could preserve the TAD-tryptophan modification upon fragmentation. We have analyzed conjugation reactions with peptides **6.3i** and **6.3j** and 3 equivalents of TAD-propanol (**2b**) in 10 X PBS buffer at pH 4 and pH 7 on a Fusion™ Lumos™ mass spectrometer equipped with an ETD module. In ETD peptide fragmentation c and z ions are observed, the settings in the Mmass software were adapted to predict c and z ions as well as the molecular ions. Note that in ETD typically (z+1)+ ions are detected instead of the z+ ions because of hydrogen migragtion[338]. The triple charged molecular ion was selected as precursor ion for all ETD experiments. The results below clearly show that the combination of electrospray ionization (ESI) and ETD fragmentation is ideally suited to preserve and identify the TAD-tryptophan modification.

Conjugation product of Val-Trp-Ser-Asn-Arg-His-Phe-Tyr-OH (VWSNRHFY, 6.3i) with TAD-propanol 2b (6.3i2b).

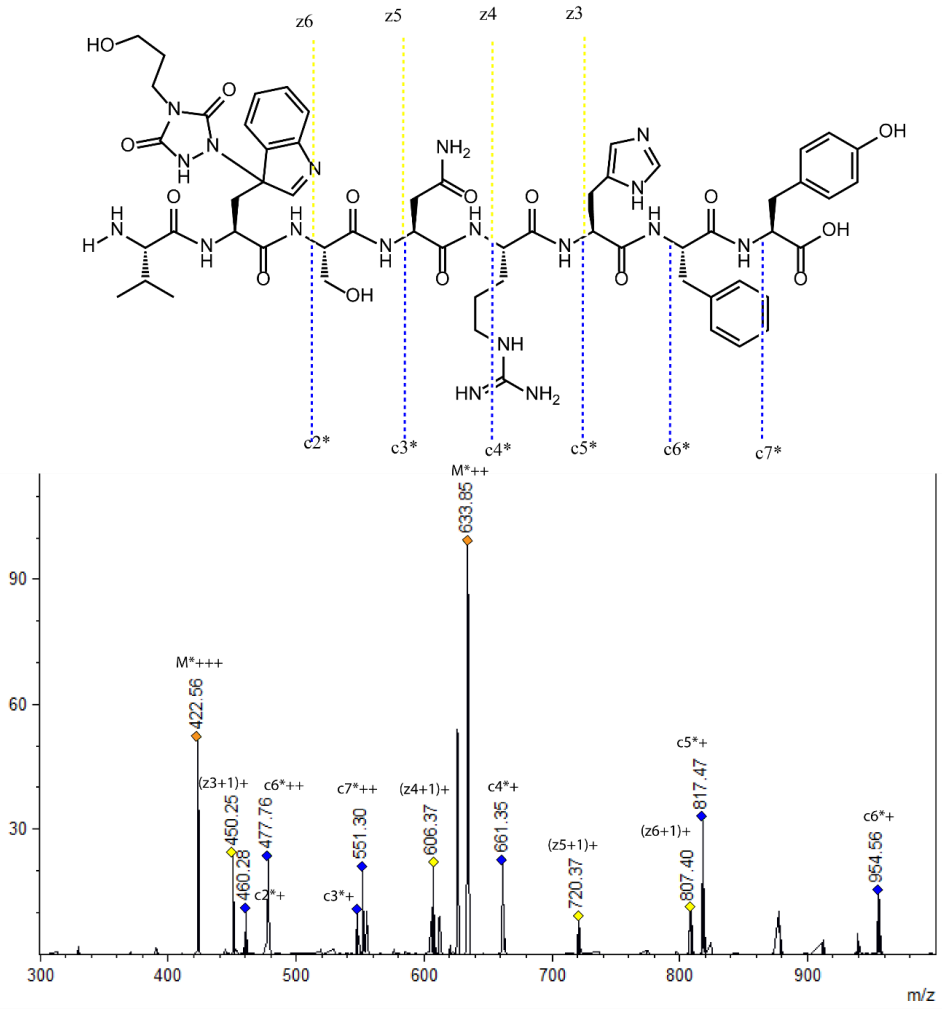


Figure 9.150 ETD MS/MS spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 4.

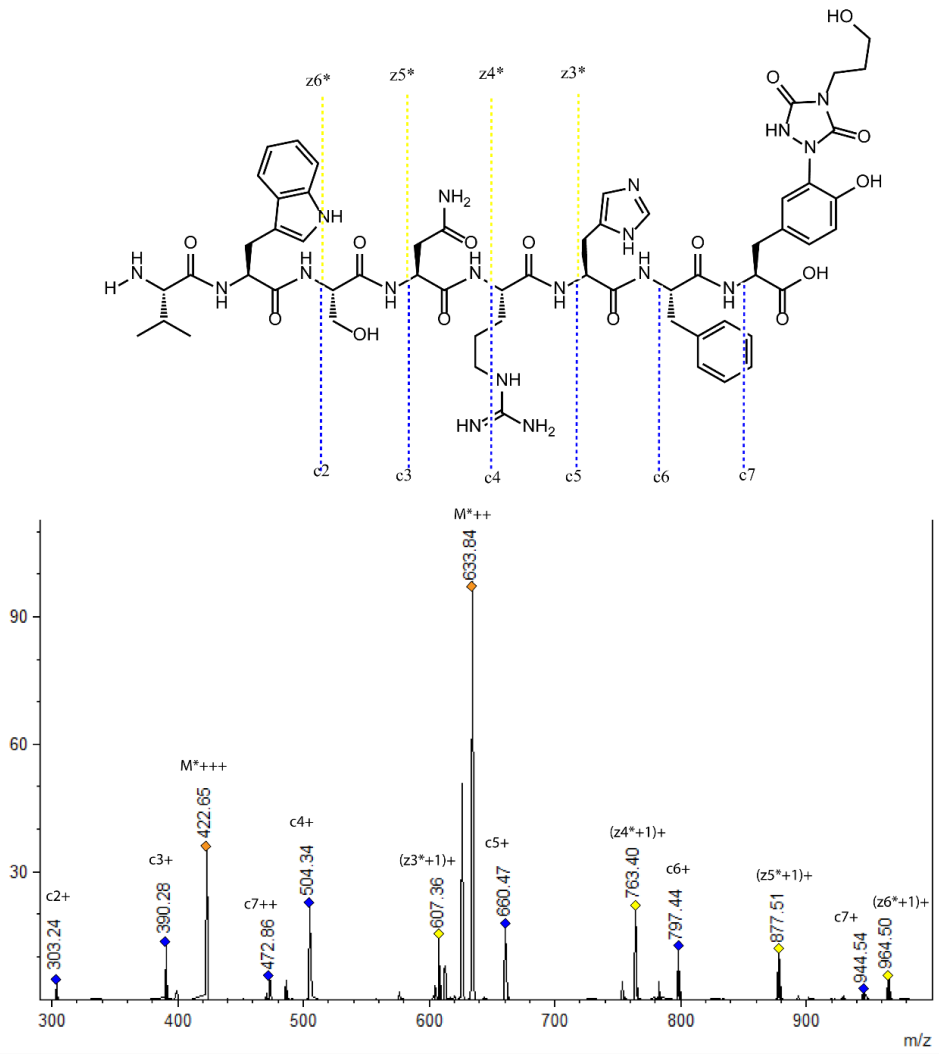


Figure 9.151 ETD MS/MS spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 7.

Conjugation product of Val-Tyr-Ser-Asn-Arg-His-Phe-Trp-OH (VYSNRHFW, 6.3j) with TAD-propanol 2b (6.3j2b).

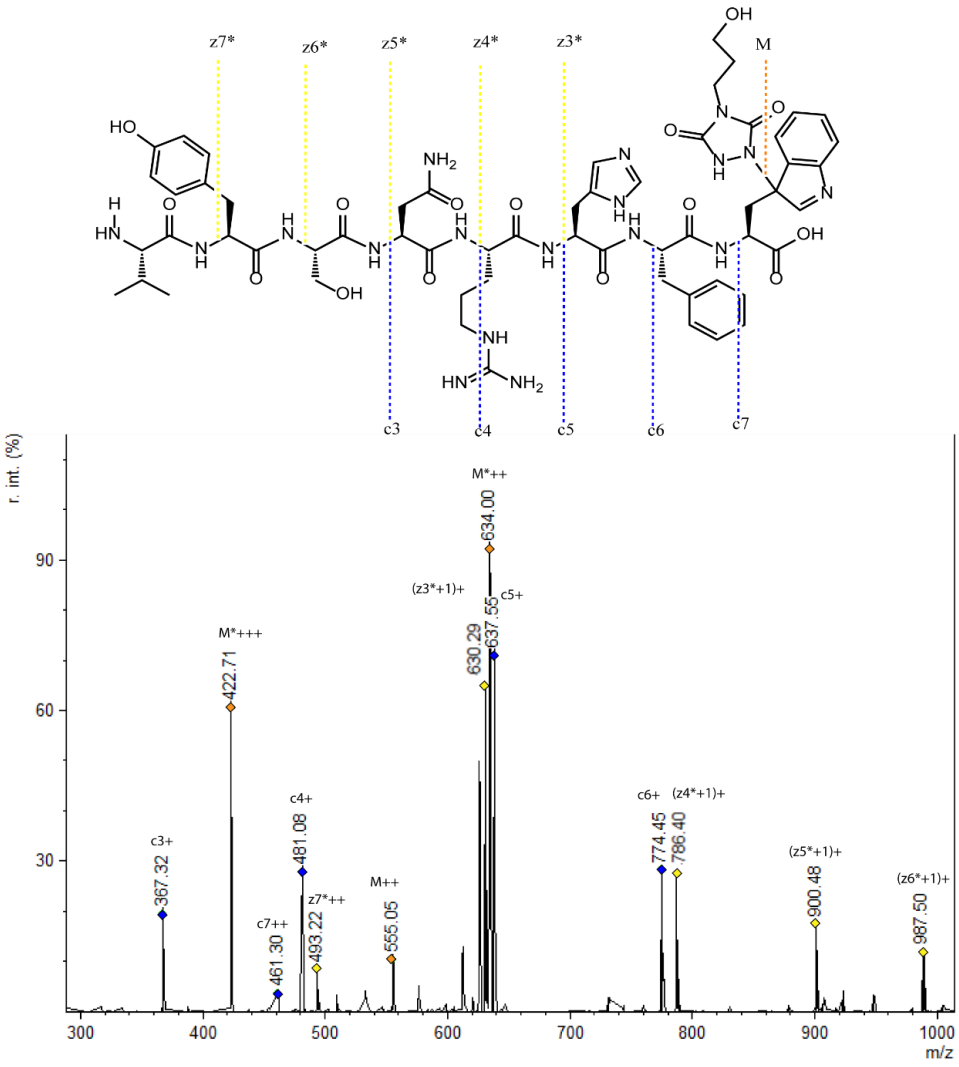


Figure 9.152 ETD MS/MS spectrum of peptide **6.3j** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 4.

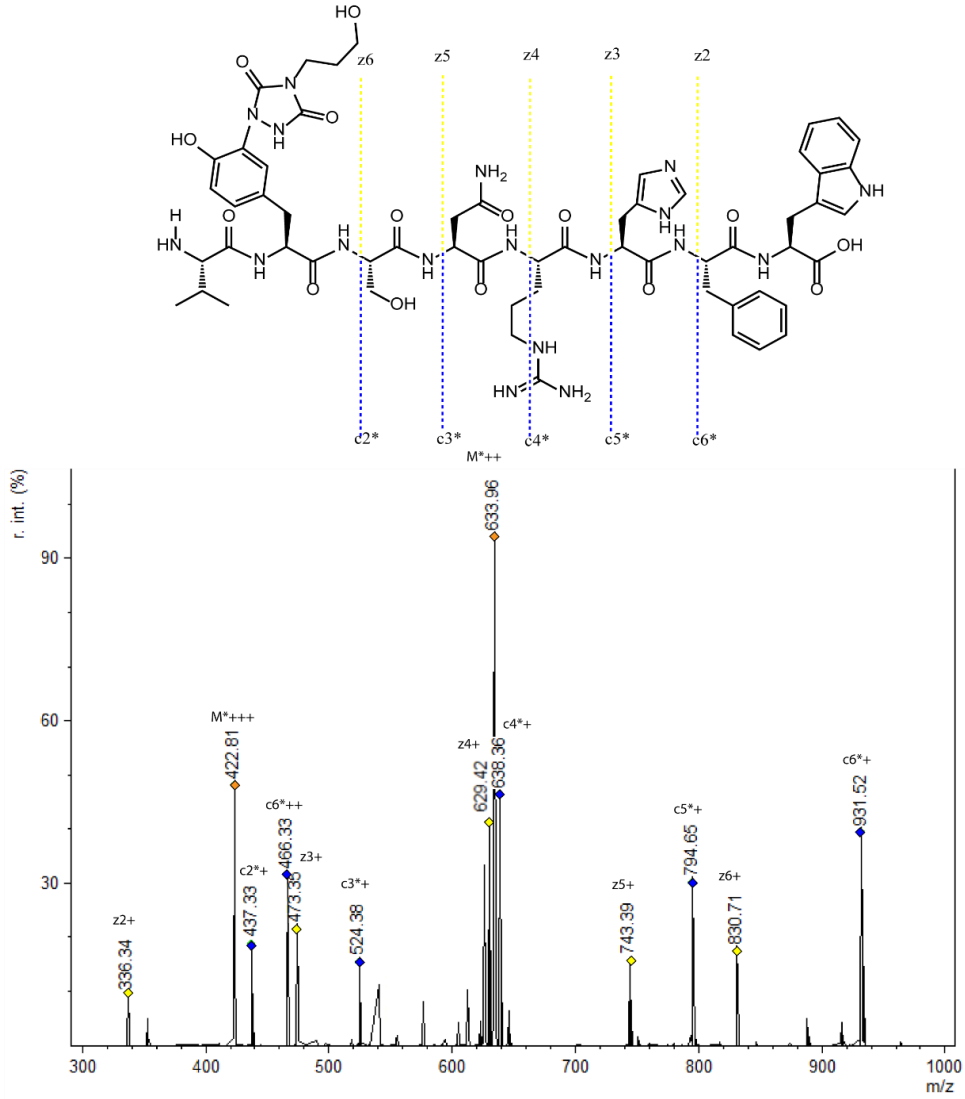


Figure 9.153 ETD MS/MS spectrum of peptide **6.3j** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 7.

9.3.3.8 LC (ESI) CID MS/MS

Since CID fragmentation is harsher compared to ETD, modifications are easily lost upon fragmentation. Furthermore, since the fragmentation energy applied is unique for the precursor ion including the modification, no further fragmentation is taking place on the precursor without the modification. Hence a low number of fragmentation ions is detected and identification cannot be established.

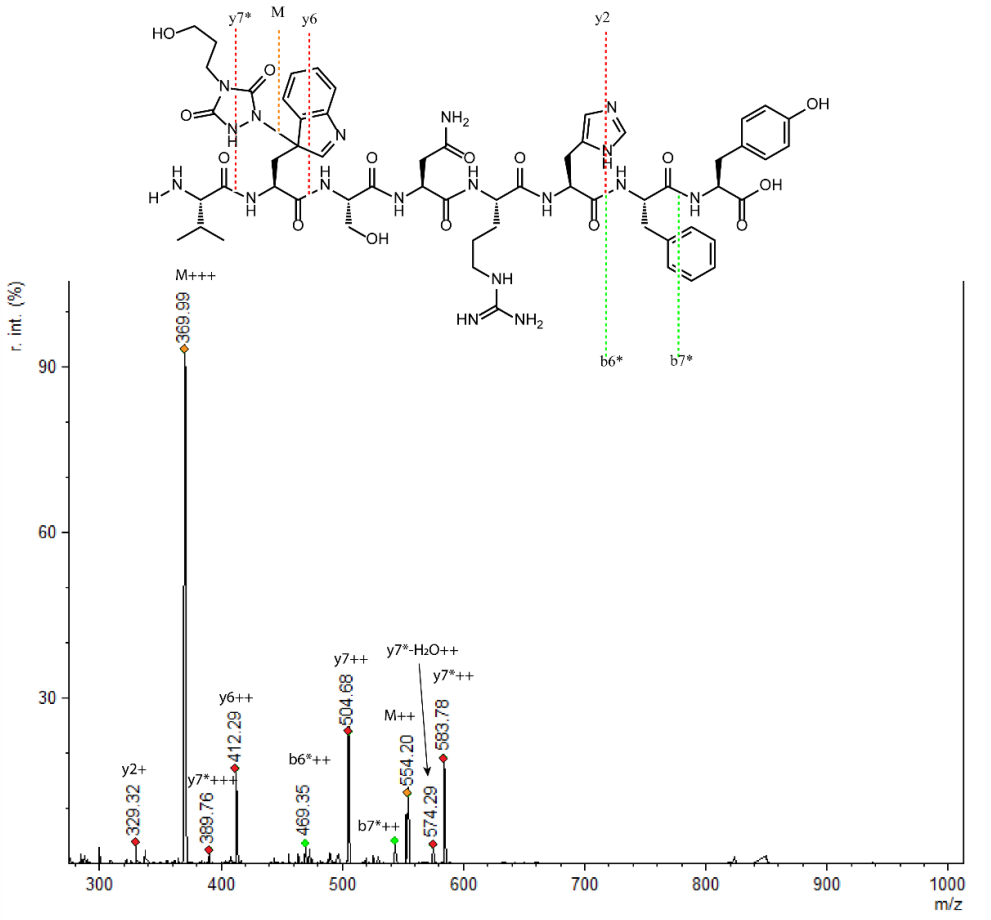


Figure 9.154 CID MS/MS spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 4.

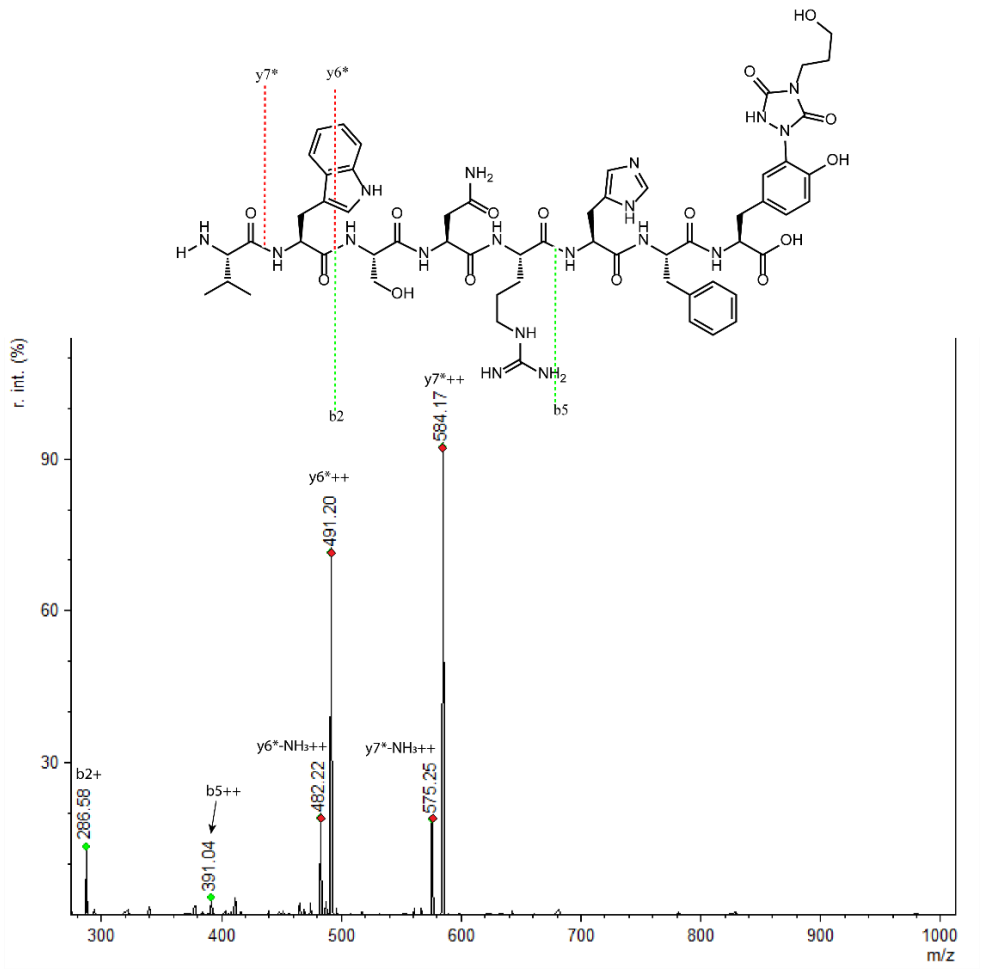


Figure 9.155 CID MS/MS spectrum of peptide **6.3i** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3i** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 7.

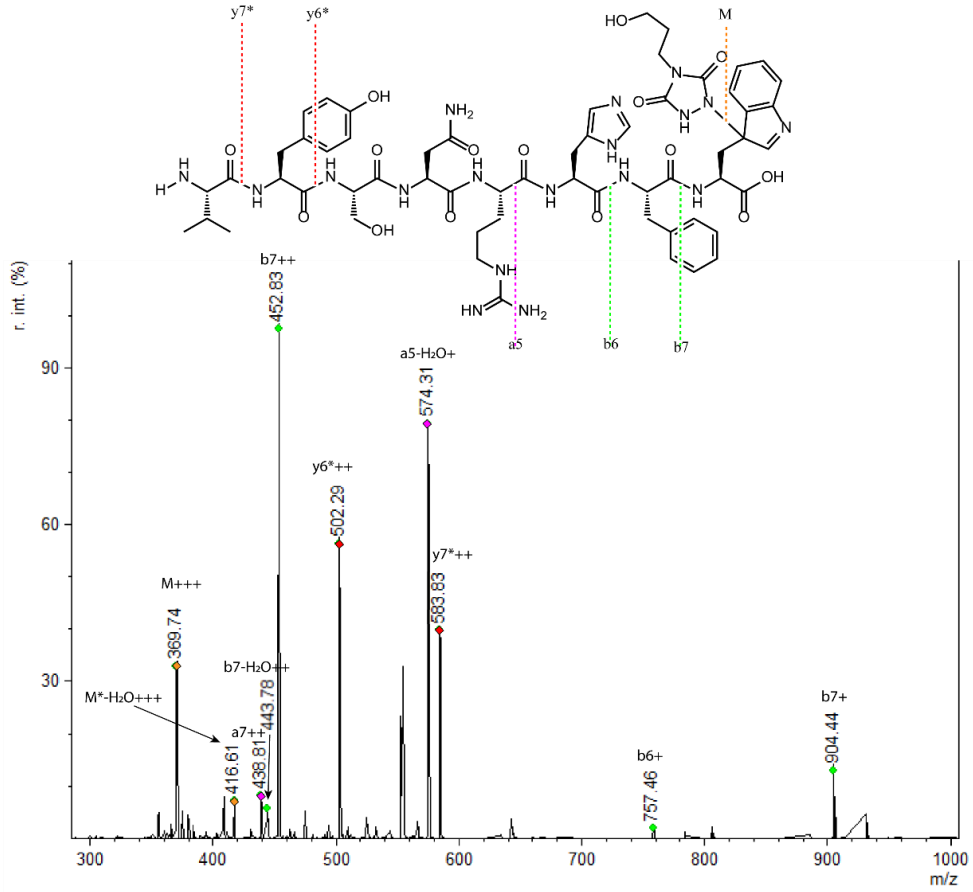


Figure 9.156 CID MS/MS spectrum of peptide **6.3j** conjugated with **2b** (precursor ion: 422.53 +++). Chemical structure of peptide **6.3j** with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 4.

9.3.3.9 LC (ESI) HCD MS/MS

Since HCD fragmentation is using nitrogen as collision gas to enable fragmentation, a modification is easily lost upon fragmentation, but unlike CID fragmentation, the fragmentation of the precursor that lost the modification is still taking place. This results in a fragmentation spectrum containing fragment ions without the modification, although the precursor mass shows the presence of the modification.

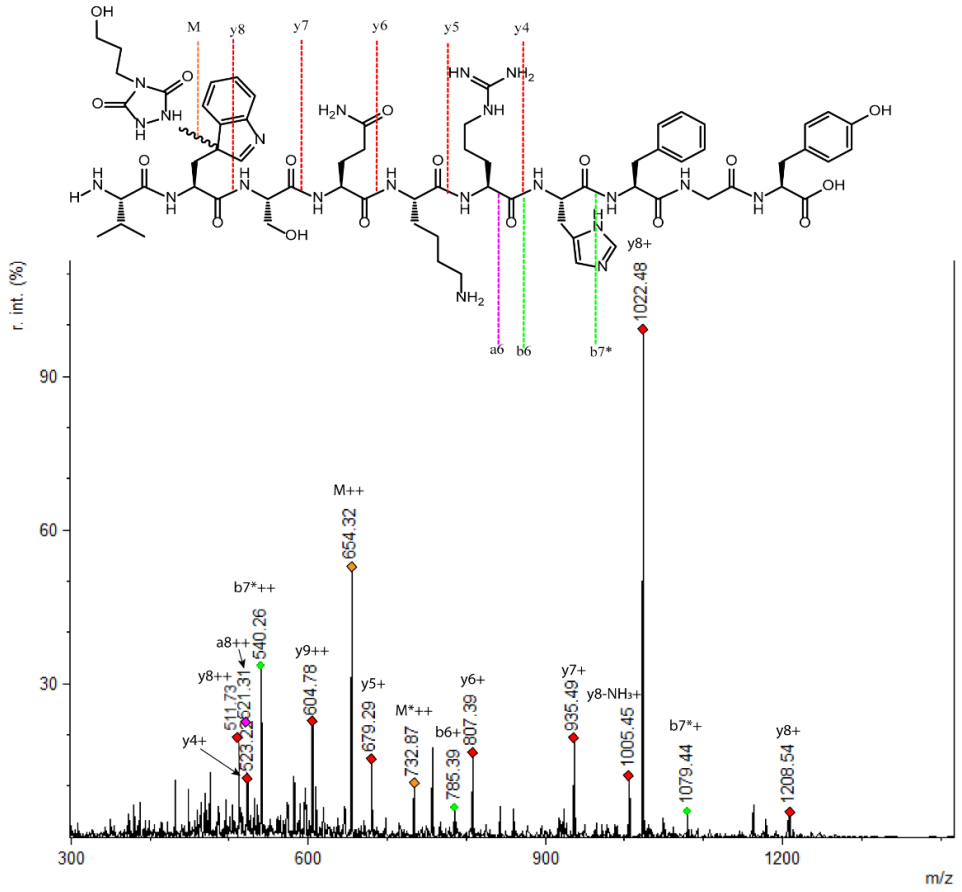


Figure 9.157 HCD MS /MS spectra of peptide **6.3k** conjugated with TAD-propanol **2b** (precursor ion: 732.86 ++). Chemical structure with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 4.

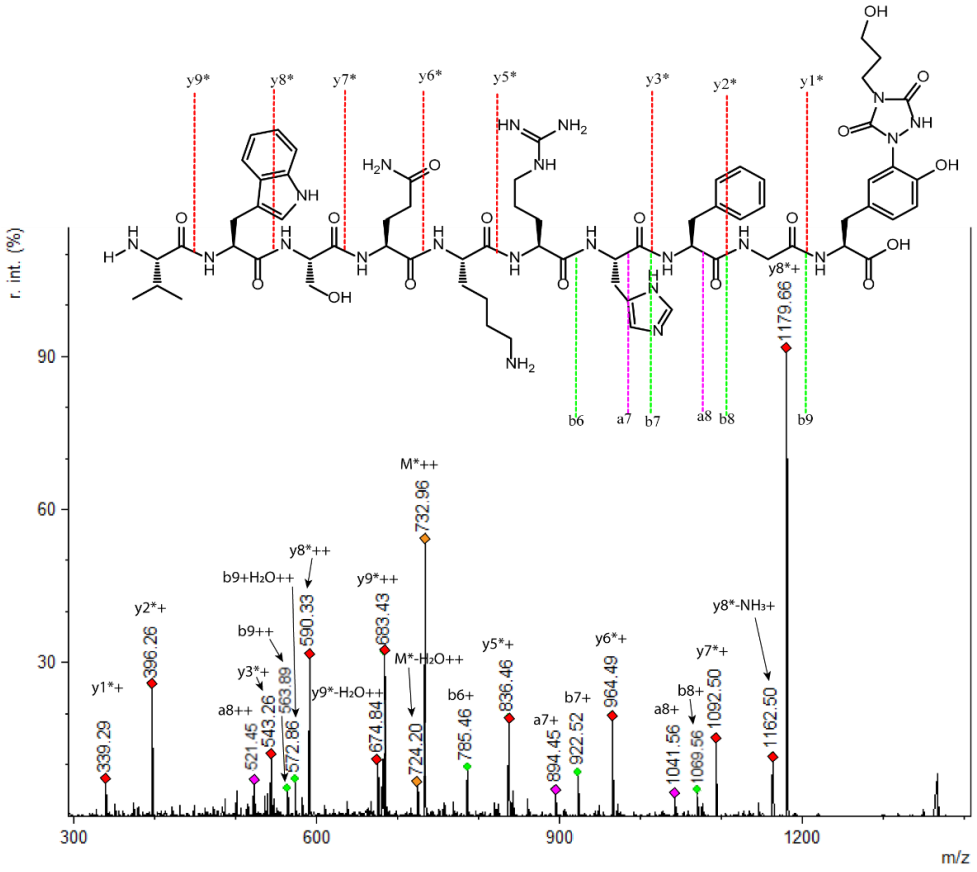


Figure 9.158 HCD MS /MS spectra of peptide **6.3k** conjugated with TAD-propanol **2b** (precursor ion: 732.86 ++). Chemical structure with detected fragment ions is shown. Fragments with "*" are TAD modified fragments. Conjugation reaction was performed in 10 X PBS pH 7.

9.3.3.10 Protein conjugation

Valentine alphabody

The valentine alphabody is a recombinantly produced protein of 11.5 kDa. The alphabody protein has a triple helix coiled coil structure.[228] In the primary sequence of this alphabody there are three tyrosine residues and zero tryptophan residues. In view of the presence of a solvent exposed cysteine residue, the alphabody occurs as a dimer under non-reducing conditions. This implies that in the alphabody dimer (connected via a disulfide linkage) there are 6 tyrosine residues. The experiments with the valentine alphabody were measured on Agilent 1100 series LC (ESI) MS equipment. LC-MS analysis of the valentine alphabody dimer by MS (ESI): m/z 22937 (calcd $[M+H]^+$ 22941).

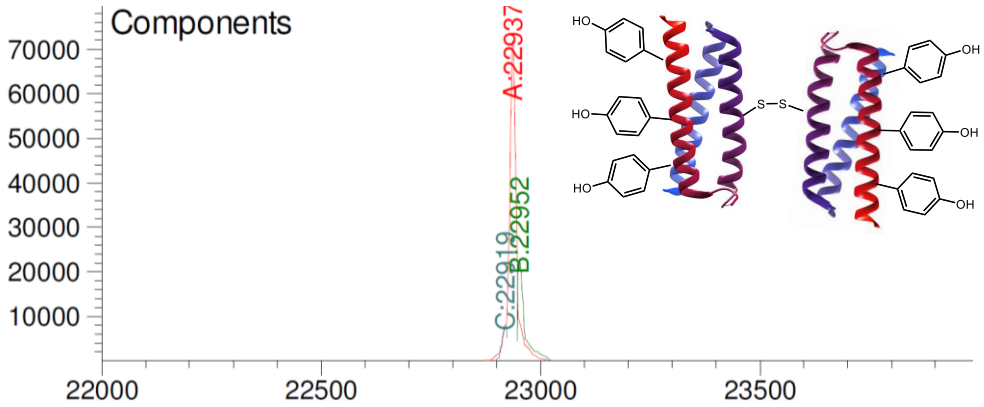


Figure 9.159 Deconvoluted mass spectrum from the LC-MS analysis of the valentine alphabody dimer. Cartoon representation of the alphabody dimer is shown, highlighting the bridging sulfur bond and the presence of the 6 tyrosine residues on the dimer. In addition to the peak corresponding to the alphabody dimer (22937 Da), also an oxidation product (22952 Da) and dehydration product (22919 Da) is observed.

A competition experiment was performed using a mixture of this alphabody dimer and the small tetrapeptide NSAW (**6.3c**). 10 µL of alphabody stock solution (4.6 mg/mL) in 10 X PBS buffer at pH 4 was added to 7 µL of 10 X PBS buffer at pH 4 and 2 µL of peptide **6.3c** 3 mM in water (3 eq of peptide compared to protein). Note that the ratio of tyrosine to tryptophan is 2 to 1. Subsequently, 3 eq. of TAD propanol **2b** were added. The reaction mixture was analysed with LC-MS (ESI).

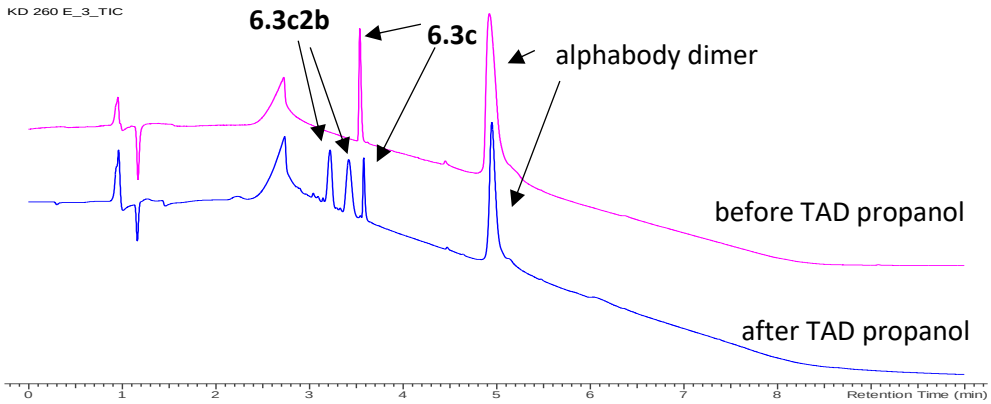


Figure 9.160 LC-chromatogram (214 nm) overlay of the competition experiment between TAD-propanol **2b** and a mixture of alphabody dimer and peptide **6.3c** in 10 X PBS pH 4. Starting mixture of **6.3c** and alphabody dimer before addition of TAD propanol (pink), after reaction (blue).

As shown in figure 9.160, peptide **6.3** is largely consumed and 2 new peaks appear as expected for a TAD-adduct on a C-terminal tryptophan (*vide supra*). Peak integration of the chromatogram demonstrated a conversion of over 80% for peptide **6.3c**. Deconvolution of the protein peak at t_R 5 min is required to examine the effect on the alphabody dimer (Figure S3.2.3). Analysis of this result revealed no formation of any alphabody conjugate at pH 4 in accord with the absence of a Trp residue in this alphabody.

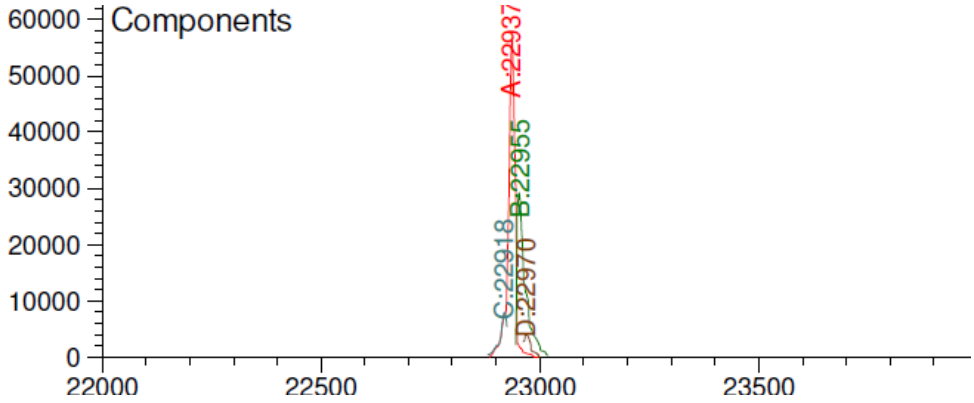


Figure 9.161 Deconvoluted mass spectrum of the peak between t_R 4.869 min and 5.085 min from the LC chromatogram in Figure 9.160. Only starting alphabody dimer is observed with a dehydration product (22918 Da) and a product with additional water (22955 Da) as well as a small amount of an product with an additional water + oxidation (22970 Da).

A competition experiment was performed using a mixture of this alphabody dimer and the small tetrapeptide NSAW (**6.3c**). 10 µL of alphabody stock solution (4.6 mg/mL) in 10 X PBS buffer at pH 7 was added to 7 µL of 10 X PBS buffer at pH 7 and 2 µL of peptide **6.3c** 3 mM in water (3 eq of peptide compared to protein). Note that the ratio of tyrosine to tryptophan is 2 to 1. Subsequently, 3 eq. of TAD propanol **2b** were added. The reaction mixture was analysed with LC-MS (ESI) (Figure 9.162).

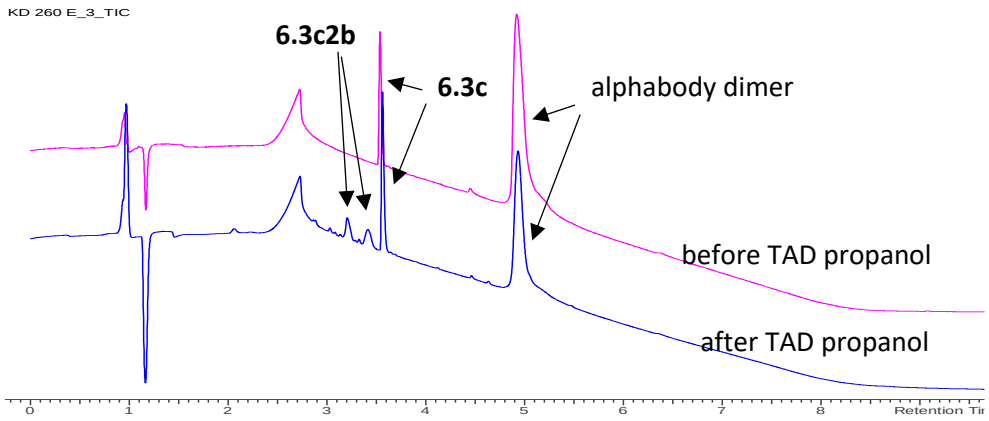


Figure 9.162 LC-chromatogram (214 nm) overlay of the competition experiment between TAD-propanol **2b** and a mixture of alphabody dimer and peptide **6.3c** in 10 X PBS pH 7. Starting mixture of **6.3c** and alphabody dimer before addition of TAD propanol (pink), after reaction (blue).

As shown in figure 9.162, peptide **6.3c** is partly consumed and 2 new peaks appear as expected for a TAD-adduct on a C-terminal tryptophan (*vide supra*). Peak integration of the chromatogram demonstrated a conversion of 30 % for peptide **6.3c**. Deconvolution of the protein peak at t_R 5 min is required to examine the effect on the alphabody dimer (Figure 9.163). Analysis of this result revealed formation of just under 30 % of alphabody TAD propanol conjugate at pH 7.

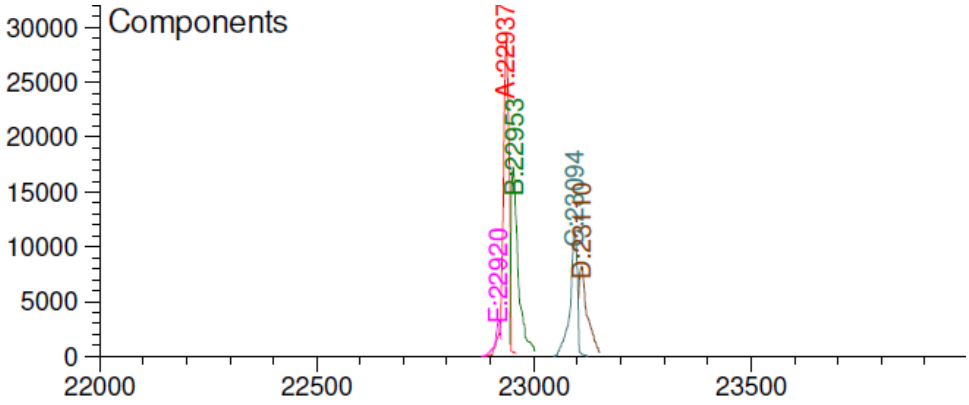


Figure 9.163 Deconvoluted mass spectrum of the peak between t_R 4.855 min and 5.234 min from the LC chromatogram in Figure S3.2.4. The expected mass for modification of the alphabody dimer with TAD-propanol **2b** is 23094 Da (found 23094 Da and oxidation product (+16 Da) 23110 Da).

Two control experiments were performed with the alphabody dimer in 10 X PBS buffer at pH 4 (figure 9.164) or pH 7 (figure 9.165) and TAD propanol (**2b**). 10 µL of alphabody stock solution (4.6 mg/mL) in 10 X PBS buffer at pH 4 or 7 was added to 9 µL of 10 X PBS buffer at pH 4 or 7 respectively. Subsequently, 8 eq. of TAD propanol **2b** with respect to each tyrosine (48 eq. for each alphabody dimer) were added. The reaction mixture was analysed with LC-MS (ESI). In the figures below the deconvoluted mass spectra from the alphabody dimer peak are presented. The data demonstrate that at pH 4 no alphabody TAD propanol (**2b**) conjugate is detected, at pH 7 43 % has at least one TAD propanol (**2b**) modification and over 10 % has two TAD propanol (**2b**) modifications.

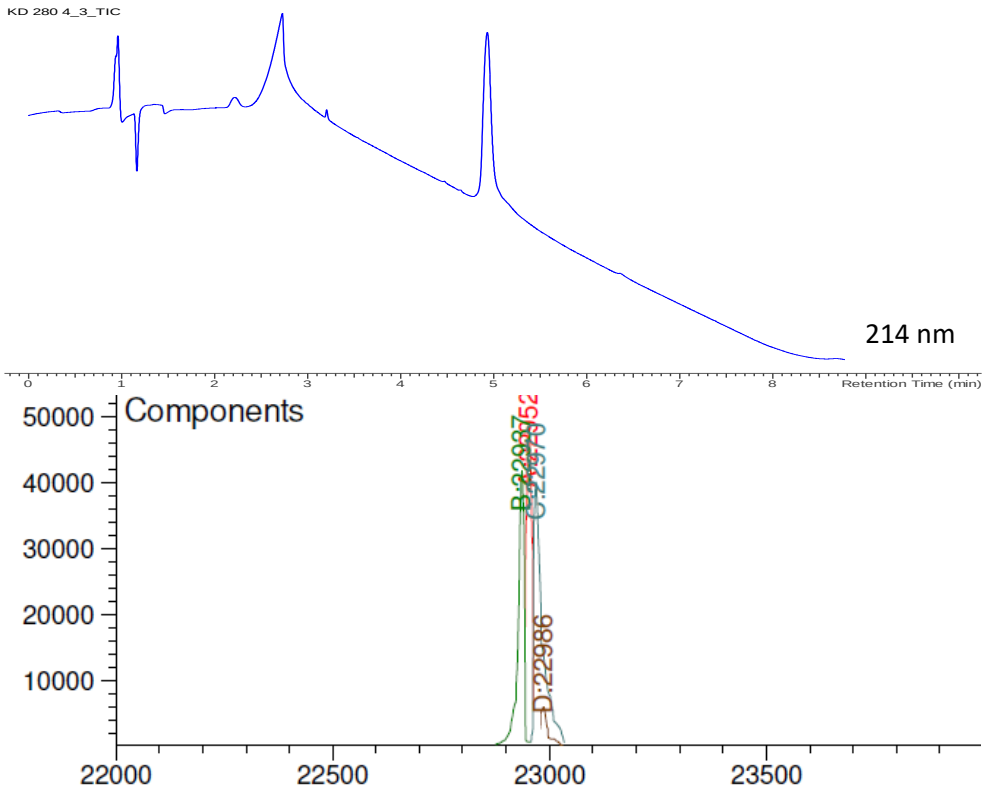


Figure 9.164 LC-chromatogram (214 nm) of the TAD propanol conjugation control experiment with the alphabody dimer in 10 X PBS buffer at pH 4 (top). Deconvoluted mass spectrum of the peak between t_R 4.856 min and 5.064 min (bottom). Only starting alphabody dimer (22937 Da) is observed with an oxidation product (22952 Da) and oxidation product with an additional water (22970 Da) as well as a small amount of an product with an additional water + double oxidation (22986 Da) no alphabody TAD conjugate is observed at pH 4.

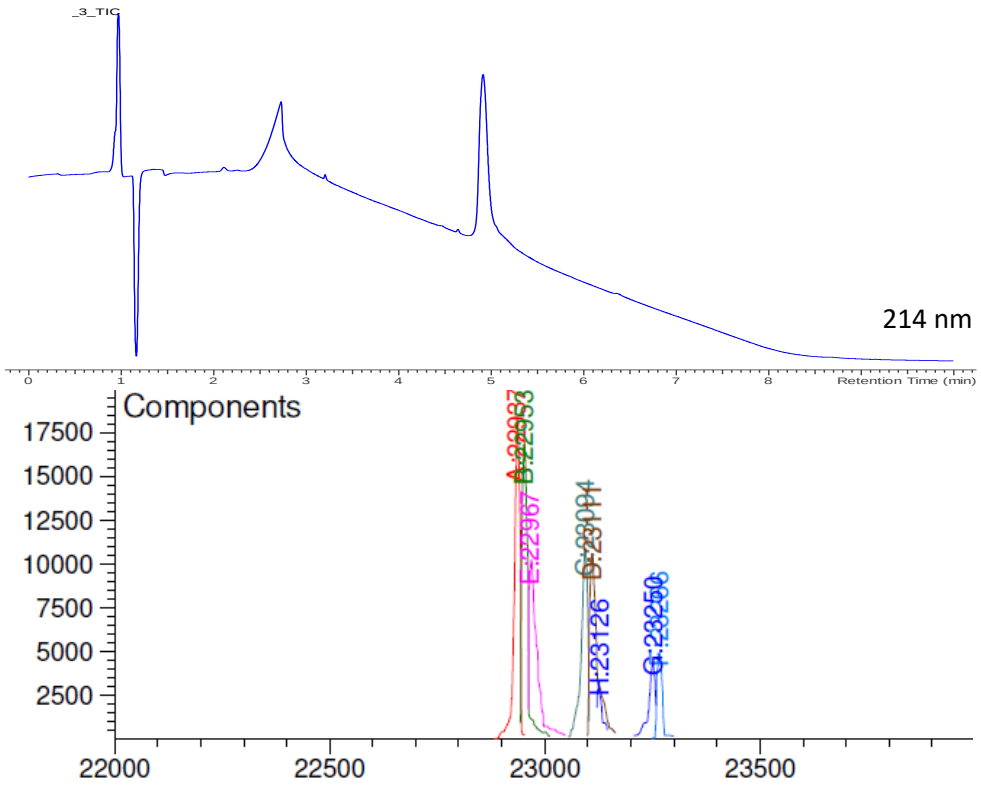


Figure 9.165 LC-chromatogram (214 nm) of the TAD propanol conjugation control experiment with the alphabody dimer in 10 X PBS buffer at pH 7 (top). Deconvoluted mass spectrum of the peak between t_R 4.780 min and 5.348 min (bottom). Starting alphabody dimer (22937 Da) is observed with an oxidation product (22953 Da) and double oxidation product (22967 Da). Additionally an alphabody dimer TAD propanol conjugate is observed (23094 Da) with its single (23111 Da) and double (23126 Da) oxidation products are observed. Lastly also an alphabody dimer with double TAD-modification is observed (23250 Da) and its corresponding oxidation product (23266 Da).

9.3.3.11 Alphabody 586D

Another model protein used in this study is the alphabody 586D, like the valentine alphabody this alphabody has a triple helix coiled coil structure and it was obtained from Complix Therapeutics. 586D contains one tryptophan and three tyrosine residues in the primary amino acid sequence, all of which are solvent accessible (as estimated from 3D models). The exact sequence of 586D cannot be disclosed for confidentiality reasons. The experiments with alphabody 586D were analyzed on a LTQ-orbitrap XL (Thermo fisher scientific) by intact protein analysis. 20 µL of the protein conjugate sample (2 µg protein) was desalted on a 7.5 cm column (made in-house, 1 mm internal diameter (I.D.), 3 µm beads, C4 Reprosil, Dr. Maisch, Germany). Proteins were eluted from the column by a linear 5 min gradient ranging from 0.1% FA, 0.05% TFA in water to 70% acetonitrile. The mass spectrometer was operated in MS mode at a resolution of 30000. MS full-scans were acquired from 600 to 4000 M/Z.

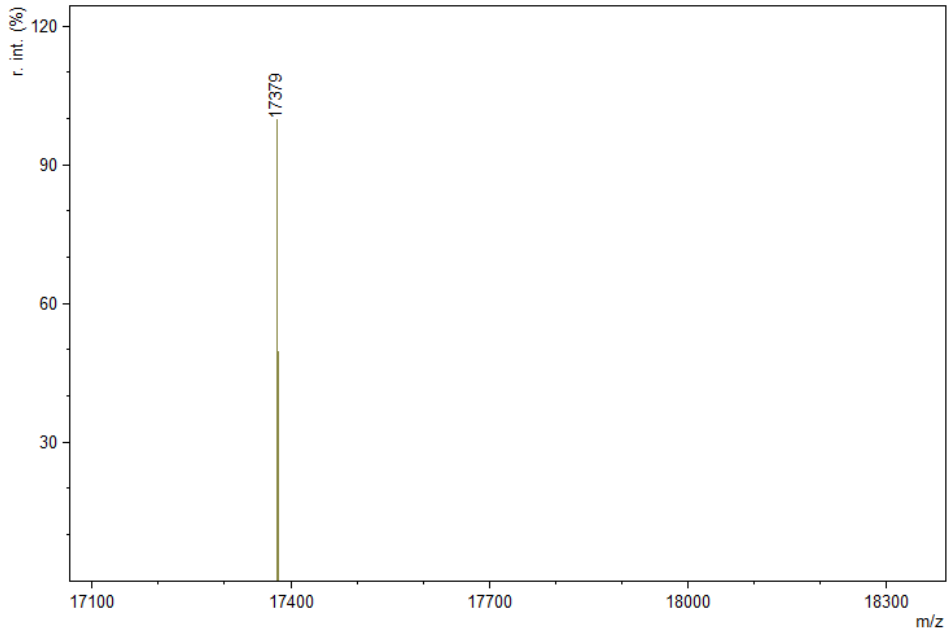


Figure 9.166 Deconvoluted mass spectrum of the analysis of the alphabody 586D starting material. The detected molecular weight is 17379 Da.

Alphabody 586D conjugation with DMEQ-TAD in 10 X PBS pH 4 (20 eq.)

2 µL of the alphabody 586D stock solution (1.16 mg/mL) in 20 mM HEPES 150 mM NaCl was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 4 µL (20 eq) of a 0.7 mM DMEQ-TAD **2d** solution in MeCN was added and mixed well by pipetting up and down for several seconds.

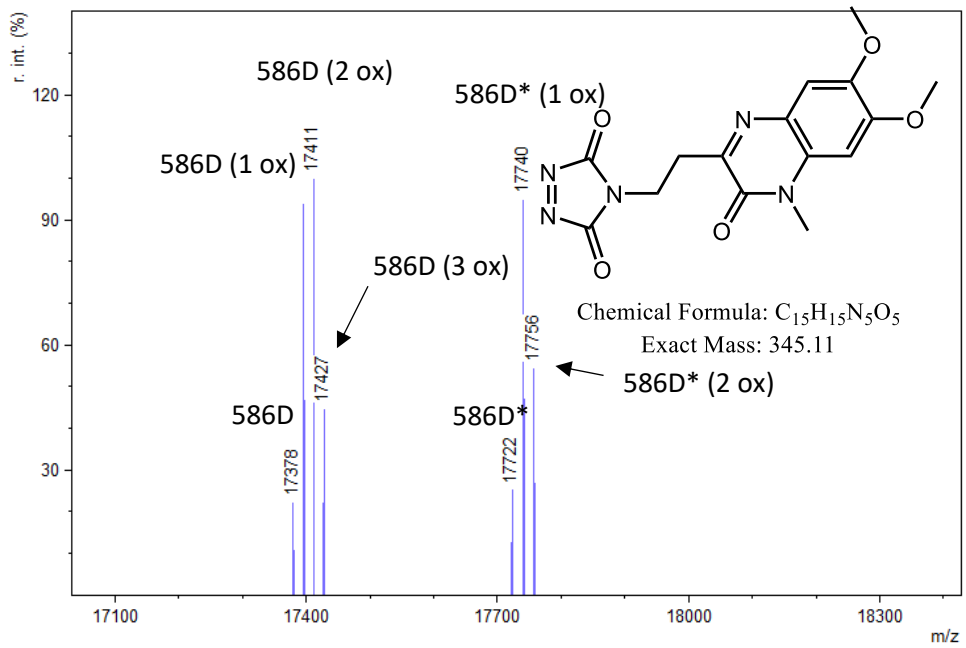


Figure 9.167 deconvoluted mass spectrum of the alphabody 586D protein after addition of DMEQ-TAD **2d** (20 eq) in 10 X PBS pH 4. We observe the starting 586D 17378 Da, with 1 oxidation (+16 Da) 17395 Da, with 2 (+32 Da) 17411 Da and with 3 oxidations (+48 Da) 17427 Da. Additionally we observe the DMEQ-TAD modified alphabody 586D also with 0, 1 and 2 oxidations corresponding to 17722 Da, 17740 Da and 17756 Da respectively.

Methionine oxidation

Upon TAD conjugation reaction we observe (M+16) oxidation products. This can be ascribed to the presence of methionine as the alphabody 586D recombinant protein contains 6 methionine residues in its primary sequence. We found that purging the buffer solution with argon prior to the conjugation reaction leads to reduced oxidation. Using this method in combination with DMEQ-TAD modification on alphabody 586D (6 methionine residues) we observed only one oxidation.

Alphabody 586D conjugation with DMEQ-TAD in 10 X PBS pH 4 (30 eq.)

To reduce methionine oxidation we first saturated the 10 X PBS pH 4 buffer. An open Eppendorf tube containing 10 X PBS pH 4 buffer solution was placed in a vial which was tightly closed with a septum. The vial was purged with argon for 2 hours. Afterwards the buffer was used immediately for protein conjugations. 2 µL of the alphabody 586D stock solution (1.16 mg/mL) in 20 mM HEPES 150 mM NaCl was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 4 µL (20 eq) of a 0.7 mM DMEQ-TAD **2d** solution in MeCN was added and mixed well by pipetting up and down for several seconds.

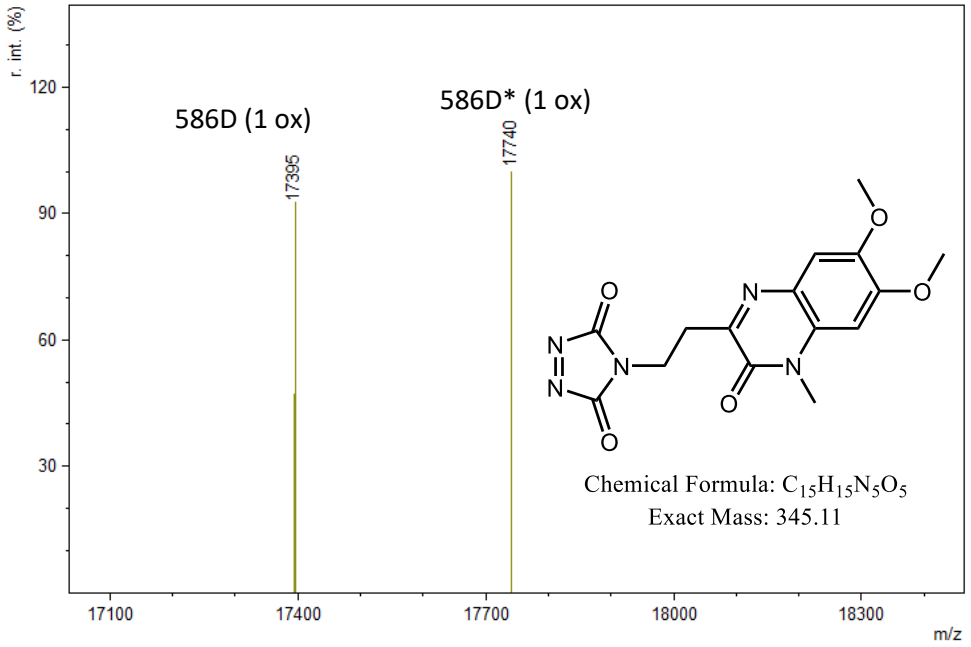


Figure 9.168 deconvoluted mass spectrum of the alphabody 586D protein after addition of DMEQ-TAD **2d** (30 eq) in 10 X PBS pH 4. Note that the buffer solution was saturated with argon prior to use. We observe the starting galectin protein with 1 oxidation (+16 Da) 17395 Da. Additionally we observe the DMEQ-TAD modified alphabody 586D also with 1 oxidation 17740 Da.

Trypsin digest of unmodified alphabody 586D and alphabody 586D conjugation with DMEQ-TAD propanol (20 eq) in 10 X PBS pH 4

2 µL of the alphabody 586D stock solution (1.16 mg/mL) in H_2O was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 4 µL of a 0.7 mM DMEQ-TAD solution in MeCN was added and mixed well by pipetting up and down for several seconds. For the unmodified sample 2µL of the alphabody 586D stock was added to 18 µL 10 X PBS pH 4. 20 µL of 0.1 M Tris buffer pH 8 was added to both samples and the pH was further adjusted to 8 with 2 µL of a 0.57 M NaOH solution. 6 µL of a 0.16 mg/mL trypsin in H_2O was added and the mixture was shaken overnight in the dark. The digests were analyzed on a Fusion Lumos (Thermo) mass spectrometer using the same LC settings as for the synthetic peptides. The mass spectrometer was operated in TopSpeed with a cycle time of 3s. MS1 scans were acquired with an AGC of 2E5 with a maximum ion time of 250 ms and a scan range of 300-1500 at a resolution of 120k. Precursor ions were selected for fragmentation through data dependent settings set to MIPS peptide mode, a charge state between 2-7, an intensity threshold of 5E3 and a dynamic exclusion of 60 s. Precursors were isolated with an isolation window of 1.2 Da in the ion routing multipole, fragmented in the HCD cell with a CE of 34% after obtaining an AGC of 12E3 with a maximum fill time of 40 ms. The fragment ions were detected though the ion trap analyzer in normal mode. In the Mascot search DMEQ-TAD modification of tyrosine and tryptophan was included as a variable modification. For the modification on tryptophan a neutral loss of 345.1073Da was taken into account for modified fragments in the HCD MS/MS spectra, since we have observed that tryptophan-TAD modifications are labile in HCD MS/MS analyses. Tolerance on MS1 level was set to 10 ppm, whereas the tolerance on MS2 level was set to 0.5 Da. The peak lists were generated through Mascot Distiller and searched against the SwissProt proteome v06_2020 restricted to human (containing 20433 sequences) concatenated with the alphabody 586D sequence.

As the sequence of alphabody 586D is confidential, we here show the analysis data for the tryptic peptide containing the tryptophan which also contains a tyrosine residue XXXXXXXYXXWXXX. This tryptic peptide was detected in the unmodified protein sample and was fragmented and annotated using HCD MS/MS and scored 93 (figure 9.169). In the sample of the alphabody 586D modified with 20 equivalents of DMEQ-TAD at pH 4, the DMEQ-TAD modified tryptic peptide XXXXXXXYXXW*XXX was detected and annotated via HCD MS/MS analysis and scored 37 (figure 9.170). The HCD analysis demonstrates that the DMEQ-TAD modification is located on the tryptophan residue. Besides the tyrosine residue in the same tryptic sequence as tryptophan, there are 2 more tyrosine residues. The tryptic peptides containing these tyrosines were detected as unmodified peptides.

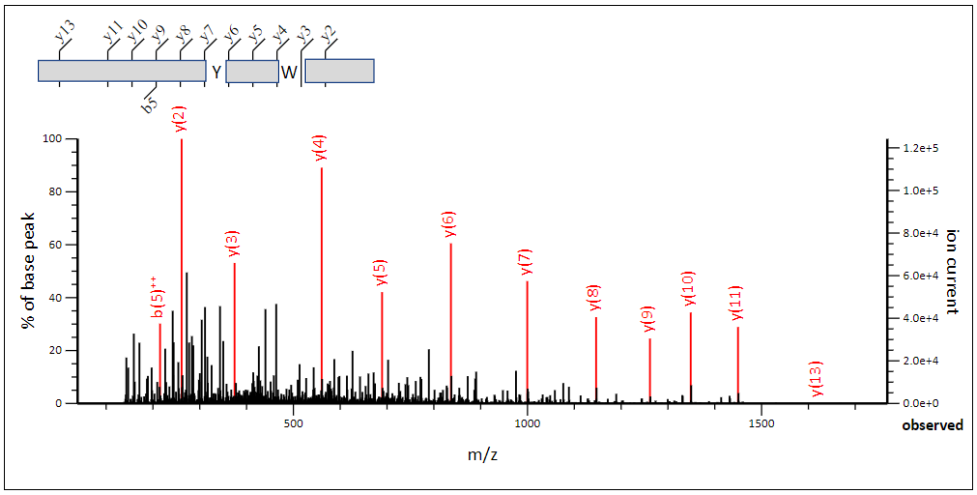


Figure 9.169 HCD MS/MS spectrum of the unmodified tryptic peptide (XXXXXXXYXXWXXX) (top).

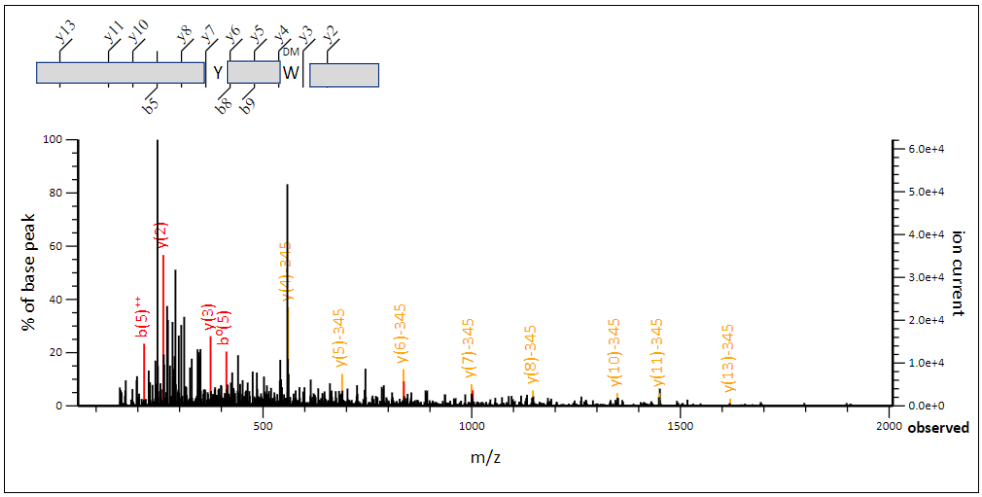


Figure 9.170 HCD MS/MS spectrum of the DMEQ-TAD modified tryptic peptide (XXXXXXXYXXW*XXX). The modification reaction was performed in 10 X PBS pH 4. The "DM" label above the tryptophan residue indicates the location of the DMEQ-TAD. Note that in this figure a fragment ion is marked with "⁰" this indicates -H_2O. The annotated peaks in yellow contain -345 in their label, this indicates the loss of DMEQ-TAD modification on tryptophan as is expected in HCD MS/MS.

Alphabody 586D DMEQ-TAD conjugate stability

The peptide conjugate stability was tested on peptide **6.3f** and TAD-propanol, **2b**. To confirm these data on the protein level the stability of conjugate alphabody 586D DMEQ-TAD, **2d** was tested via HPLC analysis at 370 nm (DMEQ absorbance). Alphabody conjugation was performed with 20 equivalents DMEQ-TAD in 10 X PBS pH 4 as described in 3.3. 100 µL of sample was diluted with 100 µL of 10 X PBS pH 9, and the pH was adjusted to a value of 7. The sample was left on the lab bench at room temperature and analysed by HPLC right after the dilution and 24 hours later. Additionally also 2 blanc samples were recorded: a blanc alphabody 586D sample to identify the alphabody retention time. The alphabody elutes around 4 minutes (figure 9.171, A: 214 nm), no absorption is detected at 370 nm for alphabody 586D (figure 9.171, B: 370 nm).

Figure 9.171 Zoom of the alphabody region of a HPLC chromatogram (A: 214 nm, B: 370 nm). Blanc sample: alphabody 586D in 10 X PBS pH 4, no DMEQ-TAD present.

Another blanc consists of an experiment with DMEQ-TAD, **2d** in 10 X PBS pH 4 but without alphabody present, to confirm that there is no absorption at 370 nm in the alphabody region in the blanc sample.

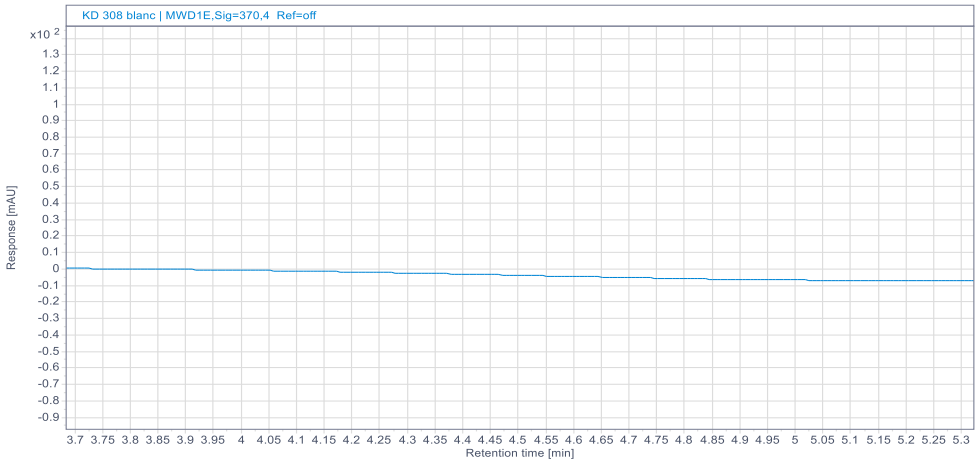


Figure 9.172 Zoom of the alphabody region of a HPLC chromatogram (370 nm). Blanc sample: DMEQ-TAD in 10 X PBS pH 4, no alphabody 586D present.

In the stability experiment the alphabody 586D DMEQ-TAD, **2d** conjugate was analysed by HPLC at the start of the experiment (after dilution into 10 X PBS pH 7) and 24 hours later. The peak area of the conjugate peak in both 370 nm chromatograms remained roughly the same over 24 hours at room temperature (figure 9.173). These findings confirm the stability of the tryptophan-TAD conjugation on proteins in 10 X PBS pH 7 at room temperature. Longer stability time measurements (up to several weeks) as we did for the peptide conjugate **6.3f2b** were not possible for the alphabody 586D conjugate due to instability of the protein itself.

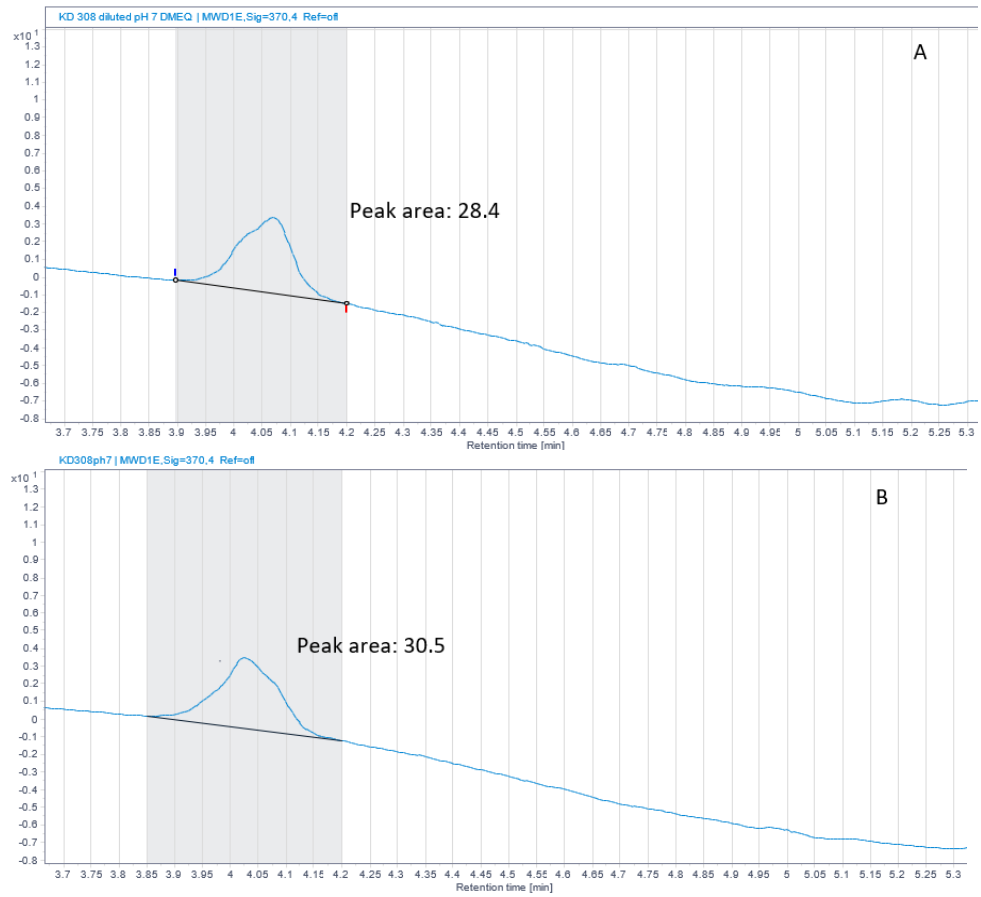


Figure 9.173 Zoom of the alphabody region of a HPLC chromatogram (370 nm). Alphabody 586D conjugated with DMEQ-TAD, **2d**. Conjugation performed in 10 X PBS pH 4. stability experiment in 10 X PBS pH 7 at room temperature. A: 370 nm chromatogram from HPLC analysis right after conjugation and dilution into PBS pH 7; B: 370 nm chromatogram from HPLC analysis after 24 hours.

9.3.3.12 Human galectin-7

A model protein used in this study is galectin-7,[327] a human protein with a molecular weight of 15.4 kDa, purchased from Merck (recombinantly expressed in *E. coli*). In the primary sequence of this protein one tyrosine and one tryptophan residue are present. Additionally there is a methionine residue present in the reported primary amino acid sequence of galectin-7[5]. In the recombinant galectin-7 protein obtained from Merck there are three unknown amino acids added[6]. The experiments with galectin-7 were analyzed on an LTQ-orbitrap XL (Thermo Fisher Scientific) through intact protein analysis in the same manner as the alphabody 586D as described above (section 3.3).

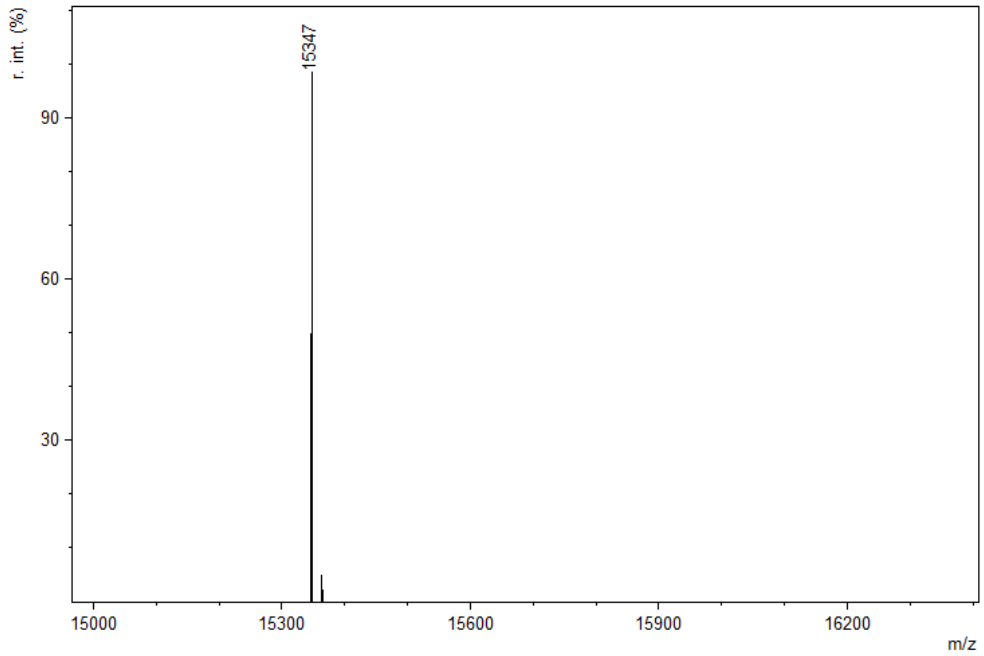


Figure 9.174 Deconvoluted mass spectrum of the analysis of the galectin-7 starting material. The detected molecular weight for the galectin-7 protein is 15347 Da.

[5] https://www.uniprot.org/uniprot/P47929

[6] Merck was not able to give further information on the three added amino acids which are present after HIS-tag removal. Most likely at least one extra methionine is present.

Galectin-7 conjugation with DMEQ-TAD in 10 X PBS pH 4 (10 eq.)

2 µL of the galectin-7 stock solution (1mg/mL) in H_2O was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 2 µL (10 eq) of a 0.7 mM DMEQ-TAD **2d** solution in MeCN was added and mixed well by pipetting up and down for several seconds.

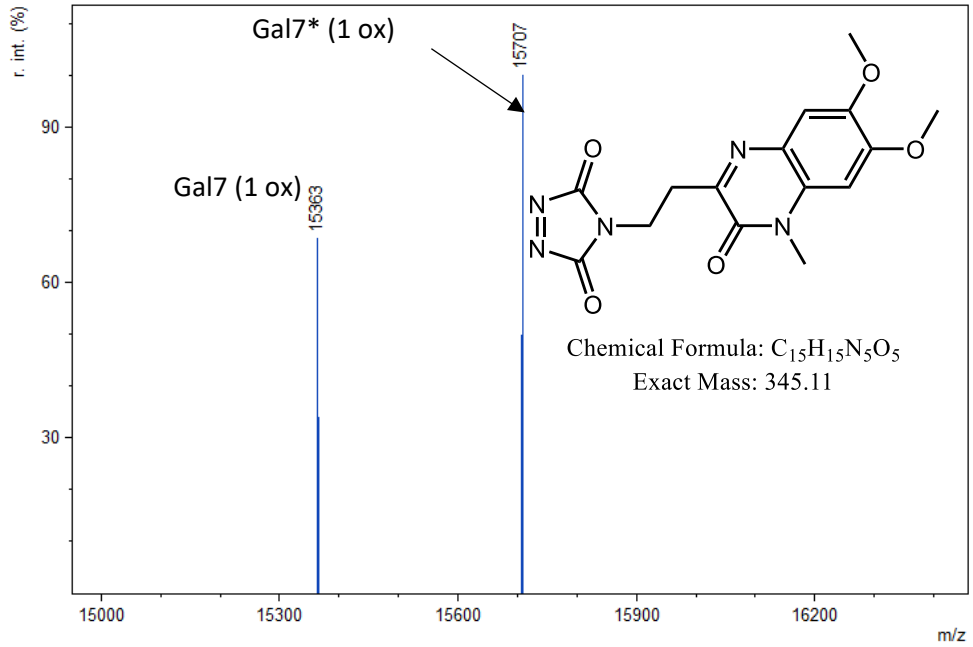


Figure 9.175 Deconvoluted mass spectrum of the galectin-7 protein after addition of DMEQ-TAD **2d** (10 eq) in 10 X PBS pH 4. We observe the starting galectin protein with an oxidation (+16 Da, *vide supra* methionine presence) at 15363 Da. Additionally we observe the DMEQ-TAD modified galectin-7 with an oxidation at 15707 Da.

Galectin-7 conjugation with TAD-propanol in 10 X PBS pH 4 (10 eq.)

2 µL of the galectin-7 stock solution (1mg/mL) in H_2O was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 2 µL (10 eq) of a 0.7 mM TAD-propanol **2b** solution in MeCN was added and mixed well by pipetting up and down for several seconds.

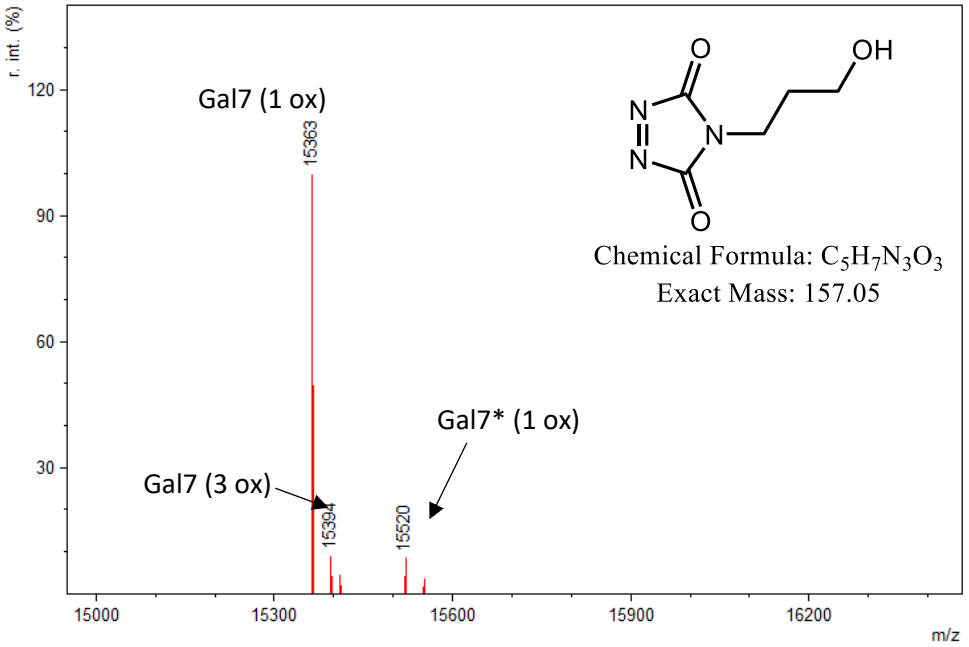


Figure 9.176 deconvoluted mass spectrum of the galectin-7 protein after addition of TAD-propanol **2b** (10 eq) in 10 X PBS pH 4. We observe the starting galectin protein with an oxidation (+16 Da, *vide supra* methionine presence) at 15363 Da and a small amount of galectin-7 with three oxidations (+48 Da, *vide supra* methionine presence) at 15394 Da. Additionally we observe the TAD-propanol modified galectin-7 also with an oxidation at 15520 Da.

Galectin-7 conjugation with PTAD-alkyne in 10 X PBS pH 4 (10 eq.)

2 µL of the galectin-7 stock solution (1mg/mL) in H_2O was added to 16 µL 10 X PBS pH 4 and mixed well by pipetting up and down. 2 µL (10 eq) of a 0.7 mM PTAD-alkyne **2c** solution in MeCN was added and mixed well by pipetting up and down for several seconds.

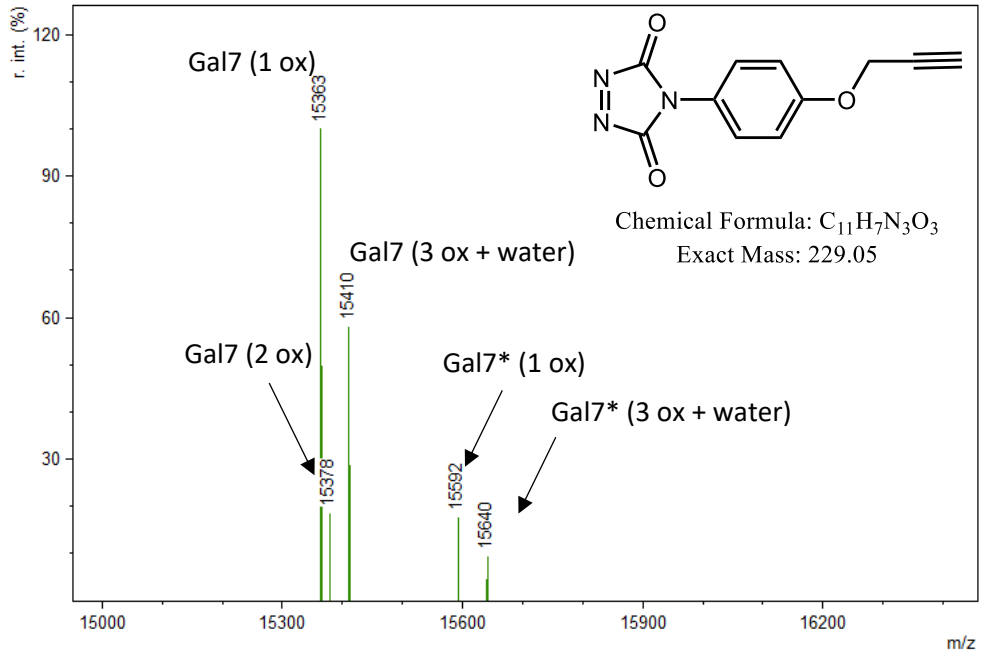


Figure 9.177 deconvoluted mass spectrum of the galectin-7 protein after addition of PTAD-alkyne **2c** (10 eq) in 10 X PBS pH 4. We observe the starting galectin protein with an oxidation (+16 Da, *vide supra* methionine presence) at 15363 Da, a double (+32 Da) at 15378 Da and triple oxidation + water (+66 Da) at 15410 Da. Additionally we observe the PTAD-alkyne modified galectin-7 with an oxidation at 15520 Da and with a triple oxidation + water at 15640 Da.

Trypsin digest of unmodified galectin-7 and galectin-7 conjugation with TAD propanol (20 eq) in 10 X PBS pH 4 and pH 7

2 µL of the galectin-7 stock solution (1mg/mL) in H_2O was added to 16 µL 10 X PBS pH 4 or 7 and mixed well by pipetting up and down. 2 µL of a 1.4 mM TAD-propanol solution in MeCN was added and mixed well by pipetting up and down for several seconds. For the unmodified galectin-7 sample 2 µL of protein stock solution was added to 18 µL of 10 X PBS buffer pH 4. 20 µL of 0.1 M Tris buffer pH 8 was added, for the sample in 10 X PBS pH 4 the pH was further adjusted to 8 with 2 µL of a 0.57 M NaOH solution. 6 µL of a 0.16 mg/mL trypsin in H_2O was added and the mixture was shaken overnight in the dark. The digests were analyzed on a Fusion Lumos (Thermo) mass spectrometer using the same LC settings as the synthetic peptides (see section 1.1). The mass spectrometer was operated in TopSpeed with a cycle time of 3s. MS1 scans were acquired with an AGC of 2E5 with a maximum ion time of 250 ms and a scan range of 300-1500 at a resolution of 120k. Precursor ions were selected for fragmentation through data dependent settings set to MIPS peptide mode, a charge state between 2-7, an intensity threshold of 5E3 and a dynamic exclusion of 60 s. Precursors were isolated with an isolation window of 1.2 Da in the ion routing multipole, fragmented in the HCD cell with a CE of 34% after obtaining an AGC of 12E3 with a maximum fill time of 40 ms. The fragment ions were detected though the ion trap analyzer in normal mode. To focus on the tyrosine and tryptophan peptide, a targeted inclusion method was used using the double and triple charged precursor ions of the TAD modified tyrosine and tryptophan peptide in the inclusion list. The dynamic exclusion option was left out for these targeted methods.

The TAD-propanol modified tryptic peptide containing tyrosine (AVVGDAQY*HHFR) was only detected in trace amounts in the galectin-7 TAD-propanol conjugation at pH 4. Peak area values for the double and triple charged TAD-modified tryptic peptide were normalized to peptides K.SSLPEGIRPGTVLR.I [7, 20] (missed 1) and R.HRLPLAR.V [111, 117] (missed 1) skyline software (figure 9.178). On the other hand the TAD-propanol modified tryptic peptide containing tryptophan was detected in both samples. Peak area values for the double and triple charged TAD-modified tryptic peptide were normalized to peptides K.SSLPEGIRPGTVLR.I [7, 20] (missed 1) and R.HRLPLAR.V [111, 117] (missed 1) skyline software (figure 9.179). In the Mascot search TAD-propanol modification of tyrosine and tryptophan was included as a variable modification. For the modification on tryptophan a neutral loss of 157.0487 Da was taken into account for modified fragments in the HCD MS/MS spectra, since we have observed that tryptophan-TAD modifications are labile in HCD MS/MS analyses (figures 9.180 and 9.182). Tolerance on MS1 level was set to 10 ppm, whereas the tolerance on MS2 level was set to 0.5 Da. The peak lists were generated through Mascot Distiller and searched against the SwissProt proteome v06_2020 restricted to human (containing 20433 sequences) concatenated with the Galectin-7 sequence. In the unmodified galectin-7 experiment, the tryptophan containing peptide was not detected.

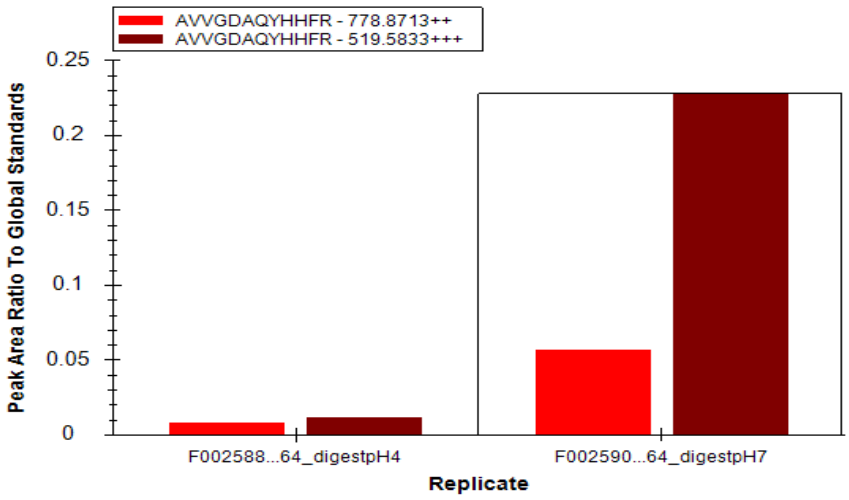


Figure 9.178 Histogram representing the peak area of the TAD-propanol modified tryptic peptide containing tyrosine (AVVGDAQY*HHFR) in the samples at pH 4 left and pH 7 right.

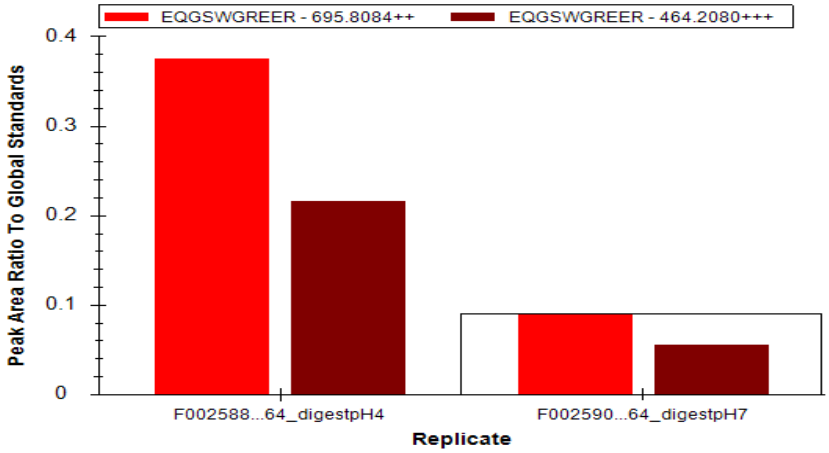


Figure 9.179 Histogram representing the peak area of the TAD-propanol modified tryptic peptide containing tryptophan (EQGSW*GREER) in the samples at pH 4 left and pH 7 right.

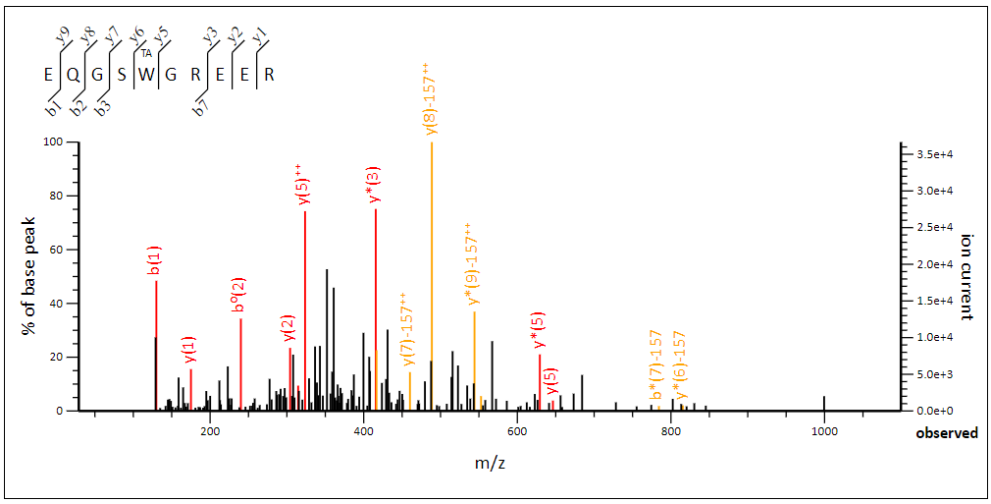


Figure 9.180 HCD MS/MS spectrum of the TAD-propanol modified tryptic peptide (EQGSW*GREER), MASCOT score 8. The modification reaction was performed in 10 X PBS pH 4. Note that in this figure some fragment ions are marked with "*" or "⁰" this indicates -NH_3 or -H_2O respectively. The "*" modification used here is not to be confused with TAD-modified peptides/proteins in the remainder of this document also indicated by "*". The annotated peaks in yellow contain -157 in their label, this indicates the loss of TAD -propanol modification on tryptophan as is expected in HCD MS/MS.

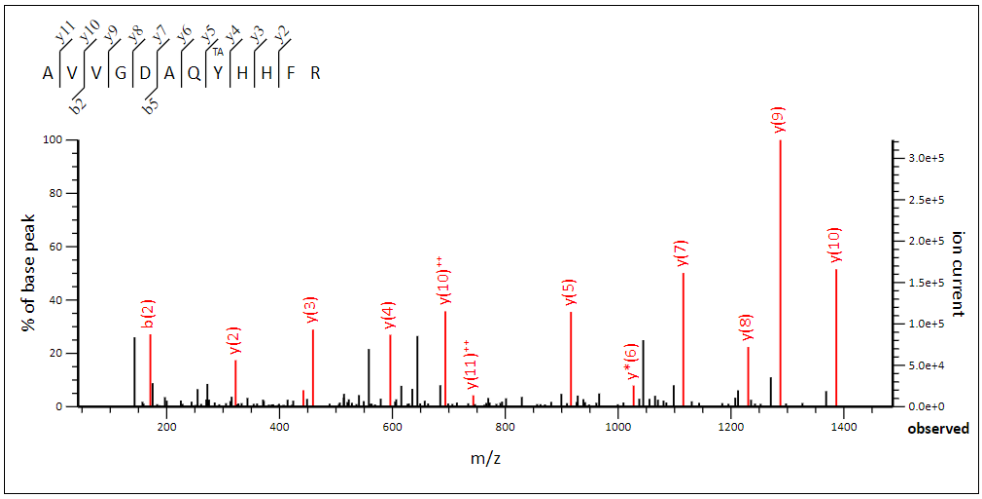


Figure 9.181 HCD MS/MS spectrum of the TAD-propanol modified tryptic peptide (AVVGDAQY*HHFR), MASCOT score 77. The modification reaction was performed in 10 X PBS pH 7. Note that in this figure some fragment ions are marked with "*"this indicates -NH_3. The "*" modification used here is not to be confused with TAD-modified peptides/proteins in the remainder of this document also indicated by "*".

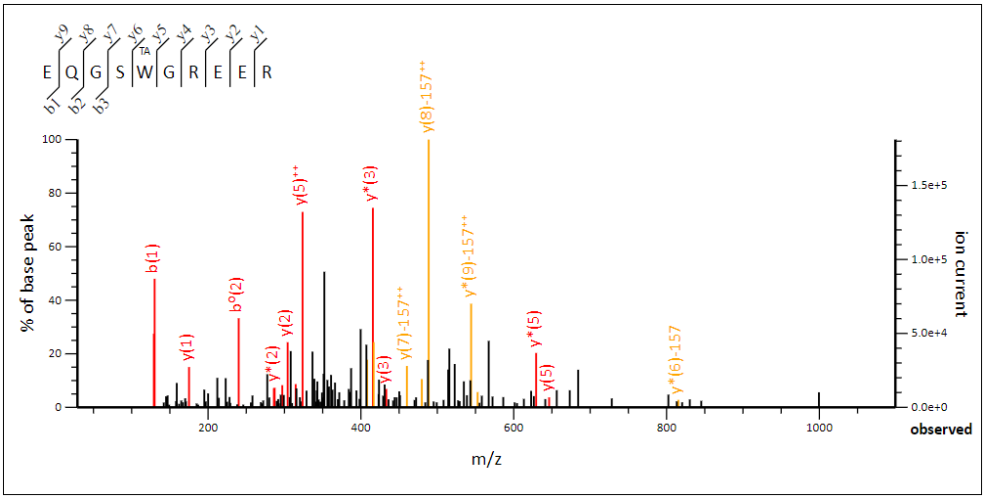


Figure 9.182 HCD MS/MS spectrum of the TAD-propanol modified tryptic peptide (EQGSW*GREER), MASCOT score 16. The TAD-propanol label is indicated by "TA" above the tryptophan residue. The modification reaction was performed in 10 X PBS pH 7. Note that in this figure some fragment ions are marked with "*" or "°" this indicates -NH_3 or -H_2O respectively. The "*" modification used here is not to be confused with TAD-modified peptides/proteins in the remainder of this document also indicated by "*".The annotated peaks in yellow contain -157 in their label, this indicates the loss of TAD -propanol modification on tryptophan as is expected in HCD MS/MS.

9.3.3.13 Small molecule Trp-TAD conjugation and NMR structural analyses of the obtained adducts

Boc-Trp-OH

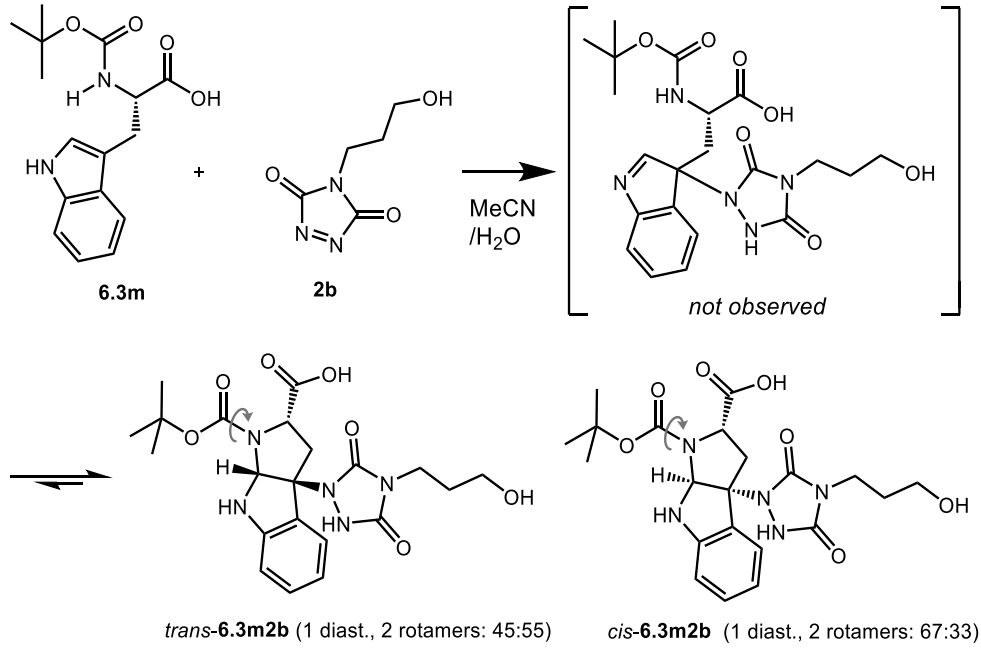


Figure 9.183 Structural representation of Boc-Trp-OH **6.3m** reaction with propanol TAD **2b**.

Compound **2b** was prepared similarly as described. 21.8 mg of trichloroisocyanuric acid (TCCA) (0.094 mmol, 0.33 eq.) was added to 5 mL MeCN together with 49 mg propanol urazole (0.31 mmol, 1.1 eq.) and the reaction was stirred for 2 hours at room temperature. Afterwards the reaction mixture was centrifuged to precipitate any compounds insoluble in MeCN and **2b** was used directly from the supernatant. Then, 85.6 mg of Boc-Trp-OH **6.3m** (0.28 mmol) was dissolved in 5mL H_2O/MeCN (1:1, , 0.056 M). Next, the freshly prepared bright purple solution of **2b** was added with a glass pipet in 5 portions of 1mL. The purple color of **2b** disappeared over several seconds following the addition of each portion to **6.3m.** LC-MS analysis of the reaction of **2b** with **6.3m** in H_2O/MeCN showed two separate product peaks of equal intensity, both corresponding to the molecular weight of the expected adduct of **6.3m** and **2b**. Both peaks were well separated and analytical samples (~5 mg) were collected using a semiprep HPLC instrument equipped with a Phenomenex Luna C18 column at 35 °C with a flow rate of 16 mL/min. The column was eluted with a gradient starting at 100% H_2O containing 0.1% TFA to 100% acetonitrile in 20 min. Both samples were lyophilized and resolubilized in DMSO-d6.

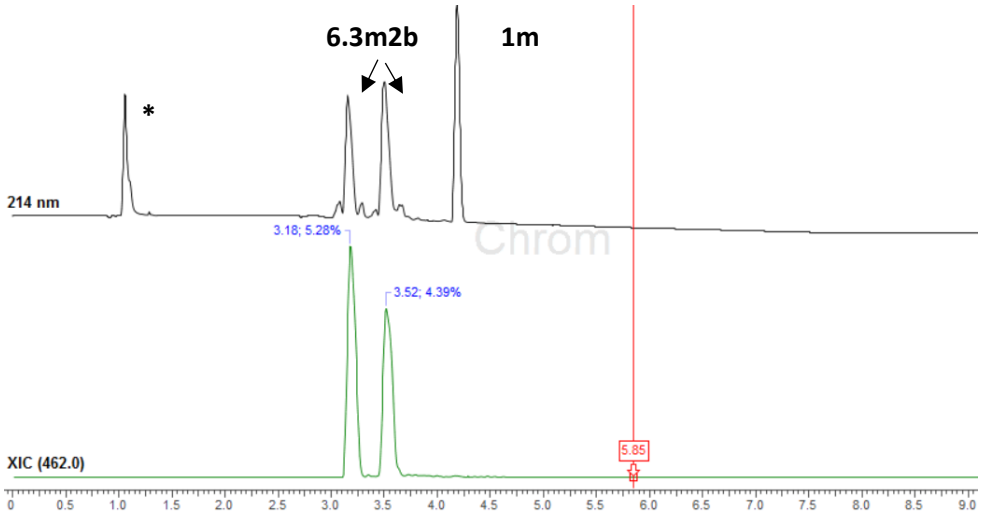


Figure 9.184 LC-chromatogram at 214 nm (top) and extracted ion chromatogram (XIC) for the **6.3m2b** ion (462 Da) (bottom). The peak eluting at around t_R = 1.1 min (marked with "*") is not peptide related.

Detailed analysis of 1D proton and 2D proton carbon heteronuclear NMR measurements indicated that the first eluting fraction corresponds to a single diastereomer of the annulated adduct **6.3m2b**, observed as 2 N-Boc-rotamers. The carboxylic acid and the urazole substituent on the pyrrolidine ring are oriented *trans* with respect to each other, while the indoline/pyrrolidine ring fusion is expected to be a *cis*-ring fusion. Similarly, the second collected fraction is most readily interpreted as two N-Boc-rotamers of the other diastereoisomer of the same adduct **6.3m2b**, where the carboxylic acid and the urazole functions are oriented in a *cis*-fashion on the newly formed pyrrolidine ring. This formation of the additional pyrrolidine ring structure from the expected 3H-indole initial adduct via a spontaneous imine/aminal conversion was previously reported in reactions involving TAD-indoles,[325] and also in organoradical type conjugations on tryptophans.[95] This equilibrium may be solvent dependent.

The isolated fraction assigned the ***trans*-6.3m2b** diastereomer formed consisted of a 45:55 mixture of Boc rotamers, as determined by integration of the tert-butyl protons which were the most well resolved signals in the 1D NMR spectrum. The listed carbon NMR data was only derived from observed cross-peaks in 2D HSQC and HMBC spectra, as the 1D carbon NMR spectra proved too complex due to scarcity of sample and the presence of rotamers:

^{1}H-NMR (400 MHz, DMSO-d_6): δ = 1.39 & 1.49 (s, 9H (55:45), OC(C*H*$_3$)$_3$), 1.59 (2H, quint(br), *J* = ~6.8 Hz, CH$_2$-C*H*$_2$-CH$_2$-OH), 2.88 & 2.91 (1H (55:45), d(br), *J* = 13.4 Hz, C*H*H-CH-CO$_2$H), 3.10 & 3.16 (1H (45:55), dd, *J* = 13.4, 9.5 Hz, CH*H*-CH-CO$_2$H), 3.32 (2H, t, *J* = 6.4 Hz, CH$_2$-C*H*$_2$-OH), 3.38 (2H, t, *J* = 7.2 Hz, N-C*H*$_2$-CH$_2$), 4.49 & 4.50 (1H (45:55), d(br), J = 9.5 Hz, C*H*-COOH), 5.72 & 5.74 (1H (45:55), s, NH-C*H*-N), 6.48 – 6.57 (band, 2H, 2 x Ar-*H*), 6.99-7.04 (band, 2H, 2 x Ar-*H*), 10.16 & 10.21 (1H(55:45), s(br), urazole-N*H*) ppm. Very broad signals are visible in the 5-6 and 11-13 ppm range, most likely belonging to OH's and NH's, integrating for about 3H in total.

^{13}C-NMR (100 MHz, DMSO-d_6, HSQC-HMBC): δ = 27.8 (C(*C*H$_3$)$_3$), 30.4 (CH$_2$*C*H$_2$CH$_2$), 35.9 (N-*C*H$_2$CH$_2$), 38.6 (*C*H$_2$CHCO$_2$H), 57.7 (*C*H$_2$OH), 58.4 (*C*HCO$_2$H), 79.0 (NH-*C*H-N),

79.4/79.7 (O*C*(CH_3)$_3$), 108.8 (=*C*H), 116.8 (=*C*H), 122.9 (=*C*H), 124.9 (C), 129.8 (=*C*H), 151.0 (C), 154.8 (*C*O), 155.8 (*C*O), 171.5 (*C*O_2H) ppm. The carbon corresponding to the carbamoyl group cannot be observed in the obtained spectra. The relative stereochemistry can be assigned based on the absence of a large scalar coupling between two vicinal protons in the pyrrolidine ring (only possible for a dihedral angle around 0°).

The isolated fraction assigned the ***cis*-6.3m2b** diastereomer structure consisted of a 67:33 mixture of Boc rotamers, as determined by integration of the tert-butyl protons which were the most well resolved signals in the 1D NMR spectrum. The listed carbon NMR data was only derived from observed cross-peaks in 2D HSQC and HMBC spectra, as the 1D carbon NMR spectra proved too complex due to scarcity of sample and the presence of rotamers:

^{1}H-NMR (400 MHz, DMSO-d_6): 1.36 & 1.47 (s, 9H (67:33), OC(C*H*$_3$)$_3$), 1.57 (2H, quint(br), *J* = ~6.6 Hz, CH_2-C*H*$_2$-CH_2-OH), 2.84 (33% of 1H, d(AB)d(br), *J* = 13.3(AB), 7.2 Hz, C*H*H-CH-CO_2H), 2.95 (67% of 1H, d(AB)d(br), *J* = 13.8(AB), 7.8 Hz, C*H*H-CH-CO_2H), 3.00 (33% of 1H, d(AB)d, *J* = 13.3(AB), 7.5 Hz, CH*H*-CH-CO_2H)3,02 (67% of 1H, d(AB)d(br), *J* = 13.8(AB), 7.6 Hz, CH*H*-CH-CO_2H), 3.31 (2H, t, *J* = 6.1 Hz, CH_2-C*H*$_2$-OH), 3.37 (2H, t, *J* = 7.3 Hz, N-C*H*$_2$-CH_2), 3.84 & 3.93 (1H (67:33), t, J = ~7.7 Hz, C*H*-COOH), 5.81 & 5.83 (1H (33:67), s, NH-C*H*-N), 6.57-6.64 (band, 2H, 2 x Ar-*H*), 7.05-7.12 (band, 2H, 2 x Ar-*H*), 10.21 & 10.29 (1H(33:67), s(br), urazole-N*H*) ppm. Very broad signals are visible in the 5-6 and 11-13 ppm range, most likely belonging to OH's and NH's, integrating for about 3H in total.

^{13}C-NMR (100 MHz, DMSO-d_6, HSQC-HMBC): δ = 27.4/27.6 (C(*C*H_3)$_3$), 30.2 (CH_2*C*$H_2$$CH_2$), 35.8 (N-*C*$H_2$$CH_2$), 39.0 (*C*$H_2CHCO_2$H), 57.7 (*C*$H_2$OH), 58.3/58.6 (*C*HC$O_2$H), 79.1/79.3 (NH-*C*H-N), 79.4/79.7 (O*C*(CH_3)$_3$), 109.0 (=*C*H), 117.4 (=*C*H), 122.7 (=*C*H), 125.3 (C), 130.0 (=*C*H), 150.1 (C), 154.7 (*C*O), 155.7 (*C*O), 172.9 (*C*O_2H) ppm. The carbon corresponding to the carbamoyl group cannot be observed in the obtained spectra. The relative stereochemistry can be assigned based on two >7 Hz scalar couplings to the methine proton in carboxyl-substituted position of the pyrrolidine ring.

N-Acyl-Trp-OMe

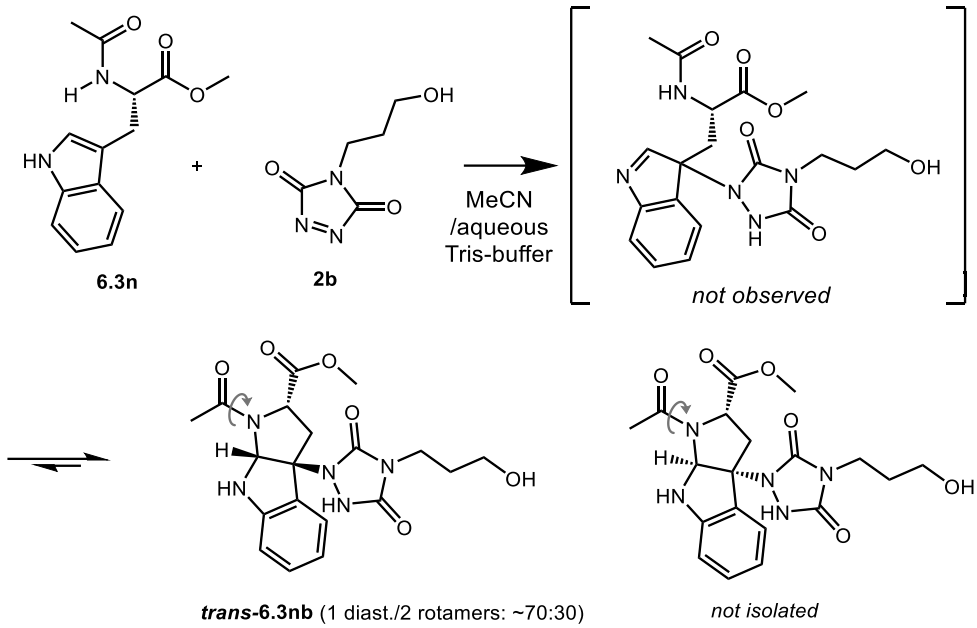


Figure 9.185 Structural representation of N-Acyl-Trp-OMe **6.3n** reaction with propanol TAD **2b**.

Compound **2b** was prepared similarly as described in S1.2. 21.8 mg of trichloroisocyanuric acid (TCCA) (0.094 mmol, 0.33 eq.) was added to 5 mL MeCN together with 49 mg propanol urazole (0.31 mmol, 1.1 eq.) and the reaction was stirred for 2 hours at room temperature. Afterwards the reaction mixture was centrifuged to precipitate any compounds insoluble in MeCN and **2b** was used directly from the supernatant. 82 mg of N-Acyl-Trp-OMe **6.3n** (0.31 mmol, 0.052 M) was dissolved in 6 mL Tris/MeCN (1:1). **2b** was added with a glass pipet in portions of 1mL (5 x). The bright purple color of **2b** disappeared after several seconds upon the addition of each portion of **2b** to **6.3n.** LC-MS analysis of the reaction of **6.3n** with **2b** in MeCN demonstrated a single broad product peak corresponding to the molecular weight of the adduct pf **6.3n** and **2b**. The reaction mixture was purified using a semiprep HPLC instrument equipped with a Phenomenex Luna C18 column at 35 °C with a flow rate of 16 mL/min. The column was eluted with a gradient starting at 100% H_2O containing 0.1% TFA to 100% acetonitrile in 20 min. HRMS (ESI): m/z calculated for **trans-6.3nb** $C_{19}H_{23}N_5O_6$ $[M+H]^+$ 418.1721, found 4181731 (figure 9.186).

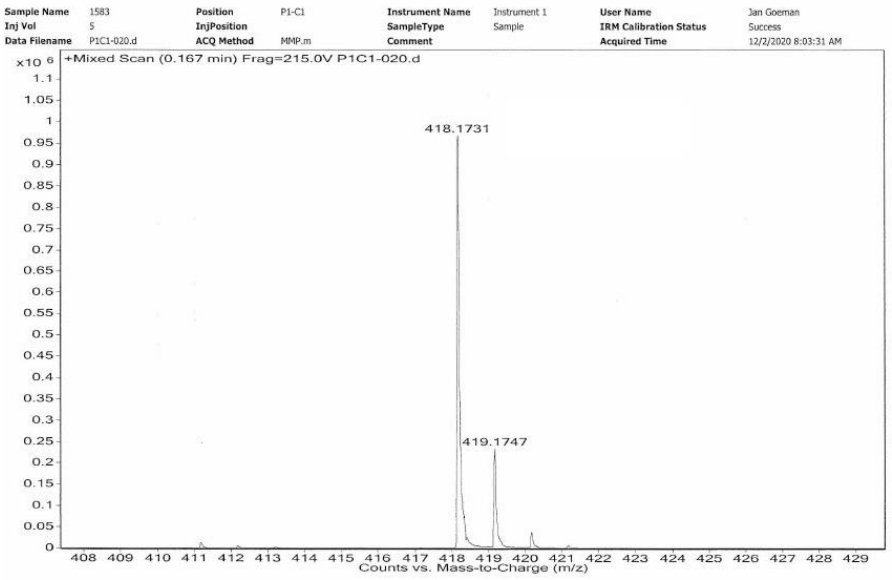


Figure 9.186 HRMS (ESI) analysis: m/z calculated for **trans-6.3n2b** $C_{19}H_{23}N_5O_6$ $[M+H]^+$ 418.1721, found 418.1731.

The collected fraction was lyophilized and resolubilized in acetonitrile-d3:D_2O (1:1). Similar to the previous characterisation of **6.3m2b**, detailed analysis of 1D proton and carbon, and 2D proton carbon heteronuclear NMR measurements, could be most readily interpreted as corresponding to the annulated adduct **6.3n2b**, but stereochemical interpretation was hampered here by the fact that the two stereoisomers were not seperable, and integration and assignments are difficult due to the possible N-acetamide rotamers or possible other isomers. The overall similarity of the major resonances (integrating for 60:25:15) allow the tentative assignment of two rotamers for the trans-diastereomer as the major product (indoline/pyrrolidine ring fused product with a trans-relation: ***trans*-6.3n2b**). The cis-diastereomer is not observed, but several minor peaks occur obscuring a full analysis. One possible explanation for this result is a fast dynamic equilibrium of aminal, hemiaminal and imine products for the cis-isomer, explaining the presence of several minor peaks in the spectrum. This could not be confirmed, and the compounds showed limited solubility in non-aqueous media solvents, further hampering analysis. It seems likely that the TAD-conjugation does not proceed stereoselectively here, but rather that it is difficult to isolate one of the two possible diastereoisomers.

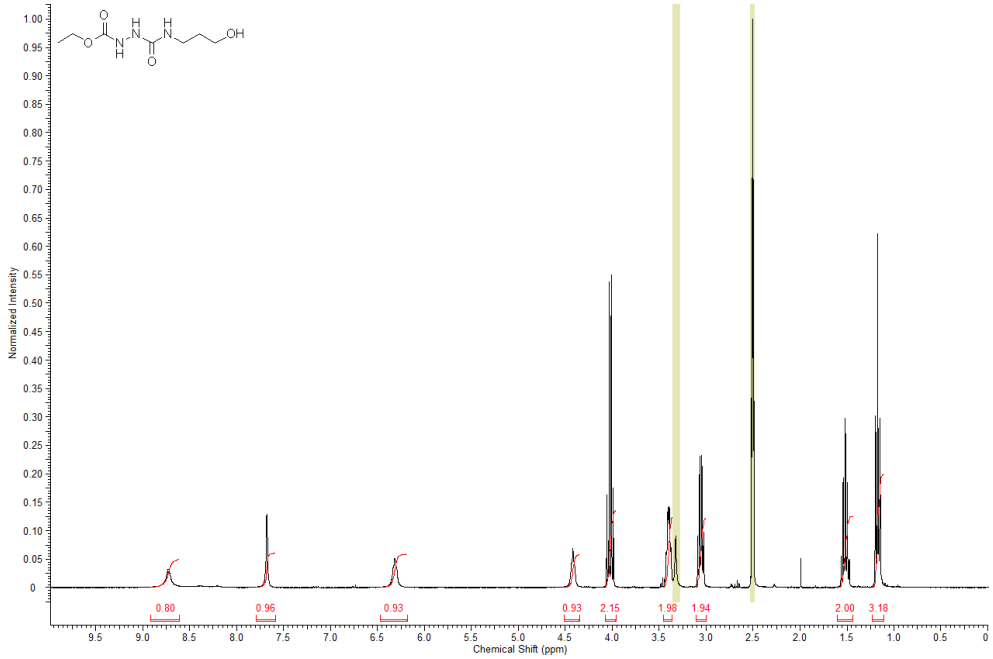
The purified analytical sample (5-10 mg) was assigned the overall ***trans*-2nb** structure, with a rotamer ratio of 70:30 of the N-acetamide, as determined by integration of the well resolved resonances in the 1D NMR spectrum. The listed NMR data is only that of the major rotamer, as analysis of the minor rotamer proved too complex due to overlap with the other minor isomers or possible hydrolysis product:

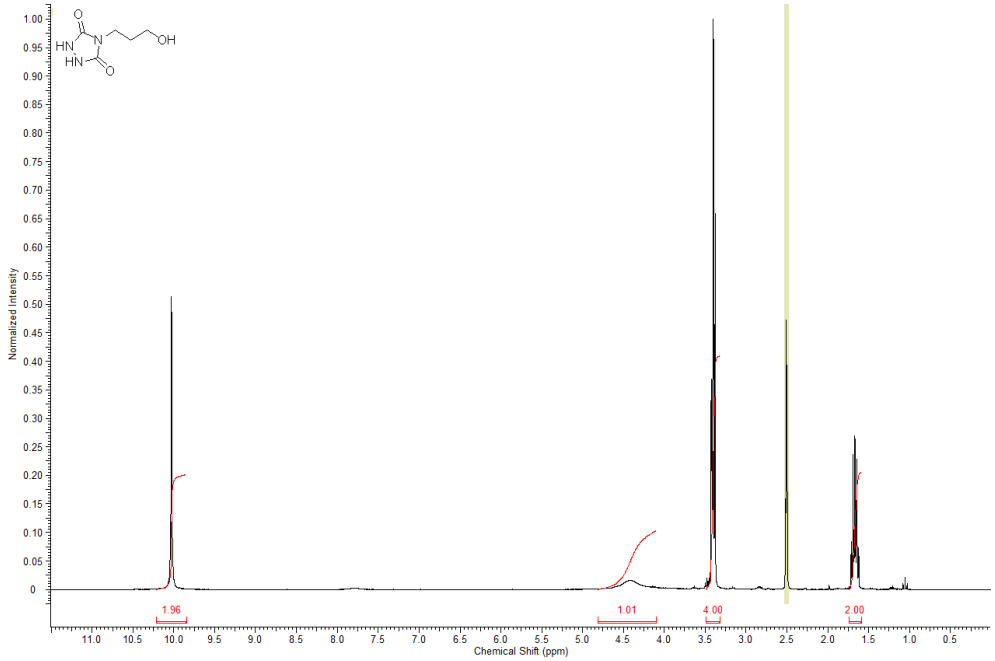
^{1}H-NMR (400 MHz, CD_3CN:D_2O 1:1): δ = 1.67 (2H, quint(br), *J* = ~6.4 Hz, CH_2-C*H*$_2$-CH_2-OH), 1.96 (3H, s, COC*H*$_3$), 3.17 (1H, dd(br), *J* = 13.0, 8.4 Hz, C*H*H-CH-CO_2D), 3.19 (3H, s, O*C*H$_3$) 3.29 (1H, d(br), *J* = 13.0 Hz, CH*H*-CH-CO_2Me), 3.28-3.34 (2H, m, CH_2-

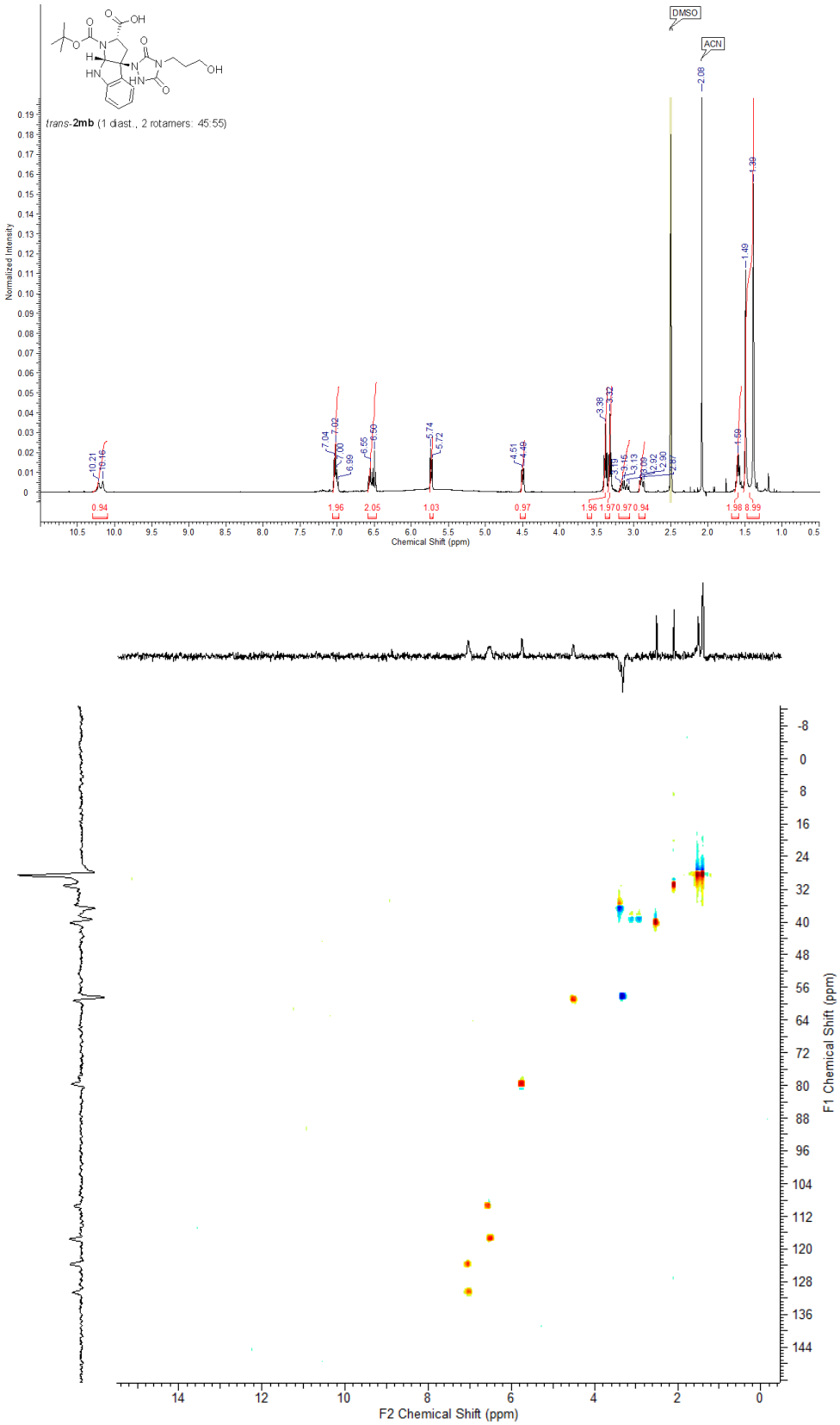
C*H*$_2$-OD), 3.39 (2H, t, *J* = 7.2 Hz, N-C*H*$_2$-CH$_2$), 4.73 (1H, d(br), J = 8.4 Hz, C*H*-CO$_2$Me), 5.76 (1H, s, NH-C*H*-N), 6.60 – 6.70 (band, 2H, 2 x Ar-*H*), 7.11-7.20 (band, 2H, 2 x Ar-*H*). The urazole and other OH and NH are not visible due to deuterium exchange with D_2O.

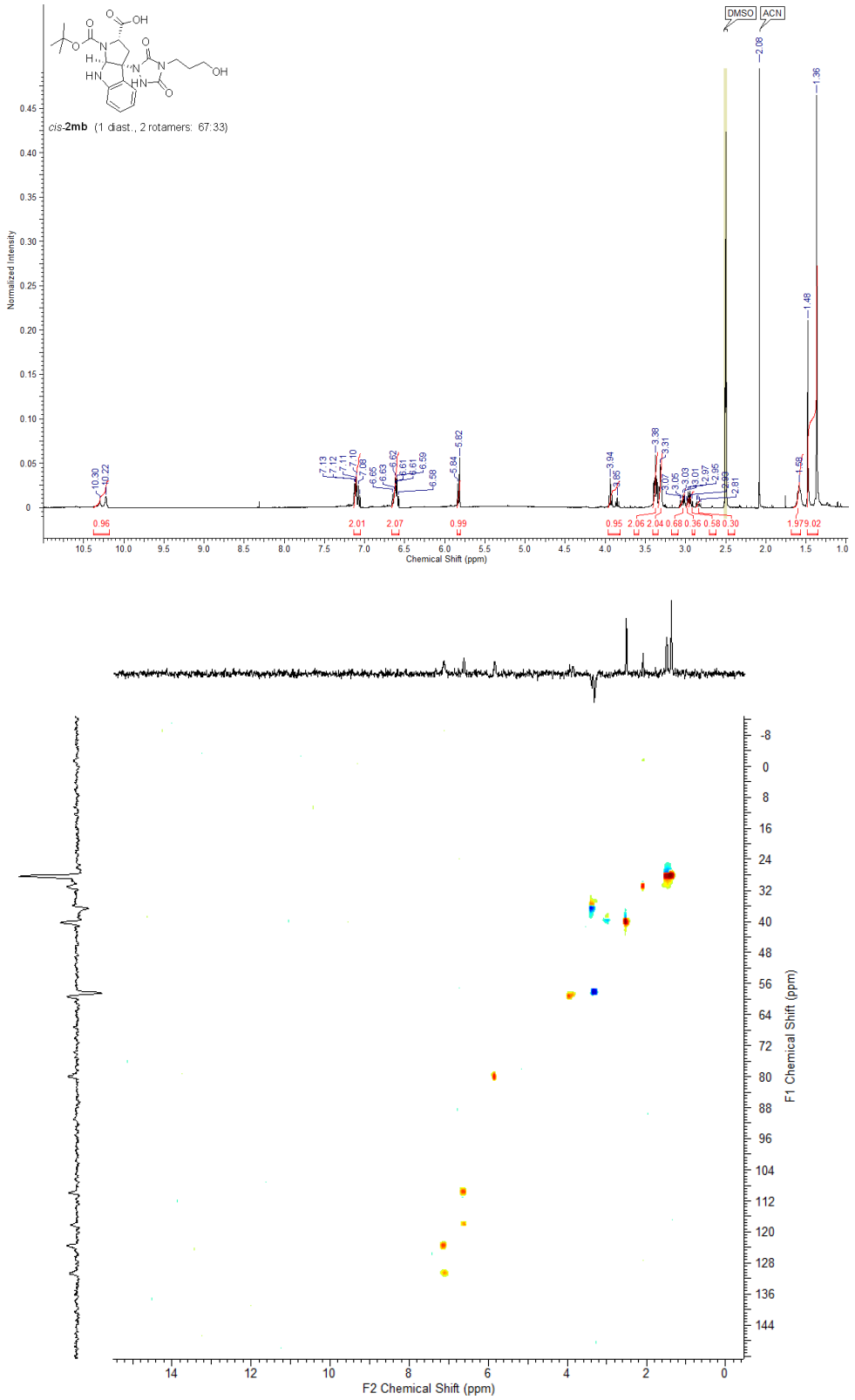
^{13}C-NMR (100 MHz, DMSO-d_6, HSQC-HMBC): δ = 21.6 (CO*C*H$_3$), 31.2 (CH$_2$*C*H$_2$CH$_2$), 36.6 (N-*C*H$_2$CH$_2$), 38.8 (*C*H$_2$CHCO$_2$H), 52.5 (O*C*H$_3$) 59.1 (*C*H$_2$OD), 61.0 (*C*HCO$_2$D), 79.6 (NH-*C*H-N), 109.0 (=*C*H), 117.4 (=*C*H), 122.7 (=*C*H), 123.0 (=*C*H), 135.2 (urazole *C*O), 136.6 (urazole *C*O), 150.5 (C), 150.9 (CH$_3$*C*O) ppm. One quaternary aromatic carbon cannot be assigned.

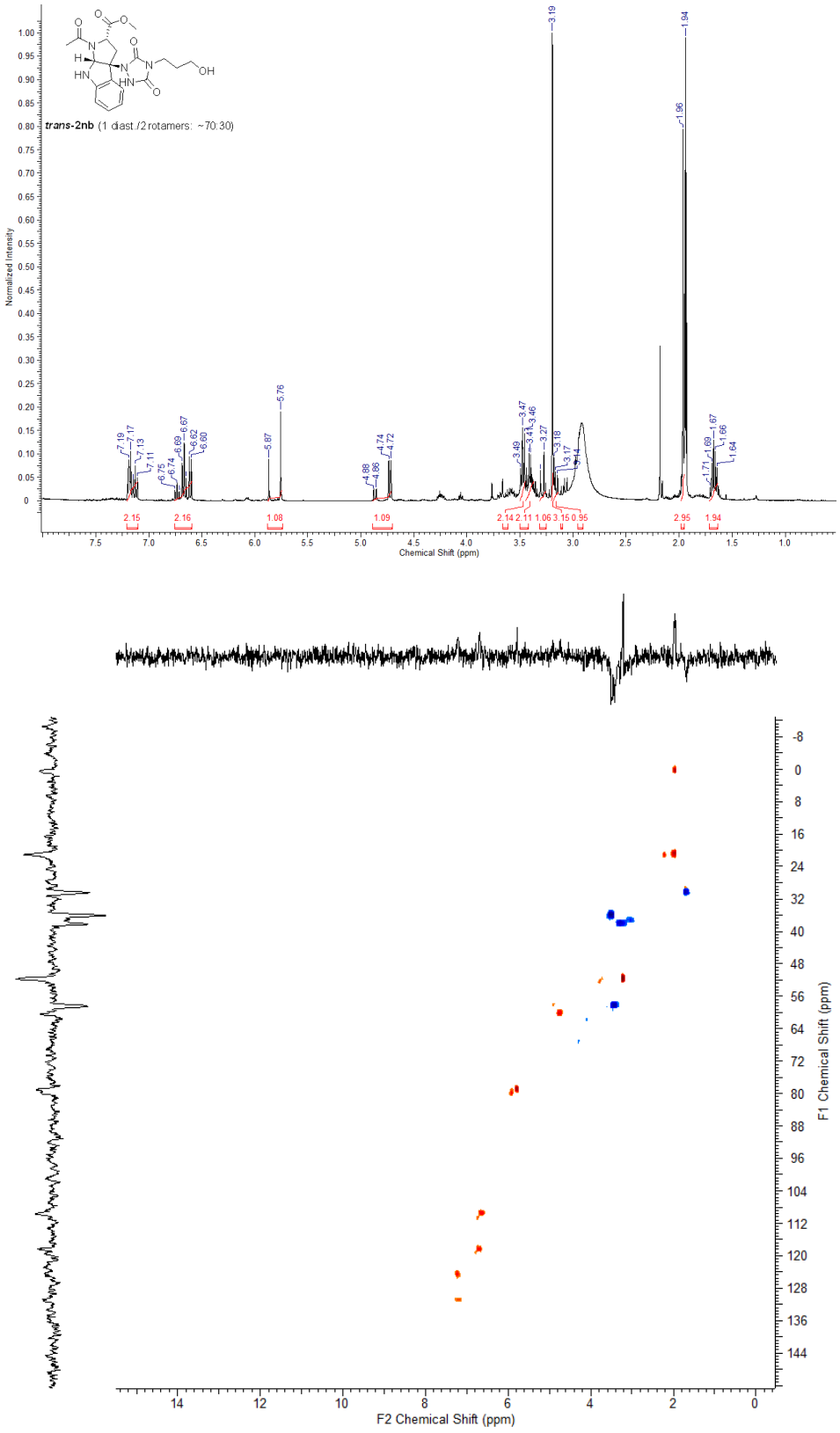
9.3.3.13 NMR spectra of compounds






trans-2mb (1 diast., 2 rotamers: 45:55)
DMSO
ACN
Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)


cis-2mb (1 diast., 2 rotamers: 67:33)
DMSO
ACN
Normalized Intensity
Chemical Shift (ppm)
F1 Chemical Shift (ppm)
F2 Chemical Shift (ppm)


trans-2nb (1 diast./2 rotamers: ~70:30)
Normalized Intensity
Chemical Shift (ppm)
F2 Chemical Shift (ppm)
F1 Chemical Shift (ppm)

10. Nederlandstalige samenvatting

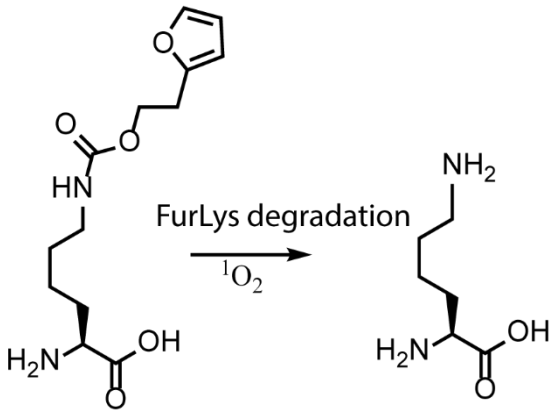
In dit werk hebben we de chemische platformen furan en triazolinedione (TAD) onderzocht voor de ontwikkeling van nieuwe plaats-specifieke methoden voor eiwitmodificatie. Plaats-specifieke eiwitmodificatie is vereist voor een breed scala aan toepassingen met als doel de eiwitfunctie te bestuderen, aan te passen en te controleren. In de loop der jaren is een substantiële verzameling van methodes ontwikkeld voor de plaats-specifieke modificatie van eiwitten; de nood om met betere en selectievere oplossingen te komen blijft echter aanwezig. De solide basis waarop dit werk is gebouwd, wordt geleverd door de furan- en TAD-chemische platformen, die in eerder werk al in verschillende toepassingen zijn bestudeerd. Een belangrijk verschil in hoe beide methodologieën werden toegepast, is gelegen in de locatie van de furan- en TAD-groep. Terwijl de furan-groep zich op het eiwit bevindt, bevindt de TAD-groep zich aan de modificerende kant. Dit verschil is terug te voeren op hun respectievelijke chemische reactiviteiten. Furan kan worden beschouwd als een gemaskeerd elektrofiel: wat betekent dat furan redelijk stabiel is in een biologische context, maar zeer reactief wordt bij oxidatie. Dit biedt de mogelijkheid om een stabiele groep in een eiwit op te nemen en maakt het mogelijk om bij oxidatie een zeer reactieve plaats in het eiwit te creëren. Het modificerende reagens kan van alles zijn, van een klein fluorescerend label tot een volledig ander eiwit. Aan de andere kant zijn TAD's reactieve verbindingen die doorgaans van korte duur zijn in biologische media, maar in staat zijn om snel te reageren met verschillende chemische groepen die van nature aanwezig zijn op eiwitten. Dit maakt TAD-moleculen minder gunstig om in eiwitten te worden opgenomen, aangezien ze waarschijnlijk zullen afbreken of nevenreacties veroorzaken voordat ze bruikbaar zouden zijn voor eiwitmodificatie. Desondanks zijn reagentia op basis van TAD zijn echter zeer geschikt als krachtige eiwitmodificerende verbindingen.

10.1 Furan chemisch platform

10.1.1 Opname van furan in een eiwit via uitbreiding van de genetische code

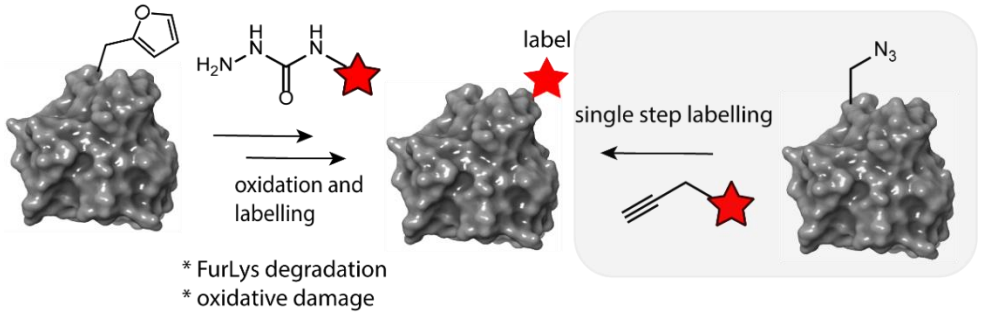
Op weg naar eiwitlabeling

Doelstelling I.1, zoals gedefinieerd in hoofdstuk drie van dit proefschrift, is gericht op de ontwikkeling van een op furan gebaseerde eiwit-labeling methode. Furan-labeling werd eerder gerapporteerd als een krachtige methode voor het labelen van peptiden met hydrazide-gefunctionaliseerde kleurstoffen. Ons doel was om de reikwijdte van deze methodologie uit te breiden naar een nieuwe en meer uitdagende klasse van substraten: eiwitten. Introductie van een furan-gemodificeerd lysine (FurLys) aminozuur via uitbreiding van de genetische code verliep efficiënt zoals gerapporteerd in de literatuur. In onze experimenten ontdekten we echter dat de carbamaat verbinding in FurLys niet stabiel was tijdens de oxidatiestap, wat leidde tot afbraak van FurLys tot lysine (figuur 10.1).



Figuur 10.1 Schematische weergave van de afbraak van FurLys tot lysine bij oxidatie met singlet zuurstof.

Bovendien werd, wanneer de oxidatieprocedure werd toegepast op niet-furan-gemodificeerde eiwitten, aanzienlijke oxidatieve schade waargenomen. Desalniettemin kan de FurLys-afbraak worden vermeden door het stabielere FurAla aminozuur te gebruiken en kunnen geoptimaliseerde oxidatieprocedures worden onderzocht. Wij zijn van mening dat de toegevoegde waarde van op furan-oxidatie gebaseerde labeling ten opzichte van de reeds beschikbare methodes voor eiwitlabeling beperkt zou zijn, aangezien er verschillende aminozuren beschikbaar zijn die bio-orthogonale reactieve delen bevatten (figuur 10.2, rechts). We hebben daarom besloten om de focus te verleggen van labeling naar crosslinking met furan gemodificeerde eiwitten.

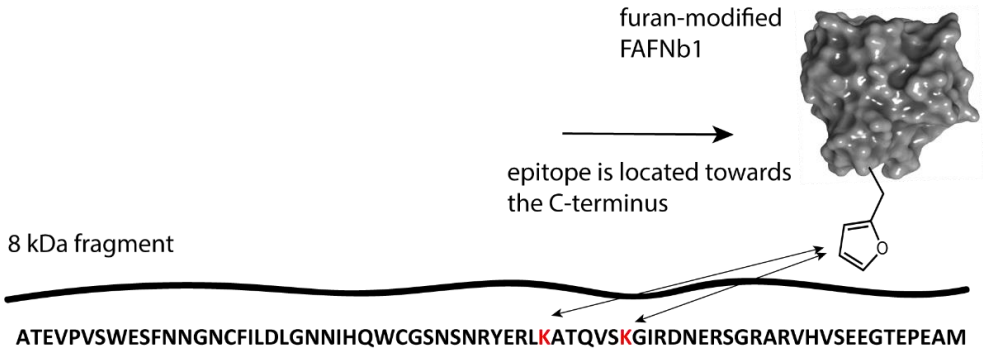


Figuur 10.2 Schematische weergave van de twee-staps benadering voor het labelen van furanlabels (links). Voorbeeld van een typische bio-orthogonale reactie in één stap, CuAAC-chemie (rechts).

Op weg naar eiwitcrosslinking

De gemaskeerde reactiviteit van furan leent zich bijzonder goed voor crosslinking toepassingen. Inderdaad, een nabij nucleofiel, gelokaliseerd op de interactiepartner van een met furan gemodificeerde POI, is in staat om snel te reageren wanneer de furan wordt geoxideerd. Dit maakt de vorming van een covalente binding mogelijk. In dit werk werd furan-chemie gebruikt in twee verschillende nanobody-target interactie modelsystemen, het gelsolin nanobody-model en het EgA1 nanobody-model zoals beschreven in doelstelling I.2. Beide modelsystemen werden geselecteerd met het oog op hun therapeutisch relevante eigenschappen.

De interactie tussen het gelsolin nanobody (FAFNb1) en een 8 kDa gelsoline fragment werd onderzocht als een eerste test. Het doel was om een furan-variant van het FAFNb1-nanobody te ontwikkelen dat in staat is om te crosslinken met het 8 kDa gelsoline fragment en zo de verdere proteolyse van dat fragment te voorkomen (waarvan wordt aangenomen dat het de oorzaak is van de ziektegerelateerde effecten). Het eerste en meest uitdagende obstakel in dit project was de afwezigheid van een kristalstructuur van de nanobody-target interactie. Met behulp van ELISA-experimenten met synthetische peptidefragmenten waren we in staat om de locatie van het nanobody-bindende epitoop verder te localiseren naar de C-terminus van het 8 kDa-peptide (figuur 10.3).

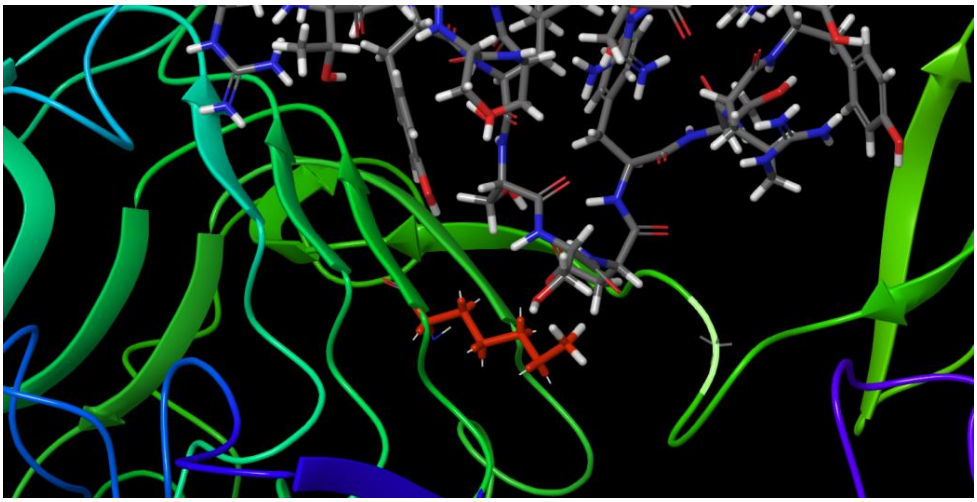


Figuur 10.3 Schematische weergave van de locatie van de met furan-gemodificeerde FAFNb1 ten opzichte van het 8kDa-fragment.

De bevestiging van deze C-terminale locatie van het epitoop geeft helaas ook aan dat de lysines die kunnen worden gebruikt voor crosslinking op het fragment van 8 kDa hoogstwaarschijnlijk te ver verwijderd zijn van het met furan gemodificeerde FAFNb1. Bovendien hebben we peptideaggregatie waargenomen in de gezuiverde peptide stalen. Hoewel we deze aggregatie niet hadden voorzien, is het niet geheel verrassend aangezien proteolyse van het fragment van 8 kDa leidt tot pathogene aggregatie, wat uiteindelijk resulteert in plaquevorming. De combinatie van deze verschijnselen en inzichten leidde tot de beslissing dat onze aandacht beter gericht is op een modelsysteem waar een kristalstructuur aanwezig is en waar een lysine aminozuur zich bevindt, op een geschikte afstand van de furan, op het nanobody-target eiwit.

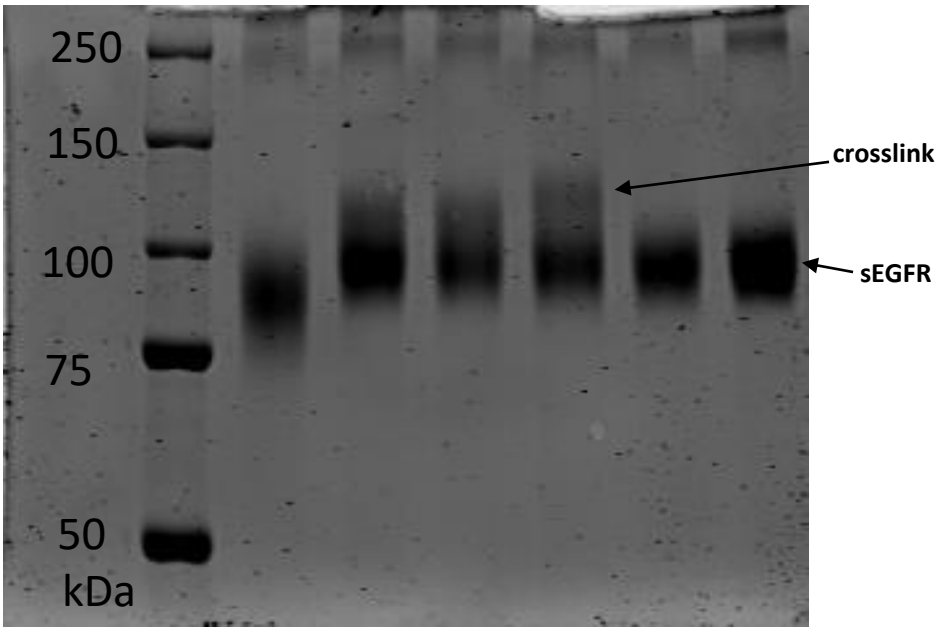
Een voorbeeld van een dergelijke nanobody-target-interactie met een bekende kristalstructuur en een nabijgelegen lysine aminozuur op het targetproteïne is de EgA1-EGFR-interactie. Deze interactie werd geïdentificeerd na een grondige literatuurstudie naar geschikte nanobody-eiwitinteracties op de eiwitdatabank. Het EgA1-nanobody werd uiteindelijk geselecteerd vanwege zijn goed bestudeerde en therapeutisch relevante doelwit. EGFR is een transmembraaneiwit en een lid van een onderfamilie van receptor tyrosinekinasen. Hoewel verminderde EGFR-signalering geassocieerd is met de ziekte van Alzheimer, is overexpressie van EGFR gelinked aan een verscheidenheid aan tumoren. Er werden vijf EgA1-nanobodies geconstrueerd, waarvan er vier een amber codon op een vooraf gedefinieerde positie hadden. De locatie van het amber codon is belangrijk voor de uiteindelijke

positie van het furan-aminozuur. Deze posities zijn geselecteerd op basis van de kristalstructuur (figuur 10.4).



Figuur 10.4 Zoom van de kristalstructuur op de interactieplaats van het EgA1 EGFR-complex. De EGFR wordt weergegeven in een cartoonweergave met het lysine aminozuur (rode stick weergavel). EgA1 wordt afgebeeld in stick-weergave.

De nanobody-varianten werden tot expressie gebracht met een furan gemodificeerd lysine derivaat dat op de amber positie was ingebouwd. *In vitro* crosslinking experimenten in aanwezigheid van het sEGFR-doeleiwit lieten veelbelovende resultaten zien, zoals kan worden afgeleid uit de SDS-PAGE-analyse (figuur 10.5).



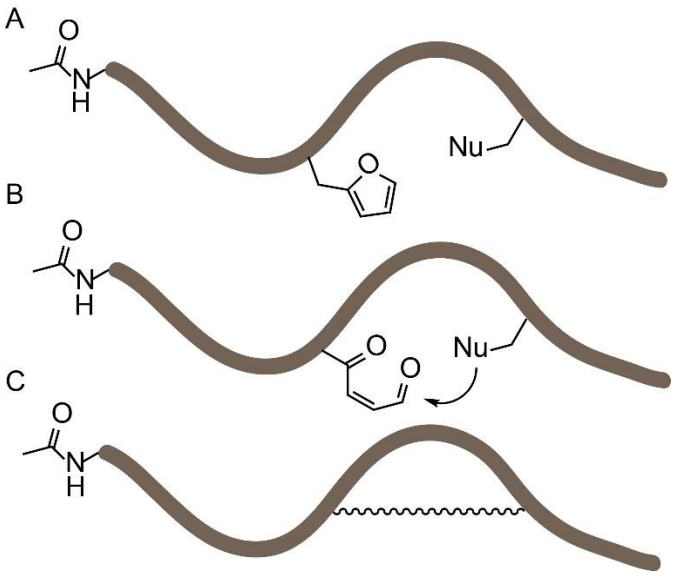
Figuur 10.5 Afbeelding van de SDS-PAGE-gel van *in vitro* equimolaire crosslinking experimenten tussen nanobody-varianten en sEGFR. Foto-oxidatie wordt bereikt door bestraling in aanwezigheid van Rose Bengal. De EDC / Sulfo-NHS-laan is een positieve controle waarbij een combinatie van EDC- en Sulfo-NHS-crosslinkers werd gebruikt.

In de SDS-PAGE-analyses wordt een extra band waargenomen die iets hoger is dan de EGFR-band voor de stalen die furan-gemodificeerde nanobodies bevatten. Aan de andere kant vertoont het monster waarin het wildtype EgA1-nanobody werd geïncubeerd met EGFR deze extra crosslinkband niet. Als positieve controle werd een combinatie van EDC en sulfo-NHS crosslinkers gebruikt, in deze baan zagen we een band lager dan EGFR zelf. Dit wordt waarschijnlijk veroorzaakt door dichte niet-specifieke crosslinking, wat een verbinding oplevert met een schijnbaar lagere molecuulmassa bij analyse via SDS-PAGE. Hoewel het FurLys aminozuur onstabiel bleek te zijn in de oxidatiestap in de labeling experimenten, lijkt het erop dat crosslink vorming met een nabij nucleofiel de afbraak van FurLys vermijdt. In combinatie met eerdere resultaten verkregen door onze groep over de endogene oxidatie van furan gemodificeerde peptiden gevolgd door crosslinking met hun celoppervlak receptor, vormen deze eerste resultaten een solide basis voor toekomstige *in cellulo* experimenten met dergelijke furan-crosslinkende nanobodies. Dit onderzoek wordt momenteel uitgevoerd door Laure Tack in het kader van het doctoraatsonderzoek dat ze in november 2020 startte.

10.1.2 Opname van furan in een peptide via herprogrammering van de genetische code

Om de reactiviteit van natuurlijke nucleofielen ten opzichte van het geoxideerde furan te onderzoeken, zoals beschreven in doelstelling I.3, werd een kleine collectie van peptiden ontworpen. De peptiden werden geproduceerd met behulp van herprogrammering van de genetische code via flexibele *in vitro* translatie (FIT). Om

de nabijheid van twee bindende eiwitten na te bootsen, bevat elk peptide een furan groep en een natuurlijk nucleofiel. Het doel is om de furan te oxideren en het staal te analyseren op mogelijke vorming van een cyclische stuctuur (figuur 10.6).



Figuur 10.6 Schematisch overzicht van de doelstelling van het flexibele *in vitro* translatie screening experiment. Een reeks peptiden werd geproduceerd door flexibele *in vitro* translatie (FIT) (A) die allemaal een furan aminozuur en een nucleofiel residu bevatten. In een tweede fase werd de furan geoxideerd (B) en kunnen nucleofielen mogelijk reageren met de nieuw gevormde keto-enal groep. De peptiden werden geanalyseerd via massaspectrometrie om eventueel gevormde cyclische peptiden (C) te identificeren.

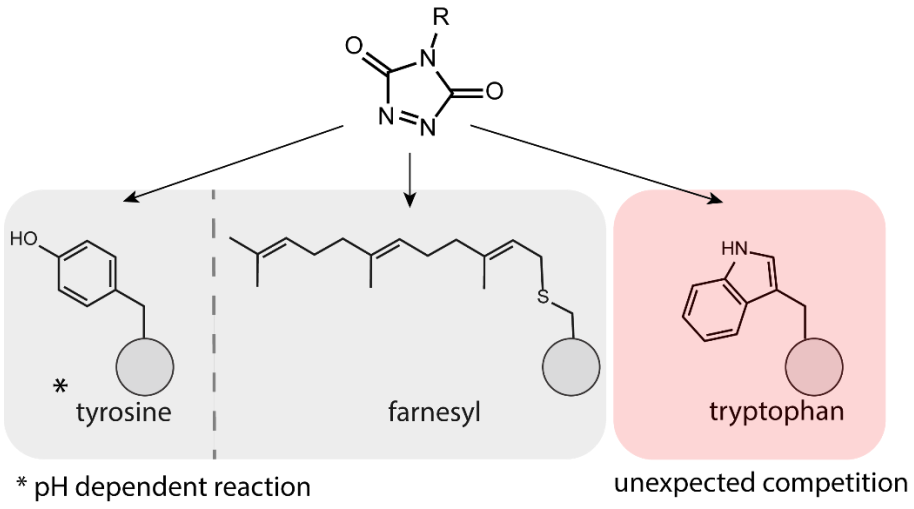
We ontdekten dat lysine in staat is om een cyclisch product te vormen, wat aangeeft dat lysine een geschikt nucleofiel is in de furan-crosslinking methode. Hoewel FIT zeer krachtig is enorm diverse bibliotheken van niet canonische aminozuren bevattende peptiden te genereren, is de hoeveelheid van elk afzonderlijk peptide doorgaans erg laag. Daarom werd SPPS-synthese gebruikt om meer peptidemateriaal te genereren voor LC-MS en structurele analyse. Bij reductie van het gevormde imine, ontdekten we dat een pyrrool wordt gevormd als het verbindende gedeelte. Naast de informatie dat lysine een geschikt nucleofiel is in de furan-crosslinking methode, hebben we gerapporteerd over deze nieuwe peptide cyclisatie methode van zijketen naar zijketen. Bovendien veronderstellen we dat de methode ook kan worden gebruikt in een kop-naar-zijketen-manier, aangezien de N-terminus doorgaans meer nucleofiel is in vergelijking met het amine van een lysine. Hoewel er geen cyclisch product werd waargenomen met andere nucleofielen, zijn er meer experimenten nodig om te bevestigen dat andere nucleofielen niet kunnen reageren met het geoxideerde furan.

De resultaten in deze sectie zijn in overeenstemming met onze gegevens over de EgA1-EGFR-crosslinking experimenten, waarbij nanobody-varianten *in silico* werden ontworpen om het furan aminozuur in de nabijheid van een nabij lysine-aminozuur op het interagerende proteine te lokaliseren.

10.2 TAD chemisch platform

10.2.1 Gericht op de natuurlijke prenylgroep

Doelstelling II.1 betreft de ontwikkeling van TAD-reagentia voor selectieve detectie van de natuurlijke prenylgroep. De reactie van TAD-tyrosine werd gerapporteerd in 2010 en is sindsdien algemeen aanvaard als een krachtige tyrosine-modificatiemethode. Daarom was een van de hoofddoelen van dit project het identificeren van condities die selectieve reactie van TAD met farnesyl in aanwezigheid van tyrosine mogelijk maken. Op basis van TAD-modificatie-experimenten met een tyrosine-bevattend peptide in verschillende buffers, verwachtten we dat het mogelijk zou zijn om selectiviteit te verkrijgen voor farnesyl in gebufferde omstandigheden onder pH 5. We waren in staat om TAD te gebruiken om selectief een gefarnesyleerd peptide te laten reageren met TAD in een competitie-experiment met een tyrosine-bevattend peptide in zuiver water. Bovendien hebben we aangetoond dat alle drie de dubbele bindingen in de farnesylgroep kunnen reageren met TAD, wat leidt tot een drievoudig gemodificeerd peptide. De mogelijkheid om meerdere labels aan een enkele prenylgroep te bevestigen, zou de gevoeligheid van een op TAD gebaseerde detectiemethode kunnen verhogen. Hoewel gerapporteerd werd dat de TAD-tryptofaanreactie traag was in buffer, zagen we verrassend genoeg een competitieve reactie van tryptofaan met TAD-reagentia (figuur 10.7).



Figuur 10.7 Schematische weergave van de natuurlijke substraten, aanwezig op eiwitten die concurreren om reactie met TAD-reagentia.

Vanwege deze tryptofaan zijreactie, zal het niet mogelijk zijn om volledige selectiviteit van TAD-reagentia voor alle 20 canonieke aminozuren te bereiken. We zijn echter van mening dat TAD-reagentia nuttig kunnen zijn voor de wetenschappelijke gemeenschap als detectiemiddelen voor geprenyleerde eiwitten. Bovendien onthulde sequentieanalyse van eiwitten uit de Ras-familie, die vanuit een oncologisch perspectief tot de meest relevante geprenyleerde eiwitten behoren, dat geen van deze eiwitten een tryptofaan aminozuur in hun sequentie heeft. De interesse in het labelen van geprenyleerde eiwitten wordt verder aangetoond door een zeer recente publicatie waarin het gebruik van TAD-reagentia voor het labelen van geprenyleerde eiwitten wordt gerapporteerd. Opmerkelijk genoeg zijn de auteurs op de hoogte van de bekende TAD-tyrosine reactie, maar rapporteren ze niet over de selectiviteit van tyrosine versus farnesyl van hun TAD-

reagens. Bovendien realiseren de auteurs zich niet dat ze in feite ook te maken hebben met concurrentie van tryptofaan (zie hierboven). Door gebufferde condities te gebruiken bij een pH lager dan 5, kunnen we tyrosine concurrentie vermijden en de labeling selectiviteit voor farnesyl aanzienlijk verbeteren.

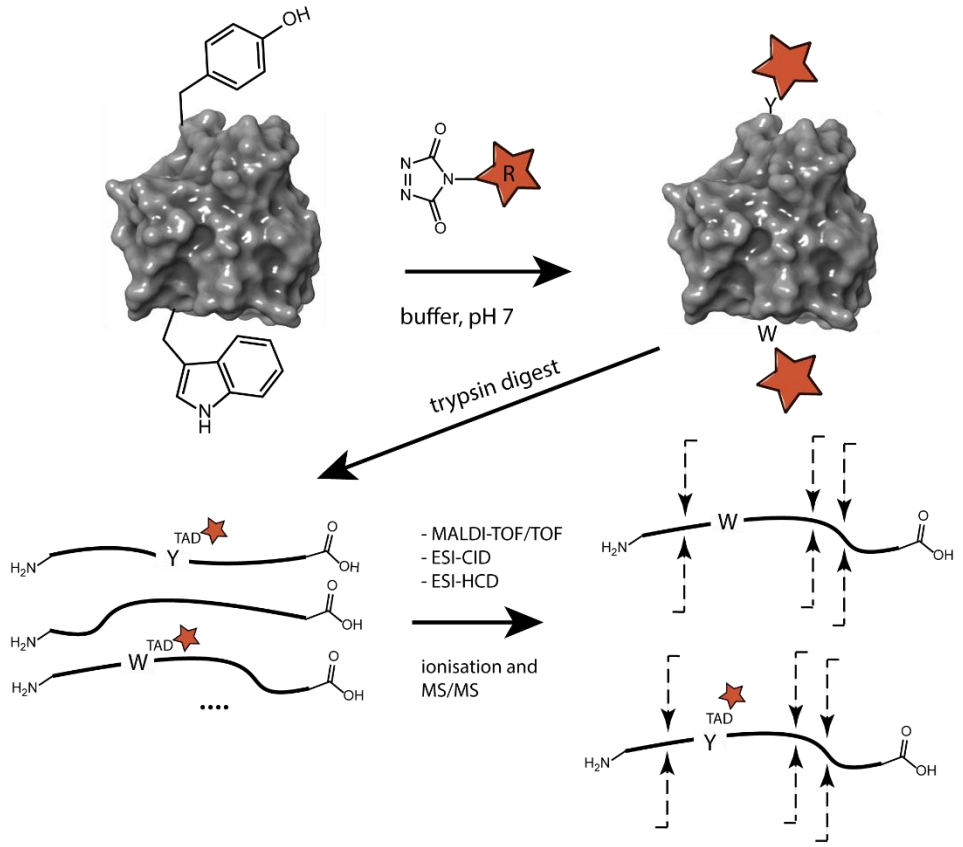
10.2.2 Triazolinedion-eiwitmodificatie: van een over het hoofd gezien off-target effect naar een tryptofaan-selectieve bioconjugatiestrategie

Triazolinediones (TAD's) zijn gebruikt voor eiwitmodificatie sinds de publicatie van de TAD-tyrosine reactie in 2010 door Barbas en collega's. De auteurs melden dat TAD's selectief reageren met tyrosine in een ene-type reactie. Later werd het reactiemechanisme voor de TAD-tyrosine bestudeerd via computationele methoden, wat leidde tot de conclusie dat de reactie feitelijk verloopt langs een elektrofiele aromatische substitutie mechanisme. Bovendien tonen de auteurs aan dat het fenolaat in plaats van het fenol de reactieve eenheid is voor reactie met TAD. Dit verbeterde begrip van de TAD-tyrosine-mechanisme is in overeenstemming met de experimentele waarnemingen dat de reactie zeer efficiënt verloopt in gebufferde omstandigheden bij pH 7. Dezelfde conclusie kan worden getrokken uit de pH-experimenten die we hebben uitgevoerd met een tyrosine-bevattend peptide en TAD-reagentia in sectie 6.2.3 van dit werk. In een eenvoudig competitie-experiment tussen tyrosine, tryptofaan en farnesyl voor TAD in water ontdekten we echter dat de TAD-tryptofaanreactie in feite allesbehalve traag was. Op dit punt bleven twee belangrijke onderzoeksvragen over. De eerste vraag is: hoe kon dit belangrijke off-target effect al meer dan tien jaar over het hoofd worden gezien door zoveel wetenschappers. En de tweede vraag is: als de tyrosine-TAD-reactie wordt geremd bij lage pH, kan er een echte aminozuur-selectieve methode worden ontwikkeld voor TAD-tryptofaan modificatie.

Hoe werd het off-target effect van TAD-reagentia op tryptofaan zo lang over het hoofd gezien?

In hun rapport over de klik-achtige reactie van TAD-tyrosine voerden Barbas en collega's een zeer slechte aminozuurselectiviteitsstudie uit. De auteurs rapporteren eigenlijk over de vorming van een tryptofaan TAD-conjugaat in een afzonderlijk LC-MS-experiment met TAD en N-acyltryptofaan methylamide. Vervolgens werd echter vastgesteld en beweerd dat de TAD-tryptofaanreactie de TAD-tyrosine reactie niet verstoorde in een competitief experiment dat vreemd genoeg alleen via NMR werd geanalyseerd. Na het eerste rapport volgden vele toepassingen en verfijningen van de TAD-tyrosine reactie. Aangezien het oorspronkelijke artikel de TAD-Tyr-benadering beschreef als een aminozuur selectieve methode voor tyrosine labeling, werd de aminozuur selectiviteit in latere werken niet in detail onderzocht. Er zijn twee aanvullende factoren die ertoe hebben bijgedragen dat dit off-target effect onopgemerkt kon blijven. De eerste is dat de TAD-modificatie op tryptofaan labiel is onder de meeste MS / MS-omstandigheden. In onze experimenten zowel op peptiden als op eiwitten ontdekten we inderdaad dat MS / MS-analyse via MALDI-TOF / TOF, ESI-CID en ESI-HCD allemaal leiden tot het verlies van de modificatie op tryptofaan tijdens fragmentatie (figuur 8.8) . We konden echter aantonen dat door gebruik te maken van de zachtere elektronenoverdrachts dissociatie (ETD)-fragmentatie de

modificatie op tryptofaan behouden blijft, waardoor ontegensprekelijk de off-target reactie van TAD-reagentia met Trp-residuen wordt bevestigd.



Figuur 10.8 Schematisch overzicht van een TAD-eiwitmodificatiereactie bij pH 7. Zowel tryptofaan- als tyrosine aminozuren zijn gemodificeerd. Het eiwit wordt gedigereerd met trypsine om de exacte modificatieplaats te bepalen. Vervolgens worden de peptiden geanalyseerd met MS/MS en hoewel de TAD-tyrosine-modificatie stabiel is tijdens MS/MS-fragmentatie, is de TAD-tryptofaanmodificatie dat niet, wat leidt tot een ogenschijnlijk aminozuur selectiviteit voor tyrosine.

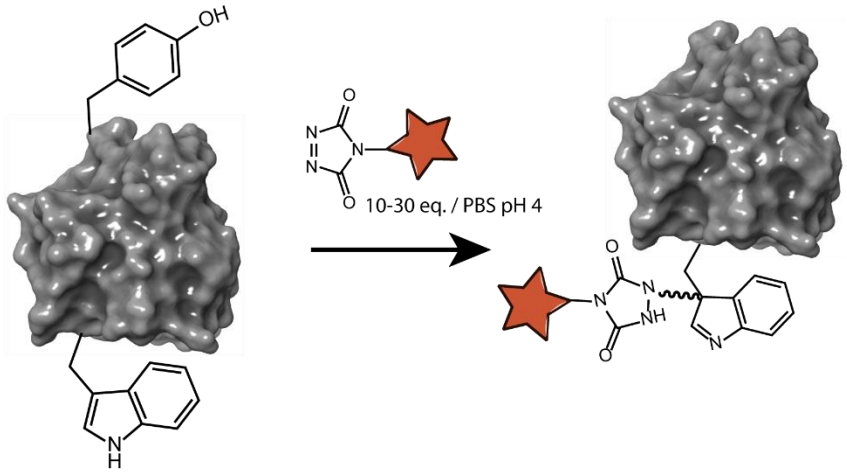
De tweede factor die heeft bijgedragen aan het feit dat het off-target effect onder de radar bleef, is de lage abundantie van tryptofaan. Met een natuurlijke abundantie van iets meer dan 1% is tryptofaan het meest zeldzame van alle 20 canonieke aminozuren.

Gegeven het feit dat alleen tijdens MS-analyse van de resulterende conjugaten de TAD-groep van de tryptofaan afvalt, terwijl het oorspronkelijke eiwit inderdaad die modificatie draagt, kan het off-target effect leiden tot verkeerde interpretaties en conclusies. Overal waar de TAD-tyrosine-modificatie wordt toegepast, van het genereren van eiwitconjugaten in antibody drug conjugates en vaccinonderzoek tot en met fundamenteel werk om de structuur van eiwitten in kaart te brengen.

Hoewel de MS / MS-labiliteit van het conjugaat en de lage abundantie van tryptofaan hebben bijgedragen aan het verbergen van het off-target effect. Eigenlijk is de TAD-tyrosine-reactie simpelweg niet aminozuur selectief, bovendien had de tryptofaan modificatie onmiddellijk moeten waargenomen worden. Naar onze mening moet er meer waarde worden gehecht aan grondig onderzoek van de aminozuur selectiviteit wanneer een nieuwe methode wordt gerapporteerd voor eiwitmodificatie.

Ontwikkeling van een op TAD gebaseerde selectieve modificatiemethode voor tryptofaan aminozuren

Door de pH-waarde onder de 5 te brengen, konden we TAD-reagentia gebruiken om tryptofaan selectief te modificeren in aanwezigheid van tyrosine aminozuren. In deze studie werden verschillende TAD-reagentia gebruikt. Deze nieuwe selectieve modificatiestrategie voor tryptofaan werd gedemonstreerd op verschillende peptiden en recombinante eiwitten (figuur 10.9).



Figuur 10.9 Schematisch overzicht van selectieve tryptofaanmodificatie met TAD-chemie in PBS-buffer bij pH 4.

Deze nieuwe methode voor tryptofaanmodificatie is gunstig in vergelijking met de huidige methoden voor selectieve tryptofaan modificatie. Francis en collega's rapporteerden een methode voor selectieve tryptofaanmodificatie in peptiden en eiwitten. De methode vind plaats bij pH 6-7 en levert in het algemeen 40-60% rendement op bij gebruik van 100 eq. van een labelmolecule. Het nadeel van deze methode is dat er een transitiemetaal nodig is (Rhodium) en dat de reactie lange reactietijden vereist. Kanai en collega's rapporteerden een transitiemetaalvrije tryptofaan conjugatie methode waarbij organoradicalen werden gebruikt om selectief te conjugeren met tryptofaan aminozuren in peptiden en eiwitten. De reactie verloopt echter onder zure omstandigheden (azijnzuur in water, pH 1-2) en de reactie wordt beëindigd door toevoeging van buffer. De auteurs hebben eiwitmodificaties aangetoond met gefunctionaliseerde verbindingen met een opbrengst van ongeveer 50%. Hoewel de reactie plaatsvindt op een transitiemetaalvrije manier, vereist het zure omstandigheden en is het niet compatibel met buffers. Onlangs rapporteerden Taylor en collega's een fotochemisch tryptofaan modificatie proces. De auteurs tonen modificaties van verschillende peptiden en een proteine: lysozym. Het nadeel van deze aanpak is dat het gebruik van UV-licht en de toevoeging van glutathion vereist en moet worden uitgevoerd in afwezigheid van zuurstof. Bovendien tonen de auteurs modificatie van korte peptiden met gefunctionaliseerde (biotine, alkyn) N-carbamoylpyridiniumzouten. Cruciaal was echter dat de modificatie van langere peptiden en een eiwit (lysozym) alleen werd gerapporteerd met een niet-

functioneel N-carbamoylpyridiniumzout. Onze op TAD gebaseerde methode voor selectieve tryptofaanmodificatie kan selectief tryptofaan in eiwitten modificeren onder gebufferde omstandigheden bij pH 4. We hebben een modificatie van meer dan 50% aangetoond met gebruik van 20-30 equivalenten fluorescerend TAD-reagens. Onze aanpak vereist geen gebruik van transitiemetalen of andere additieven, het vereist geen afwezigheid van zuurstof of UV-straling. De aanpak die we presenteren, vindt in enkele seconden plaats in gebufferd medium. Bovendien is de fluorescent gemodificeerde TAD (DMEQ-TAD) gemakkelijk commercieel verkrijgbaar. We verwachten dat onze nieuwe aanpak niet de beste zal zijn voor alle eiwitten vanwege het feit dat mild zure omstandigheden (gebufferde pH 4) vereist zijn. Maar voor de eiwitten die een kort verblijf in gebufferde omstandigheden bij pH 4 aankunnen, is onze methode verreweg het meest toepasbaar.

11. Bibliography

[1] E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Rev. Chem.* **2019**, *3*, 147–171.

[2] N. Krall, F. P. Da Cruz, O. Boutureira, G. J. L. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103–113.

[3] P. Agarwal, C. R. Bertozzi, *Bioconjug. Chem.* **2015**, *26*, 176–192.

[4] O. SHIMOMURA, F. H. JOHNSON, Y. SAIGA, *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239.

[5] R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

[6] N. C. Shaner, P. A. Steinbach, R. Y. Tsien, *Nat. Methods* **2005**, *2*, 905–909.

[7] S. Brunet, T. Sardon, T. Zimmerman, T. Wittmann, R. Pepperkok, E. Karsenti, I. Vernos, *Mol Biol Cell* **2004**, *15*, 5318–5328.

[8] E. M. Sletten, C. R. Bertozzi, *Angew. Chemie - Int. Ed.* **2009**, *48*, 6974–6998.

[9] Y. Zhang, K. Y. Park, K. F. Suazo, M. D. Distefano, *Chem. Soc. Rev.* **2018**, *47*, 9106–9136.

[10] T. Dierks, B. Schmidt, L. V. Borissenko, J. Peng, A. Preusser, M. Mariappan, K. Von Figura, *Cell* **2003**, *113*, 435–444.

[11] I. S. Carrico, B. L. Carlson, C. R. Bertozzi, *Nat. Chem. Biol.* **2007**, DOI 10.1038/nchembio878.

[12] J. S. Rush, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241.

[13] P. Wu, W. Shui, B. L. Carlson, N. Hu, D. Rabuka, J. Lee, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3000–5.

[14] S. K. Mazmanian, G. Liu, H. Ton-that, O. Schneewind, *Science (80-.).* **1999**, *285*, 760–763.

[15] J. M. Monteiro, G. Covas, D. Rausch, S. R. Filipe, T. Schneider, H. G. Sahl, M. G. Pinho, *Sci. Rep.* **2019**, *9*, 1–10.

[16] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, DOI 10.1038/nchembio.2007.31.

[17] T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, *ChemBioChem* **2008**, *9*, 802–807.

[18] S. Tsukiji, T. Nagamune, *ChemBioChem* **2009**, *10*, 787–798.

[19] K. Strijbis, E. Spooner, H. L. Ploegh, *Traffic* **2012**, *13*, 780–789.

[20] A. Sinisi, M. W. L. Popp, J. M. Antos, W. Pansegrau, S. Savino, M. Nissum, R. Rappuoli, H. L. Ploegh, L. Buti, *Bioconjug. Chem.* **2012**, DOI 10.1021/bc200577u.

[21] M. D. Witte, J. J. Cragnolini, S. K. Dougan, N. C. Yoder, M. W. Popp, H. L. Ploegh, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, DOI 10.1073/pnas.1205427109.

[22] K. T. D. P. E. Christopoulos, *Clin. Chem.* **1991**, *37*, 625–636.

[23] A. Chapman-Smith, J. E. Cronan, *Biomol. Eng.* **1999**, *16*, 119–125.

[24] M. Howarth, K. Takao, Y. Hayashi, A. Y. Ting, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7583–7588.

[25] I. Chen, M. Howarth, W. Lin, A. Y. Ting, *Nat. Methods* **2005**, *2*, 99–104.

[26] F. L. Zhang, P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241–269.

[27] M. Wang, P. J. Casey, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 110–122.

[28] N. E. Kohl, S. D. Mosser, S. J. DeSolms, E. A. Giuliani, D. L. Pompliano, S. L. Graham, R. L. Smith, E. M. Scolnick, A. Oliff, J. B. Gibbs, *Science (80-.).* **1993**, *260*, 1934–1937.

[29] J. B. Gibbs, D. L. Pompliano, S. D. Mosser, E. Rands, R. B. Lingham, S. B. Singh, E. M. Scolnick, N. E. Kohl, A. Oliff, *J. Biol. Chem.* **1993**, *268*, 7617–7620.

[30] N. Berndt, A. D. Hamilton, S. M. Sebti, *Nat. Rev. Cancer* **2011**, *11*, 775–791.

[31] Y. Kho, S. C. Kim, C. Jiang, D. Barma, S. W. Kwon, J. Cheng, J. Jaunbergs, C. Weinbaum, F. Tamanoi, J. Falck, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12479–12484.

[32] B. P. Duckworth, J. Xu, T. A. Taton, A. Guo, M. D. Distefano, *Bioconjug. Chem.* **2006**, 967–974.

[33] B. P. Duckworth, Z. Zhang, A. Hosokawa, M. D. Distefano, *ChemBioChem* **2007**, *8*, 98–105.

[34] M. Rashidian, J. M. Song, R. E. Pricer, M. D. Distefano, *J. Am. Chem. Soc.* **2012**, *134*, 8455–8467.

[35] M. Rashidian, J. K. Dozier, S. Lenevich, M. D. Distefano, *Chem. Commun.* **2010**, DOI 10.1039/c0cc03305g.

[36] D. Schumacher, J. Helma, F. A. Mann, G. Pichler, F. Natale, E. Krause, M. C. Cardoso, C. P. R. Hackenberger, H. Leonhardt, *Angew. Chemie - Int. Ed.* **2015**, *54*, 13787–13791.

[37] D. Schumacher, O. Lemke, J. Helma, L. Gerszonowicz, V. Waller, T. Stoschek, P. M. Durkin, N. Budisa, H. Leonhardt, B. G. Keller, et al., *Chem. Sci.* **2017**, *8*, 3471–3478.

[38] A. Banerjee, T. D. Panosian, K. Mukherjee, R. Ravindra, S. Gal, D. L. Sackett, S. Bane, *ACS Chem. Biol.* **2010**, *5*, 777–785.

[39] M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi, A. Y. Ting, *Nat. Biotechnol.* **2007**, *25*, 1483–1487.

[40] J. Yin, F. Liu, X. Li, C. T. Walsh, *J. Am. Chem. Soc.* **2004**, *126*, 7754–7755.

[41] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15815–15820.

[42] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86–89.

[43] A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9955–9959.

[44] A. Gautier, A. Juillerat, C. Heinis, I. R. Corrêa, M. Kindermann, F. Beaufils, K. Johnsson, *Chem. Biol.* **2008**, *15*, 128–136.

[45] B. Zakeri, M. Howarth, *J. Am. Chem. Soc.* **2010**, *132*, 4526–4527.

[46] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy, M. Howarth, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, DOI 10.1073/pnas.1115485109.

[47] X. Gao, J. Fang, B. Xue, L. Fu, H. Li, *Biomacromolecules* **2016**, *17*, 2812–2819.

[48] "https://www.uniprot.org/statistics/Swiss-Prot," **n.d.**

[49] S. B. Gunnoo, A. Madder, *ChemBioChem* **2016**, *17*, 529–553.

[50] G. Bulaj, T. Kortemme, D. P. Goldenberg, *Biochemistry* **1998**, *37*, 8965–8972.

[51] B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science (80-.).* **1998**, *281*, 269–272.

[52] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis, R. Y. Tsien, *J. Am. Chem. Soc.* **2002**, *124*, 6063–6076.

[53] T. L. Halo, J. Appelbaum, E. M. Hobert, D. M. Balkin, A. Schepartz, *J. Am. Chem. Soc.* **2009**, *131*, 438–439.

[54] C. Zhang, M. Welborn, T. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis, B. L. Pentelute, *Nat. Chem.* **2016**, *8*, 120–128.

[55] C. Zhang, P. Dai, A. A. Vinogradov, Z. P. Gates, B. L. Pentelute, *Angew. Chemie - Int. Ed.* **2018**, *57*, 6459–6463.

[56] M. C. Martos-Maldonado, C. T. Hjuler, K. K. Sørensen, M. B. Thygesen, J. E. Rasmussen, K. Villadsen, S. R. Midtgaard, S. Kol, S. Schoffelen, K. J. Jensen, *Nat. Commun.* **2018**, *9*, DOI 10.1038/s41467-018-05695-3.

[57] G. Srinivasan, C. M. James, J. a Krzycki, *Science* **2002**, *296*, 1459–1462.

[58] A. Ambrogelly, S. Gundllapalli, S. Herring, C. Polycarpo, C. Frauer, D. Sö, *Pyrrolysine Is Not Hardwired for Cotranslational Insertion at UAG Codons*, **2007**.

[59] L. Wang, A. Brock, H. Brad, P. G. Schultz, *Science (80-.).* **2001**, *292*, 498–500.

[60] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, **2001**, *254*.

[61] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441–444.

[62] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

[63] D. D. Young, P. G. Schultz, *ACS Chem. Biol.* **2018**, *13*, 854–870.

[64] L. Davis, J. W. Chin, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168–182.

[65] Y. Shimizu, T. Kanamori, T. Ueda, *Methods* **2005**, *36*, 299–304.

[66] Y. Shimizu, T. Ueda, *Methods Mol. Biol.* **2010**, DOI 10.1007/978-1-60327-331-2_2.

[67] H. Saito, D. Kourouklis, H. Suga, **2001**, *20*, 1797–1806.

[68] H. Xiao, H. Murakami, H. Suga, A. R. Ferré-D'Amaré, *Nature* **2008**, *454*, 358–361.

[69] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**, *6*, 779–790.

[70] M. Engelhard, *J. Pept. Sci.* **2016**, *22*, 246–251.

[71] P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Sci. (80-.).* **1994**, *266*, 776–779.

[72] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

[73] J. D. Warren, J. S. Miller, S. J. Keding, S. J. Danishefsky, *J. Am. Chem. Soc.* **2004**, *126*, 6576–6578.

[74] Q. Wan, S. J. Danishefsky, *Angew. Chemie - Int. Ed.* **2007**, *46*, 9248–9252.

[75] B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2000**, *2*, 1939–1941.

[76] Y. Zhang, C. Xu, H. Y. Lam, C. L. Lee, X. Li, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6657–6662.

[77] C. He, S. S. Kulkarni, F. Thuaud, J. W. Bode, *Angew. Chemie - Int. Ed.* **2015**, *54*, 12996–13001.

[78] T. C. Tsao, K. Bailey, *Biochim. Biophys. Acta* **1953**, *11*.

[79] J. E. Moore, W. H. Ward, *J. Am. Chem. Soc.* **1956**, *78*, 2414–2418.

[80] M. R. Lewis, J. E. Shively, *Bioconjug. Chem.* **1998**, *9*, 72–86.

[81] A. D. Baldwin, K. L. Kiick, *Bioconjug. Chem.* **2011**, *22*, 1946–1953.

[82] R. P. Lyon, J. R. Setter, T. D. Bovee, S. O. Doronina, J. H. Hunter, M. E. Anderson, C. L. Balasubramanian, S. M. Duniho, C. I. Leiske, F. Li, et al., *Nat. Biotechnol.* **2014**, *32*, 1059–1062.

[83] J. R. Junutula, K. M. Flagella, R. A. Graham, K. L. Parsons, E. Ha, H. Raab, S. Bhakta, T. Nguyen, D. L. Dugger, G. Li, et al., *Clin. Cancer Res.* **2010**, *16*, 4769–4778.

[84] B. Bernardim, M. J. Matos, X. Ferhati, I. Compañón, A. Guerreiro, P. Akkapeddi, A. C. B. Burtoloso, G. Jiménez-Osés, F. Corzana, G. J. L. Bernardes, *Nat. Protoc.* **2019**, *14*, 86–99.

[85] Y. Zhang, X. Zhou, Y. Xie, M. M. Greenberg, Z. Xi, C. Zhou, *J. Am. Chem. Soc.* **2017**, *139*, 6146–6151.

[86] E. De Geyter, E. Antonatou, D. Kalaitzakis, S. Smolen, A. Iyer, L. Tack, E. Ongenae, G. Vassilikogiannakis, A. Madder, *Chem. Sci.* **2021**, DOI 10.1039/d0sc05881e.

[87] D. Kalia, P. V. Malekar, M. Parthasarathy, *Angew. Chemie - Int. Ed.* **2016**, *55*, 1432–1435.

[88] M. A. Kasper, M. Glanz, A. Stengl, M. Penkert, S. Klenk, T. Sauer, D. Schumacher, J. Helma, E. Krause, M. C. Cardoso, et al., *Angew. Chemie - Int. Ed.* **2019**, *58*, 11625–11630.

[89] C. Mayer, D. G. Gillingham, T. R. Ward, D. Hilvert, *Chem. Commun.* **2011**, *47*, 12068–12070.

[90] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem. Sci.* **2011**, *2*, 1666–1676.

[91] J. J. Hill, T. L. Tremblay, C. R. Corbeil, E. O. Purisima, T. Sulea, *Sci. Rep.* **2018**, *8*, 1–14.

[92] C. M. Walko, H. J. West, *JAMA Oncol.* **2019**, *5*, 1648.

[93] J. M. McFarland, M. B. Francis, *J. Am. Chem. Soc.* **2005**, *127*, 13490–13491.

[94] J. M. Antos, J. M. McFarland, A. T. Iavarone, M. B. Francis, *J. Am. Chem. Soc.* **2009**, *131*, 6301–6308.

[95] Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, M. Kanai, *J. Am. Chem. Soc.* **2016**, *138*, 10798–10801.

[96] S. J. Tower, W. J. Hetcher, T. E. Myers, N. J. Kuehl, M. T. Taylor, *J. Am. Chem. Soc.* **2020**, *142*, 9112–9118.

[97] N. S. Joshi, L. R. Whitaker, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 15942–15943.

[98] J. M. McFarland, N. S. Joshi, M. B. Francis, *J. Am. Chem. Soc.* **2008**, *130*, 7639–7644.

[99] T. Kodadek, I. Duroux-Richard, J. C. Bonnafous, *Trends Pharmacol. Sci.* **2005**, *26*, 210–217.

[100] S. Sato, K. Nakamura, H. Nakamura, *ACS Chem. Biol.* **2015**, *10*, 2633–2640.

[101] H. Ban, J. Gavrilyuk, C. F. Barbas, *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.

[102] H. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma, C. F. Barbas, *Bioconjug. Chem.* **2013**, *24*, 520–532.

[103] D. Alvarez-Dorta, C. Thobie-Gautier, M. Croyal, M. Bouzelha, M. Mével, D. Deniaud, M. Boujtita, S. G. Gouin, *J. Am. Chem. Soc.* **2018**, *140*, 17120–17126.

[104] A. Gutteridge, J. M. Thornton, *Trends Biochem. Sci.* **2005**, *30*, 622–629.

[105] S. Jia, D. He, C. J. Chang, *J. Am. Chem. Soc.* **2019**, *141*, 7294–7301.

[106] P. N. Joshi, V. Rai, *Chem. Commun.* **2019**, *55*, 1100–1103.

[107] S. Bloom, C. Liu, D. K. Kölmel, J. X. Qiao, Y. Zhang, M. A. Poss, W. R. Ewing, D. W. C. Macmillan, *Nat. Chem.* **2018**, *10*, 205–211.

[108] Z. Zuo, D. W. C. Macmillan, *J. Am. Chem. Soc.* **2014**, *136*, 5257–5260.

[109] G. R. Delpierre, J. S. Fruton, *Proc. Natl. Acad. Sci.* **1966**, *56*, 1817–1822.

[110] N. A. McGrath, K. A. Andersen, A. K. F. Davis, J. E. Lomax, R. T. Raines, *Chem. Sci.* **2015**, *6*, 752–755.

[111] E. Jacob, R. Unger, *Bioinformatics* **2007**, *23*, 225–230.

[112] J. K. Dozier, M. D. Distefano, *Int. J. Mol. Sci.* **2015**, DOI 10.3390/ijms161025831.

[113] S. Schoffelen, M. B. van Eldijk, B. Rooijakkers, R. Raijmakers, A. J. R. Heck, J. C. M. van Hest, *Chem. Sci.* **2011**, *2*, 701–705.

[114] A. O. Y. Chan, C. M. Ho, H. C. Chong, Y. C. Leung, J. S. Huang, M. K. Wong, C. M. Che, *J. Am. Chem. Soc.* **2012**, *134*, 2589–2598.

[115] H. Faustino, M. J. S. A. Silva, L. F. Veiros, G. J. L. Bernardes, P. M. P. Gois, *Chem. Sci.* **2016**, *7*, 5052–5058.

[116] H. Ren, F. Xiao, K. Zhan, Y. P. Kim, H. Xie, Z. Xia, J. Rao, *Angew. Chemie - Int. Ed.* **2009**, *48*, 9658–9662.

[117] X. Li, L. Zhang, S. E. Hall, J. P. Tam, *Tetrahedron Lett.* **2000**, DOI 10.1016/S0040-4039(00)00592-X.

[118] T. Tamura, I. Hamachi, *J. Am. Chem. Soc.* **2019**, DOI 10.1021/jacs.8b11747.

[119] N. Stephanopoulos, M. B. Francis, *Nat. Chem. Biol.* **2011**, *7*, 876–884.

[120] D. K. Dalvie, A. S. Kalgutkar, S. C. Khojasteh-Bakht, R. S. Obach, J. P. O'Donnell, *Chem. Res. Toxicol.* **2002**, *15*, 269–299.

[121] L. A. Peterson, *Chem. Res. Toxicol.* **2013**, *26*, 6–25.

[122] D. Lu, M. M. Sullivan, M. B. Phillips, L. A. Peterson, *Chem. Res. Toxicol.* **2009**, *22*, 997–1007.

[123] S. Halila, T. Velasco, P. De Clercq, A. Madder, *Chem. Commun.* **2005**, 936–938.

[124] K. Stevens, A. Madder, *Nucleic Acids Res.* **2009**, *37*, 1555–1565.

[125] M. O. De Beeck, A. Madder, *J. Am. Chem. Soc.* **2011**, *133*, 796–807.

[126] M. Op De Beeck, A. Madder, *J. Am. Chem. Soc.* **2012**, *134*, 10737–10740.

[127] L. L. G. Carrette, T. Morii, A. Madder, *Bioconjug. Chem.* **2013**, *24*, 2008–2014.

[128] L. L. G. Carrette, E. Gyssels, J. Loncke, A. Madder, *Org. Biomol. Chem.* **2014**, *12*, 931–935.

[129] E. Gyssels, L. L. G. Carrette Dr, E. Vercruysse, K. Stevens Dr, A. Madder, *ChemBioChem* **2015**, *16*, 651–658.

[130] A. Manicardi, E. Gyssels, R. Corradini, A. Madder, *Chem. Commun.* **2016**, *52*, 6930–6933.

[131] L. L. G. Carrette, E. Gyssels, N. De Laet, A. Madder, *Chem. Commun.* **2016**, *52*, 1539–1554.

[132] K. Hoogewijs, A. Deceuninck, A. Madder, *Org. Biomol. Chem.* **2012**, *10*, 3999–4002.

[133] A. Deceuninck, A. Madder, *Chem. Commun.* **2009**, 340–342.

[134] E. Antonatou, K. Hoogewijs, D. Kalaitzakis, A. Baudot, G. Vassilikogiannakis, A. Madder, *Chem. - A Eur. J.* **2016**, *22*, 1–5.

[135] E. Antonatou, Y. Verleysen, A. Madder, *Org. Biomol. Chem.* **2017**, *15*, 8140–8144.

[136] W. Vannecke, C. Ampe, M. Van Troys, M. Beltramo, A. Madder, *ACS Chem. Biol.* **2017**, *12*, 2191–2200.

[137] B. C. Dickinson, C. J. Chang, *Nat. Chem. Biol.* **2011**, *7*, 504–511.

[138] M. Van Troys, W. Vannecke, C. Ampe, A. Madder, in *Methods Mol. Biol.*, **2019**, pp. 81–102.

[139] M. G. Brevnov, O. M. Gritsenko, S. N. Mikhailov, E. V. Efimtseva, B. S. Ermolinsky, A. Van Aerschot, P. Herdewijn, A. V. Repyk, E. S. Gromova, *Nucleic Acids Res.* **1997**, *25*, 3302–3309.

[140] C. Allmang, A. Krol, *Biochimie* **2006**, DOI 10.1016/j.biochi.2006.04.015.

[141] D. Hatfield, I. S. Choi, S. Mischke, L. D. Owens, *Biochem Biophys Res Commun* **1992**.

[142] L.-H. Fu, X.-F. Wang, Y. Eyal, Y.-M. She, L. J. Donald, K. G. Standing, G. Ben-Hayyim, *J. Biol. Chem.* **2002**.

[143] G. V. Kryukov, V. N. Gladyshev, *EMBO Rep.* **2004**, DOI 10.1038/sj.embor.7400126.

[144] G. V. Kryukov, S. Castellano, S. V. Novoselov, A. V. Lobanov, O. Zehtab, R. Guigó, V. N. Gladyshev, *Science (80-.).* **2003**, DOI 10.1126/science.1083516.

[145] J. T. Kaiser, K. Gromadski, M. Rother, H. Engelhardt, M. V. Rodnina, M. C. Wahl, *Biochemistry* **2005**, DOI 10.1021/bi051110r.

[146] X. Q. Wu, H. J. Gross, *Nucleic Acids Res.* **1994**, DOI 10.1093/nar/22.3.550-a.

[147] F. Quitterer, A. List, P. Beck, A. Bacher, M. Groll, *J. Mol. Biol.* **2012**, *424*, 270–282.

[148] C. Polycarpo, A. Ambrogelly, A. Bérubé, S. A. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Söll, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, DOI 10.1073/pnas.0405362101.

[149] M. A. Gaston, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Nature* **2011**, *471*, 647–650.

[150] J. A. Krzycki, *Curr. Opin. Chem. Biol.* **2013**, *17*, 619–625.

[151] I. U. Heinemann, A. J. Rovner, H. R. Aerni, S. Rogulina, L. Cheng, W. Olds, J. T. Fischer, D. Söll, F. J. Isaacs, J. Rinehart, *FEBS Lett.* **2012**, *586*, 3716–3722.

[152] S. M. Hancock, R. Uprety, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, DOI 10.1021/ja104609m.

[153] I. Sungwienwong, Z. M. Hostetler, R. J. Blizzard, J. J. Porter, C. M. Driggers, L. Z. Mbengi, J. A. Villegas, L. C. Speight, J. G. Saven, J. J. Perona, et al., *Org. Biomol. Chem.* **2017**, *15*, 3603–3610.

[154] L. Davis, J. W. Chin, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168–182.

[155] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science (80-.).* **2003**, *301*, 964–967.

[156] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, DOI 10.1073/pnas.172226299.

[157] D. Hanahan, *Studies on Transformation of Escherichia Coli with Plasmids*, **1983**.

[158] R. Brabham, M. A. Fascione, *ChemBioChem* **2017**, *18*, 1973–1983.

[159] J. W. Chin, *Nature* **2017**, *550*, 53–60.

[160] J. K. Takimoto, Z. Xiang, J. Y. Kang, L. Wang, *ChemBioChem* **2010**, *11*, 2268–2272.

[161] R. Huisgen, *Angew. Chem.* **1963**.

[162] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chemie - Int. Ed.* **2002**, *41*, 2596–2599.

[163] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chemie - Int. Ed.* **2001**, *40*, 2004–2021.

[164] F. Wolbers, P. ter Braak, S. Le Gac, R. Luttge, H. Andersson, I. Vermes, A. van den Berg, *Electrophoresis* **2006**, DOI 10.1002/elps.200600203.

[165] G. Wittig, R. Pohlke, *Chem. Ber.* **1961**, DOI 10.1002/cber.19610941214.

[166] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

[167] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.

[168] J. W. Wollack, B. J. Monson, J. K. Dozier, J. J. Dalluge, K. Poss, S. A. Hilderbrand, M. D. Distefano, *Chem. Biol. Drug Des.* **2014**, *84*, 140–147.

[169] J. C. T. Carlson, L. G. Meimetis, S. A. Hilderbrand, R. Weissleder, *Angew. Chemie - Int. Ed.* **2013**, *52*, 6917–6920.

[170] G. Beliu, A. J. Kurz, A. C. Kuhlemann, L. Behringer-Pliess, M. Meub, N. Wolf, J. Seibel, Z. D. Shi, M. Schnermann, J. B. Grimm, et al., *Commun. Biol.* **2019**, *2*, DOI 10.1038/s42003-019-0518-z.

[171] N. K. Devaraj, R. Weissleder, S. A. Hilderbrand, *Bioconjug. Chem.* **2008**, *19*, 2297–2299.

[172] E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chemie - Int. Ed.* **2012**, *51*, 4466–4469.

[173] A. Tuley, Y. J. Lee, B. Wu, Z. U. Wang, W. R. Liu, *Chem. Commun.* **2014**, *50*, 7424–7426.

[174] J. Luo, R. Uprety, Y. Naro, C. Chou, D. P. Nguyen, J. W. Chin, A. Deiters, *J. Am. Chem. Soc.* **2014**, *136*, 15551–15558.

[175] E. L. Vodovozova, *Biochem.* **2007**, *72*, 1–20.

[176] L. Dubinsky, B. P. Krom, M. M. Meijler, *Bioorganic Med. Chem.* **2012**, DOI 10.1016/j.bmc.2011.06.066.

[177] M. Hashimoto, Y. Hatanaka, *European J. Org. Chem.* **2008**, DOI 10.1002/ejoc.200701069.

[178] I. S. Farrell, R. Toroney, J. L. Hazen, R. A. Mehl, J. W. Chin, *Photo-Cross-Linking Interacting Proteins with a Genetically Encoded Benzophenone*, **2005**.

[179] T. K. Lindhorst, M. Märten, A. Fuchs, S. D. Knight, *Beilstein J. Org. Chem.* **2010**, *6*, 810–822.

[180] C. Chou, R. Uprety, L. Davis, J. W. Chin, A. Deiters, *Chem. Sci.* **2011**, *2*, 480–483.

[181] M. J. Schmidt, D. Summerer, *Angew. Chemie - Int. Ed.* **2013**, *52*, 4690–4693.

[182] M. J. Schmidt, A. Weber, M. Pott, W. Welte, D. Summerer, *ChemBioChem* **2014**, *15*, 1755–1760.

[183] H. Xiao, F. B. Peters, P. Y. Yang, S. Reed, J. R. Chittuluru, P. G. Schultz, *ACS Chem. Biol.* **2014**, *9*, 1092–1096.

[184] T. Passioura, H. Suga, *Chem. Commun.* **2017**, *53*, 1931–1940.

[185] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**, *6*, 779–790.

[186] T. Katoh, H. Suga, *J. Am. Chem. Soc.* **2020**, *142*, 16518–16522.

[187] T. Katoh, H. Suga, *J. Am. Chem. Soc.* **2020**, *142*, 4965–4969.

[188] T. Katoh, H. Suga, *J. Am. Chem. Soc.* **2018**, *140*, 12159–12167.

[189] K. Öjemalm, T. Higuchi, P. Lara, E. Lindahl, H. Suga, G. Von Heijne, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 10559–10564.

[190] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12297–12302.

[191] R. C. Cookson, I. D. R. Stevens, C. T. Watts, *Chem. Commun.* **1966**, 744.

[192] J. Thiele, O. Stange, *Justus Liebigs Ann. Chem.* **1894**, *283*, 1–46.

[193] K. De Bruycker, S. Billiet, H. A. Houck, S. Chattopadhyay, J. M. Winne, F. E. Du Prez, *Chem. Rev.* **2016**, *116*, 3919–3974.

[194] G. B. Butler, A. G. Williams, *J Polym Sci Polym Chem Ed* **1979**, *17*, 1117–1128.

[195] S. Billiet, K. De Bruycker, F. Driessen, H. Goossens, V. Van Speybroeck, J. M. Winne, F. E. Du Prez, *Nat. Chem.* **2014**, *6*, 815–821.

[196] M. Shimizu, T. Takahashi, S. Uratsuka, S. Yamada, *J. Chem. Soc. Chem. Commun.* **1990**, DOI 10.1039/C39900001416.

[197] M. Shimizu, S. Kamachi, Y. Nishii, S. Yamada, *Anal. Biochem.* **1991**, *194*, 77–81.

[198] K. Hoogewijs, D. Buyst, J. M. Winne, J. C. Martins, A. Madder, *Chem. Commun.* **2013**, *49*, 2927–2929.

[199] H. Ban, J. Gavrilyuk, C. F. Barbas, *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.

[200] E. Fisher, E. Fourneau, *Ber. dtsch.chem.Ges.* **1901**, *34*, 2868–2879.

[201] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, DOI 10.1021/ja00897a025.

[202] R. L. Letsinger, M. J. Kornet, *J. Am. Chem. Soc.* **1963**, DOI 10.1021/ja00902a054.

[203] C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hammers, E. B. Songa, N. Bendahman, R. Hammers, *Nature* **1993**, DOI 10.1038/363446a0.

[204] S. Muyldermans, *Annu. Rev. Biochem.* **2013**, *82*, 775–797.

[205] A. Desmyter, S. Spinelli, A. Roussel, C. Cambillau, *Curr. Opin. Struct. Biol.* **2015**, *32*, 1–8.

[206] J. B. Holz, *Transfus. Apher. Sci.* **2012**, DOI 10.1016/j.transci.2012.03.027.

[207] F. Callewaert, J. Roodt, H. Ulrichts, T. Stohr, W. J. Van Rensburg, S. Lamprecht, S. Rossenu, S. Priem, W. Willems, J. B. Holz, *Blood* **2012**, DOI 10.1182/blood-2012-04-420943.

[208] F. Peyvandi, M. Scully, J. A. Kremer Hovinga, S. Cataland, P. Knöbl, H. Wu, A. Artoni, J.-P. Westwood, M. Mansouri Taleghani, B. Jilma, et al., *N. Engl. J. Med.* **2016**, DOI 10.1056/nejmoa1505533.

[209] I. Vaneycken, J. Govaert, C. Vincke, V. Caveliers, T. Lahoutte, P. De Baetselier, G. Raes, A. Bossuyt, S. Muyldermans, N. Devoogdt, *J. Nucl. Med.* **2010**, DOI 10.2967/jnumed.109.069823.

[210] L. Huang, L. O. T. Gainkam, V. Caveliers, C. Vanhove, M. Keyaerts, P. Baetselier, A. Bossuyt, H. Revets, T. Lahoutte, *Mol. Imaging Biol.* **2008**, DOI 10.1007/s11307-008-0133-8.

[211] M. J. W. D. Vosjan, L. R. Perk, R. C. Roovers, G. W. M. Visser, M. Stigter-Van Walsum, P. M. P. Van Bergen En Henegouwen, G. A. M. S. Van Dongen, *Eur. J. Nucl. Med. Mol. Imaging* **2011**, DOI 10.1007/s00259-010-1700-1.

[212] V. Cortez-Retamozo, N. Backmann, P. D. Senter, U. Wernery, P. De Baetselier, S. Muyldermans, H. Revets, *Efficient Cancer Therapy with a Nanobody-Based Conjugate*, **2004**.

[213] T. De Meyer, S. Muyldermans, A. Depicker, *Trends Biotechnol.* **2014**, *32*, DOI 10.1016/j.tib.

[214] U. Rothbauer, K. Zolghadr, S. Tillib, D. Nowak, L. Schermelleh, A. Gahl, N. Backmann, K. Conrath, S. Muyldermans, M. C. Cardoso, et al., *Nat. Methods* **2006**, *3*, 887–889.

[215] I. Van Audenhove, J. Gettemans, *EBioMedicine* **2016**, *8*, 40–48.

[216] A. Verhelle, W. Van Overbeke, C. Peleman, R. De Smet, O. Zwaenepoel, T. Lahoutte, J. Van Dorpe, N. Devoogdt, J. Gettemans, *Mol. Imaging Biol.* **2016**, *18*, 887–897.

[217] I. Van Audenhove, C. Boucherie, L. Pieters, O. Zwaenepoel, B. Vanloo, E. Martens, C. Verbrugge, G. Hassanzadeh-Ghassabeh, J. Vandekerckhove, M. Cornelissen, et al., *FASEB J.* **2014**, *28*, 1805–1818.

[218] G. Gonzalez-Sapienza, M. A. Rossotti, S. Tabares-da Rosa, *Front. Immunol.* **2017**, *8*, 1–12.

[219] Y. Wang, Z. Fan, L. Shao, X. Kong, X. Hou, D. Tian, Y. Sun, Y. Xiao, L. Yu, *Int. J. Nanomedicine* **2016**, DOI 10.2147/IJN.S107194.

[220] E. F. Fornasiero, F. Opazo, *Bioessays* **2015**.

[221] B. Traenkle, U. Rothbauer, *Front. Immunol.* **2017**, *8*, DOI 10.3389/fimmu.2017.01030.

[222] J. A. Thorley, J. Pike, J. Z. Rappoport, in *Fluoresc. Microsc. Super-Resolution Other Nov. Tech.*, **2014**.

[223] T. De Meyer, S. Muyldermans, A. Depicker, *Trends Biotechnol.* **2014**, *32*, 263–270.

[224] J. Ries, C. Kaplan, E. Platonova, H. Eghlidi, H. Ewers, *Nat. Methods* **2012**, *9*, 582–584.

[225] I. Van Audenhove, N. Debeuf, C. Boucherie, J. Gettemans, *Biochim. Biophys. Acta - Mol. Cell Res.* **2015**, *1853*, 940–952.

[226] S. Linder, *J. Cell Sci.* **2009**, *122*, 3009–3013.

[227] D. A. Murphy, S. A. Courtneidge, *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 413–426.

[228] J. Desmet, K. Verstraete, Y. Bloch, E. Lorent, Y. Wen, B. Devreese, K. Vandenbroucke, S. Loverix, T. Hettmann, S. Deroo, et al., *Nat. Commun.* **2014**, *5*, 5237.

[229] M. J. Schmidt, D. Summerer, *Angew. Chemie - Int. Ed.* **2013**, *52*, 4690–4693.

[230] S. Billiet, K. De Bruycker, F. Driessen, H. Goossens, V. Van Speybroeck, J. M. Winne, F. E. Du Prez, *Nat. Chem.* **2014**, *6*, 815–821.

[231] W. Alexander, in *Am. Soc. Clin. Oncol. , 2010 Annu. Meet.*, **2010**, pp. 469–475.

[232] M. Hirose, Y. Yoshida, K. Horii, Y. Hasegawa, Y. Shibuya, *Arch. Oral Biol.* **2021**, *122*, 105024.

[233] W. W. Stewart, *Cell* **1978**, *14*, 741–759.

[234] K. Hoogewijs, A. Deceuninck, A. Madder, *Org. Biomol. Chem.* **2012**, *10*, 3999–4002.

[235] M. J. Davies, *Biochem. Biophys. Res. Commun.* **2003**, *305*, 761–770.

[236] K. Hoogewijs, D. Buyst, J. M. Winne, J. C. Martins, A. Madder, *Chem. Commun.* **2013**, *49*, 2927–2929.

[237] K. Meerschaert, V. De Corte, Y. De Ville, J. Vandekerckhove, J. Gettemans, *EMBO J.* **1998**, *17*, 5923–5932.

[238] H. Q. Sun, M. Yamamoto, M. Mejillano, H. L. Yin, *J. Biol. Chem.* **1999**, *274*, 33179–33182.

[239] W. Van Overbeke, A. Verhelle, I. Everaert, O. Zwaenepoel, J. Vandekerckhove, C. Cuvelier, W. Derave, J. Gettemans, *Mol. Ther.* **2014**, *22*, 1768–1778.

[240] A. de la Chapelle, R. Tolvanen, G. Boysen, J. Santavy, L. Bleeker-Wagemakers, C. P. J. Maury, J. Kere, *Nat. Genet.* **1992**, *2*, 157–160.

[241] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, C. C. F. Blake, *J. Mol. Biol.* **1997**, *273*, 729–739.

[242] K. R. Schmitz, A. Bagchi, R. C. Roovers, P. M. P. Van Bergen En Henegouwen, K. M. Ferguson, *Structure* **2013**, *21*, 1214–1224.

[243] P. H. Huang, A. M. Xu, F. M. White, *Sci. Signal.* **2009**, *2*, 1–13.

[244] R. C. Roovers, T. Laeremans, L. Huang, S. De Taeye, A. J. Verkleij, H. Revets, H. J. De Haard, P. M. P. Van Bergen En Henegouwen, *Cancer Immunol. Immunother.* **2007**, *56*, 303–317.

[245] C. J. White, A. K. Yudin, *Nat. Chem.* **2011**, *3*, 509–524.

[246] C. Morrison, *Nat. Rev. Drug Discov.* **2018**, *17*, 531–533.

[247] O. Torres, D. Yüksel, M. Bernardina, K. Kumar, D. Bong, *ChemBioChem* **2008**, *9*, 1701–1705.

[248] J. R. Kumita, D. G. Flint, G. A. Woolley, O. S. Smart, *Faraday Discuss.* **2002**, *122*, 89–103.

[249] G. Lautrette, F. Touti, H. G. Lee, P. Dai, B. L. Pentelute, *J. Am. Chem. Soc.* **2016**, *138*, 8340–8343.

[250] H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. Degrado, D. C. Greenbaum, *J. Am. Chem. Soc.* **2012**, *134*, 17704–17713.

[251] D. P. Fairlie, A. Dantas de Araujo, *Biopolymers* **2016**, *106*, 843–852.

[252] M. Chorev, E. Roubini, R. L. McKee, S. W. Gibbons, M. E. Goldman, M. P. Caulfield, M. Rosenblatt, *Biochemistry* **1991**, *30*, 5968–5974.

[253] D. Van Lysebetten, S. Felissati, E. Antonatou, L. L. G. Carrette, P. Espeel, E. Focquet, F. E. Du Prez, A. Madder, *ChemBioChem* **2018**, *19*, 641–646.

[254] R. M. Kohli, C. T. Walsh, M. D. Burkart, *Nature* **2002**, *418*, 658–661.

[255] Y. W. Kim, T. N. Grossmann, G. L. Verdine, *Nat. Protoc.* **2011**, *6*, 761–771.

[256] S. A. Kawamoto, A. Coleska, X. Ran, H. Yi, C. Y. Yang, S. Wang, *J. Med. Chem.* **2012**, *55*, 1137–1146.

[257] M. M. Madden, A. Muppidi, Z. Li, X. Li, J. Chen, Q. Lin, *Bioorganic Med. Chem. Lett.* **2011**, *21*, 1472–1475.

[258] P. A. Cistrone, A. P. Silvestri, J. C. J. Hintzen, P. E. Dawson, *ChemBioChem* **2018**, *19*, 1031–1035.

[259] C. M. Haney, M. T. Loch, W. S. Horne, *Chem. Commun.* **2011**, *47*, 10915–10917.

[260] A. J. Kamens, K. M. Mientkiewicz, R. J. Eisert, J. A. Walz, C. R. Mace, J. A. Kritzer, *Bioorganic Med. Chem.* **2018**, *26*, 1206–1211.

[261] L. R. Malins, J. N. Degruyter, K. J. Robbins, P. M. Scola, M. D. Eastgate, M. R. Ghadiri, P. S. Baran, *J. Am. Chem. Soc.* **2017**, *139*, 5233–5241.

[262] Y. Yamagishi, H. Ashigai, Y. Goto, H. Murakami, H. Suga, *ChemBioChem* **2009**, *10*, 1469–1472.

[263] Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120–129.

[264] Y. Sako, Y. Goto, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 241–249.

[265] Y. Sako, J. Morimoto, H. Murakami, H. Suga, *J. Am. Chem. Soc.* **2008**, *130*, 7232–7234.

[266] Y. Goto, K. Iwasaki, K. Torikai, H. Murakami, H. Suga, *Chem. Commun.* **2009**, 3419–3421.

[267] T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami, H. Suga, *Nat. Chem. Biol.* **2009**, *5*, 888–890.

[268] Y. Goto, H. Ashigai, Y. Sako, H. Murakami, H. Suga, *Nucleic Acids Symp. Ser.* **2006**, *50*, 293–294.

[269] K. De Bruycker, S. Billiet, H. A. Houck, S. Chattopadhyay, J. M. Winne, F. E. Du Prez, *Chem. Rev.* **2016**, *116*, 3919–3974.

[270] A. H. St. Amant, F. Huang, J. Lin, K. Rickert, V. Oganesyan, D. Lemen, S. Mao, J. Harper, M. Marelli, H. Wu, et al., *Angew. Chemie - Int. Ed.* **2019**, *58*, 8489–8493.

[271] N. T. Kamiya Yuji, Sakurai Akira, Tamura Saburo, **1978**, *83*, 135–139.

[272] S. L. Wolda, J. A. Glomset, *J. Biol. Chem.* **1988**, *263*, 5997–6000.

[273] J. F. Hancock, A. I. Magee, J. E. Childs, C. J. Marshall, *Cell* **1989**, *57*, 1167–1177.

[274] J. B. Gibbs, A. Oliff, *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 143–166.

[275] B. M. M. Grant, M. Enomoto, S.-I. Back, K.-Y. Lee, T. Gebregiworgis, N. Ishiyama, M. Ikura, C. B. Marshall, *Sci. Signal.* **2020**, *13*, eaaz0344.

[276] C. C. Palsuledesai, M. D. Distefano, *ACS Chem. Biol.* **2015**, *10*, 51–62.

[277] M. D. Resh, *Nat. Chem. Biol.* **2006**, *2*, 584–590.

[278] R. A. Schmidt, C. J. Schneider, J. A. Glomset, *J. Biol. Chem.* **1984**, *259*, 10175–10180.

[279] T. Giannakouro, J. Armstrongb, A. I. Magee, *Protein Prenylation in Schizosaccharomyces Pombe*, European Biochemical Societies, **1992**.

[280] A. William Maltese, *FASEB J.* **1990**, *4*, 3319–28.

[281] R. N. Hannoush, J. Sun, *Nat. Chem. Biol.* **2010**, *6*, 498–506.

[282] J. K. Dozier, S. L. Khatwani, J. W. Wollack, Y. C. Wang, C. Schmidt-Dannert, M. D. Distefano, *Bioconjug. Chem.* **2014**, *25*, 1203–1212.

[283] D. Das, Z. Tnimov, U. T. T. Nguyen, G. Thimmaiah, H. Lo, D. Abankwa, Y. Wu, R. S. Goody, H. Waldmann, K. Alexandrov, *ChemBioChem* **2012**, *13*, 674–683.

[284] M. K. Kim, T. S. Kleckley, A. J. Wiemer, S. A. Holstein, R. J. Hohl, D. F. Wiemer, *J. Org. Chem.* **2004**, *69*, 8186–8193.

[285] E. M. Storck, J. Morales-Sanfrutos, R. A. Serwa, N. Panyain, T. Lanyon-Hogg, T. Tolmachova, L. N. Ventimiglia, J. Martin-Serrano, M. C. Seabra, B. Wojciak-Stothard, et al., *Nat. Chem.* **2019**, *11*, 552–561.

[286] C. C. Palsuledesai, J. D. Ochocki, M. M. Kuhns, Y. C. Wang, J. K. Warmka, D. S. Chernick, E. V. Wattenberg, L. Li, E. A. Arriaga, M. D. Distefano, *ACS Chem. Biol.* **2016**, *11*, 2820–2828.

[287] A. J. DeGraw, C. Palsuledesai, J. D. Ochocki, J. K. Dozier, S. Lenevich, M. Rashidian, M. D. Distefano, *Chem. Biol. Drug Des.* **2010**, *76*, 460–471.

[288] G. Charron, L. K. Tsou, W. Maguire, J. S. Yount, H. C. Hang, *Mol. Biosyst.* **2011**, *7*, 67–73.

[289] C. C. Palsuledesai, J. D. Ochocki, T. W. Markowski, M. D. Distefano, *Mol. Biosyst.* **2014**, *10*, 1094–1103.

[290] G. Charron, M. M. H. Li, M. R. MacDonald, H. C. Hang, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11085–11090.

[291] E. M. Sletten, C. R. Bertozzi, *Angew. Chemie - Int. Ed.* **2009**, *48*, 6974–6998.

[292] J. W. Wollack, B. J. Monson, J. K. Dozier, J. J. Dalluge, K. Poss, S. A. Hilderbrand, M. D. Distefano, *Chem. Biol. Drug Des.* **2014**, *84*, 140–147.

[293] T. C. Turek, I. Gaon, M. D. Distefano, *Tetrahedron Lett.* **1996**, *37*, 4845–4848.

[294] M. L. Hovlid, R. L. Edelstein, O. Henry, J. Ochocki, A. DeGraw, S. Lenevich, T. Talbot, V. G. Young, A. W. Hruza, F. Lopez-Gallego, et al., *Chem. Biol. Drug Des.* **2010**, *75*, 51–67.

[295] J. S. Vervacke, A. L. Funk, Y. C. Wang, M. Strom, C. A. Hrycyna, M. D. Distefano, *J. Org. Chem.* **2014**, *79*, 1971–1978.

[296] Y.-C. Wang, M. D. Distefano, *Bioorg. Chem.* **2016**, *64*, 59–65.

[297] D. Kaiser, J. M. Winne, M. E. Ortiz-Soto, J. Seibel, T. A. Le, B. Engels, *J. Org. Chem.* **2018**, *83*, 10248–10260.

[298] Q.-Y. Hu, M. Allan, R. Adamo, D. Quinn, H. Zhai, G. Wu, K. Clark, J. Zhou, S. Ortiz, B. Wang, et al., *Chem. Sci.* **2013**, *4*, 3827.

[299] P. R. A. Zanon, F. Yu, P. Z. Musacchio, L. Lewald, M. Zollo, K. Krauskopf, P. Raunft, T. E. Maher, M. Cigler, C. J. Chang, et al., **n.d.**, DOI https://doi.org/10.26434/chemrxiv.14186561.v1.

[300] A. Christoforou, G. Nicolaou, Y. Elemes, *Tetrahedron Lett.* **2006**, *47*, 9211–9213.

[301] H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670–2671.

[302] I. Van Audenhove, C. Boucherie, L. Pieters, O. Zwaenepoel, B. Vanloo, E. Martens, C. Verbrugge, G. Hassanzadeh-Ghassabeh, J. J. Vandekerckhove, M. Cornelissen, et al., *FASEB J.* **2014**, *28*, 1805–1818.

[303] I. Van Audenhove, N. Debeuf, C. Boucherie, J. Gettemans, *Biochim. Biophys. Acta - Mol. Cell Res.* **2015**, *1853*, 940–952.

[304] J. L. Bos, *Cancer Res.* **1989**, *49*, 4682–4689.

[305] A. J. Krzysiak, S. A. Scott, K. A. Hicks, C. A. Fierke, R. A. Gibbs, *Bioorganic Med. Chem. Lett.* **2007**, *17*, 5548–5551.

[306] J. Zhang, J. Zhao, L. Wang, H. Hu, S. Wang, P. Yu, R. Wang, *Tetrahedron* **2021**, *81*, 131917.

[307] E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Rev. Chem.* **2019**, *3*, 147–171.

[308] P. Agarwal, C. R. Bertozzi, *Bioconjug. Chem.* **2015**, *26*, 176–192.

[309] C. Bottecchia, T. Noël, *Chem. - A Eur. J.* **2019**, *25*, 26–42.

[310] E. Baslé, N. Joubert, M. Pucheault, *Chem. Biol.* **2010**, *17*, 213–227.

[311] J. N. deGruyter, L. R. Malins, P. S. Baran, *Biochemistry* **2017**, *56*, 3863–3873.

[312] Y. Yu, L. K. Zhang, A. V. Buevich, G. Li, H. Tang, P. Vachal, S. L. Colletti, Z. C. Shi, *J. Am. Chem. Soc.* **2018**, *140*, 6797–6800.

[313] N. Kaplaneris, T. Rogge, R. Yin, H. Wang, G. Sirvinskaite, L. Ackermann, *Angew. Chemie* **2019**, *131*, 3514–3518.

[314] W. Wang, M. Ø. M. Lorion, J. Shah, A. R. Kapdi, **2018**, 14700–14717.

[315] G. L. Tolnai, J. P. Brand, J. Waser, *Beilstein J. Org. Chem.* **2016**, *12*, 745–749.

[316] Z. Ruan, N. Sauermann, E. Manoni, L. Ackermann, *Angew. Chemie Int. Ed.* **2017**, *56*, 3172–3176.

[317] A. Schischko, H. Ren, N. Kaplaneris, L. Ackermann, *Angew. Chemie - Int. Ed.* **2017**, *56*, 1576–1580.

[318] M. B. Hansen, F. Hubálek, T. Skrydstrup, T. Hoeg-Jensen, *Chem. - A Eur. J.* **2016**, *22*, 1572–1576.

[319] J. Ruiz-Rodriguez, F. Albericio, R. Lavilla, *Chem. - A Eur. J.* **2010**, *16*, 1124–1127.

[320] L. R. Malins, K. M. Cergol, R. J. Payne, *Chem. Sci.* **2014**, *5*, 260–266.

[321] H. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma, C. F. Barbas, *Bioconjug. Chem.* **2013**, *24*, 520–532.

[322] S. Vandewalle, R. De Coen, B. G. De Geest, F. E. Du Prez, *ACS Macro Lett.* **2017**, 1368–1372.

[323] M. Moinpour, N. K. Barker, L. E. Guzman, J. C. Jewett, P. R. Langlais, J. C. Schwartz, *Protein Sci.* **2020**, *29*, 1784–1793.

[324] Z. E. Potter, H. T. Lau, S. Chakraborty, L. Fang, M. Guttman, S. E. Ong, D. M. Fowler, D. J. Maly, *Cell Chem. Biol.* **2020**, *27*, 1084-1096.e4.

[325] P. S. Baran, C. A. Guerrero, E. J. Corey, *Org. Lett.* **2003**, *5*, 1999–2001.

[326] S. B. Hanay, D. F. Brougham, A. A. Dias, A. Heise, *Polym. Chem.* **2017**, *8*, 6594–6597.

[327] S. Saussez, R. Kiss, *Cell. Mol. Life Sci.* **2006**, *63*, 686–697.

[328] W. Vannecke, C. Ampe, M. Van Troys, M. Beltramo, A. Madder, *ACS Chem. Biol.* **2017**, *12*, 2191–2200.

[329] D. Kaiser, J. M. Winne, M. E. Ortiz-Soto, J. Seibel, T. A. Le, B. Engels, *J. Org. Chem.* **2018**, *83*, 10248–10260.

[330] M. Moinpour, N. Barker, L. Guzman, J. Jewett, P. Langlais, J. Schwartz, *Protein Sci.* **2020**, DOI 10.1101/2020.02.04.934406.

[331] R. Guo, H. Sakamoto, S. Sugiura, O. Minetaro, *J. Cell. Physiol.* **2007**, *2011*, 327–335.

[332] R. Xiong, S. K. Samal, J. Demeester, A. G. Skirtach, S. C. De Smedt, K. Braeckmans, *Adv. Phys. X* **2016**, *1*, 596–620.

[333] J. Liu, T. Hebbrecht, T. Brans, E. Parthoens, S. Lippens, C. Li, H. De, **2020**, *13*, 485–495.

[334] L. Wang, W. Hisano, Y. Murai, M. Sakurai, Y. Muto, H. Ikemoto, M. Okamoto, T. Murotani, R. Isoda, D. Kim, et al., *Molecules* **2013**, 8393–8401.

[335] A. Van Den Abbeele, S. De Clercq, A. De Ganck, V. De Corte, B. Van Loo, S. H. Soror, V. Srinivasan, J. Steyaert, J. Vandekerckhove, J. Gettemans, *Cell. Mol. Life Sci.* **2010**, *67*, 1519–1535.

[336] P. M. P. van Bergen en Henegouwen, K. M. Ferguson, A. Bagchi, R. C. Roovers, K. R. Schmitz, *Structure* **2013**, *21*, 1214–1224.

[337] A. Baroni, L. Vlaminck, L. Mespouille, F. Du Prez, N. Delbosc, B. Blankert, *Macromol. Rapid Commun.* **2019**, 1–5.

[338] V. Bakken, T. Helgaker, E. Uggerud, *Eur. J. Mass Spectrom.* **2004**, *10*, 625–638.