MASTERARBEIT | MASTER'S THESIS

Titel | Title

Survival Motor Neuron (SMN) Protein: Unveiling Stability Mechanisms through Phosphorylation

verfasst von | submitted by

Andreas Fraunhofer BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 863
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Biologische Chemie
Betreut von | Supervisor:	Univ.-Prof. Dr. Christian Friedrich Wilhelm Becker

Acknowledgements

I would like to express my deepest gratitude to Univ-Prof. Dr. Christian Friedrich Wilhelm Becker for the oppurtunity to work in his research group over the course of this master's thesis. I learned a lot in terms of practical and soft skills during my time at the Becker group, thanks to the expertise of Univ-Prof. Dr. Becker and the other members of the research group.

I am especially thankful to Dr. Claudia Cobos Caceres for supervising and mentoring me during the practical part of this project. Without her help and knowledge this master's thesis would not have been possible.

I am also fortunate to have had the help and support of the postdoctoral group members Dr. Tim Baldensperger, Dr. Moritz Urschbach and Dr. Somnath Mukherjee. Throughout our many conversations I gained valuable scientific input for my project.

Many thanks to the remaining members of the Becker group, Hannah Back, MSc; Stefan Moser, MSc; Dr. Leopold Dürrauer; Miriam Preissler, MSc; Mag. Gerhard Niederacher; Mag. Manuel Felkel; Dr. Aleksandr Kravchuk; Oliver Gajsek, MSc., Dr. Dominik Vogl and Laura Holzapfel, Bsc. I am fond of all the fun and interessting conversations we had and all the activities we did outside of work.

Last but not least I am thankful for my family and my girlfriend Anna Hampel for supporting and believing in me over the course of my education.

Abstract

The Survival of Motoneuron (SMN) protein is well known to cause Spinal Muscular Atrophy (SMA) if not present in sufficient concetration inside the cell. To investigate the influence of phosphorylations on serine resiudes on protein stability, this master's thesis aims to produce four different variants of the C-terminal section of the protein with different phosphorylation patterns.

To address the limitations of solid-phase peptide synthesis (SPPS) regarding peptide size, a ligation strategy was developed that segments the target sequence into smaller, more manageable peptides. These peptides were designed not only for optimal size but also to maintain similar solubility, facilitating efficient reactions during native chemical ligation (NCL). The project aims to refine both the SPPS and ligation protocols by adjusting various parameters and analyzing the results.

The target region was divided into three segments, each synthesized via SPPS and subsequently linked using NCL. The design of these fragments was influenced by the presence of a native cysteine in the sequence. The synthesis of these peptides posed significant challenges due to a high proline content. Proline complicates synthesis due to steric hindrance and the formation of a secondary amine, resulting in reduced reactivity. This challenge intensifies when consecutive prolines are present, increasing the risk of undesired insertions and producing truncated peptides that cannot proceed to further synthesis, as well as peptides with amino acid deletions.

The goal after finishing the synthesis of the different protein fragments was, to send them to a cooperation partner, which would then perform assays to assess information about protein stability and protien oligomerization processes.

Zusammenfassung

Das Survival of Motoneuron (SMN)-Protein ist bekannt dafür, Spinale Muskelatrophie (SMA) zu verursachen, wenn es nicht in ausreichender Konzentration in der Zelle vorhanden ist. Um den Einfluss von Phosphorylierungen an Serinresten auf die Stabilität des Proteins zu untersuchen, zielt diese Masterarbeit darauf ab, vier verschiedene Varianten des C-terminalen Abschnitts des Proteins mit unterschiedlichen Phosphorylierungsmustern herzustellen.

Um die Größenbeschränkungen der Festphasenpeptidsynthese (SPPS) zu überwinden, wurde eine Ligationsstrategie entwickelt, die die Sequenz in kleinere, handhabbare Peptide unterteilt. Diese Peptide wurden nicht nur in Bezug auf ihre Größe, sondern auch hinsichtlich ihrer Löslichkeit entworfen, um eine effiziente Reaktion während der nativen chemischen Ligation (NCL) zu ermöglichen. Das Projekt strebt auch die Optimierung der SPPS- und Ligationsprotokolle durch Variationen der Parameter und den Vergleich der Ergebnisse an.

Der untersuchte Bereich wurde in drei Segmente unterteilt, die jeweils mittels SPPS hergestellt und anschließend durch NCL verknüpft wurden. Das Vorhandensein eines nativen Cysteins innerhalb der Sequenz beeinflusste das Design dieser Fragmente. Die Herstellung dieser Peptide ist besonders aufgrund des hohen Prolin-Gehalts in der Sequenz herausfordernd. Prolin stellt bei der Synthese eine Schwierigkeit dar, da neben dem sterischen Hinderungseffekt die Seitenkette an das Stickstoffatom der Aminogruppe gebunden ist und somit ein sekundäres Amin bildet, wodurch Prolin im Vergleich zu anderen Aminosäuren weniger reaktiv ist. Dieses Problem wird verschärft, wenn Proline aufeinander folgen, was zu zusätzlichen Prolinen führen kann. Weitere häufige Nebenprodukte in prolinreichen Sequenzen sind verkürzte Peptide, die nach bestimmten Aminosäuren keine weiteren Kopplungsschritte durchlaufen, sowie Peptide mit fehlenden Aminosäuren.

Das Ziel nach der Synthese der verschiedenen Proteinfragmentvarianten war es, diese zu einem Kooperationspartner zu senden. Dieser würde damit Assays durchführen um Information über die Stabilität des Proteins und Oligomerizationsprozesse zu gewinnen.

List of abbreviations

ACN .....Acetonitrile
DCN .....Dichloromethane
DIC.....N,N’-Diisopropylcarbodiimide
DIPEA .....N,N-Diisopropylethylamine
DMF .....N,N-Dimethylformamide
DMS .....Dimethyl sulfide
Fmoc 9-fluorenylmethoxycarbonyl
HATU .....O-(7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
MS.....Mass spectrometry
PH .....Peptide hydrazide
Oxyma.....ethyl 2-cyano-2-(hydroxyimino) acetate
SMA .....Spinal Muscular Atrophy
SMN .....Survival of Motor Neuron
TFA .....Trifluoroacetic acid

Table of contents

1 Introduction

Spinal Muscular Atrophy

Spinal Muscular Atrophy (SMA) is an autosomal recessive disorder characterized by the degeneration of alpha motor neurons in the anterior horn of the spinal cord, leading to a decrease of muscle tissue in affected individuals. [1] This condition exhibits a diverse spectrum with three main phenotypes, which were formally classified based on variations in severity and age of onset, a consensus document on the diagnosis of children with SMA was initially developed by Wang et al. in 2007 [2] and later updated by Mercuri et al. [1] Ove the course of time a form of SMA with adult onset became recognized as type IV while the newly discovered prenatal onset of the disease would be referred to as type 0. [1] In 1995, a pivotal discovery identified the Survival of Motoneuron (SMN) protein as the gene product responsible for SMA, which leads to the disease when not present in sufficient concentrations. [3] It was discovered that SMA patients show either homozygous deletion or subtle mutations of the encoding gene, termed *SMN1*. As humans possess a paralogous gene, known as *SMN2*, which is almost identical to *SMN1* with the exception of five amino acids, it has been revealed that the amount of *SMN2* gene product is inversely correlated to disease severity. [4]

1.1 Survival of Motoneuron Protein

The SMN protein itself comprises 294 amino acid residues with a mass of approximately 38 kDa. [4] In the context of a living cell, the protein localizes to both, the cytoplasm and the nucleus. In the nucleus it is known to accumulate in nuclear structures known as Cajal bodies and Gems (Gemini of Cajal bodies). [5] On the genetic level, the *SMN1* gene consists out of nine exons, which are referred to as 1, 2a, 2b, 3, 4, 5, 6, 7 and 8. When moving from the N-terminus into the direction of the C-terminus, one encounters several functional motifs: a basic/lysine-rich domain, a Tudor domain [6], polyproline motifs [7] and a YG box. [8] Although it is sequence is almost identical to the *SMN1* gene, *SMN2* is affected by a nucleotide substitution which triggers a splicing alteration. As a result of this substitution exon 7 is skipped in the majority of cases. [9] The skip of exon 7 happens with such a high frequency that approximately 80% of all *SMN2* transcripts are present in this truncated form. [10] Consequently, the proteins produced by these transcripts are also truncated versions of the SMN protein. Notably, this truncated form exhibits reduced stability when compared to the full-length protein, rendering it susceptible to rapid degradation. [10]

1.2 Pre-mRNA splicing and small nuclear RNAs

To understand why absence of full length SMN protein causes SMA, one has to understand the process of mRNA splicing where it is involved as part of a multiprotein complex. [11] As proteins in eukaryotic organisms are encoded by exons which are interspaced by introns, a system is needed to cut out the introns and recombine the exons. This system is implemented

through a process known as pre-mRNA splicing which occurs either co- or post-transcriptionally. [12] To obtain the mature mRNA, the pre-mRNA must be assembled into a multi-component complex known as spliceosome, afterwards a two-step transesterification reaction takes place. [13] The spliceosome itself consists of five small nuclear RNAs (snRNAs; U1, U2, U4, U5, and U6), these snRNA are present in the form of small nuclear ribonucleoprotein particles (snRNPs, RNA-protein complexes) and a large number of splicing protein factors. [14] The assembly of the spliceosome together with the splicing process itself can be seen in Figure 1, the scheme was taken from a publication from Morais et al. [15]

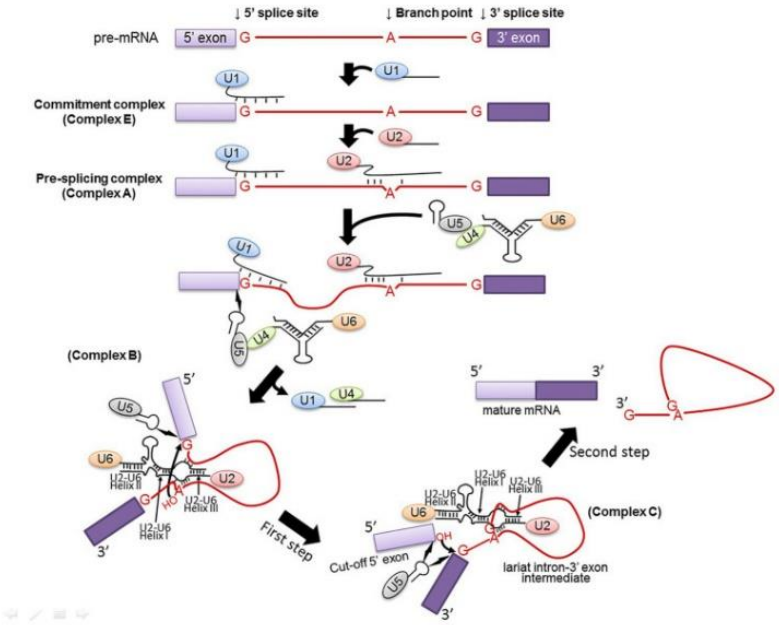

Figure 1: Spliceosome assembly [15]

During the splicing process the snRNAs on the snRNPs form hydrogen bonds with the introns which makes further interactions possible. [15] Structure wise each snRNP consists of one uridine-rich snRNA, a heptameric ring of Sm proteins (B/B', D1, D2, D3, E, F, and G), and a specific set of accessory proteins, as mentioned previously. (Faravelli et al. 2023)

1.3 The SMN Complex and snRNP formation

When we look at the formation of snRNPs, the SMN complex appears to be one of the major factors in the assembly process. When present in its functional form, the SMN protein is the core component of this multiprotein complex that processes the small nuclear RNAs, which, in turn, are needed for the splicing of pre-mRNA. [11] Other than just the SMN protein, which oligomerizes with itself, the SMN complex consists of eight other proteins, named Gemin2-8 and Unrip. [16] Although the general structure of the complex still has not been fully understood, it has become clear that the complex shows a modular composition. Hereby the proteins SMN, Gemin8 and Gemin7 are at the center of the construct and act as a binding platform for other proteins. [16] Both Gemin2 and Gemin3 as well as Gemin8 are able to bind

to SMN. Gemin8 is responsible for mediating interactions with Gemin4 and Gemin7 which in turn is able to recruit Unrip and Gemin6. [16] This leads overall to a stable structure to which Gemin5 is proposed to bind through weak interactions, mediated mainly through Gemin2. [16] The hypothesized interaction map of the complex can be seen in Figure 2. [16] During the process of snRNP formation the purpose of the SMN complex is to establish the assembly of the Sm ring onto the snRNA. [17] Originally the snRNAs are transcribed by the RNA polymerase II in the nucleus where the nascent oligomer is cleaved by the Integrator complex downstream of the 3' end. The resulting oligomer is referred to as 3' end extended pre-snRNA. [18] As the pre-snRNA is not able to leave the nucleus on its own, the help of shuttling proteins is needed for the export to the cytoplasm where the pre-snRNAs can then interact with the SMN complex. [19] Previous to the transfer of the Sm ring onto the snRNa, a complex of the Sm proteins themselves must be formed. This happens over the course of a multi-step process. The resulting structure is subsequently loaded onto the SMN complex. To make the transfer complete, the SMN complex needs to identify the snRNA. It is believed that Gemin5 is needed for this purpose as it is able to bind to some of the snRNAs features like the Sm site. [20] For the assembly of the Sm core itself, SMN,Gemin 2, Gemin3 and Gemin4 have been proven to be essential. [21] Being responsible for their stability, the Sm ring is a present in all U snRNPs and can be classified as common structural element for these kinds of snRNPs. [22] Following the ring formation, the 5' 7-methyl-guanosine cap of the snRNA gets methylated to 2,2,7-trimethyl-guanosine and the 3' end gets trimmed. [23] The snRNPs gets exported back into the nucleus after maturation process in the cytosol has been completed, afterwards they can locate to Cajal bodies. [24] Inside the Cajal bodies the snRNP undergoes further maturation. [23]

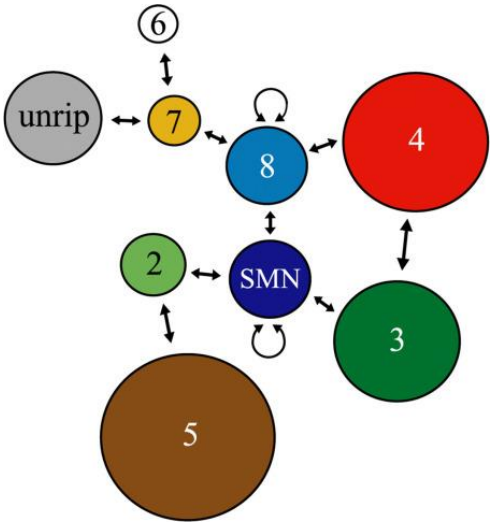

Figure 2: Proposed interaction map of the SMN complex [16]

1.4 Localization of SMN to Cajal bodies

Cajal bodies (CBs) are subnuclear structures which are named after the Spanish neurobiologist Santiago Ramón y Cajal. It has been shown that CBs are found most of the time associated with the nucleus where play an important role in RNA metabolism and formation of snRNPs. [25] CBs are dynamic structures, the total number of particles present inside a living cell varies based on the cell type and stage of development. [26] Considering the formation of these bodies, the coilin protein has been found to be a crucial factor. [27] Coilin has also been found to recruit the SMN complex and snRNPs. [28] Through the localization of the SMN protein to these structures, its involvement in RNA splicing has been postulated from early on. [29] Research has shown that during neuronal differentiation CBs in undifferentiated calls are mostly devoid of the SMN protein while over the course of differentiation the protein gets progressively recruited to the CBs. It can also be observed that during the development process, the production of SMN gets upregulated. [30] Besides CBs other particles named Gems can be found in the nucleus which contain the SMN protein but not coilin. In a similar manner differentiated cells of the nervous system show an increased amount of Gems. [30]

1.5 Approved therapies for SMA

Up to the date three therapies have been approved by the FDA for treatment of SMA, which differ in their mode of action. Namely these therapies are the antisense oligonucleotide nusinersen (Spinraza; Biogen), the gene therapy onasemnogene abeparvovec-xioi (Zolgensma; Novartis) and the small molecule risdiplam (Evrysdi; Roche). [31] In the antisense approach of nusinersen, a splicing regulatory element named intronic splicing silencer N1 (ISS-N1), located at intron 7 is targeted. [32] Through the short complementary sequence of synthetic nucleotides the splicing silencer gets deleted. [33] As a consequence the splicing pattern of SMN2 gets modified which in turn improves the phenotype. Risdiplam pursues a similar tactic as this small molecule acts as a splicing modulator. By binding to both, an exon enhancer sequence and the 5' splicing site of exon 7, at the pre-mRNA of *SMN2*, a ribonculeoprotein (RNP) complex gets stabilized. As a result exon 7 inclusion and full-length SMN protein production gets promoted. [34] In contrast Onasemnogene abeparvovec-xioi directly acts on the DNA level and replaces the monogenic defect by replacing the defect *SMN1* gene. [31] The gene is delivered in this therapy through the adeno-associated viral vector serotype 9 (AAV9), which can cross the blood-brain barrier (BBB) and efficiently transduce target cells in the CNS. [31]

1.6 Post-translational modifications (PTMs)

In eukaryotic organisms most proteins can undergo so called post-translational modification (PTMs) after they have been produced at the ribosomes. A variety of different modifications exist bust most common PTMs consists of small functional groups and complex biomolecules which are conjugated to amino acid side chains or termini by enzymes. [35] Among these

groups fall the most well-known PTMs: phosphorylation, acetylation, glycosylation, methylation and ubiquitylation. [36] Other PTMs include backbone splicing and cyclization. The main purpose of this reversible modifications is to modulate the activity of the protein and its interaction with other proteins. The location of the protein can also be affected by PTMs. [36]

1.6.1 Phosphorylation

The process of protein phosphorylation describes the addition of phosphate groups (PO_4) onto the polar side chains of amino acids, in most cases serine, threonine or tyrosine. The modification influences the structure of the protein which further can affect its properties, ranging from localization to stability and interaction with binding partners. Because of the wide range of cellular processes which are influenced by protein phosphorylations and dephosphorylations, multiple diseases can be traced back to abnormal phosphorylation patterns on proteins. [37] The transferred phosphate group comes from ATP, which is the co-substrate. [38] In total numbers phosphorylation is most common on serine, followed by threonine and tyrosine, at a relative frequency of 11.2 : 2.5 : 1 [39] The PO_4 transfer reaction is catalyzed by enzymes called protein kinases, which encompass one of the biggest gene families in eukaryotes with more than 500 known protein kinases. [37] The substrate is identified by kinases through mechanisms like subcellular localization, substrate docking interactions, and binding to scaffold proteins. An important aspect in these processes is the consensus sequence which is recognized by the kinase [40] The dephosphorylation reaction can take place in the form of a hydrolyzation of the phosphate group, mediated by phosphatase enzymes. [41]

1.6.2 Post-translational modifications on the SMN protein

In the case of the SMN protein, phosphorylations have been identified to be crucial considering localization, stability and protein function. [42] Based on the sequence of the protein, 50 amino acids which can undergo phosphorylation are present. Out of these 50 residues 26 have been verified by mass spectrometry to carry phosphorylations. [42] Just in the N-terminal Gemin2-binding domain, which is a conserved structure, five residues, namely S28, S31, T37, Y43 and S49, have been verified to be able to undergo phosphorylation. These phosphorylations have been localized at the N-terminal Gemin2-binding domain, the tudor domain as well as other regions around the N-terminus and the center of the protein. [42] Due to its limited accessibility to digestive enzymes, the C-terminus has been studied to a lesser extent regarding phosphorylations.. [43] Two of the most recently discovered amino acids to undergo phosphorylation are S290 and S292. [42] In terms of localization it has been shown in HeLa cells that cytoplasmic SMN is hyper-phosphorylated while nuclear SMN is hypo-phosphorylated. [42] For the complex formation it has been found that serine residues S28 and S31 are phosphorylated for the assembly. Within the SMN complex, the SMN protein forms an oligomeric core with the Tudor domain mediating the self-oligomerization. Additionally, the YG

box and the C-terminus also play a role in the SMN self-oligomerization process. [42] A scheme of the protein with its respective domains, published by Otter et all [16], can be seen in Figure 3

Through a series of experiments, it had been shown that a phosphorylation at the SMN N-terminus has no impact on self-oligomerization whereas C-terminal phosphorylation in close proximity to YG box affects the self-oligomerization capability. This was done by comparing mutants S28A and S31A, unable to undergo phosphorylation with the corresponding phospho-mimetics S28D and S31D. All of which oligomerized with endogenous SMN. [42] The C-terminal phospho-mimetic on the other hand S290D showed impaired oligomerization while the non-phospho mutant S290A as well as wildtype SMN did undergo oligomerization. [42] When comparing full length SMN with the SMNΔ7 produced by SMN2, it is clear that the missing exon results in lower stability. It has become clear that self-oligomerization as described above, and protein interactions, like the ones in the SMN complex, affect the stability of the protein but it is also hypothesized that phosphorylations regulate the stability. [42] Experiments with NSC34 cells revealed that the SMN phospho-mimetic S290D is less stable than its unphosphorylated counterpart. [42]

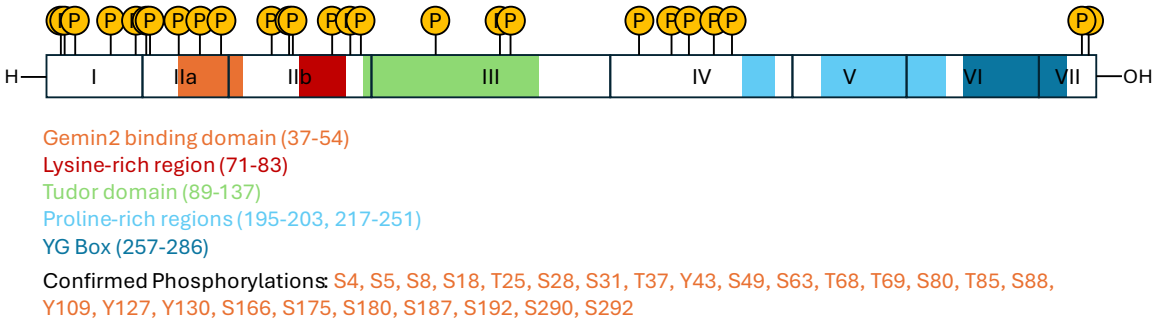

Figure 3: Scheme of the domains of the SMN protein with reported PTMs [42]

1.7 Semisynthesis as a tool for investigating PTMs

As the previous paragraph showed, PTMs can have dramatic effects on protein stability, oligomerization and functionality. As it is of great interest for researchers to study how certain modifications affect a protein of interest, various methods have been established to investigate them. The most basic tool from a chemical perspective is solid-phase peptide synthesis (SPPS), a technique which allows the site-specific incorporation of PTMs on top of scalability and automatable processes, the major drawback of SPPS is the inherent size limitations of peptides that can be produced due to a coupling efficiency which is less than 100 %. In order to circumvent this issue, a protein or an extended peptide can be segmented into several shorter peptide fragments. Through a process known as ligation, these fragments can

subsequently be coupled to each other. Different types of ligation reactions exist but over time, native chemical ligation (NCL) has emerged as the most dependable method for this purpose. [44] The reason for the success of NCL is due to the chemoselectivity, the reactions conditions, which involve solely unprotected peptides under aqueous conditions and the reliability and robustness of the reaction. In the default form, the C-terminus of the N-terminal fragment has to be converted to a thioester while the C-terminal fragments must possess a cysteine residue on the N-terminal position. [45] Fueled by its nucleophilic nature, the thiol moiety of the cysteine initiates an attack on the carbonyl carbon on the thioester, kicking of the original thiol group and forming a new intermolecular thioester intermediate. Subsequently a fast S->N acyl shift occurs in which the final product is formed. [45] A scheme of the reaction can be seen in Figure 4. [45]

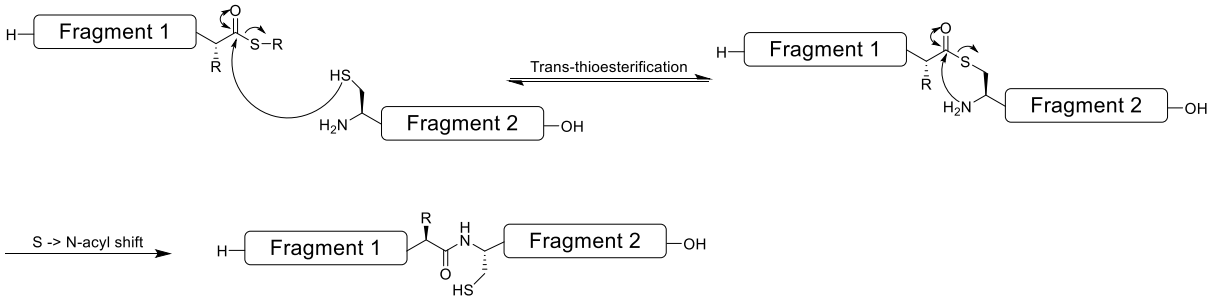

Figure 4: NCL reaction scheme [45]

The reaction is usually carried out at a neutral pH in aqueous buffer, which contains a chaotropic agent like guanidine. [45] From a kinetics perspective, the type of thioester greatly affects the ligation speed. In contrast to alkyl thioesters which react rather slow, aryl thioesters are more reactive and therefore can be used to decrease the reaction time as well as to allow kinetic control of the ligation reaction. [45] To improve the limited amount of suitable ligation sites, desulfurization reactions have been developed to convert the cysteine to an alanine which is more abundant in natural peptides. [45] Based on the desulfurization chemistry, thiolated amino acid building blocks, which can be used as an alternative to cysteine and then be converted to their native form, have been developed. [45] It has further been discovered that sulfur-based functionalities can be replaced with selenium, which is even more nucleophilic and as a consequence reacts even faster [45]. Through the study of protein splicing, an autocatalytic process, tools have been obtained for the expression of proteins with C-terminal α-thioesters. [45] This subsequently made the ligation of synthetized peptides with expressed fragments possible, a concept which is known as expressed protein ligation (EPL). [45]

1.7.1 Synthesis of peptide hydrazides

In order to be able to convert the C-terminus of a peptide into a thioester, chemical modifications have to be introduced at this C-terminal position during synthesis in order to

generate functional intermediates. The treatment of the SPPS resin with hydrazine prior to the synthesis allows the production of peptide hydrazides (PHs). To further convert a PHs to the corresponding thioester, the hydrazide is oxidized into acyl azide which then undergoes an *in situ* thiolysis to form the thioester. The conversion reaction can be performed with unprotected peptides in guanidinium buffer in a two step reaction. A scheme of the reaction, as published by Huang et al. can be seen in Figure 5. [46] Peptide hydrazides get commonly denoted by appending the term $NHNH_2$ to the peptide name.

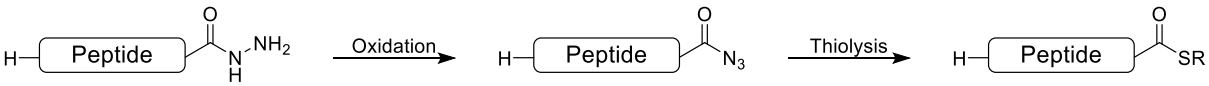

Figure 5: Processing of PH intermediates to the corresponding thioester [46]

2 Aims of the project

We aim to develop a SPPS approach for synthesizing four different variants of a specific segment within the C-terminal region of the SMN protein. These variants differ based on phosphorylations at serine residues, which are introduced during the synthesis. The serine residues under investigation are S290 and S292, chosen due to previous research that identified a connection between these phosphorylations and SMN oligomerization. [42] The target region is highlighted in red in the following amino acid sequence of the SMN protein, as encoded by the SMN1 gene.

MAMSSGGSGGGVPEQEDSVLFRRGTGQSDDSDIWDDTALIKAYDKAVASFKHALKNGDICETSGKPKTTPKRKPAKKNKS
QKKNTAASLQQWKVGDKCSAIWSEDGCIYPATIASIDFKRETCVVVYTGYGNREEQNLSDLLSPICEVANNIEQNAQENE
NESQVSTDESENSRSPGNKSDNIKPKSAPWNSFLPPPPPMPGPRLGPGKPGLKFNGPPPPPPPPPPHLLSCWLPPFPSGP
PIIPPPPPICPDSLDDADALGSMLISWYMSGYHTGYYMGFRQNQKEGRCSHSLN

The target sequence begins at P201 and extends to the terminal residue N294, resulting in a protein fragment that is 94 amino acids long and has a molecular weight of 10,106.67 Da in its unphosphorylated form. The plan was to synthesize four variants: SMN 201-294 wild type (WT), SMN 201-294 pS290, SMN 201-294 pS292, and SMN 201-294 pS290 pS292, in sufficient quantities. These variants were intended for cellular assays to investigate how terminal phosphorylation affects the peptides' oligomerization capabilities together with the other proteins of the SMN complex.

3 Materials

3.1 Water

Water deionization was carried out utilizing a Professional G7895 Aqua Purificator manufactured by Miele GmbH (Salzburg, Austria). Further purification was achieved employing

a Milli-Q Reference A+ water purification system manufactured by Merck GmbH (Vienna, Austria).

3.2 Chemicals

During this master's thesis, the solvents used—acetonitrile (ACN), N,N-Dimethylformamide (DMF), and dichloromethane (DCM)—were obtained in either high-performance liquid chromatography (HPLC) grade or peptide synthesis grade from VWR International LLC (Vienna, Austria), Sigma-Aldrich by Merck GmbH (Vienna, Austria), and Carl Roth (Karlsruhe, Germany). For peptide synthesis, three types of resin were utilized: a 2-Chlorotrityl chloride resin from Iris Biotech GmbH with a loading of 1.60 mmol/g, a ProTide Cl-TCP(Cl) resin with a loading of 0.42 mmol/g, and a preloaded Asn(Trt) TentaGel R PHB resin from Rapp Polymere GmbH with a loading of 0.19 mmol/g. The 9-fluorenylmethoxycarbonyl (Fmoc) protecting group strategy was employed, using N,N'-Diisopropylcarbodiimide (DIC), ethyl 2-cyano-2-(hydroxyimino) acetate (Oxyma), and N,N-Diisopropylethylamine (DIPEA) for activation in the automated synthesis, while O-(7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) was used for activation in the manual synthesis. For preparative HPLC, two different mobile phases were used: Solvent A, consisting of 0.10% TFA in water, and Solvent B, composed of 0.08% TFA in ACN.

4 Methods

4.1 Synthesis strategy

To address the size limitations of solid-phase peptide synthesis (SPPS), a ligation approach was developed, dividing the SMN protein sequence into smaller, more manageable peptide fragments. These fragments were designed not only to be the appropriate size but also to have similar solubility, allowing for efficient native chemical ligation (NCL). The target region was split into three segments, each synthesized using SPPS and then joined via NCL. The presence of a native cysteine in the sequence influenced the design of these fragments.

The first segment, SMN 201-230, spans from P201 to S230, while the second segment, SMN 231-249, is a 19-residue peptide covering C231 to I249. Both were synthesized as peptide hydrazides to enable NCL. For the third segment, SMN 250-294, four different variants were produced, each reflecting a distinct phosphorylation pattern. SMN 250-294 WT corresponds to the wild-type sequence without phosphorylation. SMN 250-294 pS290 contains a phosphoserine at position 290, SMN 250-294 pS292 has a phosphoserine at position 292, and SMN 250-294 pS290 pS292 features phosphorylations at both positions 290 and 292.

In the following sequence of the target compound, SMN 201-230 is highlighted in blue, SMN 231-249 in red, and SMN 250-294 in green.

H-PGPRLGPGKPGLKFNGPPPPPPPPPPHLLSCWLPPFPSGPPIIPPPPPICPDSLDDADALGSMLISWYMSGYH
TGYYMGFRQNQKEGRCSHSL-OH

After production via SPPS, the peptide hydrazides were converted to the corresponidng thioesters and coupled via NCL, the ligation scheme is depicted in Figure 6.

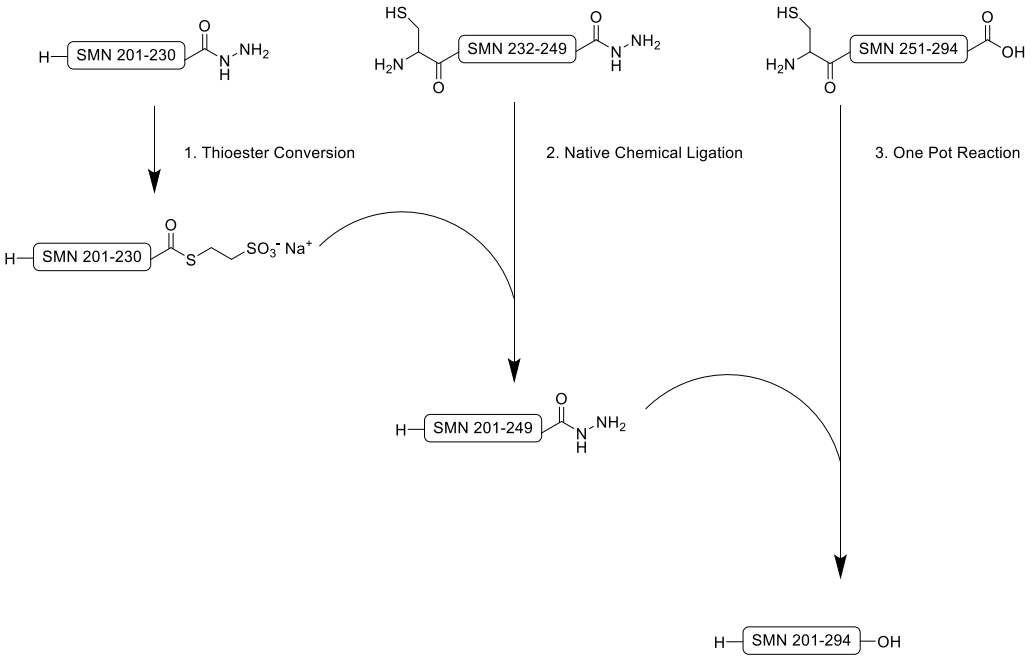

Figure 6: Scheme of the ligation strategy

4.2 Conversions of ProTide Cl-TCP(Cl) resin to the corresponding hydrazide resin

The ProTide Cl-TCP(Cl) resin had a relatively low initial loading of 0.42 mmol/g. To further reduce the loading, a 1:1 mixture of Fmoc- and Boc-protected amino acids was used. The conversion of this resin was performed three times. For the SMN 231-249-$NHNH_2$ fragment, the conversion was done twice, resulting in loadings of 0.22 mmol/g and 0.21 mmol/g. With 103 mg and 230 mg of resin, respectively, this led to total synthesis scales of 0.02 mmol and 0.05 mmol.

For the SMN 201-230-$NHNH_2$ fragment, the conversion was carried out only once. A 249 mg portion of resin was treated, achieving a loading of 0.32 mmol/g, which resulted in a synthesis scale of 0.08 mmol.

4.3 Conversion of 2-Chlorotrityl chloride resin to the corresponding hydrazide resin

The initial loading of the 2-Chlorotrityl chloride resin was 1.60 mmol/g. To reduce the loading to a level similar to that of the Cl-TCP(Cl) resin, a mixture of three equivalents of Boc-protected amino acid and one equivalent of Fmoc-protected amino acid was used. For the SMN 231-249-$NHNH_2$ fragment, the conversion was performed twice. First, 115 mg of resin was converted to hydrazide resin with a loading of 0.29 mmol/g, resulting in a synthesis scale of

0.03 mmol. In the second conversion, 400 mg of resin was treated, achieving a loading of 0.17 mmol/g, corresponding to a scale of 0.07 mmol.

For the SMN 201-230-$NHNH_2$ fragment, the conversion using the 2-Chlorotrityl chloride resin was done once. The loading of 312 mg of resin was reduced to 0.32 mmol/g, resulting in a synthesis scale of 0.10 mmol.

4.3.1 Hydrazide conversion

The SMN201-230 and SMN231-249 fragments were synthesized as peptide hydrazides, requiring a conversion step to prepare the resin. An appropriate amount of 2-Chlorotrityl chloride resin or Cl-TCP(Cl) resin was weighed out and subjected to three washes with DMF, followed by three washes with DCM, and another three washes with DMF, using approximately 5 ml of solvent for each wash. The resin was then swelled for 30 minutes in a 1:1 mixture of DMF/DCM. After swelling, the solvent was drained, and 5 ml of a solution containing 10% hydrazine monohydrate in DMF was added. The conversion reaction was carried out for 30 minutes at room temperature, after which the resin was washed with DMF. The resin was then incubated for an additional 30 minutes with another 5 ml of the 10% hydrazine solution. After this second incubation, the resin was again washed three times with DMF, three times with DCM, and three more times with DMF, using 5 ml of solvent for each wash.

Following the completion of the resin preparation, the first amino acid in the peptide sequence, protected with Fmoc and Boc groups, was coupled to the resin. Detailed information regarding the specific amino acids and their coupling is provided in the relevant sections. After the coupling reaction, the resin was washed with DCM and left to dry overnight in a desiccator. Due to the reduced loading capacity caused by the conversion process, a determination of the resin's loading was performed the following day.

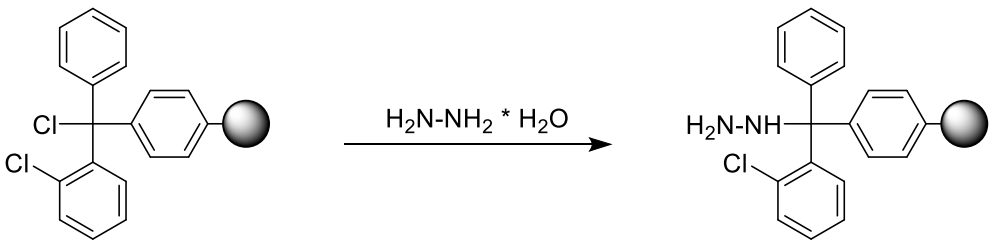

Figure 7: Scheme depicting the hydrazine conversion reaction for the 2-Chlorotrityl chloride resin

4.3.2 Loading determination

An amount of 10 mg of the dried resin was transferred into a 10 ml volumetric flask, which was then filled with 20% piperidine in DMF. The resin was thoroughly suspended and incubated for 30 minutes at room temperature. After incubation, 100 µl of the supernatant was taken and further diluted with 20% piperidine in DMF in a cuvette. The absorbance of the sample was

measured at 301 nm using a Bormac ONDA UV-30 photometer, with a blank of 20% piperidine in DMF. The loading of the resin was determined using the following equation.

$$Loading\ \left[\frac{mmol}{g}\right] = \frac{A * d}{\varepsilon * l * m} * V * 10^3$$

Equation 1: Loading determination formula

In Equation 1, A represents the UV absorbance at 301 nm. ε denotes the extinction coefficient of the dibenzofulvene adduct formed during the reaction, with a value of 7731.9 mol^{-1} L cm^{-1}. m is the mass of the resin used, measured in milligrams (mg), while d is the dilution factor, which in this case is 10. l refers to the path length of the cuvette in centimeters (cm), and V is the volume of the 20% piperidine solution used for deprotection, measured in milliliters (ml).

4.4 Synthesis of SMN 201-230-$NHNH_2$

The production of SMN 201-230-$NHNH_2$ involved a combination of manual and automated synthesis. The first 17 amino acids, up to phenylalanine, were coupled manually due to the high content of proline and other bulky amino acids in the initial portion of the sequence. Double couplings were performed for each manually coupled amino acid. It was observed that extending the coupling time for proline beyond 30 minutes did not improve the final product. Additionally, the volume of DMF used during washing significantly impacted purity, with a minimum of 40 ml of DMF required for each washing step to achieve optimal results. After the manual coupling, the resin was transferred to the automated synthesizer to complete the remaining sequence.

Peptide Sequence: H-PGPRLGPGKPGLKFNGPPPPPPPPPPHLLS-$NHNH_2$

4.5 Synthesis of SMN 231-249-$NHNH_2$

The decision was made to manually synthesize the entire SMN 231-249-$NHNH_2$ fragment due to its relatively short length of 19 residues and the challenging sequence at the beginning. The synthesis followed the default procedure, with double couplings performed throughout. To ensure purity, a minimum of 40 ml of DMF was used for each washing step. After confirming that the ligation strategy was effective in both directions, it was decided to leave the terminal cysteine unprotected.

Peptide Sequence: H-CWLPPFPSGPPIIPPPPPI-$NHNH_2$

4.6 Synthesis of SMN 250-294 and its variations

The wild-type (WT) version of the final fragment was synthesized entirely on the automated synthesizer. For the variants with phosphorylations, a hybrid approach was used, similar to the one applied for SMN 201-230-$NHNH_2$. In this case, only the first six amino acids were coupled manually due to the sensitivity of the serine phosphate group, which is prone to elimination

under microwave irradiation on the synthesizer. Since no hydrazide conversion was required, the 0.19 mmol/g preloaded Asn(Trt) TentaGel resin was utilized. The synthesis scale was set to 0.10 mmol for each peptide variant, with only single couplings performed during the manual coupling of the first six amino acids.

Peptide Sequence: H-CPDSLDDADALGSMLISWYMSGYHTGYYMGFRQNQKEGRCSHSLN-OH

4.6.1 Manual peptide synthesis

Peptides containing proline-rich sequences were synthesized using manual Fmoc SPPS. Each synthesis was carried out in a polypropylene syringe equipped with a polyethylene filter. The resin was allowed to swell in DMF for at least two hours before starting the synthesis.

Deprotection of the Fmoc protecting group was performed using a solution of 20% piperidine in DMF. The deprotection was executed in two steps: an initial treatment for seven minutes followed by a second treatment for three minutes. After deprotection, the resin was washed five times with DMF.

For the coupling step, four equivalents of the amino acid were dissolved in 3.8 equivalents of 0.5 M HATU in DMF. After dissolving the amino acids for three minutes, 8 equivalents of DIPEA were added to the solution. This mixture was incubated for an additional minute before being added to the resin. A coupling time of 30 minutes was employed. For amino acids that followed bulky building blocks, a second coupling step with the same amino acid was carried out without deprotection or additional washing.

Following the coupling step, the resin was washed five times with DMF. After the final amino acid was coupled, a final deprotection step was performed, and the resin was washed with DCM in preparation for drying under vacuum.

4.6.2 Automated peptide synthesis

In addition to manual synthesis, some sequences or portions of peptide sequences in this study were produced using an automated synthesizer. The Liberty Prime Microwave Peptide Synthesizer (CEM GmbH, Kamp-Lintfort, Germany) was employed for this purpose. The resin was first swollen in DMF for 2 hours at room temperature before being loaded onto the synthesizer.

Fmoc protecting group removal was achieved with a 20% piperidine in DMF solution, similar to the manual synthesis. This deprotection step was conducted at 90°C for 40 seconds.

For coupling, amino acids were dissolved in DMF to a concentration of 0.5 M and activated using a mixture of 2 M DIC, 0.25 M Oxyma, and 0.1 M DIPEA, all prepared in DMF. The ratio of amino acid to DIC to Oxyma relative to the scale was maintained at 5:12:5. Amino acid

coupling was carried out for 80 seconds at 105°C. Following each coupling, the resin was washed with DMF.

4.6.3 Peptide Cleavage

The completed peptide sequences were cleaved from the resin using a cleavage cocktail. For most peptides, the cocktail comprised 95% TFA, 2.5% TIS, and 2.5% H_2O. However, for the SMN 250-294 fragment, which contains methionine, a modified cocktail was used to prevent oxidation: 85% TFA, 5% TIS, 2.5% H_2O, 2.5% EDT, and 5% DMS.

Before cleavage, the resin needed to be completely dry. For each 100 mg of resin, 1 ml of the cleavage cocktail was used. The cleavage reaction was carried out for a minimum of two hours. Due to the presence of arginine in the SMN 250-294 fragment and the use of the 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) protecting group on its side chain, the reaction time was extended to a minimum of three hours.

After cleavage, the peptide was precipitated by adding the solution to a tube containing 10 ml of ice-cold ether. Following centrifugation, the ether was removed, and the precipitate was dried in air. The peptide was then re-dissolved in water; if necessary, ACN was added to aid dissolution. Finally, the solution was lyophilized to obtain the purified peptide.

4.7 Preparative RP-HPLC

Preparative reversed-phase (RP)-HPLC purification was conducted using a Waters system equipped with a Waters 2545 binary gradient module, a Waters 2489 UV/visible detector, and a Waters Fraction Collector III. Depending on the quantity of material to be purified, either a Kromasil 300-10-C4 Prep Column (21.2 x 250 mm) or a Waters 300-10 C4 Semiprep Column (10 x 250 mm) was used as the stationary phase. Solvent A and Solvent B, as previously described, served as the mobile phases.

The crude materials were dissolved in 9 mL of 6 M Guanidine HCl (GuaHCl) and filtered through a 0.2 µm PES syringe filter before being introduced into the HPLC system. Product elution was achieved using various linear gradients, tailored to the solubility of the product, which was determined in advance using analytical RP-HPLC. Collected fractions were analyzed by direct injection into an ESI-MS system.

4.8 Analytical RP-HPLC

Analytical reversed-phase (RP)-HPLC was carried out using a Dionex Ultimate 3000 HPLC system, paired with a DAD 300 detector. The analysis was performed on a Waters XBridge® Protein BEH C4 column (4.6 × 150 mm) as the stationary phase. Solvents A and B, as described for preparative HPLC, were utilized for the mobile phases.

analysis. Once the reaction was complete, it was quenched by adding 6 M guanidine solution.

4.9 Conversion of SMN 201-230 NHNH2 to the thioester

The conversion of SMN 201-230 was initially attempted using the more reactive MPAA (4-Methylmorpholine-4-oxide). However, due to issues with hydrolysis of the thioester, MPAA was replaced with MESNa (2-mercaptoethane sulfonate) in the second reaction step. For each conversion reaction, between 15 mg and 20 mg of peptide hydrazide were used. The yield of the conversion was determined to be in the range of 55% to 65%.

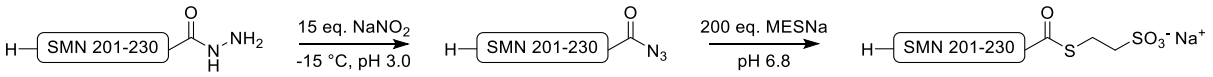

Figure 8: Scheme depicting the conversion to the 201-230 MESNa thioester

4.9.1 Thioester conversion

The conversion of peptide hydrazide to the corresponding thioester was conducted at pH 3.0 in aqueous conditions. A degassed buffer system at pH 3.0, consisting of 0.2 M NaH_2PO_4 in 6 M guanidine solution, was used. The purified peptide hydrazide was dissolved in this phosphate buffer within an Eppendorf tube containing a stirring bar, and the reaction mixture was cooled to -15°C using an ice bath.

Fifteen equivalents of $NaNO_2$ were added in the form of a 0.5 M solution. After 30 minutes, 100 equivalents of 2-mercaptoethane sulfonate (MESNa), also dissolved in the phosphate buffer, were introduced into the reaction mixture. The tube was then transferred to a water bath set at 25°C. The progress of the reaction was monitored at regular intervals using LC-MS

4.10 Ligation of SMN 201-230-TE and SMN 231-249-$NHNH_2$ to SMN 201-249-$NHNH_2$

For the native chemical ligation (NCL) of SMN 201-249-$NHNH_2$, between 10 mg and 20 mg of SMN 201-230-TE were used in conjunction with 4 mg to 9 mg of SMN 231-249-$NHNH_2$. The ratio was set so that for each equivalent of PH, 1.5 equivalents of TE were used. To enhance the reactivity of the MESNa thioester, 200 equivalents of 4-Mercaptophenylacetic acid (MPAA) were added to the reaction mixture. The yield, calculated based on the amount of PH used, ranged from a minimum of 13% to a maximum of 25%, with an average yield of approximately 20%.

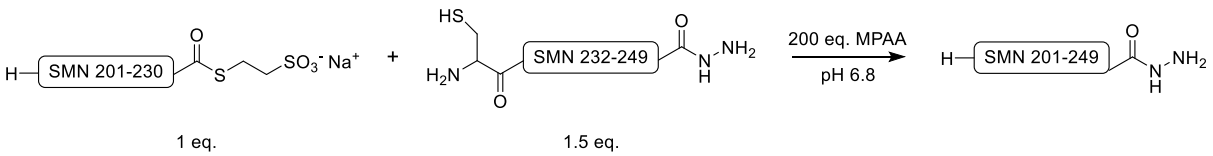

Figure 9: Scheme depicting the formation of the SMN 201-249 ligation product

4.11 One pot conversion and ligation of SMN 201-249-$NHNH_2$ and SMN 250-294 to SMN 201-294

The reaction to synthesize the final product was performed as a one-pot reaction. For the synthesis of SMN 201-294 WT, 1.73 mg of SMN 201-249-$NHNH_2$ (1.5 equivalents) and 1.17

mg of SMN 250-294 WT (1 equivalent) were used, along with 200 equivalents of MPAA. The reaction was carried out overnight. To address disulfide bond formation, a few crystals of TCEP were added to the mixture. The reaction yielded a total of 0.5 mg of final product, corresponding to a yield of 22%.

For the synthesis of SMN 250-294 pS290, 2.6 mg of SMN 201-249-$NHNH_2$ and 1.8 mg of SMN 250-294 pS290 were used, with 200 equivalents of MPAA. This maintained the same ratio as for the wild type. However, this reaction resulted in only a small amount of final product, which could not be quantified.

In the synthesis of SMN 201-294 pS292, both educts were used at a 1:1 ratio, with 2.4 mg of SMN 201-249-$NHNH_2$ and 2.6 mg of SMN 250-294 pS292. This reaction produced a mixture of final product and peptide with a total weight of 0.4 mg.

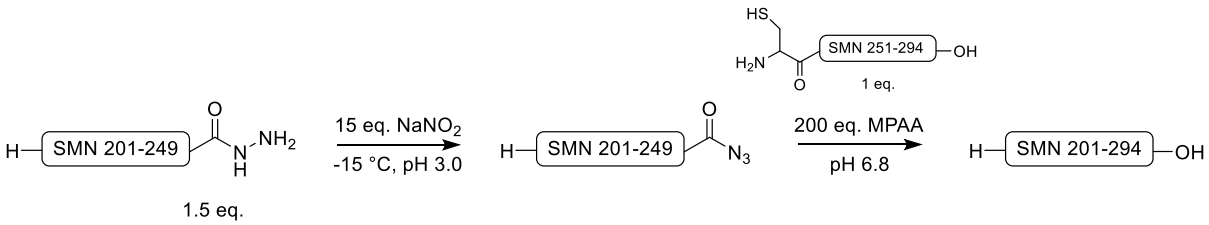

Figure 10: Formation of the final product via a one-pot reaction

4.11.1 Peptide ligation

Native chemical ligation was conducted under buffered aqueous conditions. A pH 6.8 buffer system was prepared, consisting of 0.2 M Na_2HPO_4 in 6 M guanidine solution, and was degassed under nitrogen. Both the peptide thioester and the peptide with the free cysteine were dissolved and combined in an Eppendorf tube equipped with a stirring bar. Additionally, between 100 and 200 equivalents of 4-Mercaptophenylacetic acid (MPAA) were dissolved in the ligation buffer and added to the reaction mixture. To facilitate the dissolution of MPAA, the pH of the ligation buffer was adjusted to approximately 6.8.

The ligation was carried out in a water bath at 25°C. The progress of the reaction was monitored starting from time point zero and subsequently at one-hour intervals using LC-MS analysis. Once the reaction was complete, it was quenched by adding 6 M guanidine solution.

5 Results

5.1 Peptide hydrazide synthesis

A manual synthesis of SMN 231-249-$NHNH_2$ was completed prior to this master's thesis. LCMS analysis revealed that the material contained a mixture of the desired product and a side product missing one proline residue. To assess whether these compounds could be separated, an analytical HPLC run was conducted.

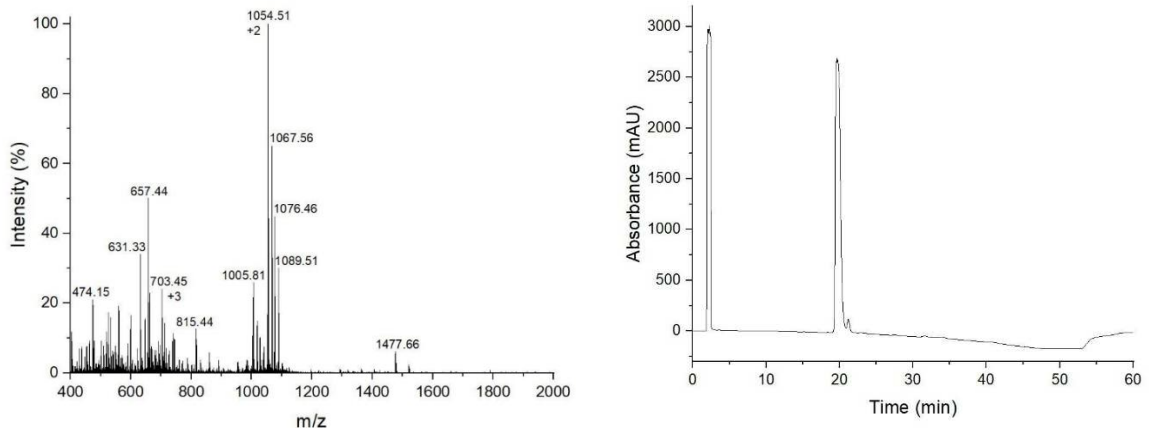

Figure 11: SMN 231-249-$NHNH_2$ crude, produced prior to the master's thesis

Based on the chromatogram, a gradient of 5-35 % B in 5 minutes, followed by a slow increase to 50 % B in 30 minutes, together with a 10-minute desalting step, was chosen. Purification was carried out on a semi-preparative scale with an input of 10 mg of the sample. Post-purification, 6.64 mg of the main pool and 0.59 mg of a side pool were recovered. However, complete removal of the six-proline side product was not achieved. The material was subsequently utilized for test ligations in an effort to optimize the synthesis of the final product. Further improvement of the synthesis process was attempted based on the new knowledge that the peptide is prone to proline insertions.

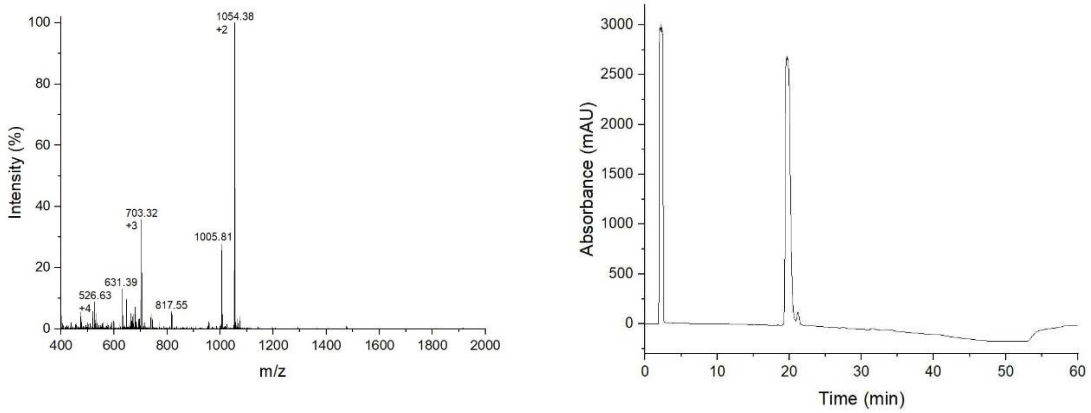

Figure 12: Purified SMN 231-249-$NHNH_2$, produced prior to the master's thesis

During the initial attempt to optimize the synthesis, a loading of 0.22 mmol/g was achieved using 103 mg of resin, corresponding to a scale of 0.02 mmol. To monitor the formation of

insertions and deletions within the five-proline sequence, test cleavages were performed after each coupled proline. Kaiser tests were also conducted to assess the completeness of the coupling steps. However, due to the poor visibility of the positive results in this detection reaction, further use of these tests was discontinued after this initial attempt. Notably, an additional proline, which got inserted during the coupling of the sixth amino acid, was detected. As a consequence, the synthesis was not continued.

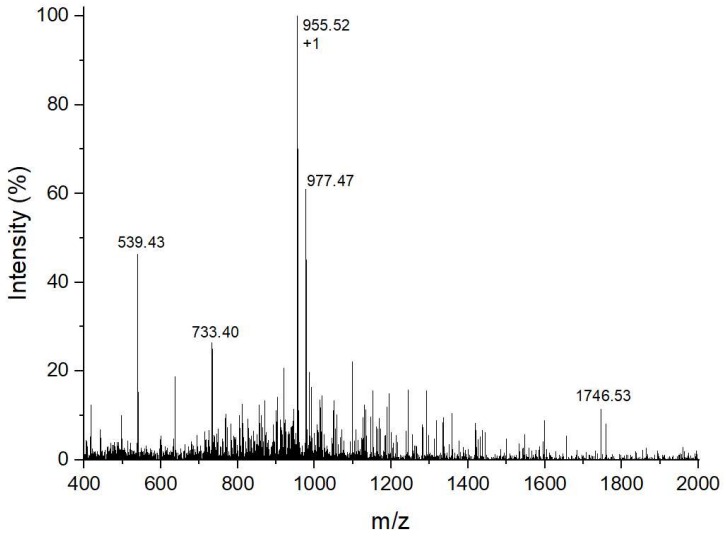

Figure 13: DI of the test cleavage after a length of six amino acids, the terminal amino acid was fmoc protected.

A second attempt was initiated using 230 mg of resin, achieving a loading of 0.21 mmol/g and a corresponding scale of approximately 0.05 mmol. In this attempt, the volume of DMF used for the washing steps was increased to 40 ml. Additionally, the syringe was manually agitated during the washing steps to ensure proper suspension of the resin. These adjustments resulted in the production of the five-proline segment with improved quality, as evidenced by the MS spectrum shown in Figure 14.

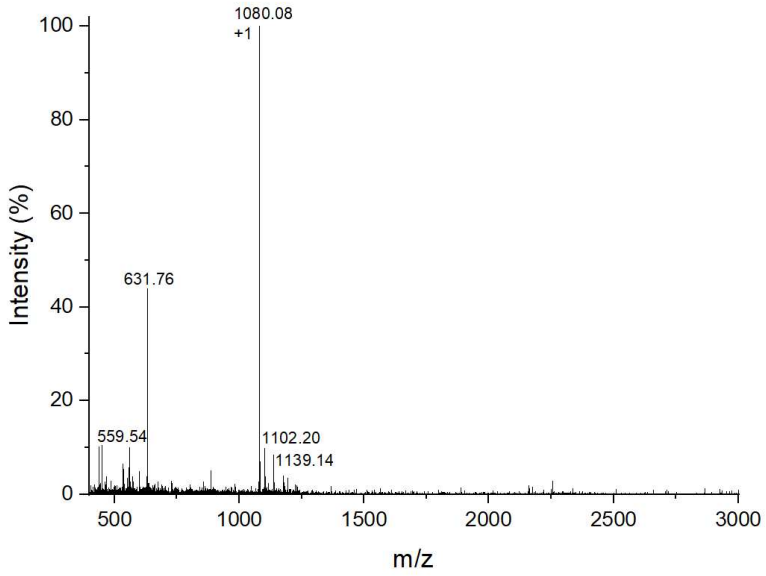

Figure 14: Test cleavage of SMN 231-249-NHNH$_2$ after eight amino acids

After confirming the absence of deletions or insertions in the initial part of the sequence, test cleavages were performed less frequently. During this attempt, it was discovered that the quality of MS spectra for cleavages involving small hydrophobic segments of the sequence could be improved by evaporating the diethyl ether in the tube and then redissolving the peptide. Following the full cleavage, the product was freeze-dried and subsequently purified the next day. For purification the same initial gradient from the first purification was used,the fractions containing the peptide were combined and freeze-dried. The following day, LCMS analysis of the peptide was conducted.

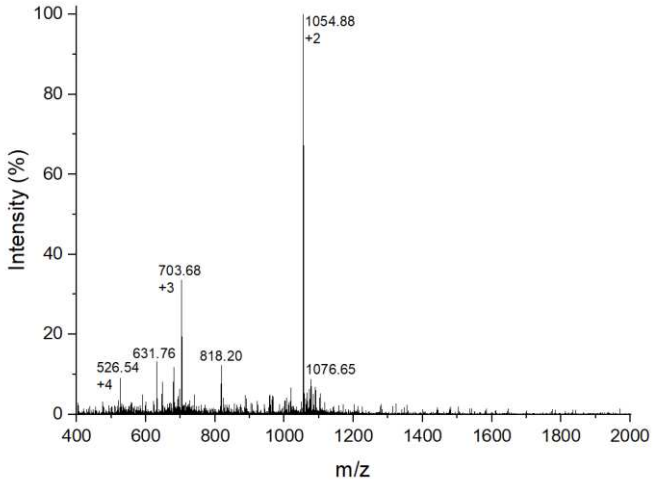

Figure 15: LCMS spectrum of the full length, Acm protected, SMN 231-249-NHNH$_2$

In this synthesis, Acm-protected cysteine was used as the terminal amino acid. Ligation trials revealed that the ligation direction did not influence the reaction speed. As a result, it was

decided to leave the terminal cysteine unprotected in future synthesis attempts to eliminate the need for a deprotection step.

To compare the outcomes of manual with automated synthesis, the same peptide was also synthesized using an automated synthesizer. In Figure 16 the analytical HPLC chromatogram before and after purification of the crude material of this attempt is shown. As can be seen, all impurities could be removed with the previously established gradient.

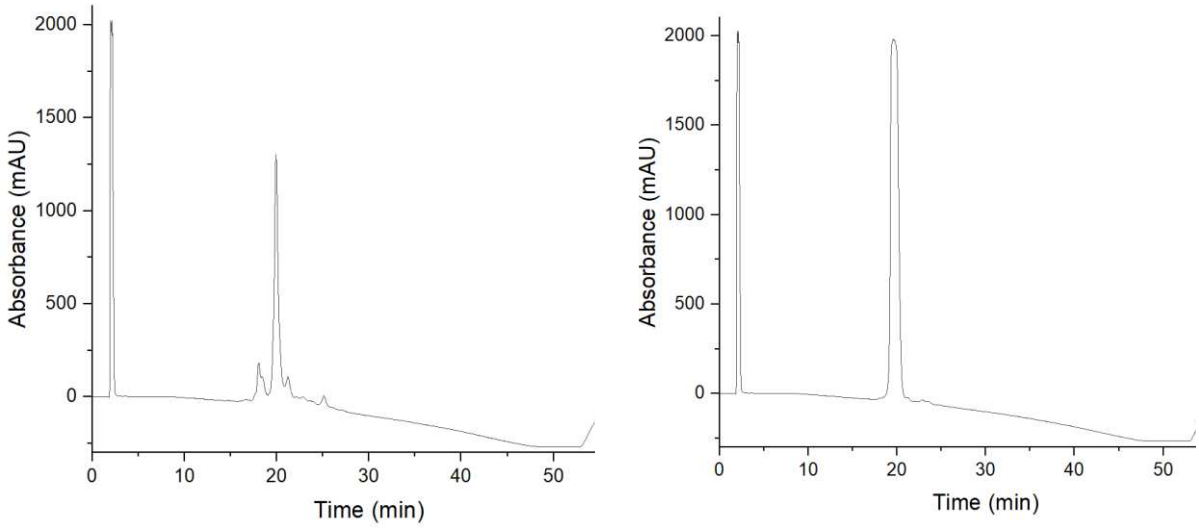

Figure 16: HPLC chromatogram of SMN 231-249-NHNH$_2$ produced by the synthesizer before and after purification

The MS spectrum of the purified material, which lacked Acm protection, is shown in Figure 17. The m/z value of 970.56 suggests the presence of a product missing one proline. This assumption is further supported by the deconvoluted spectrum, which reveals a species with a mass of 2036.68 Da, consistent with the calculated mass of the intended product, and another species with a mass of 1939.12 Da. The mass difference of 97.56 Da between these two species corresponds to the mass of a prolyl group, confirming the presence of a proline deletion.

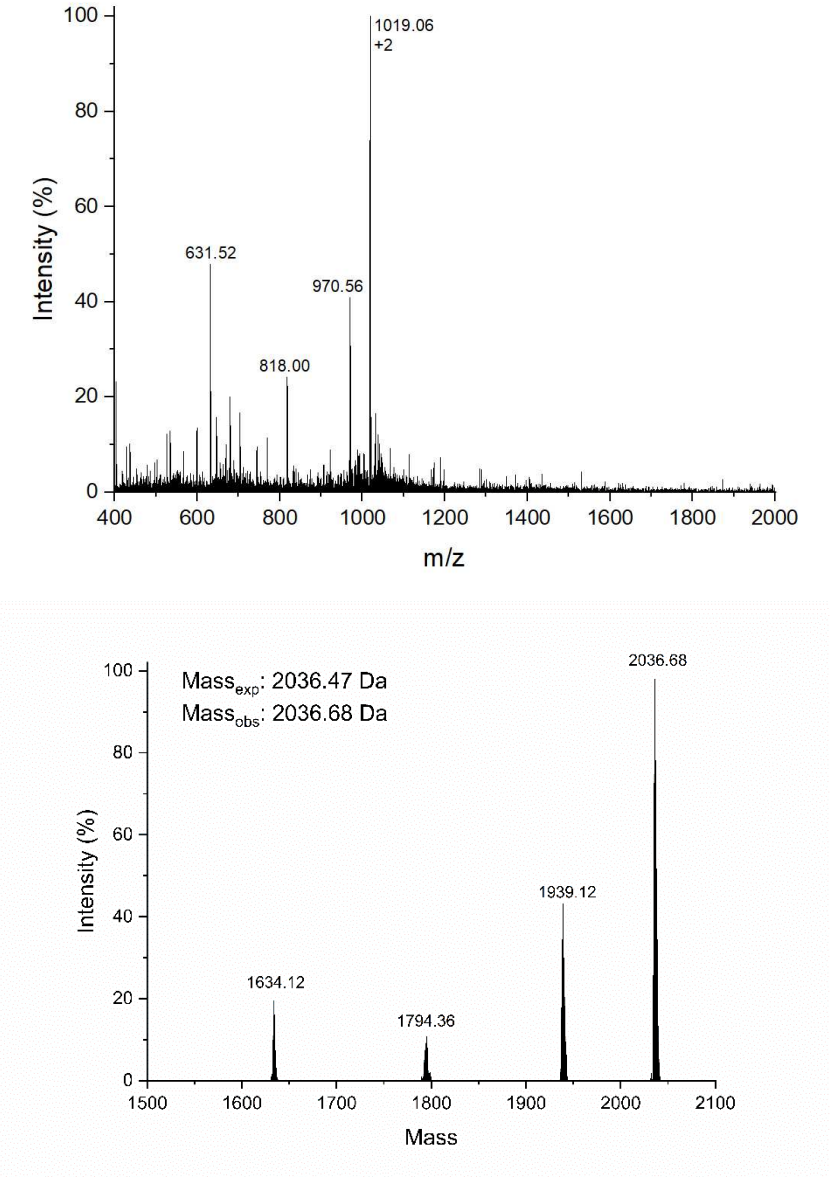

Figure 17: Full cleavage of the SMN 231-249-NHNH$_2$ peptide produced on the synthesize, the upper image shows the obtained spectrum while the image on the bottom show the deconvoluted spectrum

The differences in quality are most likely attributed to the more extensive washing steps performed during manual synthesis. It was anticipated that the automated synthesizer would yield lower quality results, as it uses only a small amount of DMF for washing. Unfortunately, due to the coelution of both species during the HPLC run, the peptide produced by the synthesizer was unsuitable for further processing. To obtain more peptide free of side products, another manual synthesis was initiated. This time, after conversion, a loading of 0.29 mmol/g was achieved, corresponding to a scale of 0.03 mmol.

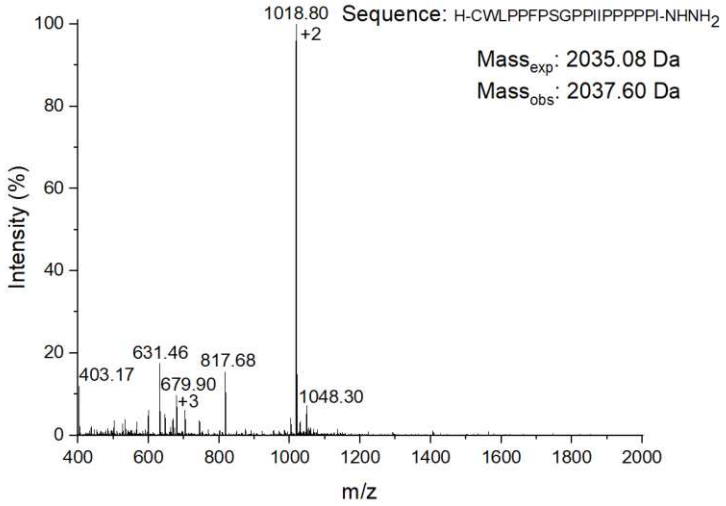

Figure 18: SMN 231-249-NHNH$_2$ produced manually

The m/z value of 631.46, observed in previous spectra, is very likely a truncation product with the sequence H-PPPPPI-NHNH$_2$. Additionally, the m/z value of 817.68 could correspond to the truncation product H-PPFPSGPPIIPPPPPI-NHNH$_2$, as the calculated m/z for the +2 species is 817.99.

The N-terminal segment posed the greatest challenge due to its ten consecutive prolines. Based on previous experience with the SMN 231-249-NHNH$_2$, it was already known that thorough washing steps are crucial for proline-rich regions. In the first attempt, a loading of 0.32 mmol/g was achieved, corresponding to a scale of 0.08 mmol. Throughout the synthesis, test cleavages were performed at regular intervals to check for side products and truncations. Before initiating the automated synthesis, a final test cleavage was conducted to verify the accuracy of the produced sequence. As can be seen in spectrum, depicted as Figure 19, a side product, containing an additional proline had likely been formed. The peak with an m/z of 925.55 indicates this species. It was decided to continue the synthesis nonetheless.

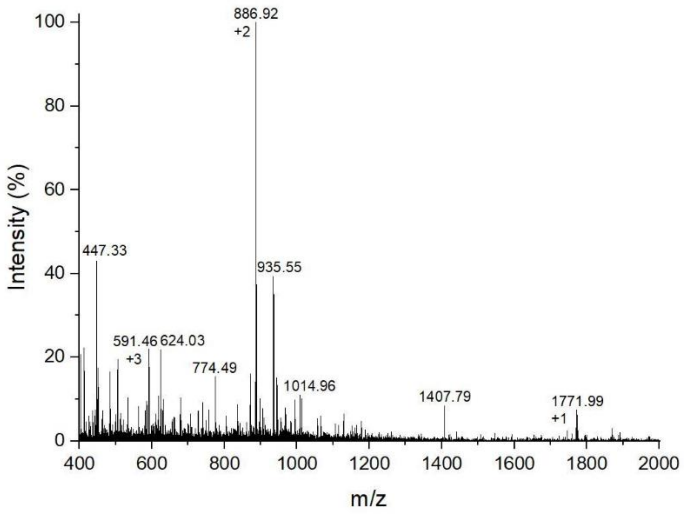

Figure 19: Test cleavage of SMN 201-230-NHNH$_2$ after 17 amino acids

After the automated synthesis, the peptide was dried in a desiccator for one and a half days. The first LC-MS run supported the theory that the crude material contained both a ten-proline species and an eleven-proline species. In this case the peak with an m/z of 1042.22 corresponds to this species.

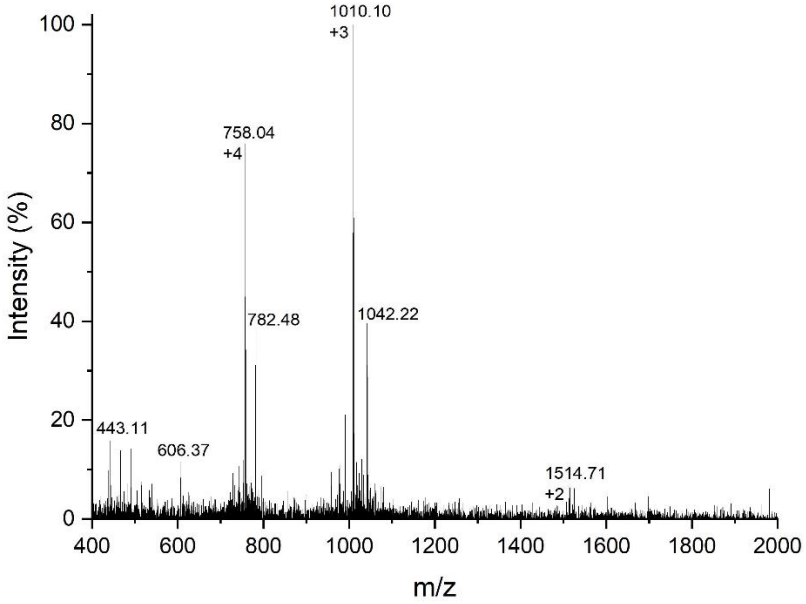

Figure 20: LCMS run of the SMN 201-230-NHNH$_2$ crude

To gather more information about the mixture, an analytical HPLC run was performed on the crude material. During this run, the eluting product peaks at 21.7 min and 23.2 min were collected and subsequently analysed through direct injections.

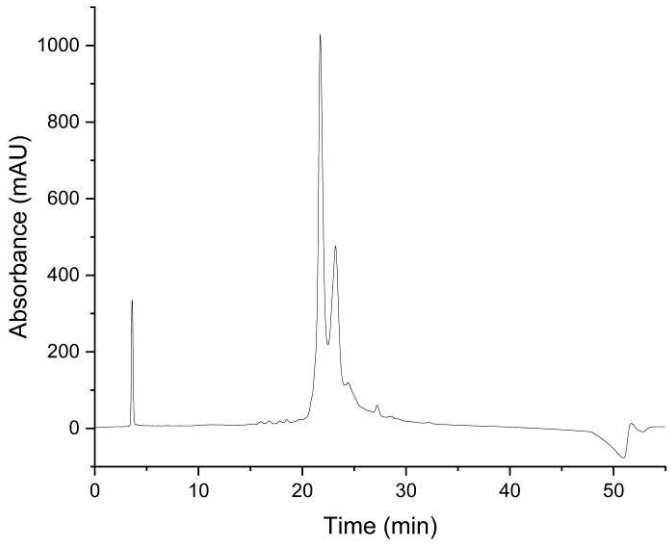

Figure 21: Analytical HPLC run of the SMN 201-230-NHNH$_2$ crude

The spectra from both peaks revealed that the ten proline and the eleven proline species eluted at different retention times, with the ten proline species belonging to the 21.7 min peak and the eleven proline species belonging to the 23.2 min peak.

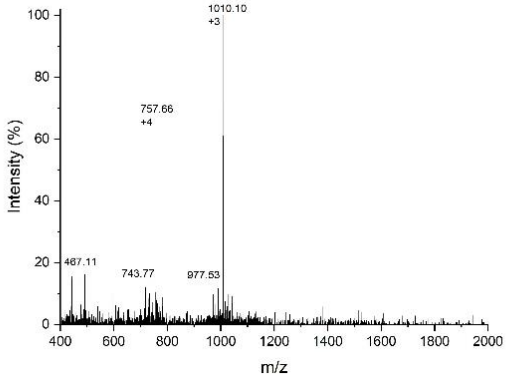

Figure 22: Direct Injections from the peak at 21.7 min

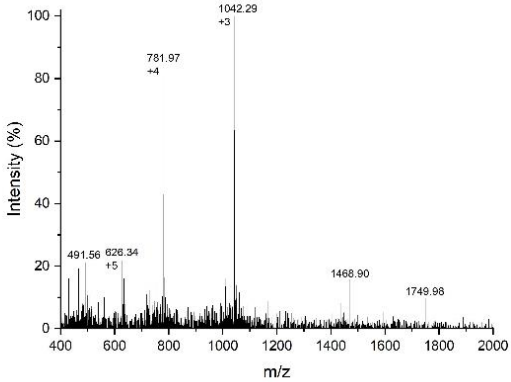

Figure 23: Direct Injection from the peak at 23.2 min

For purification a gradient from 5-35 % B in 5 minutes, followed by an increase to 55% in 30 minutes, a 10-minute desalting step was included as well. The purification was performed on prep scale. The analytical HPLC chromatogram and the LCMS spectrum of the purified product can be seen in Figure 24 and Figure 25.

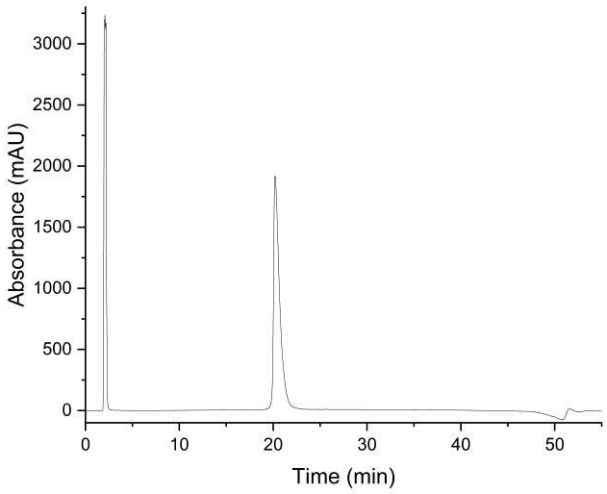

Figure 24: Analytical HPLC chromatogram of the purified SMN 201-230-NHNH$_2$

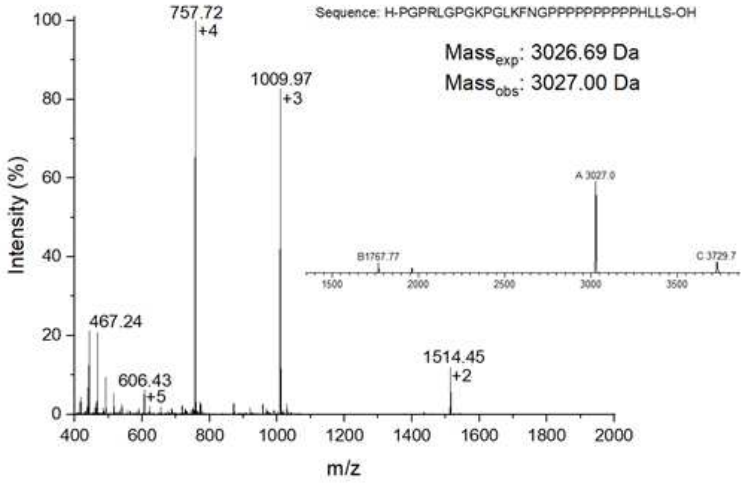

Figure 25: Final analysis of the SMN 201-230-NHNH$_2$

As shown in Figure 25, it is possible to isolate the ten-proline species from the eleven-proline species, which was unexpected given the minimal difference in polarity between them. In a second attempt, the peptide was synthesized entirely by hand. The first test cleavage was conducted after the synthesis of the ten-proline segment corresponding to a length of 14 amino acids.

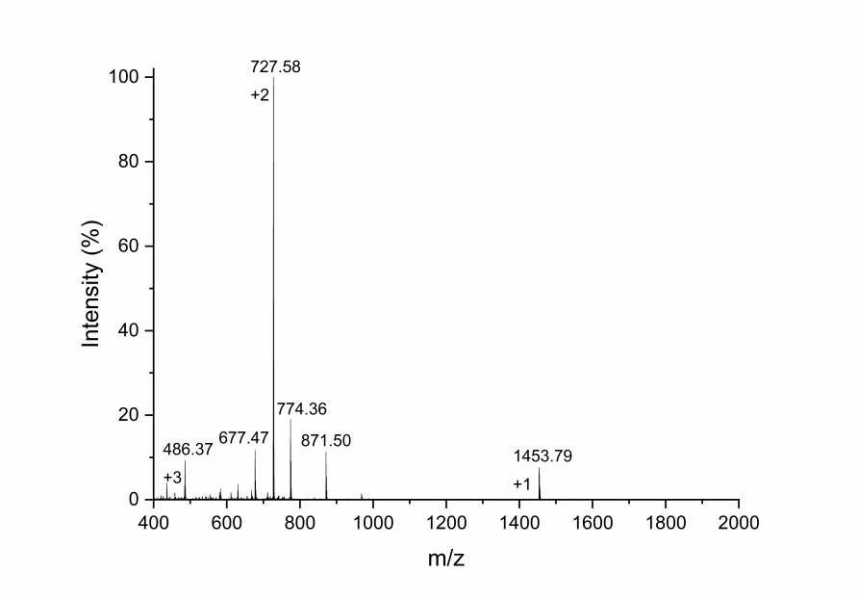

Figure 26: SMN 201-230-NHNH$_2$ after 14 AA

Throughout the entire synthesis, each amino acid was coupled using a double coupling method. Unlike the hybrid attempt, no additional proline was observed in this synthesis. Following purification, an even cleaner compound with no addition proline was obtained.

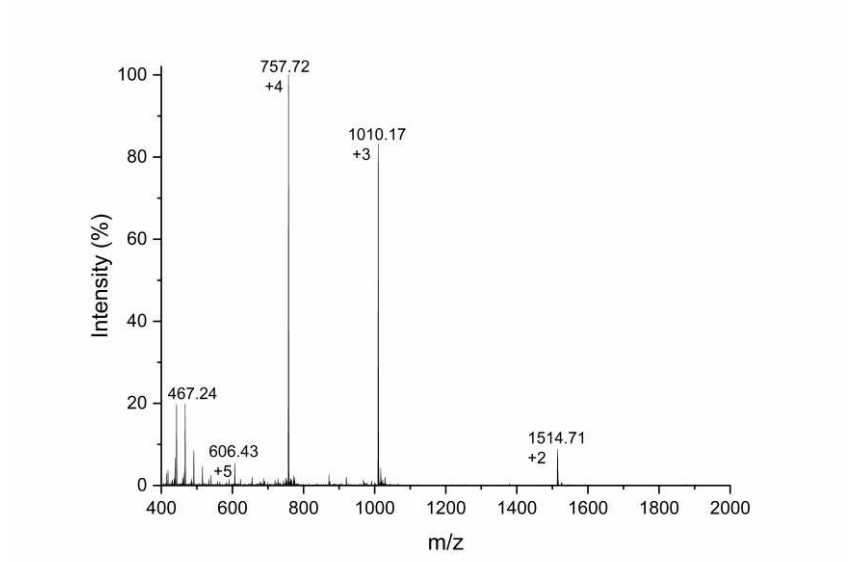

Figure 27: SMN 201-230-NHNH$_2$ full length

The remaining peptide, SMN 250-294 in its four variants, contained no proline rich regions in its sequence. Therefore, the WT version of the peptide was produced entirely on the synthesizer. The spectrum of this compound can be seen in Figure 28.

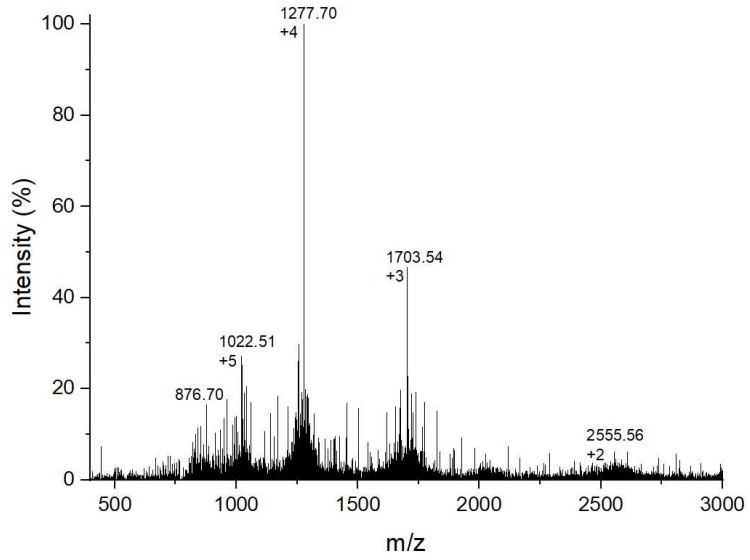

Figure 28: SMN 250-294 WT, produced on the automated synthesizer

The purification of this peptide was performed with a gradient of 5-35 % B in 5 minutes, followed by 35-57 % B in 30 minutes, a 10-minute desalting step was performed as well. The spectrum of the purified compound can be seen in Figure 29.

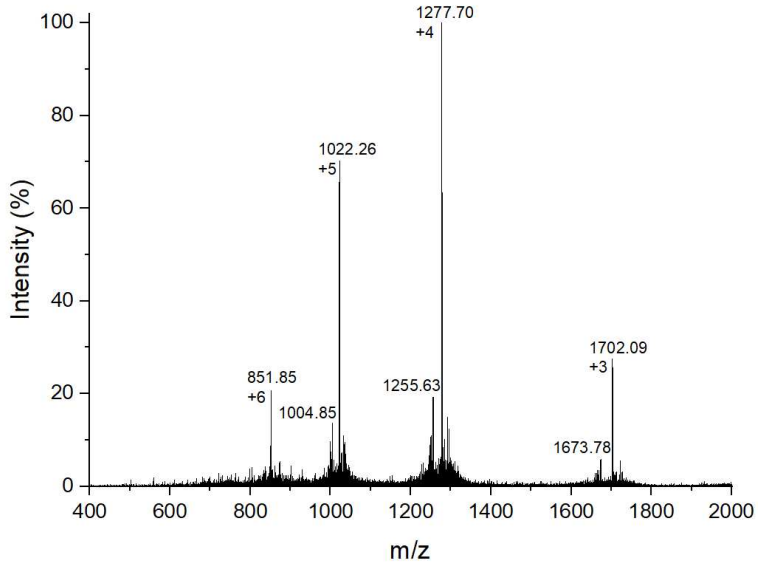

Figure 29: SMN 250-294 WT after purification

Both monophosphorylated versions of the SMN 250-294 fragment were successfully produced using a combination of manual and automated synthesis. In both attempts the first six amino acids were coupled manually through single couplings.

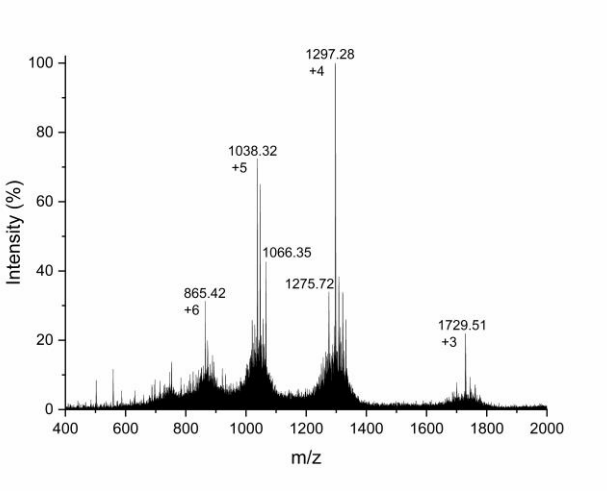

Figure 30: Crude SMN 250-294 pS290

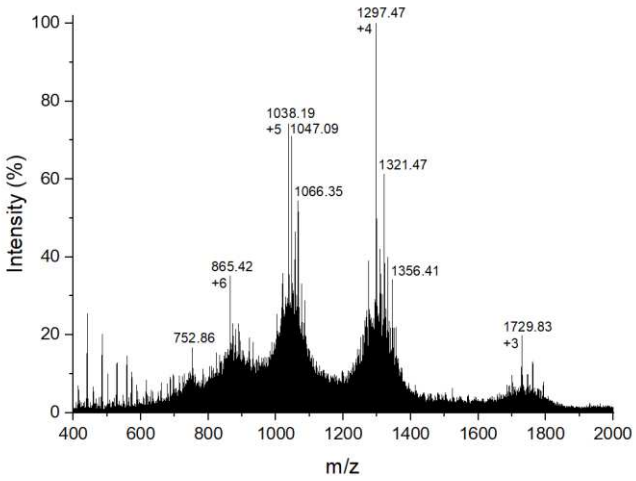

Figure 31: Crude SMN 250-294 pS292

Both peptides were purified with a gradient of 5-20 % B in 15 minutes, followed by an increase to 70 % B in 40 minutes, a 10 minute desalting step was included. The results of the purifications can be seen in the following Figures.

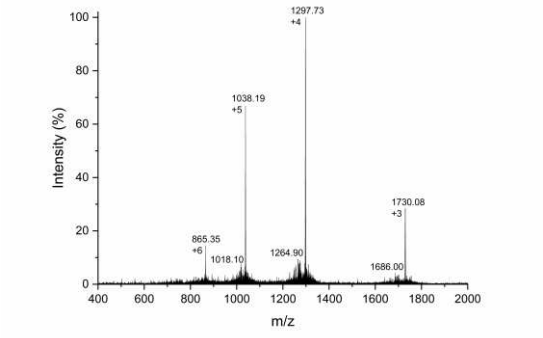

Figure 32: SMN 250-294 pS290 purified

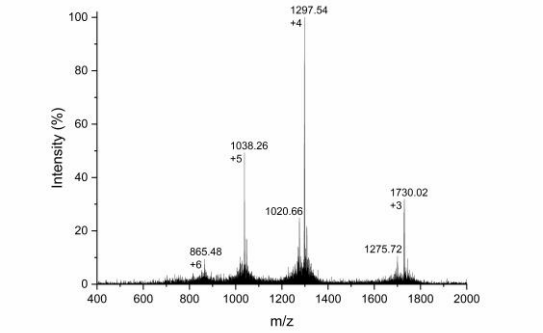

Figure 33: SMN 250-294 pS292 purified

In contrast, the spectrum obtained for the diphosphorylated peptide, which was produced in the same way as the monophosphorylated variants, revealed that an error occurred during the synthesis as no accurate mass could be identified. Since all three peptides were produced simultaneously using the same setup, it was challenging to pinpoint the specific issues that occurred during the synthesis.

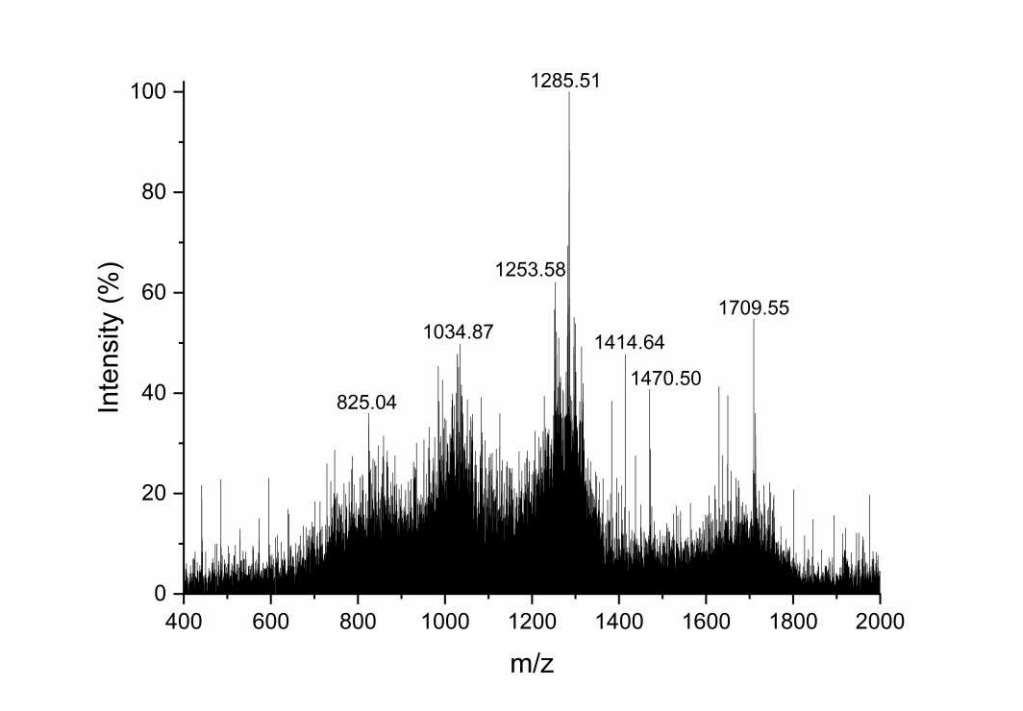

Figure 34: SMN 250-294 pS290 pS292-NHNH$_2$ after the first attempt of synthesis

To ensure the accuracy of the synthesis, the peptide was repeated using the same strategy as before, with 570 mg of resin, resulting in a scale of approximately 0.11 mmol. The first six amino acids were coupled manually. A test cleavage was conducted just before transferring the peptide to the synthesizer to confirm the correct mass. The resulting pellet was re-dissolved and loaded onto the analytical HPLC. The peaks, which eluted from the column, were collected and analyzed via direct injections.

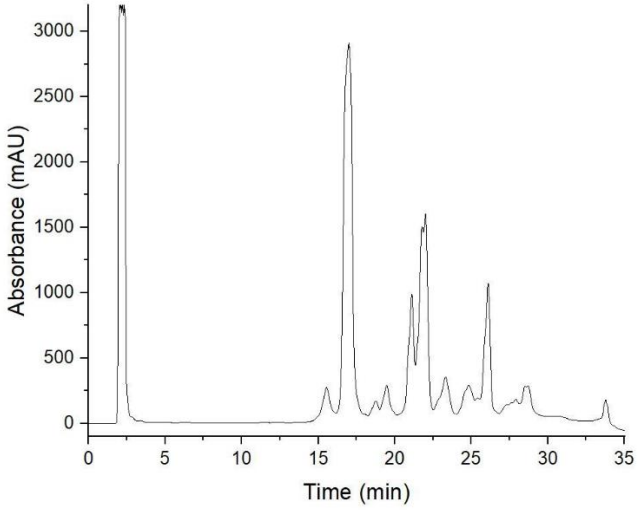

Figure 35: Chromatogram of SMN 250-294 pS290 pS292 resin before it was put onto the synthesizer

The DI from the peak at 16.90 min showed the +1 and +2 m/z of the peptide protected with Fmoc. Both spectra from the double peak at 21.80 min showed the correct mass plus one

benzyl protecting group. This is common because the phosphoserine building block is protected by a benzyl group. The double peak can potentially come from the two possible positions on which the benzyl group can remain. The remaining peak at 26.10 min corresponds to the correct mass plus two benzyl groups, indicating from a peptide where both phosphoserines remained protected.

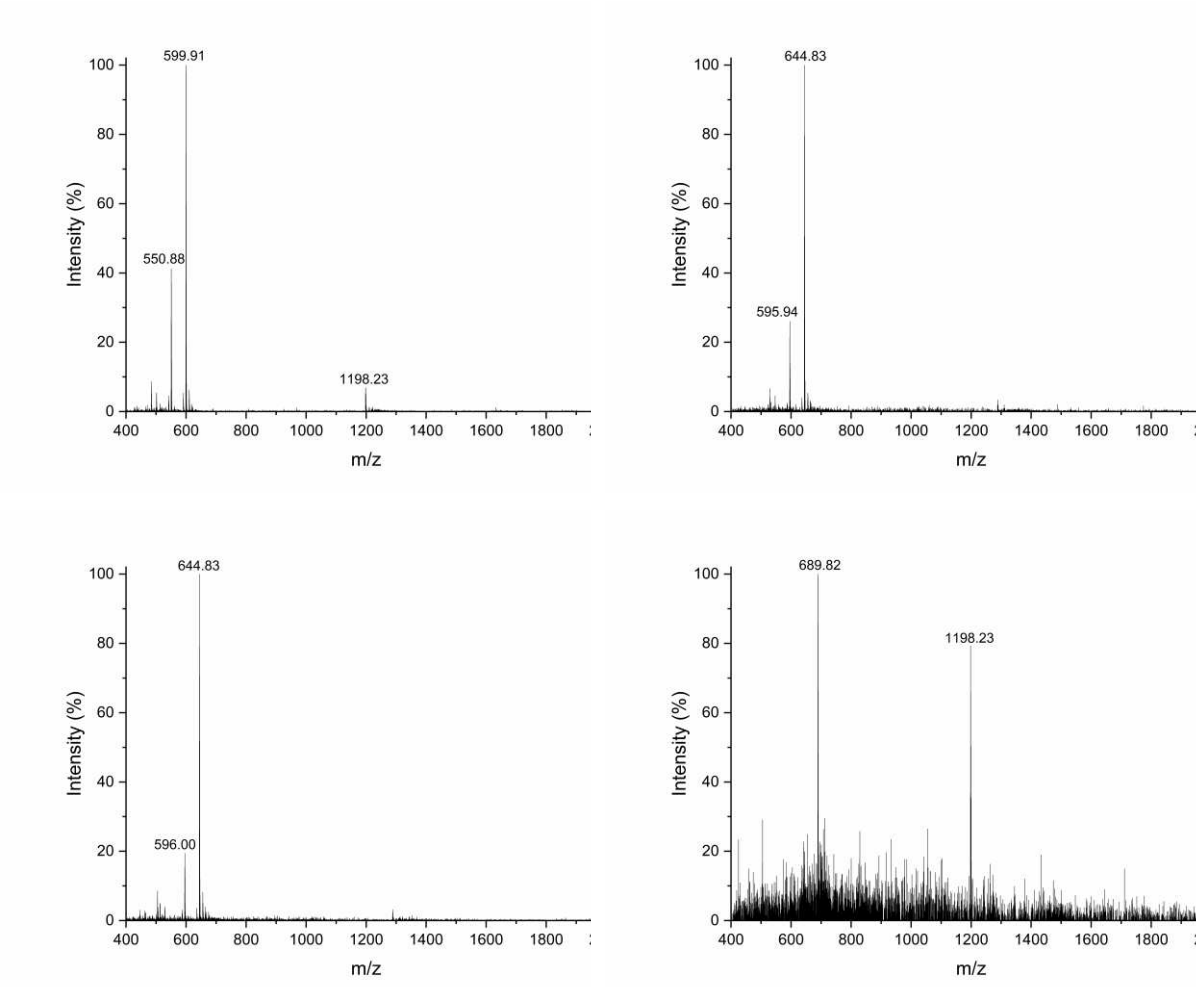

Figure 36: DIs from the peaks at 16.80 min (top left), 21.80 min (top right and bottom right) and 26.10 min (bottom right)

As the correct mass was verified the synthesis was continued on the synthesizer. Due to the presence of two phosphorylations on the peptide, it was anticipated that MALDI-TOF analysis would be necessary for accurate detection. To prepare the sample, the peptide was purified using a gradient of 5-20% solvent B over 15 minutes, followed by a gradient of 20-70% solvent B over 40 minutes. A 10-minute desalting step was also performed. This slow gradient was chosen based on an analytical HPLC run of the crude material. Following purification, an additional LC-MS run was conducted on a standard instrument.

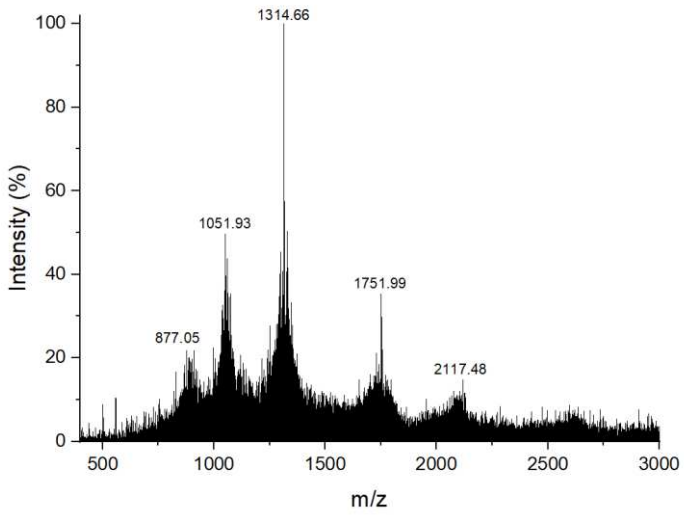

Figure 37: Crude of SMN 250-294 pS290 pS292 after purification, product could only be detected in traces

A dry sample was submitted to the mass center. The high-resolution MS spectrum obtained did not reveal the expected mass of the sample, once again raising questions about what might have gone wrong during the automated synthesis.

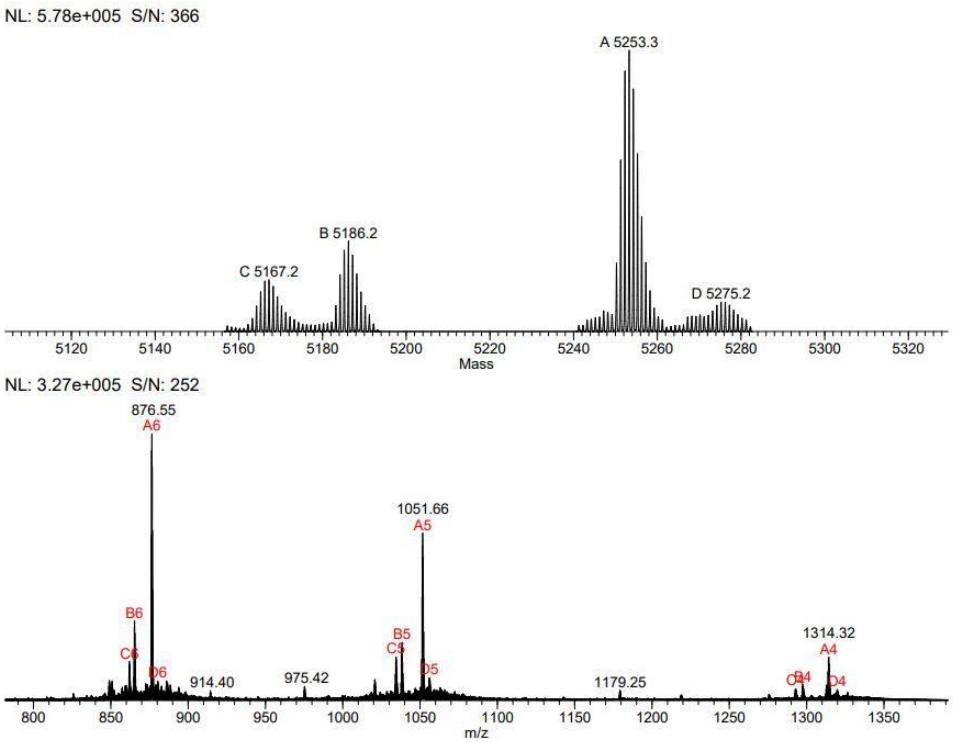

Figure 38: High resolution spectrum of the purified SMN 250-294 pS290 pS292

The main peak shows a mass of 5,253.3 Da, which is 13 Da less than the expected mass. Due to time constraints the synthesis of the dephosphorylated SMN 250-294 variant was not repeated during the master's thesis.

5.2 Selection of the ligation direction

To eliminate the need for one deprotection step on C231, the N->C ligation direction was preferred over the C->N direction. To evaluate the feasibility of both approaches, SMN 231-249 was converted to a thioester and ligated with SMN 250-294 WT. Concurrently, SMN 201-230-$NHNH_2$ was converted and ligated with SMN 231-249-$NHNH_2$. Since both reactions proceeded quickly and without issues, the N->C ligation direction was chosen for all subsequent reactions.

5.2.1 Conversion of SMN 231-249 to the corresponding thioester

The conversion reaction was carried out using 100 equivalents of MPAA as thiol for the thiolysis reaction. Once the optimal pH of 6.5 was established, the conversions proceeded quickly and reliably.

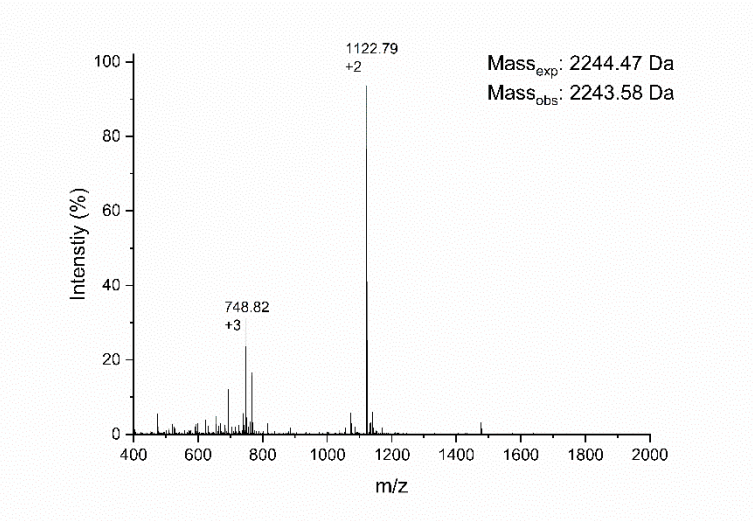

Figure 39: LCMS analysis of the SMN 231-249 MPAA thioester

5.2.2 Ligation of SMN 231-249 and SMN 250-294

For the first NCL reaction the total amount of 1.8 mg of the SMN 231-249 thioester was used together with 2.9 mg of the SMN 250-294, 8.9 mg of MPAA (equals 100 equivalents) and 3.0 mg of TCEP (equals 20 equivalents). The progress of the ligation is illustrated in the following waterfall diagram.

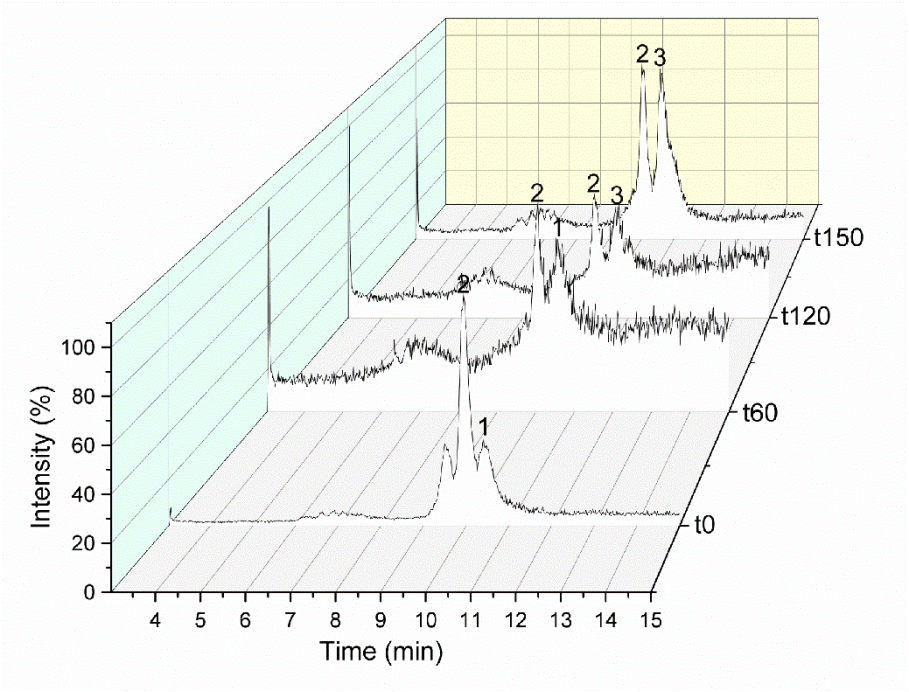

Figure 40: Time course of the ligation reaction between SMN 231-249 thioester and SMN 250-294-NHNH$_2$

The ligation product, SMN 231-294 WT (3), eluted at 10.45 minutes in the LC-MS analysis. The MS spectrum from the final LC-MS run is shown in Figure 41. By the end of the reaction, SMN 250-294 WT (1) was no longer detectable in the LC-MS, although the thioester (2) was still present in the peak at 9.80 minutes.

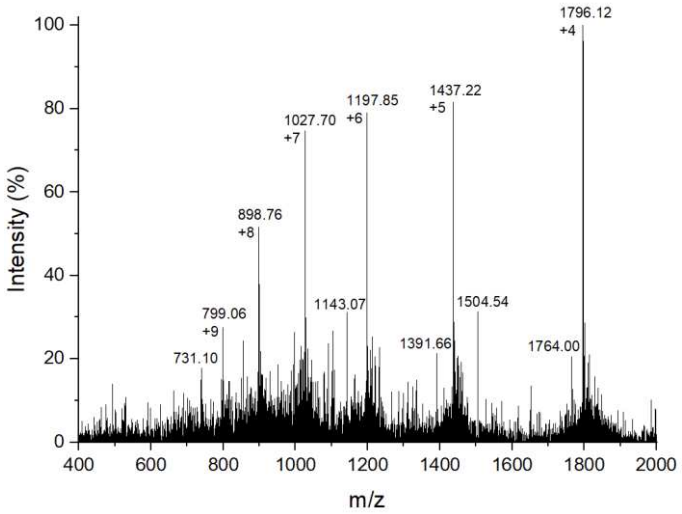

Figure 41: SMN 231-294 WT, spectrum from the 10.45 min peak of the last ligation monitoring

After quenching the reaction with an excess of 6 M guanidinium chloride solution and a spatula tip of TCEP, the reaction mixture was purified.

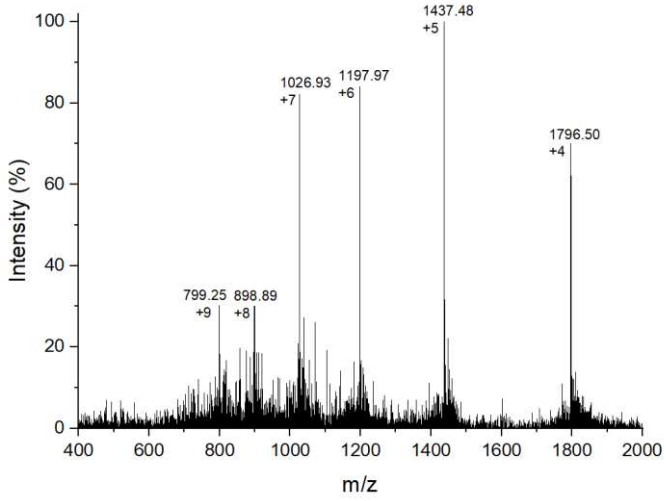

Figure 42: SMN 231-294 WT after purification

5.2.3 Conversion of SMN 201-230 to the corresponding thioester

The conversion was initially performed with 200 eq. of MPAA. After some problems concerning hydrolysis arose, the reagent was switched to MESNa. An optimal pH of 6.8 was determined.

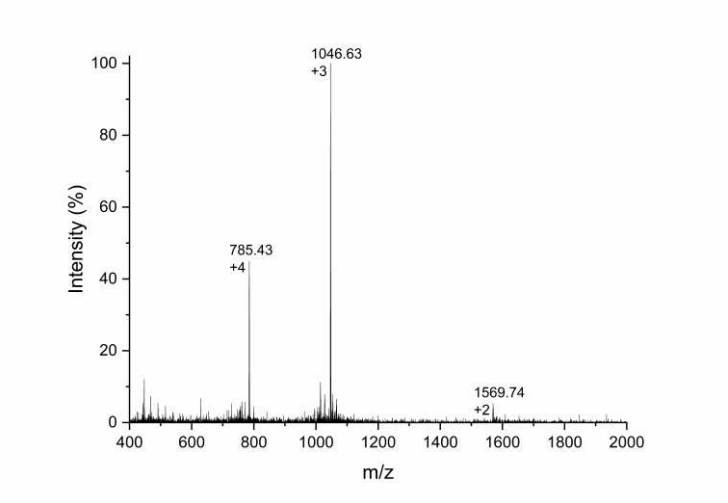

Figure 43: SMN 201-230 MESNa thioester

5.2.4 Ligation of SMN 201-230 and SMN 231-249

After a test ligation confirmed the feasibility of the SMN 201-230 and SMN 213-249 ligation, larger scale reactions were performed. For the ligation 11.38 mg of thioester were used together with 5 mg of SMN 230-249 and 80 mg of MESNa. After the purification, 1.73 mg of the ligation product were obtained.

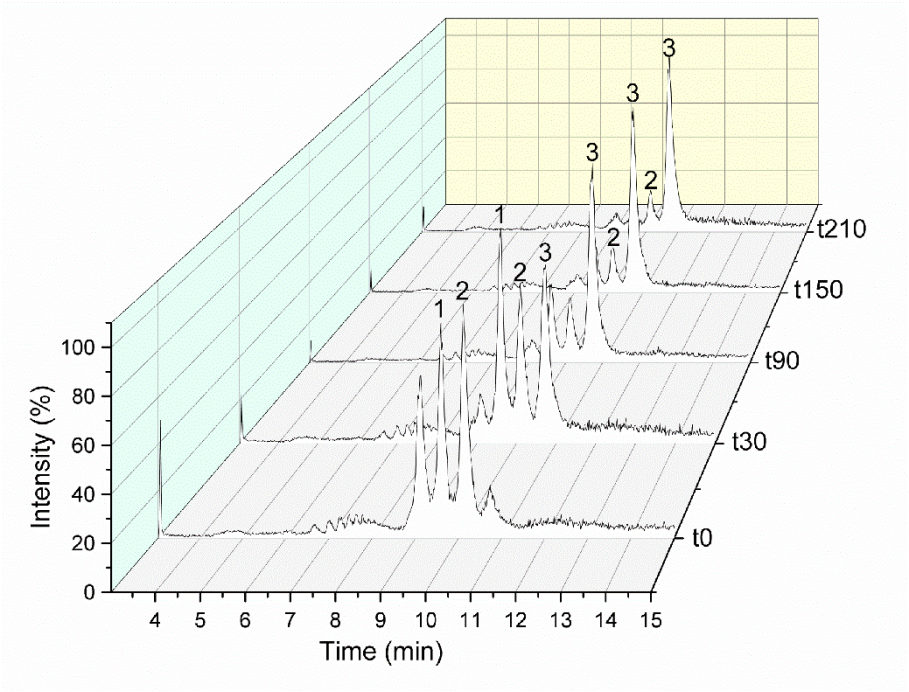

Figure 44: Time course of the ligation reaction between SMN 201-230 thioester and SMN 231-249-NHNH2

The peak at 10.40 resembles the ligation product (3) with a calculated mass of 5,032.04 Da. The TE (1) and the PH (2) are both decreasing over the time course. After the reaction was quenched, the mixture was purified with a gradient of 5-20 % B in 15 minutes followed by 20-70 % B in 55 minutes. A 10-minute desalting step was included.

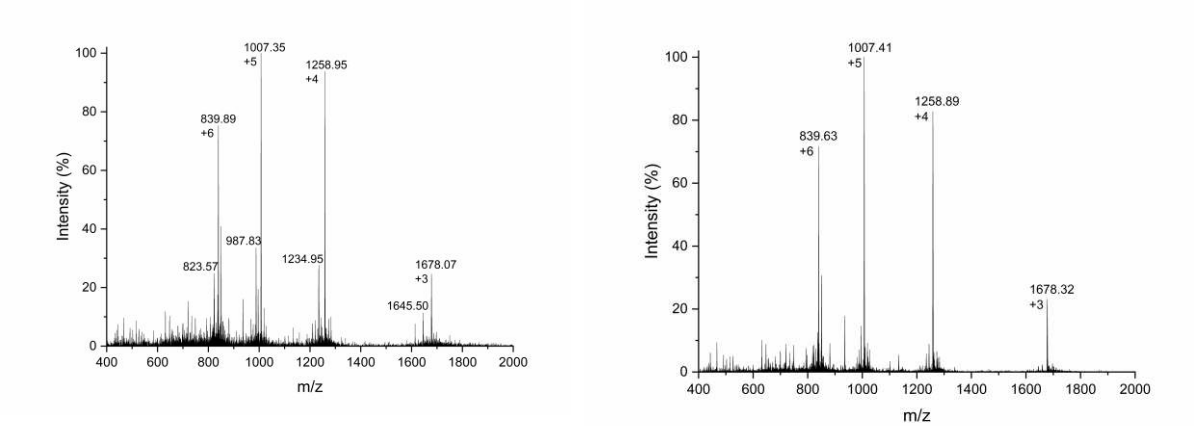

Figure 45: SMN 201-249 before and after purification

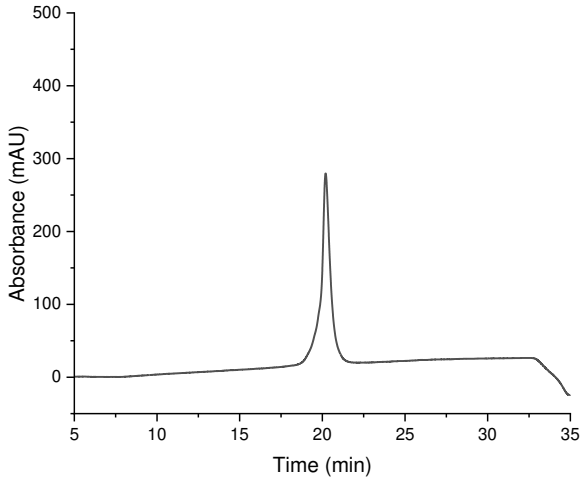

Figure 46: Analytical HPLC of SMN 201-249-NHNH2

5.3 One pot reaction to the final ligation product

In the initial attempt of the final reaction, 1.73 mg of SMN 201-249 was used in combination with 1.17 mg of SMN 250-294. During the first phase, 7 µL of $NaNO_2$ was added, followed by 7.70 mg of MPAA in the second phase. The following day, an additional small amount of TCEP was introduced, as the final product was still not visible. The subsequent LC-MS spectrum confirmed the formation of the SMN 201-294 product. After purification, 0.52 mg of SMN 201-294 was obtained. For final analysis, a small sample was sent to the mass center, where a high-resolution spectrum was recorded on the timsTOF, as shown in Figure 47.

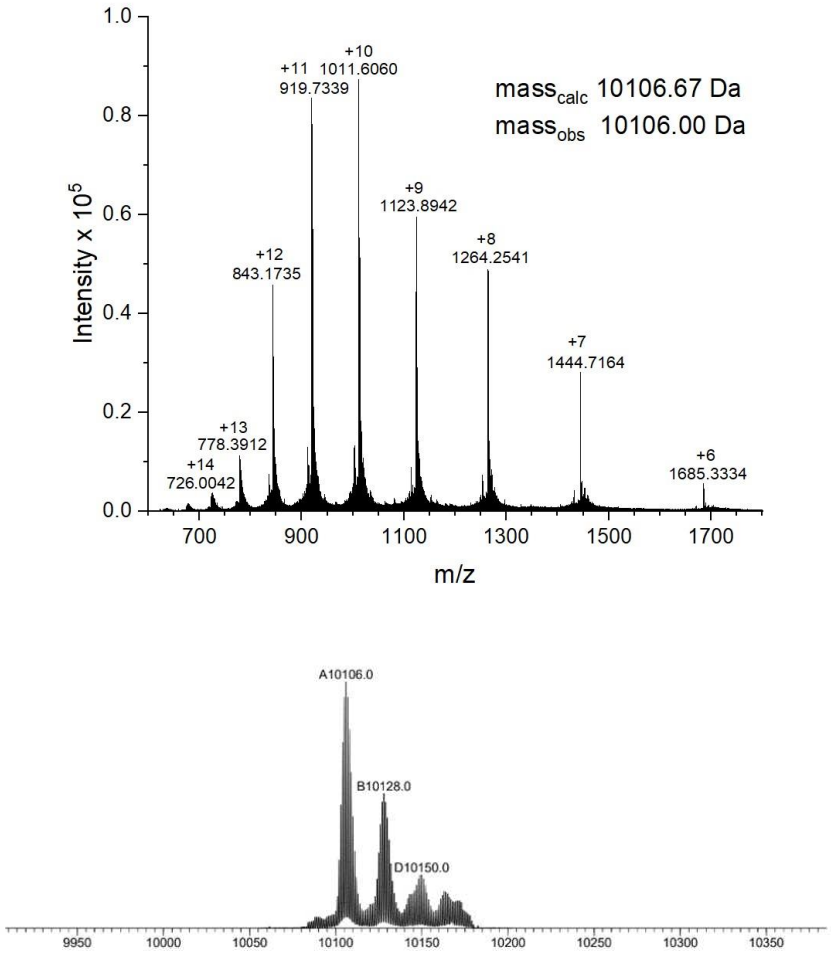

Figure 47: High resolution mass spectrum of SMN 201-294 WT

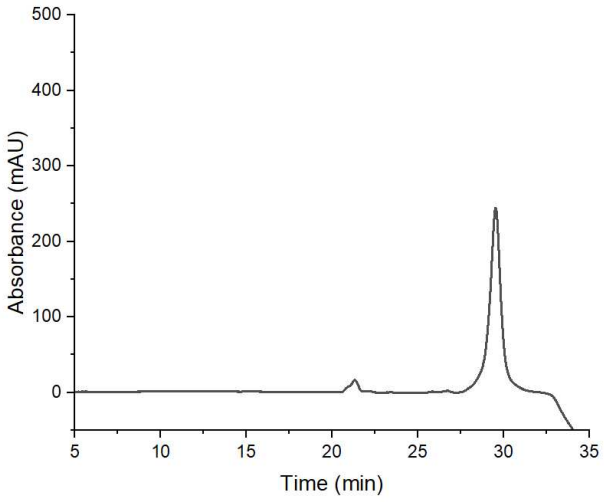

Figure 48: Analytical HPLC run of SMN 201-294 WT after purification

The reaction was repeated using SMN 250-294 pS290 to produce the SMN 201-294 pS290 product. For this attempt, 2.6 mg of SMN 201-249 and 1.8 mg of SMN 250-294 pS290 were used. However, only a small amount of the final product was produced, which coeluted with SMN 250-294 pS290. As a result, no pure material could be obtained.

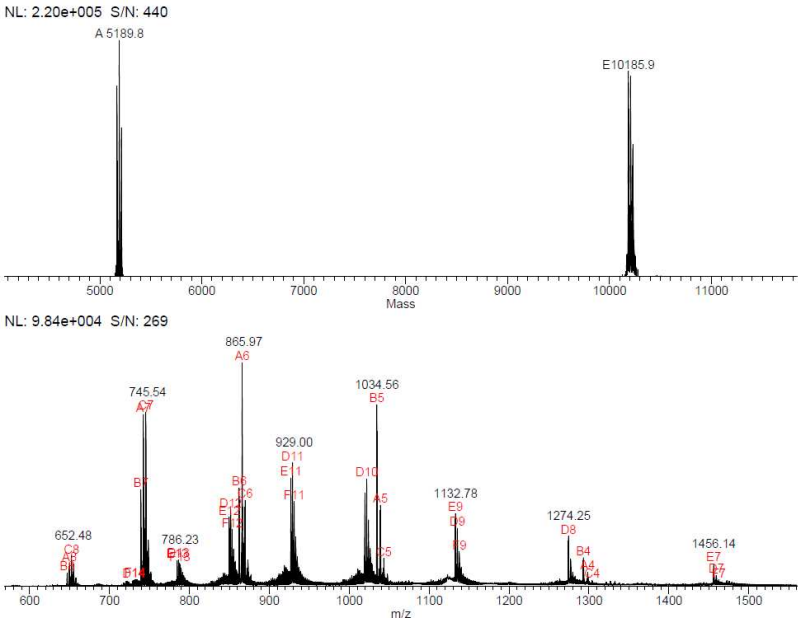

Figure 49: Spectrum of SMN 201-294 pS290

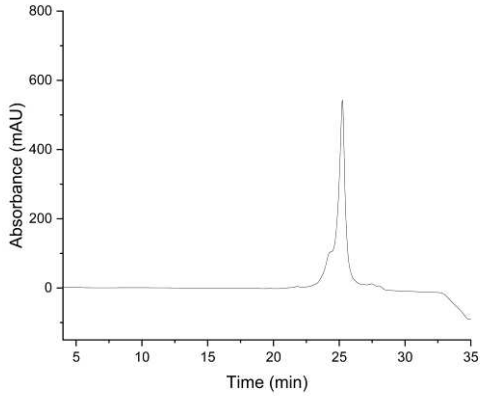

Figure 50: Ananlytical HPLC run of SMN 201-294 pS290

The third variant, SMN 201-294 pS292, was produced under the same conditions as mentioned before. To improve the problems with the purification, a slower gradient of 5-20 % B in 15 minutes, followed by an increase to 70 % B in 40 minutes was selected. A 10-minute desalting step was included. After purification, it could be seen that the material still coeluted with SMN 250-294 pS292. An analytical HPLC was performed to check if separation was possible at all. During the run the peaks were collected and analyzed with the help of the mass center.

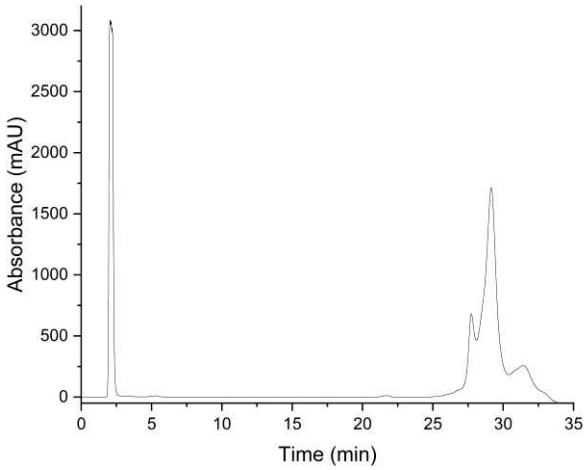

Figure 51: Analytical HPLC of SMN 201-294 pS292 after one round of purification

While the material of the first peak, as can be seen in Figure 52, contains a side product denoted as B, the spectrum of the second peak, shown in Figure 53, contains the pure product.

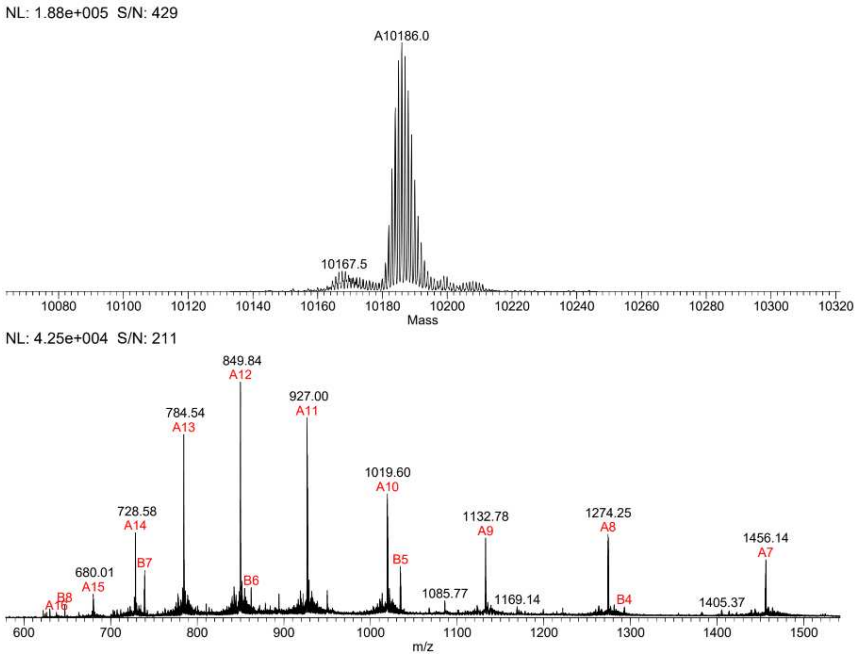

Figure 52: Spectrum of SMN 201-294 pS292, from the first peak

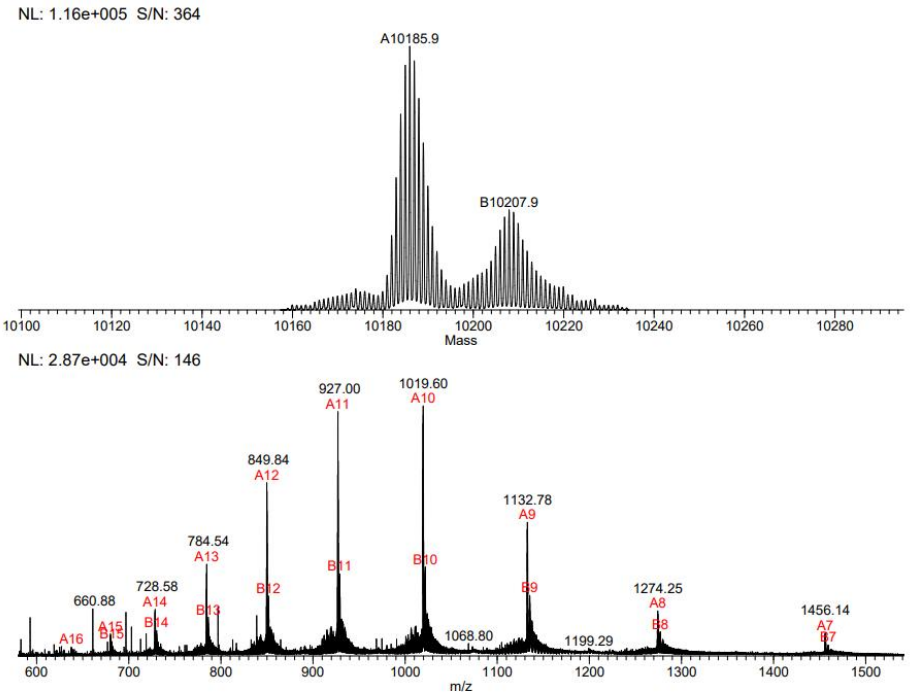

Figure 53: High resolution mass spectrum of SMN 201-294 pS292

The deconvoluted spectrum shows the compound with an observed mass of 10,185.9 Da and a second species with a mass of 10,207.9 Da, which corresponds to an adduct with sodium. The spectrum therefore verified that the product can be obtained in a sufficient purity.

5.4 Discussion of the results of the synthesis

Compared to the initial uncertainty about whether the synthesis could achieve satisfactory purity, the results turned out better than anticipated. It has been determined that the key to improving the quality of the final product is washing the peptide thoroughly during synthesis. The reason why the SMN 250-294 pS290 pS292 peptide did not work remains unclear. It is speculated that microwave irradiation during synthesis might have caused β-elimination. A research article has confirmed that piperidine can react with dehydroalanine, which is formed from phosphoserine after β-elimination. [47] This reaction could explain the observed mass difference of 13 Da. To test this hypothesis, it would be necessary to either synthesize the full eptide manually or use an automated synthesizer that does not use microwave irradiation. Another option would be to further investigate the product via mass spectrometry. A MS/MS spectrum for example could yield information about the fragmentation pattern which would allow further analysis. The problem in this case was that additional MS techniques would need knowhow and expertise for the evaluation process. For time reasons these experiments were therefore not performed.

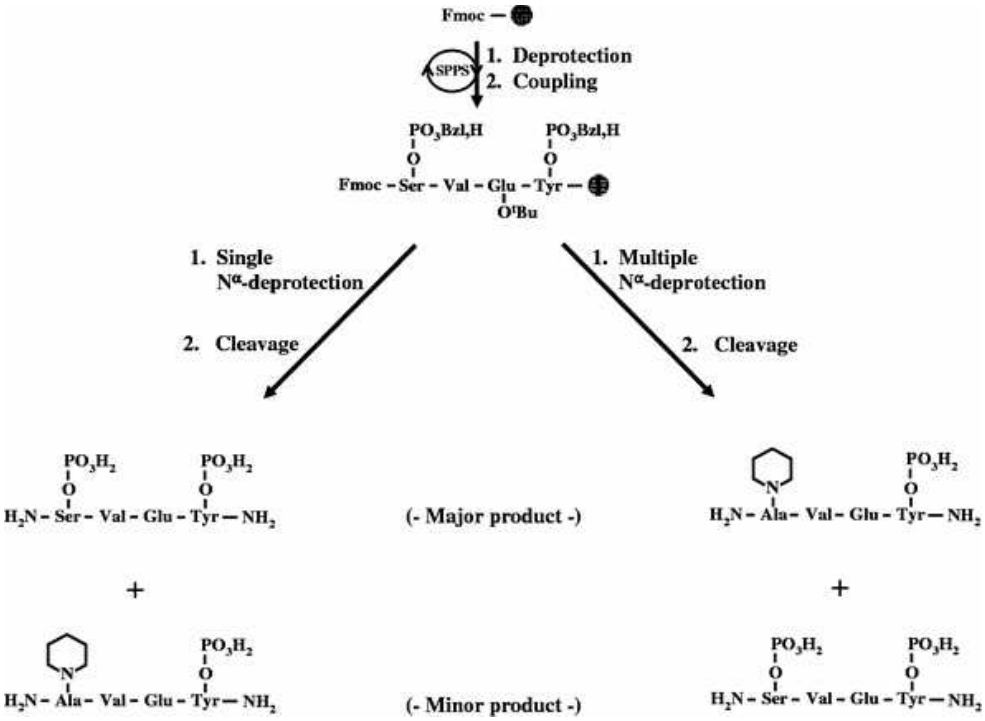

Figure 54: Scheme depicting the possible reaction path that happened during the synthesis of the diphosphorylated peptide [47]

Regarding the ligation reactions performed, the major issue encountered was the low yield after both the initial ligation and the one-pot reaction. Evidence indicates that the starting material is indeed consumed during the reaction, suggesting that the problem may lie in the product not eluting from the column effectively. To address this issue, future experiments should explore different elution conditions. Additionally, it has been observed that producing both monophosphorylated compounds is more challenging because their higher polarity shifts the retention time closer to that of the remaining segment.

6 References

[1] E. Mercuri *et al.,* "Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care," *Neuromuscular disorders : NMD*, vol. 28, no. 2, pp. 103–115, 2018, doi: 10.1016/j.nmd.2017.11.005.

[2] C. H. Wang *et al.,* "Consensus statement for standard of care in spinal muscular atrophy," *Journal of child neurology*, vol. 22, no. 8, pp. 1027–1049, 2007, doi: 10.1177/0883073807305788.

[3] S. Lefebvre *et al.,* "Identification and characterization of a spinal muscular atrophy-determining gene," *Cell*, vol. 80, no. 1, pp. 155–165, 1995, doi: 10.1016/0092-8674(95)90460-3.

[4] S. Lefebvre *et al.,* "Correlation between severity and SMN protein level in spinal muscular atrophy," *Nature genetics*, vol. 16, no. 3, pp. 265–269, 1997, doi: 10.1038/ng0797-265.

[5] Q. Liu and G. Dreyfuss, "A novel nuclear structure containing the survival of motor neurons protein," *The EMBO Journal*, vol. 15, no. 14, pp. 3555–3565, 1996, doi: 10.1002/j.1460-2075.1996.tb00725.x.

[6] P. Selenko, R. Sprangers, G. Stier, D. Bühler, U. Fischer, and M. Sattler, "SMN tudor domain structure and its interaction with the Sm proteins," *Nature structural biology*, vol. 8, no. 1, pp. 27–31, 2001, doi: 10.1038/83014.

[7] T. Giesemann *et al.,* "A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with smn in nuclear gems," *The Journal of biological chemistry*, vol. 274, no. 53, pp. 37908–37914, 1999, doi: 10.1074/jbc.274.53.37908.

[8] R. Martin, K. Gupta, N. S. Ninan, K. Perry, and G. D. van Duyne, "The survival motor neuron protein forms soluble glycine zipper oligomers," *Structure (London, England : 1993)*, vol. 20, no. 11, pp. 1929–1939, 2012, doi: 10.1016/j.str.2012.08.024.

[9] U. R. Monani *et al.,* "A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2," *Human molecular genetics*, vol. 8, no. 7, pp. 1177–1183, 1999, doi: 10.1093/hmg/8.7.1177.

[10] S. Cho and G. Dreyfuss, "A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity," *Genes & development*, vol. 24, no. 5, pp. 438–442, 2010, doi: 10.1101/gad.1884910.

[11] G. Meister, D. Bühler, R. Pillai, F. Lottspeich, and U. Fischer, "A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs," *Nature cell biology*, vol. 3, no. 11, pp. 945–949, 2001, doi: 10.1038/ncb1101-945.

[12] S. M. Berget, C. Moore, and P. A. Sharp, "Spliced segments at the 5' terminus of adenovirus 2 late mRNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 8, pp. 3171–3175, 1977, doi: 10.1073/pnas.74.8.3171.

[13] B. Ruskin, A. R. Krainer, T. Maniatis, and M. R. Green, "Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro," *Cell*, vol. 38, no. 1, pp. 317–331, 1984, doi: 10.1016/0092-8674(84)90553-1.

[14] M. S. Jurica and M. J. Moore, "Pre-mRNA splicing: awash in a sea of proteins," *Molecular cell*, vol. 12, no. 1, pp. 5–14, 2003, doi: 10.1016/S1097-2765(03)00270-3.

[15] P. Morais, H. Adachi, and Y.-T. Yu, "Spliceosomal snRNA Epitranscriptomics," *Frontiers in genetics*, vol. 12, p. 652129, 2021, doi: 10.3389/fgene.2021.652129.

[16] S. Otter, M. Grimmler, N. Neuenkirchen, A. Chari, A. Sickmann, and U. Fischer, "A comprehensive interaction map of the human survival of motor neuron (SMN) complex," *The Journal of biological chemistry*, vol. 282, no. 8, pp. 5825–5833, 2007, doi: 10.1074/jbc.M608528200.

[17] U. Fischer, Q. Liu, and G. Dreyfuss, "The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis," *Cell*, vol. 90, no. 6, pp. 1023–1029, 1997, doi: 10.1016/s0092-8674(00)80368-2.

[18] M. Hallais *et al.,* "CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing," *Nature structural & molecular biology*, vol. 20, no. 12, pp. 1358–1366, 2013, doi: 10.1038/nsmb.2720.

[19] H. Izumi, A. McCloskey, K. Shinmyozu, and M. Ohno, "p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly," *Nucleic acids research*, vol. 42, no. 6, pp. 3998–4007, 2014, doi: 10.1093/nar/gkt1365.

[20] D. J. Battle, C.-K. Lau, L. Wan, H. Deng, F. Lotti, and G. Dreyfuss, "The Gemin5 protein of the SMN complex identifies snRNAs," *Molecular cell*, vol. 23, no. 2, pp. 273–279, 2006, doi: 10.1016/j.molcel.2006.05.036.

[21] K. B. Shpargel and A. G. Matera, "Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 48, pp. 17372–17377, 2005, doi: 10.1073/pnas.0508947102.

[22] V. A. Raker, K. Hartmuth, B. Kastner, and R. Lührmann, "Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner," *Molecular and cellular biology*, vol. 19, no. 10, pp. 6554–6565, 1999, doi: 10.1128/mcb.19.10.6554.

[23] J. Pánek *et al.,* "The SMN complex drives structural changes in human snRNAs to enable snRNP assembly," *Nature communications*, vol. 14, no. 1, p. 6580, 2023, doi: 10.1038/s41467-023-42324-0.

[24] U. Fischer, V. Sumpter, M. Sekine, T. Satoh, and R. Lührmann, "Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that

binds a transport receptor independently of the m3G cap," *The EMBO Journal*, vol. 12, no. 2, pp. 573–583, 1993, doi: 10.1002/j.1460-2075.1993.tb05689.x.

[25] Z. J. Lorković and A. Barta, "Role of Cajal bodies and nucleolus in the maturation of the U1 snRNP in Arabidopsis," *PloS one*, vol. 3, no. 12, e3989, 2008, doi: 10.1371/journal.pone.0003989.

[26] M. Strzelecka, A. C. Oates, and K. M. Neugebauer, "Dynamic control of Cajal body number during zebrafish embryogenesis," *Nucleus (Austin, Tex.)*, vol. 1, no. 1, pp. 96–108, 2010, doi: 10.4161/nucl.1.1.10680.

[27] M. E. Taliansky, A. J. Love, A. Kołowerzo-Lubnau, and D. J. Smoliński, "Cajal bodies: Evolutionarily conserved nuclear biomolecular condensates with properties unique to plants," *The Plant cell*, vol. 35, no. 9, pp. 3214–3235, 2023, doi: 10.1093/plcell/koad140.

[28] M. D. Hebert, P. W. Szymczyk, K. B. Shpargel, and A. G. Matera, "Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein," *Genes & development*, vol. 15, no. 20, pp. 2720–2729, 2001, doi: 10.1101/gad.908401.

[29] I. Faravelli, G. M. Riboldi, P. Rinchetti, and F. Lotti, "The SMN Complex at the Crossroad between RNA Metabolism and Neurodegeneration," *International journal of molecular sciences*, vol. 24, no. 3, 2023, doi: 10.3390/ijms24032247.

[30] J. Navascues, M. T. Berciano, K. E. Tucker, M. Lafarga, and A. G. Matera, "Targeting SMN to Cajal bodies and nuclear gems during neuritogenesis," *Chromosoma*, vol. 112, no. 8, pp. 398–409, 2004, doi: 10.1007/s00412-004-0285-5.

[31] H. Chaytow, K. M. E. Faller, Y.-T. Huang, and T. H. Gillingwater, "Spinal muscular atrophy: From approved therapies to future therapeutic targets for personalized medicine," *Cell reports. Medicine*, vol. 2, no. 7, p. 100346, 2021, doi: 10.1016/j.xcrm.2021.100346.

[32] N. K. Singh, N. N. Singh, E. J. Androphy, and R. N. Singh, "Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron," *Molecular and cellular biology*, vol. 26, no. 4, pp. 1333–1346, 2006, doi: 10.1128/MCB.26.4.1333-1346.2006.

[33] Y. Hua, T. A. Vickers, H. L. Okunola, C. F. Bennett, and A. R. Krainer, "Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice," *American Journal of Human Genetics*, vol. 82, no. 4, pp. 834–848, 2008, doi: 10.1016/j.ajhg.2008.01.014.

[34] M. Sivaramakrishnan *et al.,* "Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers," *Nature communications*, vol. 8, no. 1, p. 1476, 2017, doi: 10.1038/s41467-017-01559-4.

[35] T. Peng, T. Das, K. Ding, and H. C. Hang, "Functional analysis of protein post-translational modifications using genetic codon expansion," *Protein science : a publication of the Protein Society*, vol. 32, no. 4, e4618, 2023, doi: 10.1002/pro.4618.

[36] A. C. Conibear, "Deciphering protein post-translational modifications using chemical biology tools," *Nature reviews. Chemistry*, vol. 4, no. 12, pp. 674–695, 2020, doi: 10.1038/s41570-020-00223-8.

[37] F. Sacco, L. Perfetto, L. Castagnoli, and G. Cesareni, "The human phosphatase interactome: An intricate family portrait," *FEBS letters*, vol. 586, no. 17, pp. 2732–2739, 2012, doi: 10.1016/j.febslet.2012.05.008.

[38] N. S. Persky *et al.,* "Defining the landscape of ATP-competitive inhibitor resistance residues in protein kinases," *Nature structural & molecular biology*, vol. 27, no. 1, pp. 92–104, 2020, doi: 10.1038/s41594-019-0358-z.

[39] G. A. Khoury, R. C. Baliban, and C. A. Floudas, "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database," *Scientific reports*, vol. 1, 2011, doi: 10.1038/srep00090.

[40] J. Mok *et al.,* "Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs," *Science signaling*, vol. 3, no. 109, ra12, 2010, doi: 10.1126/scisignal.2000482.

[41] E. Wiltshire, M. C. de Moura, D. Piñeyro, and R. S. Joshi, "Cellular and clinical impact of protein phosphatase enzyme epigenetic silencing in multiple cancer tissues," *Human genomics*, vol. 18, no. 1, p. 24, 2024, doi: 10.1186/s40246-024-00592-x.

[42] N. T. Detering, T. Schüning, N. Hensel, and P. Claus, "The phospho-landscape of the survival of motoneuron protein (SMN) protein: relevance for spinal muscular atrophy (SMA)," *Cellular and molecular life sciences : CMLS*, vol. 79, no. 9, p. 497, 2022, doi: 10.1007/s00018-022-04522-9.

[43] S. Rademacher *et al.,* "A Single Amino Acid Residue Regulates PTEN-Binding and Stability of the Spinal Muscular Atrophy Protein SMN," *Cells*, vol. 9, no. 11, 2020, doi: 10.3390/cells9112405.

[44] S. S. Kulkarni, J. Sayers, B. Premdjee, and R. J. Payne, "Rapid and efficient protein synthesis through expansion of the native chemical ligation concept," *Nat Rev Chem*, vol. 2, no. 4, 2018, doi: 10.1038/s41570-018-0122.

[45] A. C. Conibear, E. E. Watson, R. J. Payne, and C. F. W. Becker, "Native chemical ligation in protein synthesis and semi-synthesis," *Chemical Society reviews*, vol. 47, no. 24, pp. 9046–9068, 2018, doi: 10.1039/C8CS00573G.

[46] Y.-C. Huang, G.-M. Fang, and L. Liu, "Chemical synthesis of proteins using hydrazide intermediates," *National Science Review*, vol. 3, no. 1, pp. 107–116, 2016, doi: 10.1093/nsr/nwv072.

[47] T. J. Attard, N. M. O'Brien-Simpson, and E. C. Reynolds, "Identification and Suppression of β-Elimination Byproducts Arising from the Use of Fmoc-Ser(PO3Bzl,H)-OH in Peptide

Synthesis," *Int J Pept Res Ther*, vol. 15, no. 1, pp. 69–79, 2009, doi: 10.1007/s10989-008-9165-9.